DRAFT
NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review

NTP Monograph 08

September 2022

National Toxicology Program
Public Health Service
U.S. Department of Health and Human Services
ISSN: 2378-5144

Research Triangle Park, North Carolina, USA

Foreword

The National Toxicology Program (NTP), established in 1978, is an interagency collaboration within the Public Health Service of the U.S. Department of Health and Human Services. Its activities are executed through a partnership of the National Institute for Occupational Safety and Health (part of the Centers for Disease Control and Prevention), the Food and Drug Administration (primarily at the National Center for Toxicological Research), and the National Institute of Environmental Health Sciences (part of the National Institutes of Health), where this virtual program is administratively located. NTP's work focuses on the testing, research, and analysis of agents of concern to identify toxic and biological effects, provide information that strengthens the science base, and inform decisions by health regulatory and research agencies to safeguard public health. NTP also works to develop and apply new and improved methods and approaches that advance toxicology and better assess health effects from environmental exposures.

Literature-based evaluations are one means by which NTP assesses whether exposure to environmental substances (e.g., chemicals, physical agents, and mixtures) may be associated with adverse health effects. These evaluations result in hazard conclusions or characterize the extent of the evidence and are published in the NTP Monograph series, which began in 2011. NTP monographs serve as an environmental health resource to provide information that can be used to make informed decisions about whether exposure to a substance may be of concern for human health.

These health effects evaluations follow prespecified protocols that apply the general methods outlined in the "Handbook for Conducting a Literature-Based Health Assessment Using the OHAT Approach for Systematic Review and Evidence Integration."[†] The protocol describes project-specific procedures tailored to each systematic review in a process that facilitates evaluation and integration of scientific evidence from published human, experimental animal, and mechanistic studies.

Systematic review procedures are not algorithms, and the methods require scientific judgments. The key feature of the systematic review approach is the application of a transparent framework to document the evaluation methods and the basis for scientific judgments. This process includes steps to comprehensively search for studies, select relevant evidence, assess individual study quality, rate confidence in bodies of evidence across studies, and then integrate evidence to develop conclusions for the specific research question. Draft monographs undergo external peer review prior to being finalized and published.

NTP monographs are available free of charge on the NTP website and cataloged in PubMed, a free resource developed and maintained by the National Library of Medicine (part of the National Institutes of Health). Data for these evaluations are included in the Health Assessment and Workspace Collaborative.

For questions about the monographs, please email NTP or call 984-287-3211.

[†]OHAT is the abbreviation for Office of Health Assessment and Translation, which has become the Health Assessment and Translation group in the Integrative Health Assessment Branch of the Division of the National Toxicology Program at the National Institute of Environmental Health Sciences.

Table of Contents

Tables

Figures

About This Review

National Toxicology Program[1]
[1]Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Collaborators

Kyla W. Taylor, John R. Bucher, Robyn B. Blain, Christopher A. Sibrizzi, Pamela A. Hartman, Kristen Magnuson, Sorina E. Eftim, and Andrew A. Rooney

Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
Developed protocol and conducted systematic review through oversight and review of literature search, screening, data extraction, risk-of-bias assessment, and meta-analysis; wrote and edited monograph
Kyla W. Taylor, Ph.D., Project Lead
John R. Bucher, Ph.D.
Andrew A. Rooney, Ph.D.

ICF, Fairfax, Virginia, USA
Conducted literature screening, data extraction, risk-of-bias assessment, and meta-analysis; reviewed data, results, and analyses; wrote, edited, and formatted monograph; provided database and HAWC support
Robyn B. Blain, Ph.D.
Sorina E. Eftim, Ph.D.
Pamela A. Hartman, M.E.M.
Kristen Magnuson, M.E.S.M., Deputy Work Assignment Manager
Christopher A. Sibrizzi, M.P.H., Lead Work Assignment Manager

Commented [A1]: This section provides a new list of the names of all team members along with a description of tasks to which they contributed, follows our final document formatting approach, and is responsive to the [redacted] Reviewer comment below; see DocI_Monograph for detailed response:

Reviewer comment (DocI_Monograph page 11): Page 13: Should consider adding team member initials to their roles in the review.

Contributors

Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
Provided oversight for external peer review
Mary S. Wolfe, Ph.D.

Provided literature search, screening, Distiller and HAWC support
Vickie R. Walker, B.S.

Reviewed meta-analysis protocol
Suril Mehta, Dr.P.H.

Conducted a technical review of the draft monograph
Freya Kamel, Ph.D. (Retired)

Conducted literature screening
Kris Thayer, Ph.D. (currently Chemical and Pollution Assessment Division, United States Environmental Protection Agency, Washington, D.C., USA)

Division of Extramural Research and Training, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
Conducted literature screening
Heather King, Ph.D.

Kelly Government Services, Research Triangle Park, North Carolina, USA
Supported external peer review
Elizabeth A. Maull, Ph.D. (retired from NIEHS, Research Triangle Park, North Carolina, USA)

ICF, Fairfax, Virginia, USA
Provided contract oversight
David F. Burch, M.E.M., Principal Investigator
Cara Henning, Ph.D.
Jessica A. Wignall, M.S.P.H.

Reviewed data, results, and analyses
William Mendez, Ph.D.

Reviewed data, results, and analyses, conducted literature screening and risk-of-bias assessment, and provided database and HAWC support
Anna Engstrom, Ph.D.

Provided database and HAWC support
Steven Black, M.P.H.
Natalie Blanton, M.P.H.
Kaitlin Geary, B.S.
Courtney Lemeris, B.A.
Rachel McGill, B.S.
Daniel Peters, B.A.
Alessandria Schumacher, B.A.
Anna Stamatogiannakis, B.S.
River Williams, B.S.

Conducted literature screening
Carlye A. Austin, Ph.D.
Ryan Cronk, Ph.D.
Lauren Fitzharris, M.P.H.
Samantha Goodman, M.S.
Angela Hensel, B.A.
Melinda Hoang, M.P.H.
Allison Killius, B.S.
Alex Kliminsky, M.S.
Shannon McGinnis, Ph.D.

Devon Morgan, M.S.
Joei Robertson, M.Sc.
Courtney Rosenthal, M.S.
Amanda Ross, M.Ed.
Karen E. Setty, Ph.D.
Kelly Shipkowski, Ph.D.
Raquel A. Silva, Ph.D.
Parnian Soleymani, M.S.

Conducted literature screening and data extraction
Katie Duke, Ph.D.
Susan Goldhaber, M.P.H.
Revathi Muralidharan, B.S.
Jennifer Seed, Ph.D.
Codi Sharp, M.P.H.

Conducted literature screening, data extraction, and risk-of-bias assessment
Johanna Rochester, Ph.D.

Conducted literature screening and risk-of-bias assessment
Pam Ross, M.S.P.H.

Conducted data extraction
Robert Shin, M.H.S.

Conducted data extraction and risk-of-bias assessment
Tao Hong, Ph.D.
Alex Lindahl, M.P.H.
Samantha J. Snow, Ph.D., DABT

Conducted risk-of-bias assessment
Louise Assem, Ph.D.
Kristin Bornstein, Ph.D.
Jennifer Hsieh, M.S.P.H.
Yi Lu, Ph.D.

Conducted risk-of-bias assessment and reviewed data to support meta-analysis
Alexandra Goldstone, M.P.H.

Reviewed data to support meta-analysis
Jonathan Cohen, Ph.D.
Hannah Eglinton, B.A.
Cynthia J. Lin, Ph.D.

Supported meta-analysis visualizations
Rebecca Gray, M.P.H.

Coordinated peer review
Canden N. Byrd, B.S.

Designed and executed literature searches, and managed references
Michelle Cawley, M.A., M.L.S.
Jeremy S. Frye, M.S.L.S.
Nicole Vetter, M.S.L.S.

Edited and formatted monograph
Penelope Kellar, M.S.
Whitney Mitchell, B.S.
Kevin O'Donovan, B.A.

Integrated Laboratory Systems, LLC, Research Triangle Park, North Carolina, USA
Conducted a technical review of the draft monograph
Cynthia J. Willson, D.V.M., Ph.D., DACVP

Boston University, Boston, Massachusetts, USA
Reviewed the systematic review protocol
Thomas Webster, Ph.D.

Brown University, Providence, Rhode Island, USA
Reviewed the systematic review protocol
Joseph Braun, Ph.D.

Columbia University, New York, New York, USA
Reviewed the systematic review protocol
Gail Wasserman, Ph.D.

Health Research Board, Dublin, Ireland
Reviewed the systematic review protocol
Marie Sutton, Ph.D.

University of Massachusetts, Amherst, Massachusetts, USA
Reviewed the systematic review protocol
Thomas Zoeller, Ph.D.

University of Colorado, Denver, Colorado, USA
Reviewed meta-analysis protocol
Tianjing Li, Ph.D.

Peer Review

The National Toxicology Program (NTP) conducted a peer review of the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* by letter in December 2021. Reviewer selection and document review followed established NTP practices. The reviewers were charged to:

(1) Comment on the technical accuracy and whether the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* is clearly stated and objectively presented.

(2) Determine whether the scientific evidence supports the NTP's confidence ratings for the bodies of evidence regarding neurodevelopmental and cognitive health effects associated with exposure to fluoride.

NTP carefully considered reviewer comments in finalizing this monograph.

Peer Reviewers

Ethan Balk, M.D., M.P.H.
Associate Professor (Research)
Brown University School of Public Health
Providence, Rhode Island, USA

Pam Factor-Litvak, Ph.D.
Professor of Epidemiology
Columbia University Medical Center
New York, New York, USA

Erin Haynes, Dr.P.H.
Kurt W. Deuschle Professor in Preventive Medicine and Environmental Health
University of Kentucky College of Public Health
Lexington, Kentucky, USA

Julie Obbagy, Ph.D., RD
Nutritionist, Center for Nutrition Policy and Promotion
Office of Nutrition Guidance and Analysis
U.S. Department of Agriculture
Medfield, Massachusetts, USA

Heather Volk, Ph.D.
Associate Professor
Johns Hopkins Bloomberg School of Public Health
Baltimore, Maryland, USA

Publication Details

Publisher: National Toxicology Program

Publishing Location: Research Triangle Park, NC

ISSN: 2378-5144

DOI: https://ntp.niehs.nih.gov/go/785076

Report Series: NTP Monograph Series

Report Series Number: 08

Official citation: National Toxicology Program (NTP). 2022. NTP monograph on the state of the science concerning fluoride exposure and neurodevelopmental and cognitive health effects: a systematic review. Research Triangle Park, NC: National Toxicology Program. NTP Monograph 08.

Acknowledgements

This work was supported by the Intramural Research Program (ES103316, ES103317) at the National Institute of Environmental Health Sciences, National Institutes of Health and performed for the National Toxicology Program, Public Health Service, U.S. Department of Health and Human Services under contract GS00Q14OADU417 (Order No. HHSN273201600015U).

Conflict of Interest

Individuals who reviewed the systematic review protocol or meta-analysis protocol, conducted a technical review of the draft monograph, or served on the peer review panel have certified that they have no known real or apparent conflict of interest related to fluoride exposure or neurodevelopmental and cognitive health effects.

Abstract

Background: Fluoride is a common exposure in our environment that comes from a variety of sources and is widely promoted for its dental and overall oral health benefits. A 2006 evaluation by the National Research Council (NRC) found support for an association between consumption of high levels of naturally occurring fluoride in drinking water and adverse neurological effects in humans and recommended further investigation. The evidence reviewed at that time was from dental and skeletal fluorosis-endemic regions of China. Since the NRC evaluation, the number and location of studies examining cognitive and neurobehavioral effects of fluoride in humans have grown considerably, including several recent North American prospective cohort studies evaluating prenatal fluoride exposure.

In 2016, the National Toxicology Program (NTP) published a systematic review of the evidence from experimental animal studies on the effects of fluoride on learning and memory. That systematic review found a low-to-moderate level of evidence that deficits in learning and memory occur in non-human mammals exposed to fluoride.

Objective: To conduct a systematic review of the human, experimental animal, and mechanistic literature to evaluate the extent and quality of the evidence linking fluoride exposure to neurodevelopmental and cognitive effects in humans.

Method: A systematic review protocol was developed and utilized following the standardized OHAT systematic review approach for conducting literature-based health assessments. This monograph presents the current state of evidence associating fluoride exposure with cognitive or neurodevelopmental health effects and incorporated predefined assessments of study quality and confidence levels. Benefits of fluoride with respect to oral health are not addressed in this monograph.

Results: The current bodies of experimental animal studies and human mechanistic evidence do not provide clarity on the association between fluoride exposure and cognitive or neurodevelopmental human health effects.

This systematic review identified studies that assessed the association between fluoride exposure and cognitive or neurodevelopmental effects in both adults and children, which were evaluated separately. In adults, only two high-quality cross-sectional studies examining cognitive effects were available. The literature in children was more extensive and was separated into studies assessing intelligence quotient (IQ) and studies assessing other cognitive or neurodevelopmental outcomes. Eight of nine high-quality studies examining other cognitive or neurodevelopmental outcomes reported associations with fluoride exposure. Seventy-two studies assessed the association between fluoride exposure and IQ in children. Nineteen of those studies were considered to be high quality; of these, 18 reported an association between higher fluoride exposure and lower IQ in children. The 18 studies, which include 3 prospective cohort studies and 15 cross-sectional studies, were conducted in 5 different countries. Forty-six of the 53 low-quality studies in children also found evidence of an association between higher fluoride exposure and lower IQ in children.

Discussion: Existing animal studies provide little insight into the question of whether fluoride exposure affects IQ. In addition, studies that evaluated fluoride exposure and mechanistic data in humans were too heterogenous and limited in number to make any determination on biological plausibility. The body of evidence from studies in adults is also limited and provides low

confidence that fluoride exposure is associated with adverse effects on adult cognition. There is, however, a large body of evidence on IQ effects in children. There is also some evidence that fluoride exposure is associated with other neurodevelopmental and cognitive effects in children; although, because of the heterogeneity of the outcomes, there is low confidence in the literature for these other effects. This review finds, with moderate confidence, that higher fluoride exposure (e.g., represented by populations whose total fluoride exposure approximates or exceeds the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride) is consistently associated with lower IQ in children. More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ.

Commented [A2]: The following two sentences reflect revisions to clarify that our moderate confidence conclusion is primarily based on studies with total fluoride exposure that approximates or exceeds what is generally associated with consumption of optimally fluoridated water in the United States, in response to the [redacted] comments below; see DocA1_Monograph and DocB1_Monograph, respectively, for detailed response:

Reviewer comment (DocA1_Monograph, page 2): Recommendation: [redacted] requests NTP include a statement in the systematic review abstract and full text, as well as the meta-analysis, like that found in the 2020 draft monograph: "When focusing on findings from studies with exposures in ranges typically found in drinking water in the United States (0.7 mg/L for optimally fluoridated community water systems) that can be evaluated for dose response, effects on cognitive neurodevelopment are inconsistent, and therefore unclear."

Reviewer comment (DocB1_Monograph, page 2): It would be helpful if the Abstract was clear in the Discussion that the conclusion about effects on IQ in children was derived from high human exposures (higher than US exposures) without getting into more hazard conclusions or assessments.

Commented [A3]: This sentence and similar sentences throughout the monograph reflect revisions to clarify that our moderate confidence conclusion is primarily based on studies with total fluoride exposure that approximates or exceeds what is generally associated with consumption of optimally fluoridated water in the United States, in response to the [redacted] comment below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 6): Dose response: The authors correctly point out that many of the studies dealt with exposure levels considered relatively high, at least relative to the EPA drinking water standard and secondary standard. Further, in the group of studies considered to be low risk of bias, exposure was generally considered either on an arithmetic – but sometimes on a logarithmic scale (if quantitative), or based on a dichotomous variable of fluorosis, a manifestation of continuous high exposure, or whether study participants lived in an area known to have high levels of exposure. Thus, the conclusion of moderate confidence that fluoride is associated with deficits in IQ scores in children needs to be couched for these higher exposure levels, as there are few studies that provide evidence of this for exposures in the low range. This is not to say that there is no association at these lower levels, there may very well be an association; just that these results cannot be generalized to lower levels of exposure. This is true with other neurotoxins as well, for example, we know that the associations between lead and IQ scores is even steeper at the lower levels of exposure, but early studies where exposure was high were not able to discern those associations.

Preface

The National Toxicology Program (NTP) conducted a systematic review of the published scientific literature because of public concern regarding the potential association between fluoride exposure and adverse neurodevelopmental and cognitive health effects.

NTP initially published a systematic review of the experimental animal literature in 2016 that was subsequently expanded to include human epidemiological studies, mechanistic studies, and newer experimental animal literature. Because of the high public interest in fluoride's benefits and potential risks, NTP asked the National Academies of Sciences, Engineering, and Medicine (NASEM) to conduct an independent evaluation of the draft *NTP Monograph on Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* (2019 draft monograph dated September 6, 2019) and the revised draft (2020 draft monograph dated September 16, 2020), which addressed the NASEM committee's recommendations for improvement. The NASEM committee determined that, "Overall the revised monograph seems to include a wealth of evidence and a number of evaluations that support its main conclusion, but the monograph falls short of providing a clear and convincing argument that supports its assessments…." Thus, NTP has removed the hazard assessment step and retitled this systematic review of fluoride exposure and neurodevelopmental and cognitive health effects as a "state-of-the-science" document to indicate the change. This state-of-the-science document does not include the meta-analysis of epidemiological studies or hazard conclusions found in previous draft monographs; however, it provides a comprehensive and current assessment of the scientific literature on fluoride as an important resource to inform safe and appropriate use.

NTP has responded to the NASEM committee's comments on the revised draft (September 16, 2020) in a separate document (*Sup01_Monograph_NASEM_Feb_2021.pdf*) and revised relevant sections of this monograph.

Introduction

Fluoride is a common exposure in our environment from a variety of sources and is widely promoted for its dental and overall oral health benefits. Approximately 67% of the U.S. population receives fluoridated water through a community water system (CDC 2013). In other countries, fluoride supplementation has been achieved by fluoridating food products such as salt or milk. Fluoride supplementation has been recommended to prevent bone fractures (Jones et al. 2005). Fluoride also can occur naturally in drinking water. Other sources of human exposure include other foods and beverages, industrial emissions, pharmaceuticals, and pesticides (e.g., cryolite, sulfuryl fluoride). Soil ingestion is another source of fluoride exposure in young children (US EPA 2010).

The U.S. Public Health Service (PHS) first recommended that communities add fluoride to drinking water in 1962. PHS guidance is advisory, not regulatory, which means that while PHS recommends community water fluoridation as a public health intervention, the decision to fluoridate water systems is made by state and local governments. For many years, most fluoridated community water systems used fluoride concentrations ranging from 0.8 to 1.2 milligrams/liter (mg/L) (US DHHS 2015). For community water systems that add fluoride, PHS now recommends a fluoride concentration of 0.7 mg/L (equal to 0.7 parts per million [ppm]). Under the Safe Drinking Water Act, the U.S. Environmental Protection Agency (EPA) sets maximum exposure level standards for drinking water quality. The current enforceable drinking water standard for fluoride, or the maximum contaminant level (MCL), is 4.0 mg/L. This level is the maximum amount of fluoride contamination (naturally occurring, not from water fluoridation) that is allowed in water from public water systems and is set to protect against increased risk of skeletal fluorosis, a condition characterized by pain and tenderness of the major joints. EPA also has a non-enforceable secondary drinking water standard of 2.0 mg/L of fluoride, which is recommended to protect children against the tooth discoloration and/or pitting that can be caused by severe dental fluorosis during the formative period prior to eruption of teeth. Although the secondary standard is not enforceable, EPA requires that public water systems notify the public if and when average fluoride levels exceed 2.0 mg/L (NRC 2006). The World Health Organization (WHO) set a safe water guideline of 1.5 mg/L of fluoride in drinking water (first established in 1984 and reaffirmed in 1993 and 2011), which is recommended to protect against increasing risk of dental and skeletal fluorosis (WHO 2017).

As of April 2020, 1.08% of persons living in the United States (~3.5 million people) were served by community water systems (CWS) containing ≥1.1 mg/L naturally occurring fluoride. CWS supplying water with ≥1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population (~1.9 million people), and systems supplying water with ≥2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people) (CDC Division of Oral Health 2020).

~~Commonly cited h~~Health concerns ~~related~~ cited in relation to fluoride are ~~bone fractures and~~ skeletal fluorosis, lower intelligence quotient (IQ) and other neurological effects, cancer, and

Commented [A4]: The organization of the Introduction section reflects revisions to have the uses of and exposure sources to fluoride be the first topics covered, in response to the [redacted] comment below; see DocI_Monograph for detailed response:

Reviewer comment (DocI_Monograph, page 4): The scientific information presented appears technically correct and objectively presented. A few suggestions are noted below to improve clarity.
The background section could be reorganized for clarity and flow. It might be beneficial to begin the abstract and background with the pervasive use of fluoride in drinking water followed by a brief statement of the benefits. The benefits of fluoride in water has not been articulated. The benefits only need a sentence or two. The background appears to be more of a justification for the report rather than a true background of the evidence leading to the study/report.

Commented [A5]: Change made in response to the [redacted] editorial comment below; see DocB2_Monograph for detailed response:

Reviewer comment (DocB2_Monograph, page 4):~~Commonly cited health~~ Health concerns ~~related~~ cited in relation to fluoride are bone fractures and skeletal fluorosis, lower intelligence quotient (IQ) and other neurological effects, cancer, and endocrine disruption." [Text in red text inserted and red-strikethrough font deleted by the [redacted].]

Commented [A6]: Reference to 'bone fractures' removed in response to [redacted] comment below; see DocC_Monograph for detailed response:

Reviewer comment (DocC_Monograph, page 6): This statement is inflammatory. It is not a reflection of the current state of the science on this issue. However, these assertions that have been made by the Fluoride Action Network and are not evidenced-based. Ref: Osteoporos Int. 2008 Mar;19(3):257-68. Epub 2007 Aug 15.
Effects of treatment with fluoride on bone mineral density and fracture risk--a meta-analysis

P Vestergaard 1, N R Jorgensen, P Schwarz, L Mosekilde

Affiliations expand PMID: 17701094

endocrine disruption. Effects on neurological function, endocrine function (e.g., thyroid,[1] parathyroid, pineal), metabolic function (e.g., glucose metabolism), and carcinogenicity were assessed in the 2006 NRC report, *Fluoride in Drinking Water: A Scientific Review of EPA's Standards* (NRC 2006). The NRC review considered adverse effects of water fluoride, focusing on a range of concentrations (2–4 mg/L) above the current 0.7-mg/L recommendation for community water fluoridation. The NRC report concluded that the Maximum Contaminant Level Goal (MCLG), 4 mg/L, should be lowered to protect against severe enamel fluorosis and reduce the risk of bone fractures associated with skeletal fluorosis (NRC 2006). Other than severe fluorosis, NRC did not find sufficient evidence of negative health effects at fluoride levels below 4 mg/L; however, it concluded that the consistency of the results of IQ deficits in children exposed to fluoride at 2.5 to 4 mg/L in drinking water from a few epidemiological studies of Chinese populations appeared significant enough to warrant additional research on the effects of fluoride on intelligence. The NRC report noted several challenges to evaluating the literature, including deficiencies in reporting quality, lack of consideration of all sources of fluoride exposure, incomplete consideration of potential confounding, selection of inappropriate control subject populations in epidemiological studies, absence of demonstrated clinical significance of reported endocrine effects, and incomplete understanding of the biological relationship between histological, biochemical, and molecular alterations with behavioral effects.

In 2016, the National Toxicology Program (NTP) 2016, NTP published a systematic review of the evidence from experimental animal studies on the potential effects of fluoride exposure on learning and memory (NTP 2016). That systematic review found a low-to-moderate level of evidence that deficits in learning and memory occur in experimental animals exposed to fluoride. Given these findings, NTP decided to conduct additional animal studies before carrying out this full systematic review and integrate human, animal, and potentially relevant mechanistic evidence in order to reach human health hazard identification conclusions for fluoride and learning and memory effects. As the NTP (2016) report on the experimental animal evidence focused on learning and memory and developed confidence ratings for bodies of evidence by life stage of exposure (i.e., exposure during development or adulthood), this monograph also evaluates two different age groups in humans (i.e., children and adults) with a focus on cognitive neurodevelopmental effects in children and cognitive effects in adults in order to address potential differences in health impacts based on time frame of exposure (i.e., during development or during adulthood). The evaluation of experimental animal studies in this monograph has been conducted separately from the 2016 experimental animal assessment; however, like the 2016 assessment, it assessed mainly learning and memory effects in experimental animal studies to determine whether the findings inform the assessment of cognitive neurodevelopmental effects in children and cognitive effects in adults.

A committee convened by the National Academies of Sciences, Engineering, and Medicine (NASEM) reviewed earlier drafts of this monograph (September 6, 2019, and September 16, 2020) (NASEM 2020; 2021). The current document incorporates changes stemming from those reviews, and responses to the 2020 review are available as

[1]The current review has evaluated the fluoride literature with an eye toward potential thyroid effects because a large literature base has accumulated examining the interaction of fluoride with iodine uptake by the thyroid gland and consequential effects on synthesis of thyroid hormones, which are recognized to play significant roles in neurodevelopment in utero and during early childhood. This literature, along with a detailed proposed mechanism of action, was recently reviewed by Waugh (2019).

Commented [A7]: This footnote was added to increase clarity regarding the purpose of examining published data on thyroid function, in response to similar comments from two [redacted] Reviewers (note one of them provided two separate comments on this issue); see DocG_Monograph and DocI_Monograph, respectively, for detailed responses:

Reviewer comment (DocG_Monograph, page 23): A clearer statement is needed up front, ideally in the Introduction about what topics were covered by full systematic review (which are a small subset of topics of interest) and why. It's very confusing to read through repeated descriptions of topics which are not being reviewed. As an example, it's disconcerting to repeatedly see that thyroid function is an outcome of interest (without an explanation as to why this is of interest to a review of neurodevelopmental and cognitive health effects) and then to come across the statement (page 13) that "Thyroid data were … not extracted." It's difficult to pick out and follow the reasoning for excluding most topics from full evaluation. The timing of and reasoning for the decisions to focus the systematic review on just "high quality" pediatric studies is unclear.

Reviewer comment (DocG_Monograph, page 23): Of note, the Introduction and Methods do not explain why thyroid function was evaluated. This was only (partially done) on page 63.

Reviewer comment (DocI_Monograph, page 6): Thyroid function isn't mentioned until the specific aims. It should be included in background along with other possible mechanisms, if known. It is unclear why thyroid function is being evaluated as the only mechanistic pathway. A figure or illustration depicting the theoretical pathway would be helpful.

Sup01_Monograph_NASEM_Feb_2021.pdf. See Appendix B, Table B-1 for a timeline of key activities contributing to this 2022 NTP monograph, including document review activities that have occurred since 2016.

Objective and Specific Aims

Objective

The overall objective of this evaluation was to undertake a systematic review to develop NTP human health hazard identification conclusions on the association between exposure to fluoride and neurodevelopmental and cognitive effects based on assessing levels of evidence from human and non-human animal studies with consideration of the degree of support from mechanistic data. However, the NASEM Committee's reviews (NASEM 2020; 2021) of the 2019 and 2020 drafts of the monograph indicated that, "Overall the revised monograph seems to include a wealth of evidence and a number of evaluations that support its main conclusion, but the monograph falls short of providing a clear and convincing argument that supports its assessments…." For this reason, our methods were revised to remove the hazard assessment step (i.e., the section "Integrate Evidence to Develop Hazard Identification Conclusions" and the associated section "Translate Confidence Ratings into Level of Evidence for Health Effect"). In addition, a meta-analysis of the epidemiological studies examining children's IQ in relation to fluoride exposure added to the 2020 draft in response to NASEM comments (NASEM 2020) ~~will be published separately~~ was removed for further refinement in preparation for a separate publication and is not part of this document.

Therefore, the objective of this monograph is to undertake a systematic review of the literature concerning the association between fluoride exposure and neurodevelopmental and cognitive effects and to determine the level of confidence in that evidence. The assessment was based on evidence from human and non-human animal studies with consideration of mechanistic information.

Specific Aims

- Identify literature that assessed neurodevelopmental and cognitive health effects, especially outcomes related to learning, memory, and intelligence, following exposure to fluoride in human, animal, and relevant in vitro/mechanistic studies.
- Extract data on potential neurodevelopmental and cognitive health effects from relevant studies.
- Assess the internal validity (risk of bias) of individual studies using pre-defined criteria.
- Assess effects on thyroid function to help evaluate potential mechanisms of impaired neurobehavioral[2] function.
- Summarize the extent and types of health effects evidence available.

[2]The specific aim in the protocol refers to "impaired neurological function"; however, it was changed to "impaired neurobehavior function" in this document to use more precise terminology. The overall aim from the protocol remained the same for this evaluation.

Commented [A8]: Change made in response to the [redacted] comment below; see DocF_Monograph for detailed response:

Reviewer comment (DocF_Monograph, page 12): I would not presume that this will happen. Revise.

Commented [A9]: This footnote was added, and text here and throughout the monograph reflect revisions to change the term "neurologic" to "neurobehavioral" (or other appropriate text), in response to the [redacted] Reviewer comment below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 4): The term neurologic to refer to outcomes such as anxiety and aggression (and other neurobehavioral outcomes) is not quite correct. Neurologic would refer to outcomes such as tremor or other objective neurological signs. The more correct term would be neurobehavioral. I have marked some of this in the text.

- Describe limitations of the systematic review, strengths and limitations of the evidence base, identify areas of uncertainty, as well as data gaps and research needs for neurodevelopmental and cognitive health effects of fluoride.

Depending on the extent and nature of the available evidence:

- Synthesize the evidence using a narrative approach.
- Rate confidence in the body of evidence for human and animal studies separately according to one of four statements: High, Moderate, Low, or Very Low/No Evidence Available.

Methods

Problem Formulation and Protocol Development

The research question and specific aims stated above were developed and refined through a series of problem formulation steps, including:

(1) receipt of a nomination from the public in June 2015 to conduct analyses of fluoride and developmental neurobehavioral toxicity;

(2) analysis of the extent of evidence available and the merit of pursuing systematic reviews, given factors such as the extent of new research published since previous evaluations and whether these new reports address or correct the deficiencies noted in the literature (OEHHA 2011; NRC 2006; SCHER 2011);

(3) request for information in a Federal Register notice (dated October 7, 2015);

(4) consideration of comments providing a list of studies to review through Federal Register notice and public comment period from October 7, 2015, to November 6, 2015;

(5) release of draft concept titled *Proposed NTP Evaluation on Fluoride Exposure and Potential for Developmental Neurobehavioral Effects* in November 2015;

(6) presentation of draft concept at the NTP Board of Scientific Counselors (BSC) meeting on December 1–2, 2015;

(7) consideration of comments on NTP's draft concept from the NTP BSC meeting in December 2015; and

(8) consideration of input on the draft protocol from review by technical advisors.

The protocol used to conduct this systematic review was posted in June 2017 with updates posted in May 2019 and September 2020 (https://ntp.niehs.nih.gov/go/785076).[3] The protocol served as the complete set of methods followed for the conduct of this systematic review. The OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673) is a source of general systematic review methods that were selected and tailored in developing this protocol. Options in the OHAT handbook that were not specifically referred to in the protocol were not part of the methods for the systematic review.

A brief summary of the methods is presented below. Although the methods were revised to remove the hazard assessment step and meta-analysis from this document, the protocol was not further revised.

PECO Statements

PECO (**P**opulation, **E**xposure, **C**omparators and **O**utcomes) statements were developed as an aid to identify search terms and appropriate inclusion/exclusion criteria for addressing the overall research question (effects on neurodevelopmental or cognitive function and thyroid associated

[3]NTP conducts systematic reviews following prespecified protocols that describe the review procedures selected and applied from the general methods outlined in the OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673). The protocol describes project-specific procedures tailored to each systematic review that supersede the methods in the OHAT Handbook.

Commented [A10]: Footnote 3 reflects revisions to clarify that the protocol describes all the methods used for this systematic review described in the monograph in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 3): Recommendation: NTP should specify the areas where they departed from the OHAT protocol.

Commented [A11]: The following two sentences reflect revisions to further clarify the role of the OHAT handbook and the stand-alone nature of the protocol, in response to the [redacted] Reviewer comment below; see DocC_Monograph for detailed response:

Reviewer comment (DocC_Monograph, page 2): See NASEM comment about the protocol on p. 4 of Response to Fluoride NASEM Letter 10.5.2021:"Although the statement clarifies the general role of the handbook, the committee finds that it does not address the committee's previous recommendation to set the expectation for how closely the process described in the handbook will be followed in the protocol and in the eventual systematic review. For example, the handbook section "Key Questions and Analytical Framework" that guides development of the population, exposure, comparator, and outcomes (PECO) statement is not included in the fluoride protocol or the revised monograph. As the committee recommended in its previous review, NTP should treat each systematic review protocol as a stand-alone document that contains all the information necessary for understanding of the planning and conduct of the review, and these expectations should be explicitly stated in the protocol. The committee did not find that revisions of the protocol adequately addressed this recommendation.

with fluoride exposure) for the systematic review (Higgins and Green 2011). The PECO statements are listed below for human, animal, and in vitro/mechanistic studies (see Table 1, Table 2, and Table 3).

Using the PECO statements, the evaluation searched human studies, controlled exposure animal studies, and mechanistic/in vitro studies for evidence of neurodevelopmental or cognitive function and thyroid effects associated with fluoride exposure. Mechanistic data can come from a wide variety of studies that are not intended to identify a disease phenotype. This source of experimental data includes in vitro and in vivo laboratory studies directed at cellular, biochemical, and molecular mechanisms and attempt to explain how a substance produces particular adverse health effects. The mechanistic data were first organized by general categories (e.g., biochemical effects in the brain and neurons, neurotransmitters, oxidative stress) to evaluate the available information. To prioritize and consider available mechanistic data, the ~~Categories~~ categories focused on were those with more robust data at levels of fluoride more relevant to human exposure. The intent was not to develop a mechanism for fluoride induction of effects on learning and memory but to evaluate whether a plausible series of mechanistic events exists to support effects observed in the low-dose region (below approximate drinking-water-equivalent concentrations of 20 ppm for animal studies) that may strengthen a hazard conclusion if one is derived.

Commented [A12]: This sentence was revised clarify how mechanistic data were prioritized for consideration in response to the [redacted] Reviewer comment below; see DocJ_Monograph for detailed response:

Reviewer comment (DocJ_Monograph, page 5): Would it be possible to define what is meant by "Categories focused on were those with more robust data at levels of fluoride more relevant to human exposure"? Should this information be documented as part of the PECO? Was this an inclusion criteria, or just used in prioritizing or weighting the evidence in drawing conclusions?

Table 1. Human PECO (Population, Exposure, Comparator and Outcome) Statement

PECO Element	Evidence
Population	Humans without restriction as to age or sex, geographic location, or life stage at exposure or outcome assessment
Exposure	Exposure to fluoride based on administered dose or concentration, biomonitoring data (e.g., urine, blood, other specimens), environmental measures (e.g., air, water levels), or job title or residence. Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; Chemical Abstracts Service Registry Number [CASRN] 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable populations not exposed to fluoride (e.g., exposure below detection levels) or exposed to lower levels of fluoride ~~(e.g., exposure below detection levels)~~[4]
Outcomes	Neurodevelopmental outcomes, including learning, memory, intelligence, other forms of cognitive behavior, other neurological/neurobehavioral[5] outcomes (e.g., anxiety, aggression, motor activity), and biochemical changes in the brain or nervous system tissue; measures of thyroid function, biochemical changes, or thyroid tissue pathology

Commented [A13]: Change made in response to the [redacted] Reviewer comment below and footnote #4 added to clarify populations exposed to lower levels of fluoride as the comparator; see DocC_Monograph for detailed response:

Reviewer comment (DocC_Monograph, page 7): how would you determine "detection" levels if you are not measuring the dose of exposure? Not clear how comparable popltns *(sic)* of not exposed are equivalent or appropriate to use in lieu of "exposed to lower levels of fluoride"

[4]Note: The "(e.g., exposure below detection limits)" was moved after "populations not exposed to fluoride" to reflect how it was used in the literature search and elsewhere in this systematic review.

[5]The human PECO statement in the protocol refers to "neurological outcomes"; however, it was changed to "neurological/neurobehavioral outcomes" in this document to use more precise terminology for the outcomes included.

Table 2. Animal PECO Statement

PECO Element	Evidence
Population	Non-human mammalian animal species (whole organism)
Exposure	Exposure to fluoride based on administered dose or concentration and biomonitoring data (e.g., urine, blood, other specimens). Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; CASRN 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable animals that were untreated or exposed to vehicle-only treatment
Outcomes	Neurodevelopmental outcomes, including learning, memory, intelligence, other forms of cognitive behavior, other neurological/neurobehavioral[6] outcomes (e.g., anxiety, aggression, motor activity), and biochemical changes in the brain or nervous system tissue; measures of thyroid function, biochemical changes, or thyroid tissue pathology

Table 3. In Vitro/Mechanistic PECO Statement

PECO Element	Evidence
Population	Human or animal cells, tissues, or biochemical reactions (e.g., ligand binding assays)
Exposure	Exposure to fluoride based on administered dose or concentration. Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; CASRN 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable cells or tissues that were untreated or exposed to vehicle-only treatment
Outcomes	Endpoints related to neurological and thyroid function, including neuronal electrophysiology; mRNA, gene, or protein expression; cell proliferation or death in brain or thyroid tissue/cells; neuronal signaling; synaptogenesis, etc.

Literature Search

Main Literature Search

Search terms were developed to identify all relevant published evidence on developmental neurobehavioral toxicity or thyroid-related health effects potentially associated with exposure to

[6]The animal PECO statement in the protocol refers to "neurological outcomes"; however, it was changed to "neurological/neurobehavioral outcomes" in this document to use more precise terminology for the outcomes included.

fluoride by reviewing Medical Subject Headings for relevant and appropriate neurobehavioral and thyroid-related terms and by extracting key neurobehavioral and thyroid-related health effects and developmental neurobehavioral terminology from reviews and a sample of relevant studies.[7] Combinations of relevant subject headings and keywords were subsequently identified. A test set of relevant studies was used to ensure the search terms retrieved 100% of the test set. Six electronic databases were searched (see Main Literature Database Search) using a search strategy tailored for each database (specific search terms used for the PubMed search are presented in Appendix B; the search strategy for other databases are available in the protocol (https://ntp.niehs.nih.gov/go/785076). A search of PubChem indicated that sodium fluoride was not found in either the Tox21 or ToxCast databases; therefore, these databases were not included in the search. No language restrictions or publication-year limits were imposed. These six databases were searched in December 2016, and the search was regularly updated during the review process through April 1, 2019.

An additional search was conducted on May 1, 2020, where human epidemiological studies with primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence) were prioritized during screening. The review of the 2020 search results focused only on the human studies because they formed the basis of the confidence ratings (see Figure 1 for framework to assess confidence) and conclusions in the September 6, 2019, draft. A supplemental literature search of Chinese-language databases (described below) was also conducted. See Appendix B, Table B-1 for a timeline of key activities contributing to this 2022 NTP monograph, including information relevant to the timing of multiple literature searches.

Publications identified in these searches are categorized as "references identified through database searches" in Figure 2. Studies identified from other sources or manual review that satisfy the PECO criteria for inclusion ~~might impact conclusions~~ are considered under "references identified through other sources" in Figure 2. Literature searches for this systematic review were conducted independently from the literature search conducted for NTP (2016). The current literature search strategy was based on the search terms used for NTP (2016) and refined for the current evaluation, including the addition of search terms to identify human studies. Although the review process identified experimental animal studies prior to 2015, the current assessment did not evaluate these studies and relied on the NTP (2016) assessment. The focus of the literature searches for this systematic review was to identify and evaluate relevant animal studies that were published since completion of the literature searches for the NTP (2016) assessment in addition to the human and mechanistic data that were not previously evaluated.

Supplemental Chinese Database Literature Search

In order to identify non-English-language studies that might not appear in databases for the main literature search, additional searches were developed for non-English-language databases. No definitive guidance was found on the most comprehensive, highest quality, or otherwise most appropriate non-English-language databases for health studies of fluoride. Therefore, databases were chosen that identified non-English-language studies that were not captured in searches of databases from the main literature search—those previously identified from other resources (see the Searching Other Resources section below). Multiple non-English-language databases were

[7]The terms "study" and "publication" are used interchangeably in this document to refer to a published work drawn from an original body of research conducted on a defined population.

Commented [A14]: Change made in response to the [redacted] Reviewer comment below; see DocC_Monograph for detailed response:

Reviewer comment (DocC_Monograph, page 8): what is the relationship of the 11 studies identified through these means proportional to those included from the database searches? it is concerning that there were 11 studies that were NOT identified in the database search that may have been important... does this represent literature/study selection bias?

Commented [A15]: This paragraph and the next reflect revisions to the monograph that describe the additional steps the review team took to translate and extract data from all non-English language studies identified from the Chinese database searches (including some that were not previously extracted), in response to the [redacted] Reviewer comments below; see DocG_Monograph, DocH_Monograph, and DocJ_Monograph, respectively, for detailed responses:

Reviewer comment (DocG_Monograph, page 4): Regarding the overall search (for all topics) in the Chinese databases (page 10), the strategy as described is unacceptable and flies in the face of the goal of systematic review. The statements "A primary goal of the screening of the newly-retrieved human references in the supplemental search of Chinese databases was to identify null, or no-effect, studies" and "Null studies that were identified were translated and included." A plain reading of these statements suggests a high degree of bias by the researchers such that evidence of an association (not null studies) were omitted. All studies should be found and included, regarding of findings.
That said, it may be that your aim was to identify studies missed due to reporting or publication bias. If that is the case, this should be stated as the primary goal. However, it might just be better to drop this concept altogether, since the purpose of searching the non-English databases was to capture studies that have not made it into (primarily Western) databases. This is sufficient explanation.

Reviewer comment (DocH_Monograph, page 9): The supplemental search of the non-English language databases is appropriate. However, what is the rationale for saying that they were used primarily to identify null or no-effect studies? Does that mean that if a study was identified that showed an association it was not abstracted? Please be a bit more clear on this.

Reviewer comment (DocJ_Monograph, page 7): In addition, can it be assumed that any non-English paper that met criteria, regardless of outcome, would have been included? While it is understandable why the confirmatory search was done, it could be perceived as biased to only search for and include papers with null findings.

explored before two were identified, CNKI and Wanfang, that covered studies previously identified from other sources. These two Chinese electronic databases were searched in May 2020 with no language restrictions or publication year limits. Search terms from the main literature search were refined to focus on human epidemiological studies. The CNKI and Wanfang databases have character limits in the search strings; therefore, key terms were prioritized using text analytics to identify the most prevalent terms from neurodevelopmental or cognitive human epidemiological studies previously identified as relevant. Search strings were designed to capture known relevant studies that were previously identified from searching other resources without identifying large numbers of non-relevant studies [the search strategy for both databases is available in the protocol (https://ntp.niehs.nih.gov/go/785076)]. Publications retrieved were compared with publications retrieved from the main literature search, and duplicates were removed. The remaining relevant publications are categorized as "references identified through database searches" in Figure 2.

New animal and mechanistic references retrieved were scanned for evidence that might extend the information currently in the September 6, 2019, draft. Although additional studies were identified, data that would materially advance the animal and mechanistic findings were not identified; therefore, these studies were not extracted nor were they added to the draft. A primary goal of the screening of the newly retrieved human references in the supplemental search of Chinese databases was to identify studies that evaluated primary neurodevelopmental or cognitive outcomes (i.e., learning, memory, and intelligence) that may have been missed in previous searches that did not include the Chinese databases. A secondary goal was to examine whether the non-English-language studies on the Fluoride Action Network website (http://fluoridealert.org/)—a site used as another resource to identify potentially relevant studies because it is known to index fluoride publications—had been selectively presented to list only studies reporting effects of fluoride. Newly retrieved human references were reviewed to identify studies that may have been missed using previous approaches. Studies identified that evaluated primary neurodevelopmental or cognitive outcomes were included and either translated or reviewed by an epidemiologist fluent in Chinese.

Commented [A16]: This sentence reflects revisions to introduce the Fluoride Action Network and to clarify that the site was used as another resource because it is known to index fluoride publications, in response to the [redacted] Reviewer comment below; see DocJ_Monograph for detailed response:

Reviewer comment (DocJ_Monograph, page 6): This is the first time the "Flouride Action" website is mentioned (and the actual hyperlink appears in the subsequent section). It may be helpful to the reader to provide some rationale for why this website was specifically targeted.

Databases Searched

Main Literature Database Search

- BIOSIS (Thomson Reuters)
- EMBASE
- PsycINFO (APA PsycNet)
- PubMed (NLM)
- Scopus (Elsevier)
- Web of Science (Thomson Reuters, Web of Science indexes the journal Fluoride)

Supplemental Chinese Database Literature Search

- CNKI
- Wanfang

Searching Other Resources

The reference lists of all included studies; relevant reviews, editorials, and commentaries; and the Fluoride Action Network website ~~(http://fluoridealert.org/)~~ were manually searched for additional relevant publications.

Commented [A17]: Change made in response to the [redacted] comment below; see DocB2_Monograph for detailed response:

Reviewer comment (DocB2_Monograph, page 4): The URL was already noted above.

Unpublished Data

Although no unpublished data were included in the review, unpublished data were eligible for inclusion, provided the owner of the data was willing to have the data made public and peer reviewed [see protocol (https://ntp.niehs.nih.gov/go/785076) for more details].

Commented [A18]: This sentence reflects revisions to state that no unpublished data were included in the monograph, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page6): Why would you exclude conference abstracts, theses, dissertations, and other non-peer reviewed reports, but include unpublished data?

Study Selection

Evidence Selection Criteria

In order to be eligible for inclusion, studies had to satisfy eligibility criteria that reflect the PECO statements in Table 1, Table 2, and Table 3.

The following additional exclusion criteria were applied [see protocol (https://ntp.niehs.nih.gov/go/785076) for additional details]:

(1) Case studies and case reports. Although there are various definitions of 'case study' and 'case report,' the terms are used here to refer to publications designed to share health-related events on a single subject or patient with a disease, diagnosis, or specific outcome in the presence of a specific exposure (see Table 4 for study design definitions).

(2) Articles without original data (e.g., reviews, editorials, or commentaries). Reference lists from these materials, however, were reviewed to identify potentially relevant studies not identified from the database searches. New studies identified were assessed for eligibility for inclusion.

(3) Conference abstracts, theses, dissertations, and other non-peer-reviewed reports.

Commented [A19]: This new sentence was added to clarify case studies and other study design definitions in response to the [redacted] Reviewer comment below. In addition, definitions of all observational study designs were added as footnotes to Table 4; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page5): As part of eligibility criteria, what is a "case study" and how does it differ from a case report? Did you have a minimum sample size? The best [redacted] could glean is >1. Table 4 implies that case series were included. Does case series mean a single group study (all had the same exposure) or a series of cases? In either case, how are these relevant? Please make explicit what the difference is between a cohort study and a cross-sectional study. I'm assuming you require cohort studies to be longitudinal, but this should be stated explicitly. Do you have a minimum duration of follow-up to count as a longitudinal study?
All of this can/should be addressed by adding Study Design rows to Tables 1-3.

Screening Process

References retrieved from the literature search were independently screened by two trained screeners at the title and abstract level to determine whether a reference met the evidence selection criteria. Screening procedures following the evidence-selection criteria in the protocol were pilot tested with experienced contract staff overseen by NTP. For citations with no abstract or non-English abstracts, articles were screened based on title relevance (the title would need to indicate clear relevance); number of pages (articles ≤2 pages were assumed to be conference reports, editorials, or letters unlikely to contain original data); and/or PubMed Medical Subject Headings (MeSH). Using this approach, literature was manually screened for relevance and eligibility against the evidence selection criteria using a structured form in SWIFT-Active Screener (Sciome) (Howard et al. 2020). While the human screeners review studies, SWIFT-Active Screener aids in this process by employing a machine-learning software program to priority-rank studies for screening (Howard et al. 2020). SWIFT-Active Screener also refines a statistical model that continually ranks the remaining studies according to their likelihood for

inclusion. In addition, SWIFT-Active Screener employs active learning to continually incorporate user feedback during title and abstract screening to predict the total number of included studies, thus providing a statistical basis for a decision about when to stop screening (Miller et al. 2016). Title and abstract screening was stopped once the statistical algorithm in SWIFT-Active Screener estimated that 98% of the predicted number of relevant studies were identified.

Studies that were not excluded during the title and abstract screening were further screened for inclusion with a full-text review by two independent reviewers using DistillerSR® (Evidence Partners), a web-based, systematic-review software program with structured forms and procedures to ensure standardization of the process. Screening conflicts were resolved through discussion and consultation with technical advisor(s), if necessary. During full-text review, studies that were considered relevant were tagged to the appropriate evidence streams (i.e., human, animal, and/or in vitro). Studies tagged to human or animal evidence streams were also categorized by outcome as primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence); secondary neurobehavioral outcomes (anxiety, aggression, motor activity, or biochemical); or related to thyroid effects. In vitro data were tagged as being related to neurological effects or thyroid effects. Translation assistance was sought to assess the relevance of non-English studies. Following full-text review, the remaining studies were "included" and used for the evaluation.

Evaluation of SWIFT-Active Screener Results

During the initial title and abstract screening of 20,883 references using SWIFT-Active Screener, approximately 38%[8] of the studies were manually screened in duplicate to identify an estimated 98.6% of the predicted number of relevant studies using the software's statistical algorithm (13,023 references were not screened). SWIFT-Active Screener predicted that there were 739 relevant studies during the initial title and abstract screening, of which 729 were identified and moved to full-text review. The SWIFT-Active Screener statistical algorithm predicted that 10 relevant studies at the title and abstract level (10 represents 1.4% × 739 predicted relevant studies; or 739 predicted relevant studies minus 729 identified relevant studies during screening) were not identified by not screening the remaining 13,023 studies.

To further consider the impact of using SWIFT-Active Screener for this systematic review, the evaluation team assessed the SWIFT-Active screening results to gain a better understanding of the relevance of the last group of studies that was screened before 98% predicted recall (i.e., 98% of the predicted number of relevant studies were identified). The goal was to determine the likelihood of having missed important studies by not screening all of the literature. To do this, the evaluation team examined subsets of studies screened in SWIFT-Active Screener for trends and followed those studies through to full-text review for a final determination of relevance and potential impact (i.e., whether the studies had data on primary outcomes). Based on this evaluation, it was estimated that the use of SWIFT-Active Screener may have resulted in missing

[8]Howard et al. (2020) evaluated the performance of the SWIFT-Active Screener methods for estimating total number of relevant studies using 26 diverse systematic review datasets that were previously screened manually by reviewers. The authors found that on average, 95% of the relevant articles were identified after screening 40% of the total reference list when using SWIFT-Active Screener. In the document sets with 5,000 or more references, 95% of the relevant articles were identified after screening 34% of the available references, on average, using SWIFT-Active Screener.

one to two relevant human studies and one to two relevant animal studies with primary neurodevelopmental or cognitive outcomes. Therefore, the use of SWIFT-Active Screener saved considerable time and resources and is expected to miss very few potentially relevant publications.

Screening of the May 2020 Literature Search Update

For the May 1, 2020, literature search, only primary human epidemiological studies were identified for data extraction. The study screening and selection process was focused on the human studies with primary outcomes for the evaluation because they form the basis of the confidence ratings and conclusions. Animal in vivo, human secondary outcome-only, and human and animal mechanistic references were identified as part of the screening process. These studies were then scanned for evidence that might extend the information in the September 6, 2019, draft. All included studies from the May 2020 literature search update appear in Appendix C; however, other than the primary human epidemiological studies, data from the new studies were not extracted unless they would materially advance the findings.

Note that NTP is aware of a conference abstract by Santa-Marina et al. on a Spanish cohort study that looked at fluoride exposure and neuropsychological development in children (Santa-Marina et al. 2019). The evaluation team conducted a targeted literature search in April 2021 to see whether the data from this study had been published. When no publication was found, the evaluation team contacted the study authors to inquire about the publication of their data. The response from the study authors indicated that the study report was being finalized but had not yet been sent to a journal for review; therefore, it was not considered here.[9]

Commented [A20]: This paragraph and footnote 9 reflect revisions to acknowledge the potential interest in Ibarluzea et al. (2021), in response to the [redacted] Reviewer comment below. Note the Ibarluzea study is addressed here and the Aggeborn study is addressed in the response to comments document; see DocA1_Monograph for detailed response:

Reviewer comment:
Issue: New evidence Recommendation
(DocA1_Monograph, page 7): The Ibarluzea and Aggeborn & Oehman studies should be evaluated and included when assessing the evidence, similar to the 15 additional studies from the Chinese databases. [redacted] also recommends NTP include a comparison between Ibarluzea et al., 2021, and Green et al., 2019, because both studies investigate fluoride exposures at levels used for water fluoridation.

Supplemental Chinese Database Searches and Human Epidemiological Studies

Supplemental searches were conducted in non-English-language databases (CNKI and Wanfang). Of the 910 references that were identified in the supplemental Chinese database searches, 13 relevant studies published in Chinese with primary neurobehavioral or cognitive outcomes were identified during title and abstract screening (which were not identified through the main literature searches). Full texts were not found for four studies after an extensive search. The remaining nine studies for which full texts were retrieved were included and were either professionally translated or evaluated by an epidemiologist fluent in Chinese for the data extraction and quality assessment steps described below. If necessary, author inquiries were conducted in Chinese to obtain missing information relevant to the assessment of the key risk-of-bias questions described below.

[9]NTP is aware that this study was published after April 2021 (Ibarluzea et al. 2021) and, therefore, is not included in this monograph because it is beyond the dates of the literature search. Even if it had been published earlier, the study would not have contributed to the body of evidence on children's IQ because the authors assessed other neurodevelopmental or cognitive effects, specifically the association between fluoride exposure and neuropsychological development in children aged 1 year using the Mental Development Index (MDI) of the Bayley Scales of Infant Development and in children aged 4 years using the General Cognitive Index (GCI) of the McCarthy Scales of Children's Abilities (MSCA). The study will be examined as part of the NTP meta-analysis, which is being prepared as a separate report for publication.

Data Extraction

Extraction Process

Data were collected (i.e., extracted) from included studies by one member of the evaluation team and checked by a second member for completeness and accuracy. Any discrepancies in data extraction were resolved by discussion or consultation with a third member of the evaluation team.

Data Availability

Data extraction was completed using the Health Assessment Workspace Collaborative (HAWC), an open-source and freely available web-based application.[10] Data extraction elements are listed separately for human, animal, and in vitro studies in the protocol (https://ntp.niehs.nih.gov/go/785076). Data for primary and secondary outcomes, as well as thyroid hormone level data, were extracted from human studies. Studies evaluating only goiters or thyroid size were not extracted because they do not provide specific information on thyroid hormone levels that would inform whether a thyroid-mediated mechanism was involved in fluoride-associated changes in neurodevelopment. All primary outcomes and functional neurological secondary outcomes (e.g., motor activity) were extracted from animal studies identified since the NTP (2016) report. For animal mechanistic data, studies were tiered based on exposure dose (with preference given to fluoride drinking-water-equivalent exposures, which were calculated using the method described in the NTP (2016) report, of 20 ppm or less as deemed most relevant to exposures in humans), exposure duration or relevant time window (i.e., developmental), exposure route (with preference given to oral exposures over injection exposures), and commonality of mechanism (e.g., inflammation, oxidative stress, changes in neurotransmitters, and histopathological changes) were considered pockets of mechanistic data. Thyroid data were not extracted for animal studies due to inconsistency in the available data in humans. In vitro studies were evaluated, although data were not extracted from these studies as none of the findings were considered informative with respect to biological plausibility. The data extraction results for included studies are publicly available and can be downloaded in Excel format through HAWC (https://hawcproject.org/assessment/405/) (NTP 2019). Methods for transforming and standardizing dose levels and results from behavioral tests in experimental animals are detailed in the protocol(https://ntp.niehs.nih.gov/go/785076).

In 2016, NTP published a systematic review of the evidence from experimental animal studies on the potential effects of fluoride exposure on learning and memory (NTP 2016). The literature searches for the current assessment identified and evaluated relevant animal studies published since the 2016 assessment and also included human and mechanistic data that were not previously evaluated. Although literature search activities for the current assessment identified experimental animal studies prior to 2015, the current assessment did not re-evaluate animal studies published prior to 2015 because these were reviewed in the NTP (2016) assessment.

[10]HAWC (Health Assessment Workspace Collaborative): A Modular Web-based Interface to Facilitate Development of Human Health Assessments of Chemicals (https://hawcproject.org/portal/).

Commented [A21]: This sentence reflects revisions to increase clarity regarding the exclusion of topics for full evaluation, in response to related comments from [redacted] Reviewers below; see DocG_Monograph and DocI_Monograph, respectively, for detailed responses. *Note: two additional sentences were revised toward the end of this paragraph to also address this issue as described in the next comment bubble.*

Reviewer comment (DocG_Monograph, page 23): A clearer statement is needed up front, ideally in the Introduction about what topics were covered by full systematic review (which are a small subset of topics of interest) and why. It's very confusing to read through repeated descriptions of topics which are not being reviewed. As an example, it's disconcerting to repeatedly see that thyroid function is an outcome of interest (without an explanation as to why this is of interest to a review of neurodevelopmental and cognitive health effects) and then to come across the statement (page 13) that "Thyroid data were … not extracted." It's difficult to pick out and follow the reasoning for excluding most topics from full evaluation. The timing of and reasoning for the decisions to focus the systematic review on just "high quality" pediatric studies is unclear.

Reviewer comment (DocI_Monograph, page 11): Page 13: there is a statement about studies 'evaluating only goiters or thyroid size were not extracted." If so, shouldn't they be p…

Commented [A22]: The following two sentences reflect revisions to increase clarity regarding the exclusion of topics for full evaluation, and that in vitro studies were not summarized because it was considered unlikely that this literature would provide sufficient information to inform an action of fluoride on neurodevelopment, in response to comments from [redacted]Reviewers listed below; see DocG_Monograph and DocI_Monograph for detailed responses:

Reviewer comment (DocG_Monograph, page 23): A clearer statement is needed up front, ideally in the Introduction about what topics were covered by full systematic review (which are a small subset of topics of interest) and why. It's very confusing to read through repeated descriptions of topics which are not being reviewed. As an example, it's disconcerting to repeatedly see that thyroid function is an outcome of interest (without an explanation as to why this is of interest to a review of neurodevelopmental and cognitive health effects) and then…

Commented [A23]: This sentence reflects revisions to clarify that the statement applies to animal studies, in response to the [redacted]Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 7): Based on the Methods section, it appears that "high quality" pediatric studies from prior to 2015 would have been excluded in the current analyses. As written (e.g., on page 14 at the end of the Data Extraction section), it seems that older data were simply ignored (without justification). However, the Results (e.g., Figure 4) includes older studies.

Quality Assessment of Individual Studies

Risk of bias was assessed for individual studies using the OHAT risk-of-bias tool (https://ntp.niehs.nih.gov/go/riskbias) that outlines a parallel approach to evaluating risk of bias from human, animal, and mechanistic studies to facilitate consideration of risk of bias across evidence streams with common terms and categories. The risk-of-bias tool is comprised of a common set of 11 questions that are answered based on the specific details of individual studies to develop risk-of-bias ratings for each question. Study design determines the subset of questions used to assess risk of bias for an individual study (see Table 4). When evaluating the risk of bias for an individual study, the direction and magnitude of association for any specific bias is considered.

Assessors were trained with an initial pilot phase undertaken to improve clarity of rating criteria and to improve consistency among assessors. Studies were independently evaluated by two trained assessors who answered all applicable risk-of-bias questions with one of four options in Table 5 following prespecified criteria detailed in the protocol (https://ntp.niehs.nih.gov/go/785076). The criteria describe aspects of study design, conduct, and reporting required to reach risk-of-bias ratings for each question and specify factors that can distinguish among ratings (e.g., what separates "definitely low" from "probably low" risk of bias).

Key Risk-of-bias Questions

In the OHAT approach, some risk-of-bias questions or elements are considered potentially more important when assessing studies because these issues are generally considered to have a greater impact on estimates of the effect size or on the credibility of study results in environmental health studies. There are three Key Questions for observational human studies: confounding, exposure characterization, and outcome assessment. Based on the complexity of the possible responses to these questions in epidemiological studies, considerations made and methods used for evaluating the Key Questions are provided below. There are also three Key Questions for experimental animal studies: randomization, exposure characterization, and outcome assessment. In addition, for animal developmental studies, failure to consider the litter as the unit of analysis was also a key risk-of-bias concern. When there was not enough information to assess the potential bias for a risk-of-bias question and authors did not respond to an inquiry for further information, a conservative approach was followed, and the studies were rated probably high risk of bias for that question.

Risk-of-bias Considerations for Human Studies

The risk of bias of individual studies in the body of evidence was considered in developing confidence ratings. The key risk-of-bias questions (i.e., confounding, exposure characterization, and outcome assessment for human studies) are discussed in the consideration of the body of evidence. For this assessment, the key risk-of-bias questions, if not addressed appropriately, are considered to have the greatest potential impact on the results. The other risk-of-bias questions, including selection of study participants, were also considered and were used to identify any other risk-of-bias concerns that may indicate serious issues with a study that could cause it to be considered high risk of bias. No study was excluded based on concerns for risk of bias; however, the low risk-of-bias studies generally drive the ratings on confidence in the results across the

Commented [A24]: This URL was added in response to the [redacted] Reviewer comment below; see DocI_Monograph for detailed response:

Reviewer comment (DocI_Monograph, page 13): Has the OHAT been published? If so, it should be referenced. Since it's a critical tool in this review, it needs to be further described. What other QA tools are available and why weren't they used? Were the Cochrane Review recommendations for assessment of the risk of bias in research studies followed?

Commented [A25]: This sentence reflects revisions to clarify that in addition to the three key risk-of-bias questions, the impact of selection bias and other risk-of-bias questions were considered, in response to related comments from [redacted] Reviewers shown below. Also note the limitation in the *Discussion* section for high risk of bias studies "*Studies did not provide sufficient direct information (e.g., participation rates or methods for selection) to evaluate selection bias*."; see DocH_Monograph and DocJ_Monograph, respectively, for detailed response.

Reviewer comment (DocH_Monograph, page 10): The focus on confounding, exposure characterization and outcome assessment are, as indicated, the key components of evaluating observational research. The other parameter is whether the participants represent the population from which they are recruited, i.e. selection bias. In prospective cohort studies this is not an issue, as the population is really the combination of those exposed and non-exposed. For cross sectional studies, this is a bit trickier, as the participants may reflect a select group within the overall population. For studies based on national or regional registries, such as the Canadian studies, this is less of a problem, but for others there is the possibility of bias, and the direction of such bias is difficult to predict. As [redacted] looked at the studies, the vast majority do not address this issue, but [redacted] believe that it is worth a discussion or at least a mention that the possibility of selection bias is real.

Reviewer comment DocJ_Monograph, page 14): In general, it is difficult to understand how cross-sectional studies that adjusted for few, or no, confounders, employed somewhat indirect measures of fluoride exposure (or did not fully capture all sources of exposures to fluoride), or had concerns related to selection bias, were designated as "low risk of bias." If, for example, some confounders were accounted for in the design or analysis, other than statistical adjustment, it may be worth noting that on Table 6 (otherwise, it appears that many papers accounted for no confounders).

body of evidence. Human evidence was evaluated with and without high risk-of-bias studies to assess the impact of these studies on confidence in the association.

High risk-of-bias studies: Studies rated probably high risk of bias for at least two key risk-of-bias questions or definitely high for any single question are considered studies with higher potential for bias (i.e., high risk-of-bias studies) and to be of low quality. Studies could also be considered high risk of bias if rated probably high risk of bias for one key risk-of-bias question along with other concerns, including potential for selection bias and concerns with statistical methods.

Low risk-of-bias studies: The remaining studies (i.e., other than the high risk-of-bias studies) were considered to have lower potential for bias (i.e., low risk of bias) and to be of high quality. Appendix E describes strengths and limitations of the low risk-of-bias/high-quality studies identified during the assessment and clarifies why they are considered to pose low risk of bias. Details on the statistical analyses are provided in the "Other potential threats" domain in order to evaluate the adequacy of the statistical approach for individual studies.

Given the number of non-English-language studies in this assessment, the potential for the translation to introduce bias was examined as described below, and it was determined that translation of non-English-language studies did not impact evaluation of risk of bias. Thirty-two of 100 studies included in the entire human body of evidence on neurodevelopmental and cognitive effects were initially published in a foreign language (Chinese) and were either translated and published in volume 41 of the journal *Fluoride* (n = 19) or were translated by the Fluoride Action Network (n = 13) (http://fluoridealert.org/researchers/translations/complete_archive/). Most of these studies were considered to have high potential for bias due to lack of information across the key risk-of-bias questions. Therefore, in order to assess whether the lack of information relevant to key risk-of-bias concerns was the result of a loss in translation, the original Chinese publications and the translated versions of the five studies that had the most potential for being included in the low risk-of-bias group of studies were reviewed by a team member ~~fluent in~~ with Chinese as first language to determine whether the translations were accurate and whether any of the risk-of-bias concerns could be addressed (An et al. 1992; Chen et al. 1991 [translated in Chen et al. 2008]; Du et al. 1992 [translated in Du et al. 2008]; Guo et al. 1991 [translated in Guo et al. 2008a]; Li et al. 2009). For all five studies, the translations were determined to be accurate, and there was no impact of the translations on the key risk-of-bias concerns.

Confounding

Covariates were determined a priori based on factors that are associated with neurodevelopment or cognition and could be related to fluoride exposure. Covariates that were considered key for all studies, populations, and outcomes included age, sex, and socioeconomic status (e.g., maternal education, household income, marital status, crowding). Additional covariates considered important for this evaluation, depending on the study population and outcome, included race/ethnicity; maternal demographics (e.g., maternal age, body mass index [BMI]); parental behavioral and mental health disorders (e.g., attention deficit hyperactivity disorder [ADHD], depression); smoking (e.g., maternal smoking status, secondhand tobacco smoke exposure); reproductive factors (e.g., parity); nutrition (e.g., BMI, growth, anemia); iodine deficiency/excess; minerals and other chemicals in water associated with neurotoxicity (e.g., arsenic, lead); maternal and paternal IQ; and quantity and quality of caregiving environment

Commented [A26]: The following sentence was added to clarify the definition of "high risk-of-bias studies". In addition, the detailed assessments of and justifications for risk-of-bias ratings for the key studies are provided in Appendix E (Details for Low Risk-of-bias Studies) to address the potential concern of confounding and exposure classification in response to the [redacted] Reviewer comment below; see DocJ_Monograph for detailed response:

Reviewer comment: (DocJ_Monograph, page 18) Page 48-49, Assessment of Risk of Bias: While the studies noted as "low risk of bias" are certainly lower risk than the studies noted as "high risk of bias," it appears that the evidence base is still subject to a number of important risks, particularly related to confounding and exposure classification (i.e., are they "low risk" or "lower risk"?).

Commented [A27]: The following four sentences reflect revisions to clarify that all translated studies were originally published in Chinese and a review team member with Chinese as first language confirmed the accuracy of the translations, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page7): The description of the translated Chinese articles (page 15) needs to be written in the active voice to better describe who was confirming the accuracy of the translation and how. What about other languages (and what were they)?

Commented [A28]: This sentence reflects revisions to clarify that age and sex are important potential confounders regardless of life stage, in response to the [redacted] Reviewer comment below; see DocI_Monograph for detailed response:

Reviewer comment (DocI_Monograph, page 8): The approach to assess risk of bias was clearly described. A brief discussion is needed about critical confounders, including a biological exposure measure for tobacco use or exposure, such as serum cotinine, and parental IQ for the child studies If there are unique confounders for child and adult studies, this needs to be articulated. It currently appears that there are no unique confounders for child and adult.

(e.g., Home Observation Measurement of the Environment [HOME] score). To be assigned a rating of probably low risk of bias for the key risk-of-bias question regarding the confounding domain, studies were not required to address every important covariate listed; however, studies were required to address the three key covariates for all studies, the potential for co-exposures, if applicable (e.g., arsenic and lead, both of which could affect cognitive function), and any other potential covariates considered important for the specific study population and outcome. For example, studies of populations in China, India, and Mexico, where there is concern about co-exposures to high fluoride and high arsenic, were required to address arsenic. If the authors did not directly specify that arsenic exposures were evaluated, groundwater quality maps were evaluated (https://www.gapmaps.org/Home/Public) in order to identify areas of China, India, and Mexico where arsenic is a concern (Podgorski and Berg 2020). If no arsenic measurements were available for the area, the arsenic groundwater quality predictions from the global arsenic 2020 map were used (Podgorski and Berg 2020). If an area had less than 50% probability of having arsenic levels greater than 10 µg/L (the WHO guideline concentration), the area was considered not to have an issue with arsenic that needed to be addressed by the study authors~~; however, it should be noted that arsenic may be associated with neurodevelopmental effects at concentrations below 10 µg/L~~.

Commented [A29]: This sentence and text throughout the monograph reflect revisions to change the word 'confounder' to 'covariate', in response to several related comments from the same [redacted] Reviewer listed below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 5): RE: confounding and covariates. Recent thinking regarding confounding requires the use of directed acyclic graphs to define variables which are theoretically confounders (based on previous literature). Thus, some clarification is needed on how the set of three important confounders were selected, i.e. sex, child age and a measure of socioeconomic status. Indeed, based on literature from other potential neurotoxins (e.g. lead, polychlorinated biphenyls, phthalates) it seems a ...

Commented [A30]: The removed text had initially been added in response to the [redacted] Reviewer comment below; however, as we were unable to appropriately support the statement with a reference, this statement has been removed. See DocH_Monograph for detailed response.

Reviewer comment (DocH_Monograph, page 10): For confounding, please see [redacted] remarks above. [redacted] do think that biological sex needs to be considered an effect modifier as in other studies of neurotoxins and neurodevelopmental outcomes. Further, as indicated later in the monograph, at ...

Exposure

Fluoride ion is rapidly absorbed from the gastrointestinal tract and is rapidly cleared from serum by distribution into calcified tissues and urinary excretion (IPCS 2002). There is general consensus that the best measures of long-term fluoride exposure are bone and/or tooth measurements, and other than measures of dental fluorosis, these were not performed in any of the studies reviewed in this document. Prolonged residence in an area with a given fluoride content in drinking water has been considered in many studies as a proxy for long-term exposure.

Commented [A31]: This paragraph was added in response to two related comments on toxicokinetics from the same [redacted] Reviewer listed below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 7): In general the report is comprehensive and includes all necessary material. Hence, [redacted] have no major additions or deletions. [redacted] have one small addition, which would be a discussion of the toxicokinetics of fluoride – this is necessary because the half life is relatively short, and a spot measure ...

Exposure was assessed using a variety of methods in the human body of evidence. Studies provided varying levels of details on the methods used and employed different exposure characterization methods to group study subjects into exposed and reference groups. Exposure metrics included spot urine (from children or mothers during at least one trimester of gestation), serum, individual drinking water, intake from infant formula, estimated total exposure dose, municipal drinking water (with residence information), evidence of dental or skeletal fluorosis, area of residence (endemic versus a non-endemic fluorosis area, with or without individual validation of exposure), burning coal (with or without fluoride), and occupation type.

Commented [A32]: This sentence was added in response to the [redacted] Reviewer comments on serum fluoride below; see DocI_Monograph for detailed responses:

Reviewer comment (DocI_Monograph, page 7): A brief discussion of serum fluoride needs to be included – similar to the urinary fluoride description (page 16).

Reviewer comment (DocI_Monograph, page 12): Add a brief section on serum fluoride levels. Urinary fluoride levels is fully described, but serum has been omitted.

Urinary fluoride levels measured during pregnancy and in children include all ingested fluoride and are considered a valid measure to estimate total fluoride exposure (Villa et al. 2010; Watanabe et al. 1995); however, the type and timing of urinary sample collection are important to consider. Urinary fluoride is thought to reflect recent exposure but can be influenced by the timing of exposure (e.g., when water was last consumed, when teeth were last brushed). When compared with 24-hour urine samples, spot urine samples are more prone to the influence of timing of exposure and can also be affected by differences in dilution; however, many studies attempted to account for dilution either by using urinary creatinine or specific gravity. Good correlations between 24-hour samples and urinary fluoride concentrations from spot samples adjusted for urinary dilution have been described (Zohouri et al. 2006). Despite potential issues with spot urine samples, if authors made appropriate efforts to reduce the concern for bias (e.g.,

Commented [A33]: This sentence was added to address the best measures for assessing long-term fluoride exposure, in response to the [redacted] Reviewer comment below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 16): Because the urine and serum biomarkers of fluoride represent relatively recent exposure, it is difficult to infer that the associations are from cumulative exposure without laying out the assumptions, i.e. long term residential history, similar habits of toothpaste use, etc.

accounting for dilution), studies that used this metric were generally considered to have probably low risk of bias for exposure.

Analytical methods to measure fluoride in biological or water samples also varied, some of which included atomic absorption, ion-selective electrode methods, colorimetric methods, or the hexamethyldisiloxane microdiffusion method. Individual-level measures of exposure were generally considered more accurate than group-level measures; however, using group-level measures (e.g., endemic versus non-endemic area) in an analysis was less of a concern if the study provided water or urinary fluoride levels from some individuals to verify that there were differences in the fluoride exposure between groups. Studies that provided results by area and also reported individual urinary or serum fluoride concentrations or other biochemical measures, including dental fluorosis in the children or urinary levels in mothers during pregnancy, were considered to have probably low risk of bias. Ideally, these studies would still need to consider and adjust for area-level clustering; however, these concerns are captured in evaluations of other potential threats to internal validity.

Commented [A34]: This sentence was added in response to the [redacted] Reviewer comment on clustering below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 11): A further concern with exposure assessment brought up in the previous review concerns the issue of clustering with regard to exposure. The authors of the monograph do a very nice job of addressing this issue as it was raised in the prior review, but pointing to the sensitivity analyses. [redacted] only concern remaining is that this is mentioned up front when the exposure characterization is discussed in the methods.

Outcome

Studies included in this evaluation used a wide variety of methods to measure IQ and other cognitive effects. Measures of IQ were generally standardized tests of IQ; however, for these standardized methods to be considered low potential for bias, they needed to be conducted in the appropriate population or modified for the study population. Because results of many of the tests to measure neurodevelopment and cognitive function can be subjective, it was important that the outcome assessors were blind to the fluoride exposure when evaluating the results of the tests. If the study reported that the assessor was blind to the exposure, this was assumed to mean that the outcome assessor did not have any knowledge of the exposure, including whether the study subjects were from high-fluoride communities. If cross-sectional studies collected biomarker measurements at the time of an IQ assessment, this was considered indirect evidence that the outcome assessor would not have knowledge of the fluoride exposure unless there was also potential for the outcome assessor to have knowledge of varying levels of fluoride by study area. In cases wherein the study did not specify that the outcome assessors were blind, the study authors were contacted and asked whether the outcome assessors were, in fact, blind to exposure. When authors responded and indicated that outcome assessors were blind to exposure or that it was not likely that they would have had knowledge of exposure, this was considered direct or indirect evidence, respectively, that blinding was not a concern for those studies.

Any discrepancies in ratings between assessors were resolved by a senior technical specialist and through discussion when necessary to reach the final recorded risk-of-bias rating for each question along with a statement of the basis for that rating. Members of the evaluation team were consulted for assistance if additional expertise was necessary to reach final risk-of-bias ratings based on specific aspects of study design or performance reported for individual studies. Study procedures that were not reported were assumed not to have been conducted, resulting in an assessment of "probably high" risk of bias. Authors were queried by email to obtain missing information, and responses received were used to update risk-of-bias ratings.

Table 4. OHAT Risk-of-bias Questions and Applicability by Study Design

Risk-of-bias Questions	Experimental Animal[a]	Human Controlled Trials[b]	Cohort[c]	Case-control[d]	Cross-sectional[e,c]	Case Report/Case Series[f]
1. Was administered dose or exposure level adequately randomized?	X	X				
2. Was allocation to study groups adequately concealed?	X	X				
3. Did selection of study participants result in the appropriate comparison groups?			X	X	X	
4. Did study design or analysis account for important confounding and modifying variables?			X	X	X	X
5. Were experimental conditions identical across study groups?	X					
6. Were research personnel blinded to the study group during the study?	X	X				
7. Were outcome data complete without attrition or exclusion from analysis?	X	X	X	X	X	
8. Can we be confident in the exposure characterization?	X	X	X	X	X	X
9. Can we be confident in the outcome assessment (including blinding of outcome assessors)?	X	X	X	X	X	X
10. Were all measured outcomes reported?	X	X	X	X	X	X
11. Were there no other potential threats to internal validity?	X	X	X	X	X	X

[a]Experimental animal studies are controlled exposure studies. Non-human animal observational studies can be evaluated using the design features of observational human studies such as cross-sectional study design.
[b]Human Controlled Trials are studies in humans with controlled exposure (e.g., randomized controlled trials, non-randomized experimental studies).
[c]Cohort studies are observational studies in humans that examine a cohort prospectively or retrospectively over time. Although cohort studies may include longitudinal analyses, it is not a prerequisite of the cohort study design.
[d]Case-control studies are observational studies in humans that compare exposures of individuals who have a specific health effect or disease with exposures of controls who do not have the effect or disease. Controls generally come from the same population from which the cases were derived.
[e]Cross-sectional studies are observational studies in humans that examine the relationship between exposures and outcomes or health effects assessed contemporaneously. ~~e~~Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).
[f]A case report (or case study) is a descriptive study of a single individual or small group in which the study of an association between an observed effect and a specific environmental exposure is based on clinical evaluations and histories of the individual(s). A case series study in environmental epidemiology is designed to share health-related events on a collection of case reports on subjects with the same or similar health outcome(s) and environmental exposure(s).

Commented [A35]: Footnote c was revised and new footnotes d, e, and f were added to provide definitions of all observational study designs, in response to the [redacted] Reviewer comments below; see DocG_Monograph and DocJ_Monograph, respectively, for detailed response:

Reviewer comment (DocG_Monograph, page 5): As part of eligibility criteria, what is a "case study" and how does it differ from a case report? Did you have a minimum sample size? The best [redacted] could glean is >1. Table 4 implies that case series were included. Does case series mean a single group study (all had the same exposure) or a series of cases? In either case, how are these relevant? Please make explicit what the difference is between a cohort study and a cross-sectional study. [redacted] assuming you require cohort studies to be longitudinal, but this should be stated explicitly. Do you have a minimum duration of follow-up to count as a longitudinal study?
All of this can/should be addressed by adding Study Design rows to Tables 1-3.

Reviewer comment (DocJ_Monograph, page12): In general, it would be helpful to the reader to describe the included study designs earlier in the respective results sections (i.e., along with the total #s of included articles, describing the # of cross-sectional, prospective cohort, etc is important).

Commented [A36]: Footnote c on cohort studies was revised in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 14): It is unclear whether you are using the term prospective (cohort study) to mean prospective (as opposed to retrospective, it's correct meaning) or longitudinal (as opposed to cross-sectional). Please use the correct term. In any case, the reader needs to know both whether studies were prospective or retrospective and whether studies were longitudinal or cross-sectional.

Answers to the risk-of-bias questions result in one of the following four risk-of-bias ratings:

Table 5. The Four Risk-of-bias Rating Options

Symbol	Description
++	***Definitely Low*** *risk of bias*: There is direct evidence of low risk-of-bias practices.
+	***Probably Low*** *risk of bias*: There is indirect evidence of low risk-of-bias practices, **OR** it is deemed that deviations from low risk-of-bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias.
– NR	***Probably High*** *risk of bias*: There is indirect evidence of high risk-of-bias practices (indicated with "–"), **OR** there is insufficient information provided about relevant risk-of-bias practices (indicated with "NR" for not reported). Both symbols indicate probably high risk of bias.
—	***Definitely High*** *risk of bias*: There is direct evidence of high risk-of-bias practices.

Organizing and Rating Confidence in Bodies of Evidence

Health Outcome Categories for Neurodevelopmental and Cognitive Effects

After data were extracted from all studies, the health effects results within the category of neurodevelopmental or cognitive effects were grouped across studies to develop bodies of evidence or collections of studies with data on the same or related outcomes. The grouping of health effect results was not planned a priori. The vast majority of the human studies evaluated IQ in children as the single outcome; therefore, the discussion of cognitive neurodevelopmental effects in children focuses on IQ studies with supporting information from data on other endpoints. Cognitive function in adults was evaluated separately. Consistent with the NTP (2016) assessment, the primary focus within the animal study body of evidence was on animal studies with endpoints related to learning and memory.

Considerations for Pursuing a Narrative or Quantitative Evidence Synthesis

This evaluation provides only a narrative review of the data; however, heterogeneity within the available evidence was evaluated to determine whether a quantitative synthesis (i.e., meta-analysis) would be appropriate. Choi et al. (2012) and Duan et al. (2018) conducted meta-analyses and found that high fluoride exposure was associated with lower IQ scores. Choi et al. (2012) was able to determine a risk ratio for living in an endemic fluorosis area but was unable to develop a dose-response relationship. Duan et al. (2018) reported a significant non-linear dose-response relationship between fluoride dose and intelligence with the relationship stated as most evident with exposures from drinking water above 4 mg/L (or 4 ppm) fluoride. Duan et al. (2018) found similar results as Choi et al. (2012) for the standardized mean difference; however, the majority of the available studies in both analyses compare populations with high fluoride exposure to those with lower fluoride exposure (with the lower exposure levels frequently in the range of drinking water fluoridation in the United States). The meta-analysis conducted in

Commented [A37]: Change made in response to the [redacted] Reviewer comment below; see DocF_Monograph for detailed response:

Reviewer comment (DocF_Monograph, page 13): Would revise because it is being submitted for publication and also [redacted] would not link it to further informing this since there are still questions about the meta-analysis.

association with this systematic review further informs this issue and will be ~~published seperately~~refined in preparation for a separate publication.

Confidence Rating: Assessment of Body of Evidence

The quality of evidence for neurodevelopmental and cognitive function outcomes was evaluated using the GRADE system for rating the confidence in the body of evidence (Guyatt et al. 2011; Rooney et al. 2014). More detailed guidance on reaching confidence ratings in the body of evidence as "high," "moderate," "low," or "very low" is provided in the protocol (https://ntp.niehs.nih.gov/go/785076). In brief, available human and animal studies on a particular health outcome were initially grouped by key study design features, and each grouping of studies was given an initial confidence rating by those features. Starting at this initial rating (see column 1 of Figure 1), potential downgrading of the confidence rating was considered for factors that decrease confidence in the results (see column 2 of Figure 1). Potential upgrading of the confidence rating was considered for factors that increase confidence in the results (see column 3 of Figure 1). Short descriptions of the factors that can decrease or increase confidence in the body of evidence for human studies are provided below [see protocol (https://ntp.niehs.nih.gov/go/785076) for additional details related to the human body of evidence, as well as considerations for experimental animal studies].

Factors to Consider for Potential Downgrading

- Risk of bias: Addresses whether the body of evidence did not account for critical factors in study quality or design, including confounding bias, selection bias, exposure assessment, and outcome assessment. Consideration for downgrading the confidence rating is based on the entire body of evidence, and the evidence is downgraded when there is substantial bias across most studies that could lead to decreased confidence in the results and when the studies without substantial bias could not support the confidence rating. Individual studies are evaluated for risk of bias based on a set of criteria (as discussed above); magnitude and direction of the bias are also considered.
- Unexplained inconsistency: Addresses inconsistencies in results across studies of similar populations and design that can be determined by assessing similarity of point estimates and extent of overlap between confidence intervals or more formally through statistical tests of heterogeneity. Sensitivity analysis can be used to assess the impact of specific variables on the outcome. Inconsistencies that can be plausibly explained by characteristics of the studies (e.g., sex-associated differences) are typically not used to support a downgrade. A downgrade would only be applied when there is an inconsistency that cannot be explained and results in reduced confidence in the body of evidence.
- Indirectness: Addresses generalizability and relevance to the objective of the assessment. As outlined in the Objective and consistent with the population specified in the PECO statement, this systematic review evaluated the extent and quality of the evidence linking fluoride exposure to neurodevelopmental and cognitive effects in humans without restriction as to age, sex, geographic location, or life stage at exposure or outcome assessment. Furthermore, the review did not exclude subjects exposed in occupational settings. All exposure levels and scenarios encountered in

human studies are considered direct (i.e., applicable, generalizable, and relevant to address the objective of the assessment); therefore, a downgrade for indirectness would not be applied to bodies of evidence from human studies.

- Imprecision: Addresses confidence associated with variability in quantitative measures such as effect sizes. Typically, 95% confidence intervals are used as the primary method to assess imprecision, but considerations can also be made on whether studies were adequately powered. Meta-analyses can also be used to determine whether the data are imprecise. When a meta-analysis is not appropriate or feasible, imprecision can be based on variability around the effect estimate. A downgrade would occur if the body of evidence was considered to be imprecise based on a meta-analysis, or if serious or very serious imprecision was consistently present in the body of evidence. A downgrade is especially likely if imprecision raised questions as to whether an overall effect was significant.
- Publication bias: Addresses evidence of biased publication practices. Downgrade if one strongly detects publication bias. Publication bias is difficult to detect but may be evident if major sections of the research community are not publishing (e.g., absence of industry, academic, or government studies) on a topic or if there are multiple instances wherein data from conference abstracts are never published in peer-reviewed journals. In addition, there are methods included in conducting a meta-analysis to detect whether there is potential for publication bias, including the use of fit-and-trim models, which help identify how publication bias may affect the results of the meta-analysis. Although a meta-analysis is not included in this systematic review, there are two published meta-analyses (Choi et al. 2012; Duan et al. 2018) in addition to the one associated with this systematic review (manuscript in progress) that can be used to address publication bias.

Factors to Consider for Potential Upgrading

- Large magnitude of effect: Factors to consider include the outcome being measured and the dose or exposure range assessed. The confidence can be upgraded if the body of evidence is suggestive of a large magnitude of effect. GRADE provides guidance on what can be considered a large magnitude of effect based on relative risk (i.e., suggests one upgrade in confidence if relative risk is greater than 2 and two upgrades in confidence if greater than 5). However, not all studies provide data as a risk estimate, and smaller changes, such as increases in blood pressure, may have greater impact on health at the population level. Consideration for an upgrade is not based on a single study, and what constitutes a large magnitude of effect will depend on the outcome and the potential public health impact.
- Dose response: Patterns of dose response are evaluated within and across studies. Confidence in the body of evidence can be increased when there is sufficient evidence of a dose-response pattern across multiple studies.
- Consistency: Does not apply in this evaluation. The consideration of a potential upgrade for consistency is primarily for non-human animal evidence in which it would be applied to address increased confidence based on an observation of consistent effects across multiple non-human animal species. For human evidence, this factor would generally not be applied. Human studies are instead evaluated for

issues of consistency that could result in downgrading confidence for unexplained inconsistency (see "Factors to Consider for Potential Downgrading" above).

- Consideration of residual confounding: Applies to observational studies and refers to consideration of unmeasured determinants that are likely to be distributed unevenly across groups. Residual confounding can push results in either direction, but confidence in the results is increased when the body of evidence is biased by factors that counter the observed effect and would cause an underestimation of the effect. Confounding that would cause an overestimation of the effect is considered under the risk-of-bias considerations for decreasing confidence.

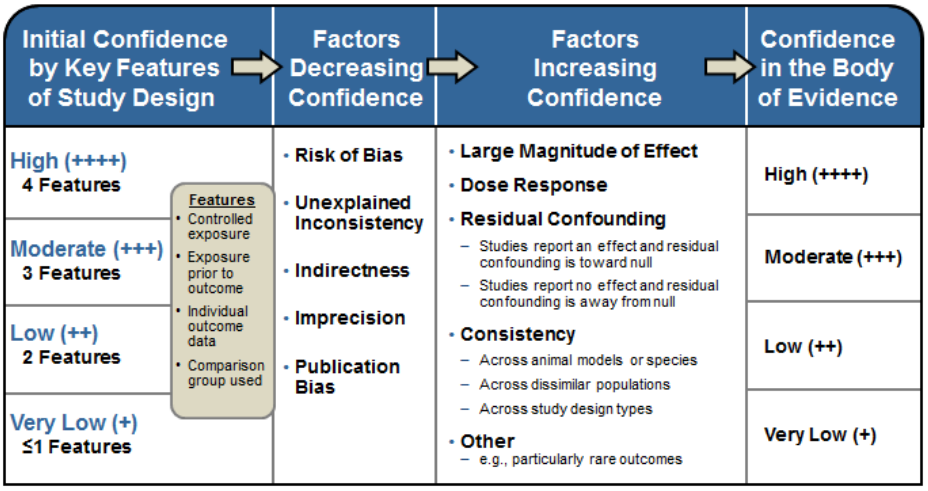


Figure 1. Assessing Confidence in the Body of Evidence

Confidence ratings were assessed by the evaluation team for accuracy and consistency, and discrepancies were resolved by consensus and consultation with technical advisors as needed. Confidence ratings for the primary outcomes are summarized in evidence profile tables for each outcome.

Results

Literature Search Results

The electronic database searches retrieved 25,450 unique references with 11 additional references[11] identified by technical advisors or obtained by manually searching the Fluoride Action Network website or reviewing reference lists of published reviews and other included studies. During title and abstract screening, 1,036 references were moved to full-text review and 24,425 were excluded (11,402 by manual screening for not satisfying the PECO criteria and 13,023 based on the SWIFT-Active Screener algorithm). Among the 1,036 references that underwent full-text review, 547 studies were considered PECO-relevant (see Appendix C for list of included studies). A few studies assessed data for more than one evidence stream (human, non-human mammal, and/or in vitro), and several studies assessed more than one type of outcome (e.g., primary and secondary outcomes). Included studies break down as follows:

- 167 human studies (84 primary only; 13 secondary only; 5 primary and secondary; 8 primary and thyroid; 2 secondary and thyroid; and 55 thyroid only);
- 339 non-human mammal studies (7 primary only; 186 secondary only; 67 primary and secondary; 6 primary, secondary, and thyroid; 4 secondary and thyroid; and 69 thyroid only); and,
- 60 in vitro/mechanistic studies (48 neurological and 12 thyroid).

Additional details on the screening results are provided in Appendix C. These screening results are outlined in a study selection diagram that reports numbers of studies excluded at each stage and documents the reason for exclusion at the full-text review stage (see Figure 2) [using reporting practices outlined in ~~Moher et al. (2009)~~ Page et al. (2021)].

Commented [A38]: This footnote was added to further describe why and how studies were identified from other sources in response to related comments from [redacted] Reviewers listed below; see DocG_Monograph and DocJ_Monograph, respectively, for detailed responses:

Reviewer comment (DocG_Monograph, page 9): Page 22, can you provide a brief explanation for why the 15 additional identified references were missed by your literature searches?

Reviewer comment (DocJ_Monograph, page 11): Identifying 15 references through other sources seems somewhat high. Was there a need to adjust the original search strategy to capture those references?

Commented [A39]: Change made in response to the [redacted] comment below; see DocB1_Monograph for detailed response:

Reviewer comment (DocB1_Monograph, page 5): The authors are using an old version of the PRISMA flow diagram - The 2020 PRISMA flow diagram can be found here: https://prisma-statement.org/prismastatement/flowdiagram.aspx

[11]These 11 studies (9 human and 2 animal studies) were not identified through the electronic database searches, as they were not indexed in any of the electronic databases searched. Note that the supplemental search of non-English-language databases was designed in part to identify non-English-language studies that are not indexed in traditional bibliographic databases such as PubMed. It was successful in this goal, as multiple studies that were initially only identified through "other sources" were subsequently captured in the supplemental Chinese database search, leaving only 11 as identified through other sources.

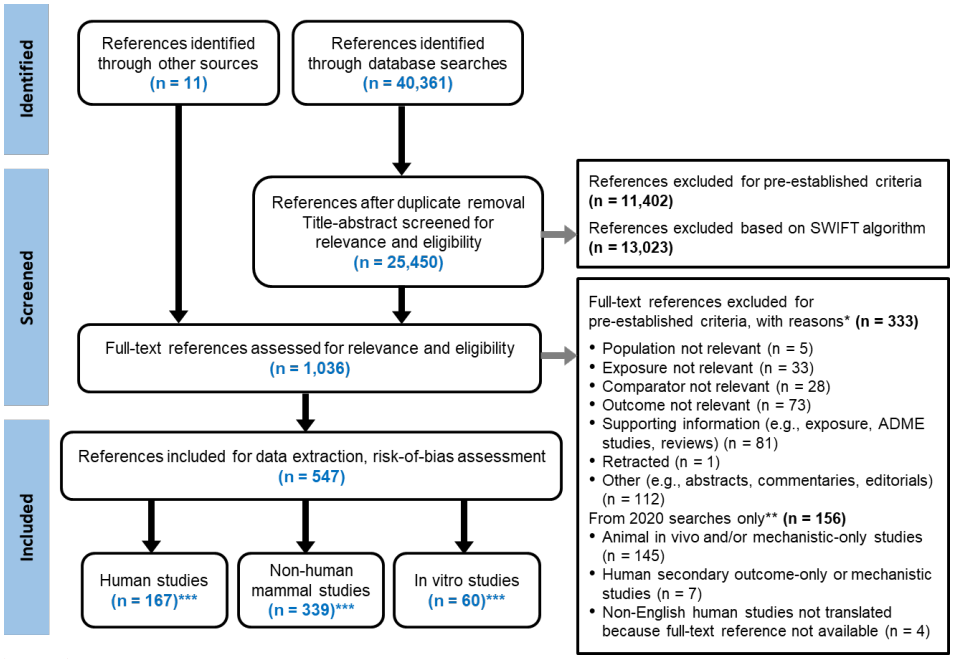


Figure 2. Study Selection Diagram[a]

[a]An interactive reference flow diagram is available here: https://hawcproject.org/summary/visual/assessment/405/Figure-2/.
*~~Includes s~~Studies from all literature searches conducted during the review excluded at the full-text level for pre-established criteria; see the Methods section for extraction and search update information. Studies may have been excluded for more than one reason; the first reason identified was recorded.
**~~Includes all studies from all~~ Studies excluded from the 2020 literature searches ~~not otherwise excluded~~ for reasons other than pre-established criteria; see the Methods section for extraction and search update information.
***Publications may contain more than one evidence stream, so the numbers will not total the 547 included studies.

Commented [A40]: Figure 2 has been revised for accuracy, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 10): The numbers of abstracts in Figure 2 do not align with the text.

Commented [A41]: Footnotes * and ** were revised to provide additional clarity on the 333 excluded studies. Although the reviewer was not correct in their suggestion it pointed out the need for clarification in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 11): The top line of the excluded box should state that the 333 were from the original (pre-2020) search.

Human Neurodevelopmental and Cognitive Data

The body of literature that evaluates the association between fluoride exposure and neurodevelopmental and cognitive effects in humans is relatively robust with a large number of studies (n = 100) that cover a wide array of endpoints (see Figure 3). Seventy-two human studies investigated IQ in children. Additional studies evaluated learning and memory (n = 9 studies) or other cognitive developmental effects (e.g., total neurobehavioral scores and total mental capacity index in children, cognitive impairment in adults; n = 15 studies).[12] For this review, the evidence in children and adults was evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood.

[12]Some studies are included in more than one endpoint category (e.g., IQ and other cognitive developmental effects); therefore, these counts are not mutually exclusive.

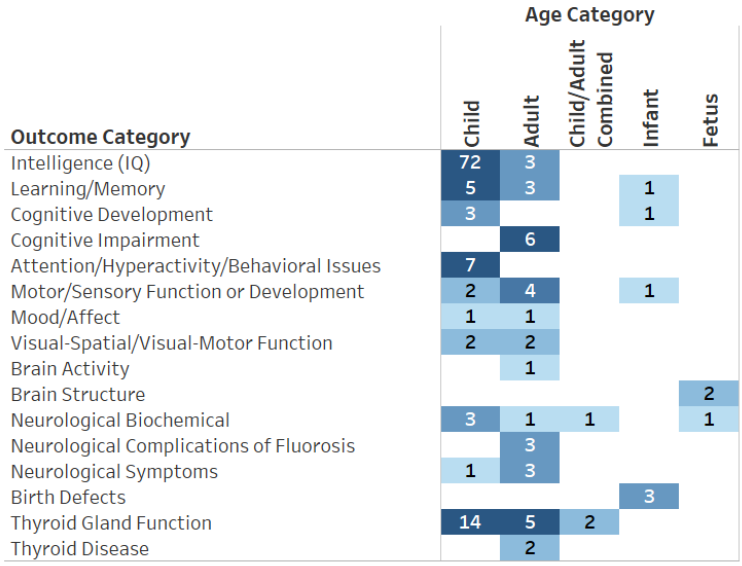


Figure 3. Number of Epidemiological Studies by Outcome and Age Categories[a]

[a]Interactive figure and additional study details are available at https://public.tableau.com/app/profile/ntp.visuals/viz/FluorideTableauDashboards/ReadMe.
Choi et al. (2015) used subtests of the omnibus IQ test reported by the authors as Wechsler Intelligence Scale for Children-Revised (WISC-IV) to evaluate visuospatial abilities (using block design) and executive function (using digit span). These endpoints are included in the intelligence (IQ) outcome category as they are subsets of the IQ tests.
Three additional publications based on subsamples (i.e., 50–60 children) of the larger Yu et al. (2018) cohort were identified (Zhao et al. 2019; Zhao et al. 2020; Zhou et al. 2019) and are not included in the counts of this figure.

Because the majority of studies evaluated intelligence, the following section focuses on IQ effects in children followed by separate discussions on other measures of cognitive function and neurobehavioral effects in children and cognitive effects in adults. Studies that evaluated mechanistic data in humans, including effects on the thyroid, are discussed in the Mechanistic Data in Humans section. Note that a few studies were identified on congenital neurological malformations and neurological complications of fluorosis; however, they are not considered further due to the limited number of studies and the heterogeneity of outcomes evaluated in those studies.

Commented [A42]: This sentence reflects revisions that a few studies on these other health outcomes were identified, which is more in line with other statements in the *Results* section, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 11): On page 24, [redacted] have trouble with the Results statement "Congenital neurological malformations and neurological complications of fluorosis are not considered further due to the limited number of studies and the heterogeneity of outcomes evaluated in these studies." This belongs in the Methods, complete with a full explanation for criteria used to or not to report/consider.

IQ in Children

Seventy-two epidemiological studies were identified that evaluated the association between fluoride exposure and children's IQ. Nineteen of the 72 IQ studies were determined to have low potential for bias (i.e., were of high quality). Looking across the literature, there has been a progression over the years in the quality of studies conducted to assess the association between fluoride exposure and IQ in children, with more recent studies including better study designs, larger sample sizes, and more sophisticated statistical analysis. Older studies often had limitations related to study design or methods, and most of the high risk-of-bias studies (i.e.,

studies of low quality) were published prior to the 2006 NRC evaluation of fluoride in drinking water. In contrast, 18 of the low risk-of-bias studies were published after the 2006 NRC evaluation of fluoride in drinking water, and over half of those were published between 2015 and 2020 (Figure 4).

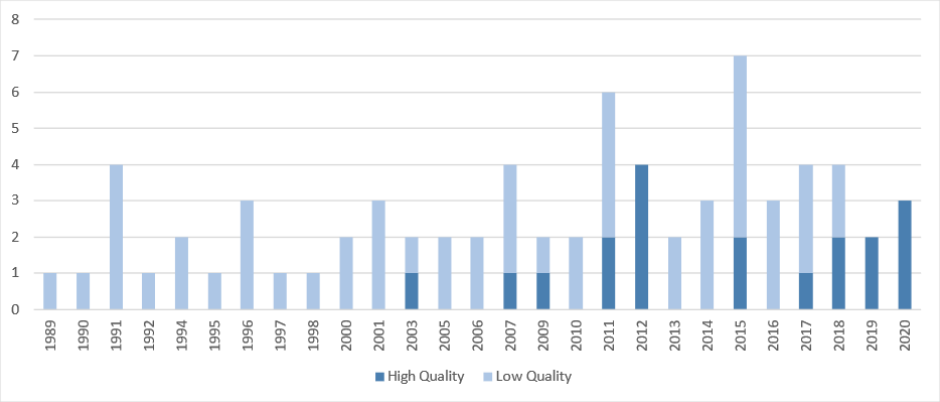


Figure 4. Number of High- and Low-quality Studies of Fluoride Exposure and IQ in Children by Year of Publication

Several characteristics of recent studies contribute to higher study quality in the overall body of literature on children's IQ and fluoride, including:

- Demonstration that exposure occurred prior to outcome assessment (an important factor when considering confidence in study results; see Figure 1) either by study design (e.g., for prospective cohort studies) or analysis (e.g., prevalence of dental fluorosis in children, limiting study populations to children who lived in the same area for long periods of time).
- Improved reporting of key study details that are necessary to evaluate study quality and allow for a more precise analysis of risk of bias.
- Increased consideration of key covariates (e.g., socioeconomic status) including potential co-exposures (e.g., arsenic or lead intake).
- Increased use of individual-level exposure measures (urine or water) as well as prenatal fluoride exposure to assess either individual-level fluoride exposure or—if still using group-level data—to confirm that regions being compared had differences in fluoride exposure.
- Utilization of more sophisticated sampling techniques for the study populations (e.g., stratified multistage random sampling).
- Application of more sophisticated regression approaches (e.g., piecewise linear regression models, multi-level regression with random effects, or generalized additive models for longitudinal measurements of fluoride).

- For studies using individual-level exposure measures, application of more sophisticated regression techniques to account for clustering at the cohort level by using cohort as a fixed or random effect and by accounting for numerous covariates that capture the cohort effect.

In addition, newer studies represent more diverse study populations across several countries (Figure 5), whereas all identified peer-reviewed studies that were published prior to 2006 took place in a single country (China). The majority of high-quality, low risk-of-bias studies exhibit these important study design and analysis characteristics, as discussed further in subsequent sections.

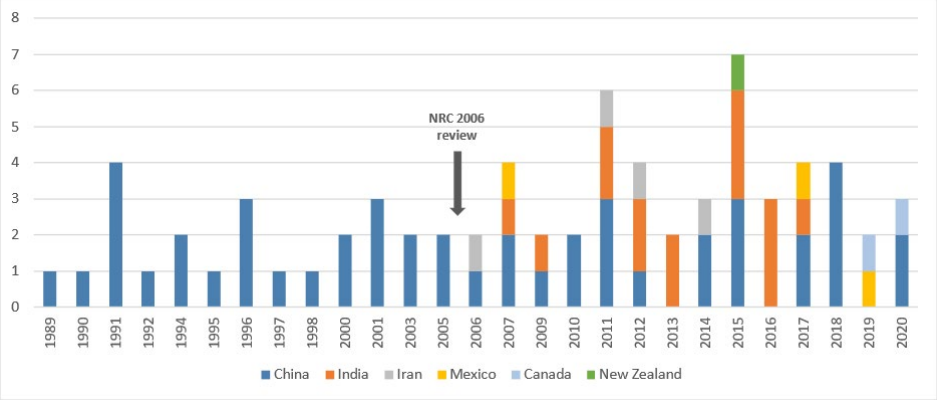


Figure 5. Number of Studies of Fluoride Exposure and IQ in Children by Country and Year of Publication

All available studies were considered in this evaluation; however, review of the body of evidence focused on the high-quality, low risk-of-bias studies for two main reasons. First, there are fewer limitations and greater confidence in the results of the high-quality studies. Second, there are a relatively large number of high-quality studies (n = 19), such that the body of evidence from these studies could be used to evaluate confidence in the association between fluoride exposure and changes in children's IQ. Therefore, the remainder of the discussion on IQ in children focuses on the 19 studies with low risk of bias. The high risk-of-bias studies are discussed briefly relative to their overall support of findings from the low risk-of-bias studies.

Low Risk-of-bias IQ Studies

Overview of Studies

Nineteen studies (3 ~~longitudinal~~ prospective cohort and 16 cross-sectional studies) with low potential for bias evaluated the association between fluoride exposure and IQ in children (see Quality Assessment of Individual Studies section for methods on determining which studies pose low risk of bias). These IQ studies were conducted in 15 study populations across 5 countries

Commented [A43]: This sentence reflects revisions to include information on counts of studies per study design, in response to the [redacted] Reviewer comment below; see DocJ_Monograph for detailed response:

Reviewer comment (DocJ_Monograph, page 12): In general, it would be helpful to the reader to describe the included study designs earlier in the respective results sections (i.e., along with the total #s of included articles, describing the # of cross-sectional, prospective cohort, etc is important).

Commented [A44]: The word "longitudinal" was previously added here and two other places in the monograph in response to the [redacted] Reviewer comment below; however, upon further consideration, we decided it would be more clear and consistent to add definitions of all observational study designs as footnotes to Table 4. Therefore, the new footnotes now clarify that cohort studies are observational studies in humans that examine a cohort prospectively or retrospectively over time, and although cohort studies may include longitudinal analyses, it is not a prerequisite of the cohort study design. See DocG_Monograph for detailed response.

Reviewer comment (DocG_Monograph, page 14): It is unclear whether you are using the term prospective (cohort study) to mean prospective (as opposed to retrospective, it's correct meaning) or longitudinal (as opposed to cross-sectional). Please use the correct term. In any case, the reader needs to know both whether studies were prospective or retrospective and whether studies were longitudinal or cross-sectional.

Commented [A45]: This sentence reflects revisions to refer the reader back to the *Methods* section that describes the risk-of-bias assessment for human studies, in response to the [redacted] Reviewer comment below; see DocK_Monograph for detailed response:

Reviewer comment (DocK_Monograph, page 4): It might be useful to have reminder, or reference back to the section in the text where the risk of bias information for human and animal studies is described in the methods (page 18), prior to presentation of the low risk of bias results for humans (page 28) and animals (page 67).

and included more than 7,000 children. Specifically, of the 19 low risk-of-bias studies of IQ in children:

- ten were conducted in four areas of China on seven study populations,[13]
- three were conducted in three areas of Mexico on three study populations,
- two were conducted in Canada using the same study population,
- three were conducted in three areas of India on three study populations, and
- one was conducted in Iran.

Most studies measured fluoride in drinking water (n = 15) and/or urine (child or maternal) (n = 15). Two studies measured fluoride in serum. The IQ studies used a variety of tests to measure IQ. Because IQ tests should be culturally relevant, the tests used often differed between studies, reflecting adjustments for the range in populations studied (e.g., western vs. Asian populations). In some cases, different IQ tests were used to study similar populations. Overall, these studies used IQ tests that were population- and age-appropriate.

Table 6 provides a summary of study characteristics and key IQ and fluoride findings for the 19 low risk-of-bias studies (organized by country and then by year). Several of these studies conducted multiple analyses and reported results on multiple endpoints. The purpose of the table is to summarize key findings (independent of whether an association is indicated) from each study and is not meant to be a comprehensive summary of all results from each study. For each study, results are summarized for each exposure measure assessed, but results from multiple analyses using the same exposure measure may not be presented for all studies unless multiple analyses yielded conflicting results. See Appendix E for additional information on each study in Table 6, including strengths and limitations, clarifications for why studies are considered to pose low risk of bias, and information regarding statistical analyses, important covariates, exposure assessment, and outcome assessment.

[13]In this document, "study population" refers to a defined population on which an original body of research was conducted. The published work drawn from that original body of research is often referred to as a "study." IQ studies that report on the same study populations are identified in Table 6.

Commented [A46]: This sentence was revised to clarify that Table 6 was organized by country and then by year in response to the [redacted] Reviewer comment below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 12): In general, studies designated as "low risk of bias" were interpreted correctly. I have a few suggestions as to how to clarify many of the points made.
While the results are generally consistent (table 6) it would be useful to present the results based on the exposure metric used. For example, studies using fluoride concentrations in "high" and "low" areas could be grouped together to illustrate the change in IQ points. Additionally, the actual IQ test used could also be used to group studies within exposure metric. There are clear differences in the scoring for the Raven and the WASI/WPPSI, for example and these are hard to tell from the presentation.

Commented [A47]: This sentence, and several sentences throughout the monograph, reflect revisions to replace 'effect' with 'association', in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 16): Also note that observational studies (with rare exceptions) do not provide evidence of an "effect", only of an association. Please use the term judiciously or not at all.

Commented [A48]: This sentence was added to direct readers to Appendix E for additional information such as statistical methods not provided in Table 6 in response to the [redacted] Reviewer comment below; see DocI_Monograph for detailed response:

Reviewer comment (DocI_Monograph, page 17): Table 6 could include the following: 1) statistical methods; 2) confounders, particularly exposure to other known neurotoxicants, and how they were measured; 3) might rename 'Assessment timing' to age of participants or just combine the information with the location/subject's column

Table 6. Studies on IQ in Children[a]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
China					
Xiang et al. (2003a)[d]	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children [512]	Drinking water Mean (SD): 0.36 (0.15) (control), 2.47 (0.79) (high fluoride) mg/L Children's urine Mean (SD): 1.11 (0.39) (control), 3.47 (1.95) (high fluoride) mg/L Village of residence (non-endemic vs. endemic fluorosis)	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significant dose-related association of fluoride on IQ score based on drinking water quintile levels with significantly lower IQ scores observed at water fluoride levels of 1.53 mg/L or higher; % of subjects with IQ <80 was significantly increased at water levels 2.46 mg/L or higher; significant inverse correlation between IQ and urinary fluoride (Pearson correlation coefficient of −0.164); mean IQ scores for children in non-endemic region (100.41 ± 13.21) significantly higher than endemic region (92.02 ± 13.00) No statistical adjustment for covariates
Ding et al. (2011)	Cross-sectional Inner Mongolia (Hulunbuir City)/elementary school children [331]	Children's urine Range: 0.1–3.55 mg/L Drinking water (reported but not used in analyses) Mean (SD): 1.31 (1.05) mg/L	Children (ages 7–14 years)	IQ: Combined Raven's Test for Rural China	Significant association between urinary fluoride and IQ score (each 1-mg/L increase was associated with a decrease in IQ score of 0.59 points; 95% CI: −1.09, −0.08) Adjusted for age
Xiang et al. (2011)[d]	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children [512]	Children's serum Mean (SD): 0.041 (0.009) (control), 0.081 (0.019) (high fluoride) mg/L	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significant linear trend across quartiles of serum fluoride and children's IQ score <80 (adjusted ORs for Q1 and Q2; Q1 and Q3; and Q1 and Q4, respectively: 1; 2.22 [95% CI: 1.42, 3.47]; and 2.48 [95% CI: 1.85, 3.32]); significant associations at ≥0.05 mg/L serum fluoride Adjusted for age and sex

Commented [A49]: This text and text throughout the monograph reflect revisions to change the term 'gender' to 'sex', in response to the [redacted] Reviewer comment below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 4): Use of the term "gender" to denote sex differences is not in line with current usage. Gender refers to the socially constructed variable, while sex refers to the biological variable. Please change.

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Wang et al. (2012)[d]	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children [526]	Children's total fluoride intake Mean (SD): 0.78 (0.13) (control), 3.05 (0.99) (high fluoride) mg/day Village of residence (non-endemic vs. endemic fluorosis) Drinking water (reported for villages but not used in analyses) Mean (SD): 0.36 (0.11) (control), 2.45 (0.80) (high fluoride) mg/L	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significantly lower mean IQ in the endemic versus non-endemic regions, as reported in Xiang et al. (2003a); when high-exposure group was broken into four exposure groups based on fluoride intake, a dose-dependent decrease in IQ and increase in % with low IQ observed; significant correlation between total fluoride intake and IQ ($r = -0.332$); for IQ <80, adjusted OR of total fluoride intake per 1-mg/(person/day) was 1.106 (95% CI: 1.052, 1.163) Adjusted for age and sex
Choi et al. (2015)	Cross-sectional Mianning County/1st grade children [51]	Drinking water GM: 2.20 mg/L Children's urine GM: 1.64 mg/L Severity of fluorosis (Dean Index)	Children (ages 6–8 years)	IQ: WISC-IV (block design and digit span)	Compared to normal/questionable fluorosis, presence of moderate/severe fluorosis significantly associated with lower total (adjusted $\beta = -4.28$; 95% CI: −8.22, −0.33) and backward (adjusted $\beta = -2.13$; 95% CI: −4.24, −0.02) digit span scores; linear associations between total digit span and log-transformed urinary fluoride (adjusted $\beta = -1.67$; 95% CI: −5.46, 2.12) and log-transformed drinking water fluoride (adjusted $\beta = -1.39$; 95% CI: −6.76, 3.98) observed but not significant; forward digit span had similar results as backward and total but was not statistically significant; block design (square root transformed) not significantly associated with any measure of fluoride exposure Adjusted for age and sex, parity, illness before 3 years old, household income last year, and caretaker's age and education

Commented [A50]: This text and text throughout the monograph reflect revisions to update units of change in effect estimate per change in fluoride exposure or add cutoffs for categorical outcomes, in response to the [redacted] Reviewer comment below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 4): The estimated regression coefficients from the studies need to be presented more clearly. For example, many times there is no reference, e.g. increase (or decrease) in score per 1 mg/L F in urine. Further, for the presentation of odds ratios, it is not clear what the dichotomous (or categorical) outcome variable is (e.g. IQ below 50). These suggestions are for clarity as well as for correctness.

Commented [A51]: This summary of the study results, and text throughout the monograph reflect revisions to change the term 'correlation' to 'association' when a regression coefficient was used, in response to the [redacted] Reviewer comment below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 4): In general the scientific information presented, including the data in tables and figures, is technically correct and clearly and objectively presented. Specific comments regarding the general evaluation of studies:
The use of the term "correlation" is confusing ([redacted] have marked this several times in the document and tables, but there are also other occurrences). Correlation is generally used to denote a correlation coefficient (either Pearson or Spearman); however, [redacted] believe it has been used to denote the estimated regression coefficients (more on this below). [redacted] would recommend changing the terminology for clarity.

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Zhang et al. (2015b)	Cross-sectional Tianjin City (Jinnan District)/school children [180]	Drinking water Mean: 0.63 (control), 1.40 (endemic fluorosis) mg/L (SD not reported) Children's urine Mean (SD): 1.1 (0.67) (control), 2.4 (1.01) (endemic fluorosis) mg/L Children's serum Mean (SD): 0.06 (0.03) (control), 0.18 (0.11) (endemic fluorosis) mg/L	Children (ages 10–12 years)	IQ: Combined Raven's Test for Rural China	Significant correlation between IQ score and children's serum fluoride ($r = -0.47$) and urinary fluoride ($r = -0.45$); significant difference in mean IQ score for high-fluoride area (defined as >1 mg/L in drinking water; 102.33 ± 13.46) compared with control area (109.42 ± 13.30); % of subjects with IQ <90 significantly increased in high-fluoride area (28.7%) vs. low-fluoride area (8.33%); not significantly correlated with water fluoride Adjusted for age and sex, if applicable
Cui et al. (2018)	Cross-sectional Tianjin City (districts Jinghai and Dagang)/school children [323]	Children's urine Median (Q1–Q3): 1.3 (0.9–1.7) mg/L (boys), 1.2 (0.9–1.6) mg/L (girls)	Children (ages 7–12 years)	IQ: Combined Raven's Test for Rural China	Significant association between IQ score and log-transformed urinary fluoride (adjusted $\beta = -2.47$; 95% CI: −4.93, −0.01) Adjusted for age, mother's education, family member smoking, stress, and anger

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Yu et al. (2018)[e,f]	Cross-sectional Tianjin City (7 towns)/children [2,886]	Drinking water Mean (SD): 0.50 (0.27) (normal), 2.00 (0.75) (high) mg/L Children's urine Mean (SD): 0.41 (0.49) (normal), 1.37 (1.08) (high) mg/L	Children (ages 7–13 years)	IQ: Combined Raven's Test for Rural China	Significant difference in mean IQ scores in high water fluoride areas (>1.0 mg/L; 106.4 ± 12.3 IQ) compared to the normal water fluoride areas (≤1.0 mg/L; 107.4 ± 13.0); distribution of the IQ scores also significantly different (p = 0.003); every 0.5-mg/L increase in water fluoride was associated with a decrease of 4.29 in IQ score (95% CI: −8.09, −0.48) when exposure was between 3.40 and 3.90 mg/L; no significant association between 0.2 and 3.40 mg/L; every 0.5-mg/L increase in urinary fluoride was associated with a decrease of 2.67 in IQ score (95% CI: −4.67, −0.68) between 1.60 and 2.50 mg/L but not at levels of 0.01–1.60 mg/L or 2.50–5.54 mg/L. Adjusted for age and sex, maternal education, paternal education, and low birth weight
Cui et al. (2020)	Cross-sectional Tianjin City (all districts)/school children (potentially some overlap with Cui et al. (2018)) [498]	Children's urine <1.6–≥2.5 mg/L	Children (ages 7–12 years)	IQ: Combined Raven's Test	Decreasing mean (± SD) IQ score with increasing urinary fluoride levels (statistical significance not reached based on a one-way ANOVA) <1.6 mg/L: 112.16 ± 11.50 1.6–2.5 mg/L: 112.05 ± 12.01 ≥2.5 mg/L: 110 ± 14.92 No statistical adjustment for covariates

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Wang et al. (2020b)[c]	Cross-sectional Tianjin City (villages not specified)/school children [571]	Drinking water Mean (SD): 1.39 (1.01) mg/L Children's urine Mean (SD): 1.28 (1.30) mg/L	Children (ages 7–13 years)	IQ: Combined Raven's Test for Rural China	Significant associations between IQ and water and urinary fluoride concentrations in boys and girls combined based on both quartiles and continuous measures (water: 1.587 decrease in IQ score per 1-mg/L increase; urine: 1.214 decrease in IQ score per 1-mg/L increase); no significant effect modification of sex Adjusted for age and sex, BMI, maternal education, paternal education, household income, and low birth weight
Mexico					
Rocha-Amador et al. (2007)	Cross-sectional Moctezuma and Salitral in San Luis Potosi State and 5 de Febrero of Durango State /elementary school children [132]	Drinking water Mean (SD): 0.8 (1.4), 5.3 (0.9), 9.4 (0.9) mg/L (3 rural areas) Children's urine Mean (SD): 1.8 (1.5), 6.0 (1.6), 5.5 (3.3) mg/L (3 rural areas)	Children (ages 6–10 years)	IQ: WISC-Revised Mexican Version	Significant associations between log-transformed fluoride and IQ scores (full-scale IQ adjusted βs of −10.2 [water] and −16.9 [urine]; CIs not reported); arsenic also present, but the association with arsenic was smaller (full-scale IQ adjusted βs of −6.15 [water] and −5.72 [urine]; CIs not reported) Adjusted for blood lead, mother's education, SES, height-for-age z-scores, and transferrin saturation

Commented [A52]: This summary of the study results reflects revisions to make clear that the association between arsenic exposure and children's IQ was smaller in magnitude than the association between fluoride exposure and children's IQ, in response to the [redacted] Reviewer comment below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 13): At times, associations are presented as different when other covariates are controlled. [redacted] presume that these assessments were made by inspection of the results in the studies, but should either be backed up with statistical testing or admitted that they were made by inspection. For example, in table 6 the study by Rocha-Amador, et al states that the estimated associations between fluoride and the full scale IQ (WISC) were smaller when arsenic was controlled, the estimated betas are given, but there is no indication whether the differences are statistically different.

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Bashash et al. (2017)	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [299] IQ analysis [211]	Maternal urine during pregnancy Mean (SD): 0.90 (0.35) mg/L Children's urine Mean (SD): 0.82 (0.38) mg/L	Children (ages 6–12 years)	IQ: WASI-Spanish Version	Significantly lower child IQ score per 0.5-mg/L increase in maternal urinary fluoride (adjusted β = −2.50; 95% CI: −4.12, −0.59); no significant association with children's urine Adjusted for sex, gestational age; weight at birth; parity (being the first child); age at outcome measurement; and maternal characteristics, including smoking history (ever smoked during the pregnancy vs. nonsmoker), marital status (married vs. not married), age at delivery, education, IQ, and cohort
Soto-Barreras et al. (2019)	Cross-sectional Chihuahua/school children [161]	Children's urine Range: 0.11–2.10 mg/L Drinking water Range: 0.05–2.93 mg/L Fluoride exposure dose (summary statistics not reported) Fluorosis index (summary statistics not reported)	Children (ages 9–10 years)	IQ: Raven's Colored Progressive Matrices	No significant difference in urinary fluoride, drinking water fluoride, fluoride exposure dose, or fluorosis index in subjects across different IQ grades No statistical adjustment for covariates

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Canada					
Green et al. (2019)[g]	Cohort (prospective) 10 cities/Maternal-Infant Research on Environmental Chemicals (MIREC) [512] Non-fluoridated [238] Fluoridated [162] Boys [248] Girls [264]	Maternal urine during pregnancy Mean (SD): 0.51 (0.36) mg/L (0.40 [0.27] mg/L in non-fluoridated areas and 0.69 [0.42] mg/L in fluoridated areas) Maternal fluoride intake during pregnancy Mean (SD): 0.54 (0.44) mg/day (0.30 [0.26] and 0.93 [0.43] mg/day, respectively) Drinking water Mean (SD): 0.31 (0.23) mg/L (0.13 [0.06] and 0.59 [0.08] mg/L, respectively)	Children (ages 3–4 years)	IQ: full-scale, performance, and verbal using Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	Significantly lower full-scale IQ (adjusted β = −4.49; 95% CI: −8.38, −0.60) and performance IQ (adjusted β = −4.63; 95% CI: −9.01, −0.25) per 1-mg/L increase in maternal urinary fluoride in boys but not girls (adjusted β = 2.40; 95% CI: −2.53, 7.33 and adjusted β = 4.51; 95% CI: −1.02, 10.05, respectively) or boys and girls combined (adjusted β = −1.95; 95% CI: −5.19, 1.28 and adjusted β = −1.24; 95% CI: −4.88, 2.40, respectively); significantly lower full-scale IQ (adjusted β = −3.66; 95% CI: −7.16, −0.15) per 1-mg increase in maternal fluoride intake (no sex interaction); significantly lower full-scale IQ (adjusted β = −5.29; 95% CI: −10.39, −0.19) per 1-mg/L increase in water fluoride concentration (no sex interaction); no significant associations observed between measures of fluoride and verbal IQ Adjusted for sex, city, HOME score, maternal education, race, and prenatal secondhand smoke exposure

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Till et al. (2020)[g]	Cohort (prospective) 10 cities/ MIREC [398] Non-fluoridated [247] Fluoridated [151] Breastfed as infants [200] Formula-fed as infants [198]	Drinking water Mean (SD) For breastfed infants: 0.13 (0.06) mg/L in non-fluoridated areas and 0.58 (0.08) mg/L in fluoridated areas For formula-fed infants: 0.13 (0.05) mg/day in non-fluoridated areas and 0.59 (0.07) mg/L in fluoridated areas Infant fluoride intake Mean (SD) For breastfed infants: 0.02 (0.02) mg/day in non-fluoridated areas and 0.12 (0.07) mg/day in fluoridated areas For formula-fed infants: 0.08 (0.04) mg/day in non-fluoridated areas and 0.34 (0.12) mg/day in fluoridated areas Maternal urine during pregnancy	Children (ages 3–4 years)	IQ: full-scale, performance, and verbal using Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	Drinking water Breastfed infants: Lower (not significant) full-scale IQ (adjusted β = −1.34, 95% CI: −5.04, 2.38) per 0.5-mg/L increase in water fluoride concentration; significantly lower performance IQ (adjusted β = −6.19, 95% CI: −10.45, −1.94) Formula-fed infants: Significantly lower full-scale IQ (adjusted β = −4.40, 95% CI: −8.34, −0.46) per 0.5-mg/L increase in water fluoride concentration; significantly lower performance IQ (adjusted β = −9.26, 95% CI: −13.77, −4.76) Infant fluoride intake Breastfed: No results reported Formula-fed: Lower (not significant) full-scale IQ (adjusted β = −2.69, 95% CI: −709, 3.21) per 0.5-mg/L increase in fluoride intake from formula; significantly lower performance IQ (adjusted β = −8.76, 95% CI: −14.18, −3.34) Maternal urine during pregnancy+

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
		Mean (SD) Breastfed: 0.42 (0.28) mg/L in non-fluoridated areas and 0.70 (0.39) mg/L in fluoridated areas Formula-fed: 0.38 (0.27) mg/L in non-fluoridated areas and 0.64 (0.37) mg/L in fluoridated areas			Lower (not significant) full-scale IQ (adjusted β = −1.08, 95% CI: −1.54, 0.47) per 0.5-mg/L increase in maternal urinary fluoride++; lower (not significant) performance IQ (adjusted β = −1.31, 95% CI: −3.63, 1.03)++ Lower (not significant) performance IQ (adjusted β = −1.50, 95% CI: −3.41, 0.43) per 0.5-mg/L increase in maternal urinary fluoride+++; significantly lower full-scale IQ (adjusted β = −2.38, 95% CI: −4.62, −0.27)+++ No association between verbal IQ scores and any measure of fluoride exposure +Maternal urinary fluoride analyzed as covariate in the drinking water and infant fluoride intake from formula models and not in an individual model ++After additional adjustment for drinking water and breastfeeding status +++After additional adjustment for infant fluoride intake from formula All models adjusted for maternal education, maternal race, age at IQ testing, sex, HOME total score, and secondhand smoke status in the child's home (separate analysis also adjusted for mother's urinary fluoride)

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
India					
Sudhir et al. (2009)	Cross-sectional Nalgonda District (Andhra Pradesh)/school children [1,000]	Drinking water Level 1: <0.7 mg/L Level 2: 0.7–1.2 mg/L Level 3: 1.3–4.0 mg/L Level 4: >4.0 mg/L	Children (ages 13–15 years)	IQ: Raven's Standard Progressive Matrices	Significant increase in mean and distributions of IQ grades (i.e., increase in proportion of children with intellectual impairment) with increasing drinking water fluoride levels No statistical adjustment for covariates
Saxena et al. (2012)	Cross-sectional Madhya Pradesh/school children [170]	Drinking water ≥1.5 mg/L (high fluoride group) Children's urine Range: 1.7–8.4 mg/L	Children (age 12 years)	IQ: Raven's Standard Progressive Matrices	Significant correlations between IQ grade and water (r = 0.534) and urinary (r = 0.542) fluoride levels; in adjusted analyses, significant increase in mean IQ grade (i.e., increase in proportion of children with intellectual impairment) with increasing urinary fluoride; no significant differences in the levels of urinary lead or arsenic in children with the different water fluoride exposure levels Covariates included in the analysis were not reported
Trivedi et al. (2012)	Cross-sectional Kachchh, Gujarat/school children (6th and 7th grades) [84]	Mean (SE) Low-fluoride villages: drinking water: 0.84 (0.38) mg/L Children's urine: 0.42 (0.23) mg/L High fluoride villages: drinking water: 2.3 (0.87) mg/L Children's urine: 2.69 (0.92) mg/L	Children (ages 12–13 years)	IQ: questionnaire prepared by Professor JH Shah (97% reliability rating)	Significantly lower mean IQ score in high fluoride villages (92.53 ± 3.13) compared to the low-fluoride villages (97.17 ± 2.54); differences significant for boys and girls combined, as well as separately No statistical adjustment for covariates

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Iran					
Seraj et al. (2012)	Cross-sectional Makoo/school children [293]	Drinking water Mean (SD): 0.8 (0.3) (normal), 3.1 (0.9) (medium), 5.2 (1.1) (high) mg/L	Children (ages 6–11 years)	IQ: Raven's Colored Progressive Matrices	Significant association between water fluoride and IQ score (adjusted β = −3.865 per 1-mg/L increase in water fluoride); CIs not reported); significantly higher mean IQ score in normal area (97.77 ± 18.91) compared with medium (89.03 ± 12.99) and high (88.58 ± 16.01) areas Adjusted for age, sex, child's education level, mother's education level, father's education level, and fluorosis intensity

ANOVA = analysis of variance; GM = geometric mean; HOME = Home Observation Measurement of the Environment; IQ = intelligence quotient; Q1, Q3 = first and third quartiles; SD = standard deviations; WASI = Wechsler Abbreviated Scale of Intelligence (Spanish version); WISC-IV = Wechsler Intelligence Scale for Children-Revised (as reported by Choi et al. 2015).

[a]Includes low risk-of-bias studies.

[b]Associations between IQ and fluoride levels were reported quantitatively, when possible. For studies with multiple analyses and results, the table summarizes key findings and is not a comprehensive summary of all findings. Results also indicate when a study reported no association between IQ and fluoride, provided as a qualitative statement of no association.

[c]See Figure A-1 through Figure A-8 for additional study results.

[d]Xiang et al. (2003a), Xiang et al. (2011), and Wang et al. (2012) are based on the same study population.

[e]Yu et al. (2018) and Wang et al. (2020b) are based on the same study population.

[f]Three additional publications based on a subsample (i.e., 50–60 children) of the larger Yu et al. (2018) cohort were identified (Zhao et al. 2019; Zhao et al. 2020; Zhou et al. 2019); however, these publications focused on mechanistic considerations and are not included in the study totals for IQ because the main study by Yu et al. (2018) is considered a better representation of the IQ results.

[g]Green et al. (2019) and Till et al. (2020) are based on the same study population.

Summary of Results

Overall Findings

The results from 18 of the 19 high-quality (low risk-of-bias) studies (3 ~~longitudinal~~ prospective cohort studies from 2 different study populations and 15 cross-sectional studies from 13 different study populations) that evaluated IQ in children provide consistent evidence that higher fluoride exposure is associated with lower IQ scores (see "Summary of IQ Results" in Table 6) (Bashash et al. 2017; Choi et al. 2015; Cui et al. 2018; Ding et al. 2011; Green et al. 2019; Rocha-Amador et al. 2007; Saxena et al. 2012; Seraj et al. 2012; Sudhir et al. 2009; Till et al. 2020; Trivedi et al. 2012; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Xiang et al. 2011; Yu et al. 2018; Zhang et al. 2015b). Only one study (Soto-Barreras et al. 2019) did not observe an association between fluoride exposure and IQ; however, results were not provided in a manner that allowed for a direct comparison with other low risk-of-bias studies (see Appendix E for details). A strength of the findings across 18 of 19 low risk-of-bias studies was the consistent association between higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximated or exceeded the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] and lower IQ scores among studies of varying study designs, exposure measures, and study populations. In studies that analyzed the sexes separately (n = 5 studies with 2 studies reporting on the same study population), consistent findings of lower IQ associated with fluoride exposure were generally reported for both sexes. There is some indication of differential susceptibility between sexes, but ultimately, due to too few high-quality studies that analyzed exposure and outcome by sex separately and a lack of consistent findings that one sex is more susceptible, it is unclear whether one sex is more susceptible to the effects of fluoride exposure than the other. The body of evidence from the 19 low risk-of-bias studies is described in further detail below. Prospective cohort studies are discussed first, as this study design can establish a temporal relationship between exposure and outcome, which would contribute to demonstrating causality and, therefore, providing the strongest evidence for an association between fluoride exposure during development and IQ in children.

Results by Study Design – Prospective Cohort Studies

As noted above, three ~~longitudinal~~ prospective cohort studies, conducted in Mexico and Canada, were identified and considered to reflect a low risk for bias. All three prospective cohort studies found an association between increasing maternal or child fluoride exposure and lower IQ in children (Bashash et al. 2017; Green et al. 2019; Till et al. 2020). Two of the studies (Green et al. 2019; Till et al. 2020) were based on the same Canadian study population, but one evaluated prenatal fluoride exposure and the other evaluated postnatal fluoride exposure. Green et al. (2019) included maternal urinary fluoride, maternal fluoride intake, and water fluoride concentrations, while Till et al. (2020) used fluoride intake from formula or water concentrations in formula-fed versus breastfed infants. Multiple analyses were conducted in each prospective study, and results by analysis for the three prospective studies are discussed below. In summary, although not every analysis found a statistically significant association, together the three studies provided consistent evidence that increasing maternal fluoride levels were associated with lower IQ scores in the children.

In the Early Life Exposures in Mexico to Environmental Toxicants cohort, Bashash et al. (2017) observed a statistically significant association (p-value = 0.01) between lower IQ scores in children and prenatal fluoride exposure measured by maternal urinary fluoride (measured during

Commented [A53]: The following two sentences reflect revisions to clarify that while Green *et al. (2019)* and Till *et al.* (2020) use the same study population, the exposure measures used are different between the two publications, thus warranting consideration as separate studies, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 13): A small point, but [redacted] think the description of 19 studies somewhat exaggerates the size of the body of evidence, since these studies were conducted in 15 study populations. For example, on page 36, it is unclear why the two articles by Green and Till should get double the weight (2 vs. 1 study) simply because the authors chose to publish 2 (vs. 1) articles.

all three trimesters and included if at least one measurement was available). An increase of 0.5 mg/L of maternal urinary fluoride was associated with a 2.5-point decrease in IQ score [95% CI: −4.12, −0.59] in boys and girls combined (see Figure A-8). This study also reported an inverse association between IQ level and children's urinary fluoride levels (single spot urine sample); however, this specific result did not achieve statistical significance (a 0.5-mg/L increase of child urinary fluoride was associated with a 0.89-point decrease in IQ score [95% CI: −2.63, 0.85]) (Bashash et al. 2017).

In the Maternal-Infant Research on Environmental Chemicals cohort, consisting of 10 cities in Canada, Green et al. (2019) also reported inverse associations between IQ scores in children and multiple measures of prenatal fluoride exposure, including maternal urinary fluoride, maternal fluoride intake, and water fluoride concentrations. Green et al. (2019) observed a statistically significantly lower IQ for boys associated with maternal urinary fluoride averaged across trimesters (4.49-point decrease in IQ score [95% CI: −8.38, −0.60; p-value = 0.02] per 1-mg/L increase in maternal urinary fluoride); however, results were not significant in boys and girls combined (1.95-point decrease in IQ [95% CI: −5.19, 1.28]) and were positive but not significant in girls (2.40-point increase in IQ [95% CI: −2.53, 7.33]). Other measures of prenatal exposure (maternal fluoride intake or water fluoride concentrations) were associated with lower IQ scores in boys and girls combined; the authors found no significant effect measure modification between child sex and fluoride exposure in these analyses so they did not report boys and girls separately (Green et al. 2019). Specifically, when evaluating the association between estimated maternal fluoride intake based on maternal water and beverage consumption during pregnancy and IQ in children, a 1-mg increase in daily maternal consumption of fluoride during pregnancy was associated with a significantly decrease in IQ score of 3.66 points in boys and girls combined (95% CI: −7.16, −0.15; p-value = 0.04). Similarly, based on drinking water ~~fluoride~~ concentrations ~~for pregnant women from fluoridated areas (mean water fluoride levels of 0.59 ± 0.08 mg/L) versus pregnant women from~~ and ~~non-fluoridated areas (mean water fluoride levels of 0.13 ± 0.06 mg/L),~~ a 1-mg/L increase of fluoride in drinking water was ~~were~~ associated with a significant 5.29-point decrease in IQ score ~~per 1-mg/L increase in fluoride~~ in both boys and girls combined (95% CI: −10.39, −0.19; p-value <0.05) (Green et al. 2019).

In a study of the same study population as Green et al. (2019) that used fluoride intake from formula or water concentrations in formula-fed versus breastfed infants, Till et al. (2020) observed significantly lower performance IQ scores with higher fluoride regardless of the comparison used (p-values ≤0.004). They did not observe any association with verbal IQ, and full-scale IQ was only significantly lower in formula-fed infants using water fluoride concentrations as the exposure measure (p-value = 0.03). Breastfed infants and fluoride intake from formula also showed inverse associations but were not significant.

Taken together, the three prospective cohort studies (based on two North American study populations) indicate consistency in results across different types of analysis and across two study populations that higher fluoride exposure during development is associated with lower IQ scores.

Results by Study Design – Cross-sectional Studies

As with the prospective cohort studies, the cross-sectional studies reported a consistent association between fluoride exposure and lower IQ scores in children. Fifteen of the 16 low risk-of-bias cross-sectional studies [i.e., all with the exception of Soto-Barreras et al. (2019)]

Commented [A54]: This sentence reflects revisions to clarify if results were statistically significant, in response to the [redacted] Reviewer comment below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, Page 13): Please note when the result is not statistically significant and likely due to small sample sizes (e.g. discussion of the Green et al paper on page 37). Also for that paper, the results seem to be different by biological sex, an example of effect modification that would be expected for a neurotoxin.

Commented [A55]: Change made in response to the [redacted] Reviewer comment below; see DocB2_Monograph for detailed response:

Reviewer comment (DocB2_Monograph, Page 5): Is this actually observed (a 1 mg/L difference in fluoride concentrations leading to a 5.29 point decrease in IQ), or is this a predicted hypothetical effect from a model? If this is a modeled result rather than an observed result, should this be stated more clearly? Same comment may apply broadly.

consistently demonstrate that exposure to fluoride is associated with lower IQ scores. Fourteen of these 15 studies [with the exception of Cui et al. (2020)] reported significant associations.

Cross-sectional studies can have limitations, as the study design often cannot ensure that exposure preceded outcome. This uncertainty reduces confidence in study findings compared with prospective cohort studies—which, by design, establish that exposure occurred prior to outcome—and is captured in the outcome assessment. In some cases, cross-sectional studies do provide indicators of prior exposure (e.g., prevalence of dental fluorosis, limiting study populations to subjects who lived in the same area for long periods of time). Evidence that exposure occurred prior to the outcome of interest increases the confidence in results and any potential association reported in these studies. Of the 16 low risk-of-bias cross-sectional studies, 12 established that exposure preceded the outcome assessment (Choi et al. 2015; Ding et al. 2011; Rocha-Amador et al. 2007; Saxena et al. 2012; Seraj et al. 2012; Soto-Barreras et al. 2019; Sudhir et al. 2009; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Xiang et al. 2011; Yu et al. 2018). Five studies from different study populations indicated that a large portion of the exposed children had dental fluorosis (ranging from 43% to 100%) at the time of assessment (Choi et al. 2015; Ding et al. 2011; Seraj et al. 2012; Sudhir et al. 2009; Yu et al. 2018). Because dental fluorosis occurs when fluoride is consumed during enamel formation (usually during the first 6–8 years of life), the presence of dental fluorosis suggests that exposures to fluoride occurred prior to the outcome assessment. Nine studies from six study populations (including Yu et al. (2018) and Sudhir et al. (2009) listed above) excluded subjects who had not lived in the study area for a specified period of time, sometimes since birth (Rocha-Amador et al. 2007; Saxena et al. 2012; Soto-Barreras et al. 2019; Sudhir et al. 2009; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Xiang et al. 2011; Yu et al. 2018). Because these areas were generally known to be fluoride-endemic for long periods of time, it can generally be assumed that in these nine studies, exposure occurred prior to the outcome. Taken together, 12 cross-sectional studies from 9 study populations provide indicators of prior exposure.

Results by Study Design – Cross-sectional Study Variations

Overall, the cross-sectional studies consistently provide evidence that higher fluoride exposure is associated with lower IQ scores in children. Several cross-sectional studies conducted multiple analyses (e.g., reported results for multiple exposure metrics, endpoints, subpopulations). Although some of these variations are heterogeneous and are not comparable across studies, the consistency of the results across multiple metrics contributes to the confidence in the data. Table 6 summarizes key results for each of the low risk-of-bias cross-sectional studies, and a few examples of the within-study variations in results are provided below.

Nine cross-sectional studies (from six study populations) assessed the association between IQ and multiple exposure measures (Choi et al. 2015; Rocha-Amador et al. 2007; Saxena et al. 2012; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Xiang et al. 2011; Yu et al. 2018; Zhang et al. 2015b). Lower IQ was consistently observed across exposure measures in these studies; however, Choi et al. (2015), a small pilot study (n = 51), did not achieve statistical significance in all results by exposure measure. Specifically, the authors reported a consistent association between all fluoride exposure measures assessed (drinking water, children's urine, and severity of fluorosis) and digit span measures (subtest of the WISC-IV omnibus IQ test); however, results were only statistically significant when fluoride exposure was based on moderate or severe dental fluorosis in children (see Figure A-7). Choi et al. (2015) also observed

Commented [A56]: This sentence, and several sentences throughout the monograph, were revised to further distinguish between the comparison group and the group(s) exposed to higher levels of fluoride, in response to the [redacted] Reviewer comment below; see DocI_Monograph for detailed response:

Reviewer comment (DocI_Monograph, page 12): One key feature for confidence rating is 'comparison group used.' This needs to be discussed further since fluoride exposure may be pervasive in water supplies. If so, in studies including a comparison group, include the comparison and how it was determined. Cross-sectional studies using biomarkers as continuous variables can be very strong.

Commented [A57]: Change made in response to the [redacted] comment below; see DocF_Monograph for detailed response:

Reviewer comment (DocF_Monograph, pages 5 - 11) : High?

Commented [A58]: This sentence reflects revisions to clarify that the consistency of the results across multiple metrics contributes to the confidence in the body of evidence, rather than "increase", in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 14): The Results about the cross-sectional studies (page 38) state that "the consistent results across multiple metrics increase our confidence in the data." Based on the appropriate description in the Methods on Page 21, upgrading based on Consistency "does not apply in this evaluation".

some variation in results by outcome assessed (i.e., square root transformed block design and digit span [forward, backward, and total]). It was the only cross-sectional study that did not provide a full IQ score but instead provided results by specific subtests. The study authors consistently observed an inverse association between fluoride exposure and results from the digit span subtest (which specifically assesses executive function); however, results from the block design (square root transformed), a subtest of the WISC-IV omnibus IQ test that specifically assesses visuospatial function, was not associated with fluoride exposure. Note that Rocha-Amador et al. (2009) also assessed visuospatial function, and the authors reported a significant association (p-value <0.001) between fluoride exposure and decreased visuospatial constructional ability using the Rey-Osterrieth Complex Figure (ROCF) Test. Ultimately, too few studies were identified that reported results by subtest of omnibus IQ tests or assessed domains other than IQ (e.g., visuospatial function) to examine or explain the variation by outcome observed in Choi et al. (2015). The only other studies that provided a breakdown of the full IQ score were the prospective cohort studies by Green et al. (2019) and Till et al. (2020), which provided results for full-scale IQ as well as results for performance and verbal IQ. In both of these studies, lower verbal IQ was not associated with fluoride exposure, but lower performance and full-scale IQ were associated with fluoride exposure. There are too few studies to evaluate whether there is a specific aspect of IQ testing that is affected by exposure to fluoride, but the studies nonetheless consistently provide evidence that fluoride exposure is associated with lower IQ.

Yu et al. (2018) reported an overall association between lower IQ and higher fluoride exposure across multiple analyses but observed some variation in IQ results by urinary exposure level. The authors reported inverse associations between IQ and children's medium- and high-range urinary fluoride levels (1.60–2.50 mg/L and 2.50–5.54 mg/L, respectively), although change in IQ score was greater in the medium-range group (2.67 points decrease [95% CI: −4.67, −0.68]) for every 0.5-mg/L increase of urinary fluoride than in the high-range group (0.84 points decrease [95% CI: −2.18, 0.50]) (see Figure A-7). No association was reported at low-range urinary fluoride levels (0.01–1.60 mg/L). Note that Yu et al. (2018) also reported an inverse association between IQ and drinking water fluoride levels at 3.40–3.90 mg/L (4.29-point decrease in IQ score [95% CI: −8.09, −0.48]) for every 0.5-mg/L increase in water fluoride; a 0.04-point decrease in IQ score [95% CI: −0.33, 0.24] was observed for 0.5-mg/L increase in water fluoride at levels of 0.20–3.40 mg/L). The variation by exposure level in urine could not be verified in the analysis of drinking water exposures because there were only two water exposure groups (low and high). In a second study (Wang et al. 2020b), authors conducted a categorical analysis using urinary fluoride quartiles with reported betas per quartile. As observed in Yu et al. (2018), there were decreasing trends in IQ within each quartile; however, unlike Yu et al. (2018), Wang et al. (2020b) observed a larger decrease in IQ with each increasing urinary quartile and observed similar results using water fluoride quartiles (Wang et al. 2020b). Note that Wang et al. (2020b) cannot be compared directly to Yu et al. (2018) for evaluation at the higher exposure levels because the two studies do not use the same categorical exposure ranges. Although additional studies may have looked at different exposure levels, none of these studies provided results in the same manner as Yu et al. (2018) and Wang et al. (2020b) (i.e., betas by exposure category). Instead, these other studies provided an overall beta or mean IQ scores by exposure level. Despite the noted variations among these studies, the overall results still consistently support an association between fluoride exposure and lower IQ.

Two studies (Cui et al. 2018; Zhang et al. 2015b) observed associations between lower IQ in children and exposure to fluoride, with variations in results in subpopulations of children with different polymorphisms (see Figure A-7). These were the only two studies that considered polymorphism as a sub-analysis. Cui et al. (2018) observed a significant association between log-transformed children's single spot urinary fluoride and lower IQ scores (2.47-point decrease in IQ scores [95% CI: −4.93, −0.01; p-value = 0.049] per ln-mg/L increase in urinary fluoride), and the association was strongest in subjects with a TT polymorphism (compared with children with a CC or CT polymorphism) in the dopamine receptor D2 (DRD2) gene (12.31-point decrease in IQ score [95% CI: −18.69, −5.94; p-value <0.001] per ln-mg/L increase in urinary fluoride), which, according to the authors, probably resulted in a reduced D2 receptor density (Cui et al. 2018). Similarly, Zhang et al. (2015b) observed a significant association between lower IQ scores and children's single spot urinary fluoride (2.42-point decrease in IQ scores [95% CI: −4.59, −0.24; p-value = 0.030] per 1-mg/L increase in urinary fluoride), and the association was strongest in subjects with a val/val polymorphism (compared with children who carried the heterozygous or homozygous variant genotypes [met/val or met/met]) in the catechol-O-methyltransferase (COMT) gene (9.67-point decrease in IQ score [95% CI: −16.80, −2.55; p-value = 0.003] per 1-mg/L increase in urinary fluoride).

Overall, the cross-sectional studies consistently support a pattern of findings that higher fluoride exposure is associated with lower IQ scores in children. Slight within-study variations occur that may be associated with study variables such as IQ domains or subsets of IQ tests in a few studies that conducted multiple analyses, but these variations are heterogenous and cannot be further explored with the available studies. Despite these few variations, the overall evidence of an association with lower IQ is apparent.

Commented [A59]: This sentence, and several sentences throughout the monograph, reflect revisions to use the terms 'effect,' 'association,' and 'correlation' most appropriately, in response to the [redacted] Reviewer comment below; see DocJ_Monograph for detailed response:

Reviewer comment (DocJ_Monograph, page 17): Page 39, last full paragraph includes the sentence, "Despite these few variations, the overall evidence of an effect on IQ is apparent." This reviewer suggests editing the word "effect" to "association" or "correlation," given that the included studies are all observational.

Exposure Measure and Study Population Factors

Low risk-of-bias studies provide consistent evidence that higher fluoride exposure is associated with lower IQ scores across studies using different exposure measures. In addition to water fluoride levels, studies measured fluoride exposure using single serum samples in children (Xiang et al. 2011; Zhang et al. 2015b), single spot urine samples in children (Cui et al. 2018; Ding et al. 2011; Rocha-Amador et al. 2007; Saxena et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Yu et al. 2018; Zhang et al. 2015b), and prenatal maternal urinary measures (Bashash et al. 2017; Green et al. 2019), all of which were demonstrated to be consistently associated with lower IQ scores (see Figure A-6, Figure A-7, and Figure A-8). Urine levels encompass all sources of fluoride exposure and provide a better measure of the totality of exposure. As noted previously, even though some studies measured single spot samples, which may not be representative of peak exposure, these studies generally provided evidence that fluoride exposure had been occurring for some time. The consistency in the results across studies that used different measures of fluoride exposure and different life stages at which fluoride was measured strengthens the body of evidence.

The low risk-of-bias studies consistently provide evidence that higher fluoride exposure is associated with lower IQ scores across studies of different study populations. These 19 high-quality studies represent diverse populations (n = 15 study populations) across 5 countries. Eighteen of the 19 studies conducted in Canada (n = 2), China (n = 10), India (n = 3), Iran (n = 1), and Mexico (n = 2) provide evidence that exposure to fluoride is associated with lower IQ scores; 1 study conducted in Mexico did not observe an association but reported results in a

manner that did not allow for a direct comparison with the other studies (see Appendix E for details). The overall consistency in the study results across study populations adds strength to the body of evidence.

Exposure Levels

As described in this section, the body of evidence for studies assessing the association between fluoride exposure and IQ in children consistently provides evidence of an association between higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] and lower IQ in children; however, there is less certainty in the evidence of an association in populations with lower fluoride exposures. In the September 6, 2019, draft of this monograph, NTP conducted a qualitative analysis of children's IQ studies that 1) evaluated lower fluoride exposures (<1.5 mg/L) in drinking water and/or urine and 2) provided information to evaluate dose response (i.e., provided three or more fluoride exposure groups or a dose-response curve in their publication) in the lower fluoride exposure range. Nine low risk-of-bias studies met these criteria, which includes the three prospective cohort studies discussed in this section. Based on the qualitative review of these studies, the evidence of an association between fluoride exposure below 1.5 mg/L and lower IQ in children appeared less consistent than results of studies at higher exposure levels.

A draft quantitative dose-response meta-analysis was prepared and included in the September 16, 2020, draft monograph (NTP 2020). This meta-analysis is undergoing further refinement in preparation for separate publication and may further inform a discussion on the association between fluoride exposure levels and IQ in children.

Sex Considerations

Recent literature suggests that adverse neurodevelopmental effects of early-life exposure to fluoride may differ depending on timing of exposure and sex of the exposed subject. In a review of the human and animal literature, Green et al. (2020) concluded that, compared with females, male offspring appear to be more sensitive to prenatal but not postnatal exposure to fluoride, with several potential sex-specific mechanisms.

Sex differences were examined in five of the low risk-of-bias studies (in four study populations) (Green et al. 2019; Trivedi et al. 2012; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a). In general, sex differences were difficult to assess for trends within different study populations because few studies in the body of evidence analyzed exposure and stratified results by sex. Although these five studies reported IQ scores separately for boys and girls, only two of these studies analyzed fluoride exposure for boys and girls separately (Green et al. 2019; Wang et al. 2020b), which is essential for evaluating whether a differential change in IQ by sex may be related to higher susceptibility in one sex or higher exposure in that sex. The remaining three studies stratified results by sex (Trivedi et al. 2012; Wang et al. 2012; Xiang et al. 2003a), but the analyses were based on area-level exposure data (e.g., low-fluoride village compared with high fluoride village) and not drinking water or urinary fluoride concentrations. In the five studies that reported results by sex separately, consistent findings of lower IQ associated with higher fluoride exposure were generally reported for both sexes. There was some variation in the results between sexes across study populations and exposure measures, but there is insufficient

Commented [A60]: This paragraph was added to clarify that recent publications had reported sex differences and describe potential sex-specific mechanisms in response to the [redacted] Reviewer comment below; see DocJ_Monograph for detailed response:

Reviewer comment (DocJ_Monograph, page 18): Page 40, "Gender considerations": Is there some biological plausibility that there would be sex differences in the relationship between fluoride exposure and neurocognitive outcomes. The term "susceptibility" is used several times, but it is unclear what that means. It seems to imply a biological reason, but it is unclear whether mechanistic evidence is supportive of that (or if gender differences actually represent some sort of residual confounding).

Commented [A61]: Change made in response to the [redacted] Reviewer comment below; see DocF_Monograph for detailed response:

Reviewer comment (DocF_Monograph, pages 5 - 11) : High?

evidence to determine whether one sex is more susceptible to the effects of fluoride exposure than the other.

Green et al. (2019) observed a significant inverse association between maternal urinary fluoride levels and IQ scores in boys (p-values ≤0.04) but not girls in a Canadian population. Green et al. (2019) did not find any sex differences in the association between IQ and water fluoride concentrations. Wang et al. (2020b) evaluated Chinese boys and girls separately and combined and observed statistically significant decreasing trends in IQ in all groups by urinary fluoride quartiles (p-values for trend ≤0.035) (see Figure A-7). Similarly, when evaluated as a continuous variable, spot urinary fluoride levels (per 1-mg/L increase) were significantly associated with lower IQ scores in girls (−1.379 [95% CI: −2.628, −0.129; p-value = 0.031]), boys (−1.037 [95% CI: −2.040, −0.035; p-value = 0.043]), and in the sexes combined (−1.214 [95% CI: −1.987, −0.442; p-value = 0.002]). According to water fluoride quartiles, Wang et al. (2020b) found that there was a significant trend in the sexes combined, although the decreasing trend in boys and girls separately did not achieve statistical significance (p-values = 0.077 and 0.055, respectively). When water fluoride levels were evaluated as a continuous variable (per 1-mg/L increase), there were significant associations with lower IQ scores in girls (−1.649 [95% CI: −3.201, −0.097]; p-value = 0.037), boys (−1.422 [95% CI: −2.792, −0.053; p-value = 0.042]), and the sexes combined (−1.587 [95% CI: −2.607, −0.568]; p-value = 0.002).

The remaining three studies that reported results by sex-based comparisons of areas of high and low urinary or water fluoride did not report exposure levels separately for boys and girls, which decreases the utility of the data to evaluate differential susceptibility by sex. Trivedi et al. (2012) observed significantly lower IQ in children in high fluoride Indian villages compared with low-fluoride villages with decreases observed in boys and girls separately or combined (p-values ≤0.05) (see Figure A-2). Xiang et al. (2003a) and Wang et al. (2012) provide data on the same study population in China. There was a significantly lower IQ in the high fluoride area compared with the low-fluoride area in boys and girls separately and in the sexes combined (p-values <0.01), although the difference was greater in girls. Because fluoride exposure was not analyzed for boys and girls separately, it is unclear whether the greater change in IQ scores in girls could be attributed to higher susceptibility to fluoride exposure or differences in fluoride exposure by sex.

In summary, it is unclear whether one sex is more susceptible to the effects of fluoride exposure than the other due to the limited number of studies that analyzed exposure and outcome by sex and the lack of a consistent pattern of findings that one sex is more susceptible. Green et al. (2019) did not observe an association between maternal urinary fluoride levels and IQ scores in girls but did observe a significant association in boys. Although this is an indication of higher sensitivity in boys in this analysis, the authors did not detect this sex difference using other measures of prenatal exposure (maternal fluoride intake or water fluoride concentrations). Wang et al. (2020b) and Trivedi et al. (2012) reported statistically significant associations in both boys and girls without indication that one sex may be more susceptible. Although Xiang et al. (2003a) and Wang et al. (2012) reported a greater change in IQ in girls than boys, the studies used area-level exposure data, and the authors did not determine whether fluoride exposure differed in boys versus girls. Therefore, it is unclear whether this differential result by sex is an indication of higher susceptibility in girls or whether it could be explained by a difference in exposure by sex. Overall, there are too few studies that analyzed exposure and outcome by sex separately to properly evaluate whether there is differential susceptibility to fluoride exposure by sex, and

results from the five low risk-of-bias studies that do evaluate sex differences indicate that there is no consistent difference by sex across the different study populations.

Summary of Key Findings for Low Risk-of-bias Children's IQ Studies

In summary, the high-quality studies (i.e., studies with low potential for bias) consistently demonstrate lower IQ scores with higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)]. The consistency in association is observed among studies of varying study designs, exposure measures, and study populations. Although some studies that conducted multiple analyses observed within-study variations in results (e.g., differences between subsets of IQ tests), these variations were unique to individual studies and did not detract from the overall consistency in the findings that higher fluoride is associated with lower IQ scores.

High Risk-of-bias IQ Studies

The results from 53 studies with high potential for bias that evaluated IQ in children also consistently provide supporting evidence of decrements in IQ associated with exposures to fluoride. Forty-six of the 53 studies reported an association between high fluoride exposure and lower IQ scores in children.

Risk of Bias for IQ Studies in Children

The confidence in the human body of evidence was based on studies with the lowest potential for bias. A total of 19 studies on IQ in children had little or no risk-of-bias concerns, representing a relatively large body of evidence for low risk-of-bias studies (i.e., 15 study populations across 5 countries evaluating more than 7,000 children). These 19 studies are considered low risk of bias because they were rated probably low or definitely low risk of bias for at least two of the three key risk-of-bias questions and did not have any other risk-of-bias concerns that would indicate serious issues with the studies. Thirteen of the 19 studies were rated definitely low or probably low risk of bias for all risk-of-bias questions, and the remaining 6 studies were rated probably high risk of bias for a single question that was judged to have minimal impact on overall potential for bias. None of the 19 studies had a rating of definitely high risk of bias for any question. Risk-of-bias ratings for individual studies for all questions are available in Figure D-1 through Figure D-4, with risk-of-bias ratings for IQ studies in children available in Figure D-5 through Figure D-8 and Appendix E. Although the low risk-of-bias studies had minimal or no concerns, the studies with high overall potential for bias had a number of risk-of-bias concerns, including potential confounding, poor exposure characterization, poor outcome assessment, and, in many cases, potential concern with participant selection. The key risk-of-bias questions are discussed below.

Confounding for IQ Studies in Children

Low Risk-of-bias Studies

As discussed above, there are 19 studies considered to have low risk of bias when assessed across all risk-of-bias domains. Sixteen of the 19 low risk-of-bias studies [i.e., all with the exception of Cui et al. (2020), Ding et al. (2011), and Soto-Barreras et al. (2019)] were considered to have low potential for bias due to confounding because the authors addressed the three key covariates for all studies (i.e., age, sex, and socioeconomic status) through study design

Commented [A62]: This sentence reflects revisions to additional context for the use of the WHO Guidelines for Drinking -Water Quality value of 1.5 mg F/L, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 15): Missing is a clear description or analysis across studies of what constitutes higher exposure levels that are associated with lower IQ. Page 40 starts a description of Exposure Levels, but lacks any quantitative description of high (or low) exposure. While [redacted] understand that a better analysis may arise from the future meta-analysis, there should be enough data in Table 6 to allow a more coherent summary of exposure level thresholds analyzed.
The *Summary of Key Findings for Low Risk-of-bias Children's IQ Studies* on Page 42 (and again on page 48) describes higher exposure as ⩾1.5 mg/L, but other than a mention of the 2016 report, this threshold is not described or presented in the Results. To the reader, this threshold is unsupported by the included studies.

or analysis. Other important covariates, including health factors, smoking, and parental characteristics, were also addressed in many of the low risk-of-bias studies (see Figure 6).

Co-exposures to arsenic and lead were not considered a concern in 18 of 19 low risk-of-bias studies [i.e., all except for Soto-Barreras et al. (2019)] because the studies addressed the potential co-exposures, the co-exposures were not considered an issue in the study population, or the impact of the potential bias on the results was not a concern. Fifteen of 19 low risk-of-bias studies either addressed potential bias related to co-exposure to arsenic through study design or analysis or co-exposure to arsenic was unlikely in the study area. All 15 studies observed an association between lower IQ and higher fluoride exposure. Co-exposure to arsenic was not accounted for in the remaining four low risk-of-bias studies and was the main potential concern in these studies; however, three of these studies (Wang et al. 2012; Xiang et al. 2003a; Xiang et al. 2011) were still considered low risk of bias for confounding because although arsenic was observed in the water in the low-fluoride (and not the high-fluoride) comparison areas, which would bias the association toward the null, an association was still observed. In this case, the lack of adjustment for arsenic strengthens the evidence for an association and does not represent a potential concern. The other study did not address arsenic co-exposure and, as noted above, was conducted in an area that had potential for arsenic exposure to occur (Soto-Barreras et al. 2019); it is also the only low risk-of-bias study that did not observe an association between lower IQ and higher fluoride exposure (see Appendix E for further discussion of the risk-of-bias concern regarding arsenic for this study). Although Soto-Barreras et al. (2019) did not discuss arsenic, there is no direct evidence that arsenic was present in the study area. Fourteen studies accounted for co-exposure to lead through study design or analysis, and all observed an association between lower IQ and fluoride exposure. Five studies did not consider co-exposure to lead; however, for all of these studies, co-exposure to lead was considered unlikely to have an impact in these study populations as there was no evidence that lead was prevalent or occurring in relation to fluoride (Cui et al. 2018; Cui et al. 2020; Soto-Barreras et al. 2019; Till et al. 2020; Trivedi et al. 2012).

There is considerable variation in the specific covariates considered across the 19 low risk-of-bias studies. The consistency of results across these studies suggests that confounding is not a concern in this body of evidence. Each of the 18 low risk-of-bias studies that observed an association between fluoride and IQ (see Summary of Results section above) considered a unique combination of covariates. The findings of these studies consistently provide evidence of an association between lower IQ in children and exposure to fluoride regardless of the inclusion or absence of consideration of any one or combination of covariates of interest. For example, maternal or family member smoking was addressed in 7 of the 19 low risk-of-bias studies, and this did not appear to affect the conclusions. All 7 studies that accounted for smoking found evidence of an association between fluoride exposure and lower IQ scores as did 11 of the 12 studies that did not account for smoking. Similarly, all 16 studies that addressed the three key covariates (age, sex, SES) (16 of 16 studies) and two of the three studies that did not fully account for them also found evidence of an association between fluoride exposure and lower IQ scores. In summary, when considering the impact of each covariate (or combinations of covariates) on the consistency of results, no trends are discernable that would suggest that bias due to confounding has impacted or would explain the consistency in findings across the body of evidence that fluoride exposure is associated with lower IQ in children.

Commented [A63]: Change made in response to the [redacted] Reviewer comment below; see DocF_Monograph for detailed response to related reviewer comment:

Reviewer comment (DocF_Monograph, pages 5 - 11) : High?

Commented [A64]: The following two sentences reflect revisions to further explain why the concern over co-exposure to arsenic in Soto-Barreras et al. (2019) would not result in the study being considered high risk of bias overall, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 17): It is unclear why Soto Barreras is considered to be low risk of bias (overall) if they did not account for arsenic in a high-exposure area. This seems like a major flaw. [redacted] did not find any description in the main part of the results (pages 28-41) that discuss this study and why it's included. Although, there's the unclear, unreferenced statement (page 36) that "Only one study did not observe evidence of an association between fluoride exposure and IQ; however, results were not provided in a manner that allowed for a direct comparison with other low risk-of-bias studies."

Commented [A65]: Change made in response to the [redacted] Reviewer comment below; see DocF_Monograph for detailed response to related reviewer comment:

Reviewer Comment (DocF_Monograph, pages 5 - 11) : High?

Five of the low risk-of-bias studies confirmed the robustness of the results by conducting sensitivity analyses (Bashash et al. 2017; Green et al. 2019; Till et al. 2020; Wang et al. 2020b; Yu et al. 2018), and none of the sensitivity analyses adjusting for additional covariates found meaningful shifts in the association between fluoride exposure and IQ or other measures of cognitive function. Bashash et al. (2017) found that adjusting for HOME score increased the association between maternal urinary fluoride and children's IQ. Green et al. (2019) reported that adjusting for lead, mercury, manganese, perfluorooctanoic acid, and arsenic concentrations did not substantially alter the associations with IQ. Sensitivity analyses by Yu et al. (2018) that adjusted for covariates (including age, sex, and socioeconomic status) did not find differences in the results compared with the primary analyses. Wang et al. (2020b) found the results of the sensitivity analysis to be the same as the results from the primary analysis. Till et al. (2020) observed that adjusting for maternal urinary fluoride levels, as a way to consider postnatal exposure, had little impact on the results.

Among the 19 low risk-of-bias studies, three were identified that have potential for bias due to confounding (Cui et al. 2020; Ding et al. 2011; Soto-Barreras et al. 2019). This was mainly due to a lack of details on covariates considered key for all studies (i.e., age, sex, and SES). See Appendix E for further discussion of the risk-of-bias concerns regarding confounding for individual studies. Although these three studies have some potential for bias due to confounding, they are considered to be low risk of bias overall because they have low potential for bias for the other two key risk-of-bias questions (exposure characterization and outcome assessment), and no other major concerns for bias were identified. Consistent with the 16 studies that adequately addressed confounding, two of these three studies also provide evidence of an association between fluoride exposure and lower IQ scores in children.

Taken together and considering the consistency in the results despite the variability across studies in which covariates were accounted for, bias due to confounding is not considered to be a concern in the body of evidence. The potential for the consistency in results to be attributable to bias due to confounding in the 19 low risk-of-bias studies is considered low.

Study (Location)[a]	Potential Covariates Considered[b]: Subject Characteristics: Age	Sex	Race/Ethnicity	Health Factors[c]	Other Exposures: Arsenic	Smoking	Iodine	Lead	Other[c]	Socioeconomic Factors: SES[d]	Caregiving Environment (e.g., HOME score)	Parental Characteristics: Demographics[c]	Reproductive Factors[c]	Health Factors[c]	IQ	Other[c]	Notes	Reported Association with Fluoride[e]
Overall RoB Rating for Confounding: Probably Low																		
Bashash 2017 (Mexico)	√	√	–	–	√	√	–	√	√	√	√	√	√	–	√	√	Other exposures: Hg, Ca Demographics: maternal age Reproductive: parity, birth order, birth weight, gestational age at delivery Other: cohort	Yes
Choi 2015 (China)	√	√	–	√	√	–	–	√	–	√	–	√	√	√	–	√	Health: subject Fe deficiency, illnesses before age 3, medical history of subject and caretakers Demographics: parental age Reproductive: parity Other: residential history	Yes
Cui 2018 (China)	√	√	√	√	√	√	√	–	–	√	–	√	√	√	–	√	Health: subject BMI, stress/anger/anxiety/depression, psychological trauma, having a cold, in relatives: thyroid diseases, cancer, mental retardation Demographics: maternal age Reproductive: abnormal birth Other: alcohol consumption, proximity to factory, physical activity, various dietary factors, environmental noise	Yes
Green 2019 (Canada)	√	√	√	–	√	√	–	√	√	√	√	√	√	–	–	√	Other exposures: Hg, Mn, PFOA Demographics: parental age, pre-pregnancy BMI Reproductive: parity, weeks of gestation, birth weight, maternal chronic condition during pregnancy Other: alcohol consumption, birth country, voiding interval in urine sampling, breastfeeding duration	Yes
Rocha-Amador 2007 (Mexico)	√	√	–	√	√	–	–	√	–	√	–	–	–	–	–	–	Health: subject height and weight by age, transferrin saturation	Yes
Saxena 2012 (India)	√	√	–	√	√	–	√	√	–	√	–	–	–	–	–	√	Health: subject height for age ratio, weight for height ratio, medical history Other: residential history	Yes
Seraj 2012 (Iran)	√	√	–	–	√	–	√	√	–	√	–	–	–	–	–	√	Other: fluorosis intensity	Yes
Sudhir 2009 (India)	√	√	–	–	√	–	–	√	–	√	–	–	–	–	–	√	Other: staple food consumed, liquids routinely consumed, aids used for oral hygiene maintenance (fluoridated or nonfluoridated)	Yes
Till 2020 (Canada)	√	√	√	–	√	√	–	–	–	√	√	–	–	–	–	√	Other: city	Yes
Trivedi 2012 (India)	√	√	–	–	√	–	√	–	–	√	–	–	–	–	–	–		Yes
Wang 2012 (China)	√	√	–	√	–	–	√	√	–	√	–	–	–	√	–	√	Health: medical conditions Other: transportation, natural environment, and lifestyle	Yes
Wang 2020b (China)	√	√	–	√	√	√	√	√	–	√	–	–	√	–	–	√	Health: subject BMI, exclusions based on diseases affecting intelligence, history of trauma or neurological disorders, positive screening test history Reproductive: low birth weight Other: alcohol consumption	Yes
Xiang 2003 (China)	√	√	–	–	–	–	√	√	–	√	–	–	–	–	–	–		Yes
Xiang 2011 (China)	√	√	–	–	–	–	√	√	–	√	–	–	–	–	–	–		Yes
Yu 2018 (China)	√	√	–	√	√	√	√	√	√	√	–	–	√	–	–	√	Health: subject BMI Reproductive: disease history during pregnancy, delivery conditions Other: dental fluorosis prevalence, consanguineous marriage	Yes
Zhang 2015b (China)	√	√	–	√	√	–	√	√	√	√	–	–	–	–	–	√	Health: subject physical and mental health status Other: thyroid hormone levels, residential history, having knowledge of fluorosis, COMT genotype	Yes
Overall RoB Rating for Confounding: Probably High																		
Cui 2020 (China)	–	√	–	√	√	√	√	–	–	√	–	√	√	√	–	√	Health: stress/anger/anxiety, having a cold, in relatives: mental retardation Demographics: maternal age Reproductive: abnormal birth, smoking and drinking during pregnancy Other: thyroid hormone levels	Yes[f]
Ding 2011 (China)	√	–	–	–	√	–	√	√	–	–	–	–	–	–	–	–		Yes
Soto-Barreras 2019 (Mexico)	√	√	–	–	–	–	–	–	–	√	–	–	–	–	–	–		No

Figure 6. Important Covariates Considered in Low Risk-of-bias IQ Studies Conducted in Children

[a]Includes all low risk-of-bias IQ studies in children. Studies are organized as those with an overall risk-of-bias rating for confounding as probably low (green) followed by those with an overall risk-of-bias rating for confounding as probably high (yellow).
[b]Covariates represented here are those considered important for this evaluation. Depending on the specific study population, individual covariates may be considered a potential confounder, effect measure modifier, and/or co-exposure. See study details provided in HAWC (NTP 2019) for information on additional covariates.
Factors outlined in blue are key covariates for all studies (subject age, subject sex, SES) and arsenic (which is of particular importance to some study populations).

[A] √ indicates that a covariate was considered. Examples of what it means for a covariate to be "considered": it was adjusted for in the final model, it was considered in the model but not included in the final model because it did not change the effect estimate, it was reported to have the same distribution in both the exposed and unexposed groups, it was reported to not be associated with the exposure or outcome in that specific study population. For arsenic, a √ might also be used when arsenic was not expected to be an issue because there is no evidence to indicate that the co-exposure was prevalent or occurring in relation to fluoride. See risk-of-bias explanations in Appendix E [or HAWC (NTP 2019) for details. A hyphen (-) indicates that the factor was not considered.

[c]See the "Notes" column for additional details.

[d]Covariates considered measures of SES include SES scaled scores, household/family income, child education, caretaker/parental education, and occupation/employment.

[e]Extent of reported associations varies by study. "Yes" indicates that study authors provided evidence of an association between lower IQ scores and fluoride exposure.

[f]Study reported lower IQ scores with increasing fluoride exposure, but the results did not achieve statistical significance.

High Risk-of-bias Studies

Most high risk-of-bias studies (n = 53) considered important covariates to some degree through study design or analysis; however, when considering the full scale of potential concerns of bias due to confounding, all but three of these studies were rated probably or definitely high risk of bias. The majority of high risk-of-bias studies accounted for one or two of the three covariates considered key for all studies (age, sex, SES) but did not address all three and did not address other covariates considered important for the specific study population and outcome. Potential confounding related to important co-exposures (e.g., arsenic) was often not addressed in high risk-of-bias studies. In studies in which there was high exposure to fluoride via drinking water with high naturally occurring fluoride or from the use of coal-containing fluoride, most researchers did not account for potential exposures to arsenic, which is commonly found in coal and drinking water in fluoride-endemic areas of China and Mexico.

Despite the lack of adequate consideration of key covariates in the vast majority of high risk-of-bias studies, the results across most of these studies (46 of 53) consistently provide evidence of an association between fluoride exposure and IQ, supporting the results observed in the low risk-of-bias studies. This finding suggests that confounding is likely less of a concern for the body of evidence as a whole than for any individual study. Although the high risk-of-bias studies may have more potential for bias due to confounding compared with the low risk-of-bias studies, the consistent IQ findings across high and low risk-of-bias studies indicate that the results cannot be explained solely by potential bias due to confounding.

Exposure Characterization in IQ Studies

Low Risk-of-bias Studies

In general, there were few, if any, risk-of-bias concerns regarding exposure characterization in the low risk-of-bias studies. These studies mainly had individual exposure data based on urine or water measures with appropriate analyses. Although there are concerns related to using urine samples (see the Risk-of-bias Considerations for Human Studies section for details), the evidence suggests that urinary fluoride is a reasonable measure of exposure (Villa et al. 2010; Watanabe et al. 1995). Using three methods to account for urine dilution, Till et al. (2018) reported that adjusted risk estimates did not differ from unadjusted estimates. Analyzing the same study population as Till et al. (2018), Green et al. (2019) found that adjusting for time of urine collection or time of collection since last void during pregnancy did not substantially affect associations with IQ results in either boys or girls. In addition, adjusting maternal urinary fluoride for creatinine did not substantially alter the observed association (Green et al. 2019). To provide a more accurate and sensitive measurement of maternal urinary fluoride than a single measurement provides, Green et al. (2019) included only participants with valid fluoride

Commented [A66]: This footnote reflects revisions to clarify what we mean by "consider", in response to related comments from [redacted] Reviewers listed below; see DocG_Monograph and DocJ_Monograph, respectively, for detailed response:

Reviewer comment (DocG_Monograph, page 18): It's unclear what is meant (in Table 6 and scattered throughout the Results) that there was "No statistical adjustment for confounders" but then in Figure 6 (and also in the text) all studies "consider" the potential confounders age, sex, and SES.

Reviewer comment (DocJ_Monograph, page 14): In general, it is difficult to understand how cross-sectional studies that adjusted for few, or no, confounders, employed somewhat indirect measures of fluoride exposure (or did not fully capture all sources of exposures to fluoride), or had concerns related to selection bias, were designated as "low risk of bias." If, for example, some confounders were accounted for in the design or analysis, other than statistical adjustment, it may be worth noting that on Table 6 (otherwise, it appears that many papers accounted for no confounders).

For example, Xiang, 2003a did not statistically adjust for any confounders. They did report some findings in relation to some of the confounders, but not to the extent that [redacted] would perceive them to have been fully accounted for. ...

Commented [A67]: This footnote was added to list all measures related to SES that were considered in the low risk-of-bias IQ-in-children studies, in response to related comments from [redacted] Reviewers listed below; see DocH_Monograph, DocI_Monograph, and DocK_Monograph, respectively, for detailed responses:

Reviewer comment (DocH_Monograph, page 5): RE: confounding and covariates. Recent thinking regarding confounding requires the use of directed acyclic graphs to define variables which are theoretically confounders (based on previous literature). Thus, some clarification is needed on how the set of three important confounders were selected, i.e. sex, child age and a measure of socioeconomic status. Indeed, based on literature from other potential neurotoxins (e.g. lead, polychlorinated biphenyls, phthalates) it seems as though child sex would be an effect modifying variable, not a confounder (child sex would not be related, for example to exposure status under any definition of confounding). Variables such as arsenic or lead exposure would be co- ...

Commented [A68]: This footnote was added to provide further clarity that parental education is captured under SES in Figure 6, in response to the [redacted] Reviewer comment below; see DocK_Monograph for detailed response:

Reviewer comment (DocK_Monograph, page 4): Additionally, it might be helpful to identify a limited set of confounder as required for evaluation. For example, those included in Figure 6 do not include all described in Table 6, and in fact m[a]y not present an important one: parental educational attainment.

measurements at all trimesters in their analysis. Other studies also measured urinary fluoride multiple times throughout pregnancy (Bashash et al. 2017). Some studies demonstrated correlations between urinary fluoride and fluoride in drinking water, fluorosis, or estimated dose based on drinking water concentrations and consumption (Choi et al. 2015; Ding et al. 2011; Green et al. 2019; Saxena et al. 2012; Yu et al. 2018; Zhang et al. 2015b). Till et al. (2018) demonstrated that there was a linear association between urinary fluoride concentrations in pregnant women and drinking water fluoride concentrations regardless of method used to correct for urine dilution or whether adjustments were made for dilution. Bashash et al. (2017) excluded exposure outliers and found that doing so did not substantively change the results. Taken together, these studies suggest that urinary fluoride is a reasonable measure of exposure despite some potential issues.

All but one low risk-of-bias study was rated probably or definitely low risk of bias for exposure assessment. Seraj et al. (2012) had potential exposure misclassification and was rated probably high risk of bias for exposure assessment. Villages were categorized as normal (0.5–1 ppm), medium (3.1 ± 0.9 ppm), or high (5.2 ± 1.1 ppm) based on average fluoride content in drinking water in varying seasons over a 12-year period. Mild fluorosis observed in children in the normal fluoride level group indicates that there may have been higher exposure in this group at some point in the past; however, this would bias the results toward the null, and the children in the normal fluoride group had a significantly higher IQ score compared with the medium and high fluoride groups (p-value = 0.001). There were also significant associations between lower IQ scores and fluorosis intensity (p-value = 0.014) and water fluoride concentration when evaluated as a continuous variable (p-values <0.001). Although there is potential for exposure bias, the apparent exposure misclassification and inclusion of children with higher fluoride exposure in the normal group indicate that the association may be greater than what was observed in this study.

High Risk-of-bias Studies

A frequent, critical limitation among the high risk-of-bias studies was lack of information regarding exposure or poor exposure characterization. Many of the high risk-of-bias studies compared only subjects living in two regions with differing levels of fluoride exposure, and although most of them did provide some differentiation in levels of fluoride between the areas, limited or no individual exposure information was reported. Among studies that provided drinking water levels of fluoride in two areas being compared, sufficient information to determine whether the individual study subjects were exposed to these levels was often not reported. Some studies also lacked information on fluoride analysis methods and timing of the exposure measurements. In some cases (n = 3), study areas that were considered endemic for dental and/or skeletal fluorosis were compared with non-endemic areas, or high-fluoride areas were compared with low-fluoride areas, with no other information provided on fluoride levels in the areas (Li et al. 2003 [translated in Li et al. 2008c]; Ren et al. 1989 [translated in Ren et al. 2008]; Sun et al. 1991). Although living in an area endemic for fluorosis could be an indicator of exposure, these studies did not specify whether the study subjects themselves had fluorosis. Another study used only dental fluorosis as a measure of fluoride exposure in subjects who were all from an endemic area with similar drinking water fluoride levels (Li et al. 2010). In one case, multiple sources of fluoride exposure were assessed separately without properly controlling for the other sources of exposure, which could bias the results (Broadbent et al. 2015). Broadbent et al. (2015) assessed fluoride exposure in three ways: use of community water in a fluoridated area

versus a non-fluoridated area, use of fluoride toothpaste (never, sometimes, always), or use of fluoride tablets prior to age 5 (ever, never). The same children were used for each analysis without accounting for fluoride exposure through other sources. For example, there were 99 children included in the non-fluoridated area for the community water evaluation, but there is no indication that these 99 children were not some of the 139 children that had ever used supplemental fluoride tablets or the 634 children that had always used fluoride toothpaste. Therefore, comparing fluoridated areas to non-fluoridated areas without accounting for other sources of exposure that might occur in these non-fluoridated areas would bias the results toward the null.

Outcome Assessment for IQ Studies

Low Risk-of-bias Studies

The low risk-of-bias studies have few concerns regarding outcome assessment. All 19 low risk-of-bias studies used appropriate methods for measuring IQ in the study population being assessed, and blinding of outcome assessors was not a concern in 18 of the 19 studies [i.e., all low risk-of-bias studies except Sudhir et al. (2009)]. Fourteen of these 18 studies reported blinding of the outcome assessors, or correspondence with the study authors confirmed that it was not likely an issue. For the remaining 4 of the 18 studies, it was assumed that the outcome assessors were most likely blind because exposure was assessed via urine or drinking water obtained at the same time as the outcome assessment in the general population studies. One IQ study (Sudhir et al. 2009) had concerns for potential bias in the outcome assessment due to lack of information to determine whether blinding at the time of the outcome assessment was a concern (see Appendix E for details).

High Risk-of-bias Studies

Among the studies with high risk of bias, the main limitation in the outcome assessment was the lack of reporting on blinding of the outcome assessor (i.e., whether the outcome was assessed without knowledge of exposure). Although there is little concern that the children's knowledge of their own exposure would bias the way they took the IQ tests, there is potential for bias if the tests were administered by an interviewer, or if the scoring of results could be subjective (e.g., drawing tests), and the interviewer or scorer had knowledge of the children's exposure. Most of the studies did not provide sufficient information on the person scoring or administering the tests or other information on the assessment methods to alleviate concerns for potential interviewer or reviewer bias.

High risk-of-bias studies were mainly carried out in two separate populations without information provided that the tests were conducted in a central location. In many cases, the methods indicated that the tests were conducted at the schools in the study area (indicating that there was likely knowledge of exposure). In some cases, the outcomes were not considered sensitive measures (e.g., Seguin Form Board Test to test for IQ), or the test was not considered appropriate for the study population (e.g., a test validated in a western population was used on a rural Chinese population).

Confidence Assessment of Findings on IQ in Children

We conclude that there is moderate confidence in the body of evidence that higher fluoride exposure is associated with lower IQ in children. This confidence rating was reached by starting

Commented [A69]: This sentence, and many sentences throughout the monograph, reflect revisions to further clarify the studies to which we are referring, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 18): The Outcome Assessment for IQ Studies section (page 48) is unclear. This problem occurs in much of the write up, where it is unclear what studies are being referred to. It states that 18 of 19 studies were low risk ("used appropriate methods for measuring IQ"), but does not indicate which study did not use appropriate methods or what the problem is. At the end of the paragraph there's a sentence about Sudhir not reporting blinding, but the paragraph starts by saying that "blinding of outcome assessors was not a concern).

Commented [A70]: This paragraph reflects revisions to rearrange the text to start with the initial confidence rating and how initial confidence rating is determined, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 20): A clearer statement, up front, is needed that the starting point for confidence is "moderate" and why this is the case. [redacted] think you're trying to say this with "The initial moderate confidence rating in the body of evidence" on page 48, but this sentence is unclear. [redacted] still unsure if "initial" here means where the GRADE confidence rating starts before assessing the evidence. Why start at moderate? The Methods section does not describe this concept.

with an initial confidence rating based on key study design features of the body of evidence and then considering factors that may increase or decrease the confidence in that body of evidence. The initial moderate confidence rating is based on 15 of the 19 low risk-of-bias studies that have 3 of the 4 key study design features shown in Figure 1 (i.e., exposure occurred prior to outcome, individual-based outcomes were evaluated, and a comparison group was used). Three of these studies were prospective cohort studies, and 12 were cross-sectional studies that provided evidence of long-term, chronic fluoride exposure prior to outcome measurement.

There are nine factors to consider for increasing or decreasing the confidence in the body of evidence (provided in Figure 1). Discussion of each of these factors in the body of evidence on fluoride exposure and IQ in children is presented below.

- **Risk of bias**: Only studies that were considered to have low risk of bias were included in the moderate confidence rating; therefore, there was no downgrade for risk-of-bias concerns.
- **Unexplained inconsistencies:** The ~~data are~~direction of the association is consistent in the majority of studies, and there was no downgrade for this factor. Eighteen of the 19 low risk-of-bias studies reported associations between higher fluoride levels and lower IQ scores in children. These studies were conducted in 5 different countries on more than 7,000 children from 15 different study populations. There is consistency in ~~results~~ the direction of the association across prospective and cross-sectional study designs. There is also consistency in ~~results~~ the direction of the association across studies using different fluoride exposure measures, including urinary and drinking water fluoride. The one study that did not observe an association did not provide results in a comparable manner and therefore this body of evidence is not considered to have unexplained inconsistencies.
- **Indirectness:** IQ in humans is a direct measure of the association of interest; therefore, no adjustment in confidence is warranted.
- **Imprecision:** There is no evidence of serious imprecision that would warrant a downgrade. Eighteen low risk-of-bias studies reported lower IQ with higher fluoride, and no issues with imprecision were identified to challenge the significance of the ~~effect~~ response estimates.
- **Publication bias:** There is no strong evidence of publication bias; therefore, no downgrade was applied for publication bias. Two published meta-analyses (Choi et al. 2012; Duan et al. 2018) did not indicate strong evidence of publication bias. The draft meta-analysis conducted by NTP in the September 16, 2020, draft monograph found no publication bias among the low risk-of-bias studies (NTP 2020). Among high risk-of-bias studies, adjusting for publication bias using the trim-and-fill analysis estimated that, in the absence of publication bias, the inverse direction of association and statistical significance remained, thus indicating that there was no need to downgrade for publication bias.
- **Large magnitude of effect size:** Although some individual studies indicated a large magnitude of effect size, the magnitude of effect was not the same across all studies. Therefore, the overall data would not support an upgrade due to a large magnitude of

Commented [A71]: This bullet reflects revisions to further characterize the consistency in results in response to the [redacted] Reviewer comment below; see DocJ_Monograph for detailed response:

Reviewer comment (DocJ_Monograph, page 19): Page 48-49, Assessment of Unexplained Inconsistencies: While there is some consistency in findings suggesting that increased exposure to fluoride is associated with lower IQ, many studies reported mixed results (generally reporting a mix of inverse associations and null findings). How were these mixed findings taken into consideration when evaluating unexplained inconsistencies?

Commented [A72]: Change made in response to the [redacted]Reviewer comment below; see DocC_Monograph for detailed response:

Reviewer comment (DocC_Monograph, page15): Imprecision clearly evident upon visual inspection of Figures A1 through A5, which frequently shows wide and overlapping confidence intervals. Therefore, this discussion item should be revised.

Commented [A73]: This sentence was revised to clarify that serious imprecision as described in the protocol is required to downgrade, which was not found for this body of evidence, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, Page 12): Regarding the GRADE assessment of imprecision, please provide more detail about your thresholds between precise and imprecise. For example, what 95% CI would indicate imprecise?

- **Dose response:** Evidence of an exposure-response relationship that could justify an upgrade to the confidence in the body of evidence is not presented in this monograph. While the overall findings qualitatively appear less clear in the lower exposure range, many of the studies that provide data to evaluate exposure response were judged to be high risk of bias. The meta-analysis conducted in association with this systematic review further informs this issue and will be ~~published~~ ~~seperately~~refined in preparation for a separate publication.
- **Residual confounding:** Xiang et al. (2003a), Xiang et al. (2011), and Wang et al. (2012) studied the same population where arsenic occurred in the area with low fluoride but did not occur in the area with high fluoride. This would have biased the results toward the null, but there were significantly lower IQ scores in the area with high fluoride. The remaining studies do not provide enough information to consider whether residual confounding occurred for the body of evidence. Note that parental IQ has the potential to be an important factor when considering residual confounding based on likely correlations between parental IQ and children's IQ; however, there is not sufficient evidence that parental IQ is associated with water fluoride content. Taken together, the overall data would not support an upgrade due to residual confounding.
- **Consistency:** The consideration of a potential upgrade for consistency in the methods is primarily for non-human animal evidence, where it would be applied to address increased confidence for consistent effects across multiple non-human animal species. For human evidence, it is generally not applied, and the data would only be considered in deciding whether to downgrade for unexplained inconsistency. Therefore, no upgrade is applied for consistency.

As described above, there are no changes in confidence rating based on any of the possible upgrade or downgrade factors. The magnitude of effect size and the overall strength and quality of the human literature base provide moderate confidence in the body of evidence that higher exposure to fluoride is associated with lower IQ in children (see the Discussion section for strengths and limitations of the evidence base). Note that additional, well-designed prospective cohort studies with individual-level exposure data and outcome measures could provide increased confidence in the association between fluoride exposure and lower IQ in children.

Other Neurodevelopmental or Cognitive Effects in Children

Low Risk-of-bias Studies

Overview of Studies

Nine low risk-of-bias studies (three prospective cohort and six cross-sectional studies) evaluated the association between fluoride exposure and cognitive neurodevelopmental effects other than IQ in children. These nine studies were conducted in multiple study populations in three countries, specifically:

- three were conducted in three areas of China on three study populations,
- four were conducted in two areas of Mexico on three study populations, and
- two were conducted in Canada using the same study population.

Commented [A74]: This bullet reflects revisions to limit the description of the dose-response relationship, since the dose-response meta-analysis was removed from the monograph, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 19): Similarly, the description around "Dose-response" on page 49 is not clearly supported by the text of the Results section. There is no clear dose-response section of the Results where related findings are described. The Results text mostly summarizes as "high" or "exposure" or in some instances association with a 1-mg/L increase or the equivalent.

Commented [A75]: Change made in response to the [redacted] Reviewer comment below; see DocF_Monograph for detailed response:

Reviewer comment (DocF_Monograph, page 13): Would revise because it is being submitted for publication and also [redacted] would not link it to further informing this since there are still questions about the meta-analysis.

Commented [A76]: This bullet reflects revisions to remove text on the consistency across studies, as this was addressed in the preceding Unexplained Inconsistencies bullet, in response to the [redacted] Reviewer comment below; see
DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 19): The "Consistency" section on page 49 should not discuss the consistency across studies. This was addressed in Unexplained Inconsistencies on the prior page. Do not confuse the two issues for the reader.

Commented [A77]: This sentence reflects revisions to include information on counts of studies per study design, in response to the [redacted] Reviewer comment below; see DocJ_Monograph for detailed response:

Reviewer comment (DocJ_Monograph, page 12): In general, it would be helpful to the reader to describe the included study designs earlier in the respective results sections (i.e., along with the total #s of included articles, describing the # of cross-sectional, prospective cohort, etc is important).

There is considerable heterogeneity across studies, particularly in the different health outcomes evaluated and ages assessed. Most studies measured fluoride in the drinking water or urine (child or maternal) with one study using severity of dental fluorosis as an exposure measure in addition to drinking water and children's urine. Two of the studies were conducted on infants, with one evaluating effects within 72 hours of birth (Li et al. 2004 [translated in Li et al. 2008a]) and the other evaluating effects at 3 to 15 months of age (Valdez Jimenez et al. 2017). The remaining studies were conducted in children of varying ages, ranging from 4 to 17 years. Other cognitive neurodevelopmental outcomes assessed include neurobehavioral effects in infants, learning and memory impairment, and learning disabilities such as attention deficit hyperactivity disorder (ADHD). Few studies measured the same health outcomes, used the same outcome assessment methods, or evaluated the same age groups.

Table 7 provides a summary of study characteristics and key findings related to other cognitive neurodevelopmental outcomes and fluoride exposure for the nine low risk-of-bias studies. The different tests conducted and the populations on which the tests were conducted are also indicated in Table 7. Several of these studies conducted multiple analyses and reported results on multiple endpoints. The purpose of the table is to summarize key findings (independent of whether an association was found) from each study and is not meant to be a comprehensive summary of all results. For each study, results are summarized for each exposure measure assessed. Results from multiple analyses using the same exposure measure may not all be presented unless conflicting results were reported. See Appendix E for additional information on studies in Table 7, including strengths and limitations, clarifications for why they are considered to pose low risk of bias, and information regarding statistical analyses, covariates, exposure assessment, and outcome assessment.

Commented [A78]: This sentence was added to direct the reader to Appendix E for information on statistical methods and other suggestions for Table 6 in response to the [redacted] Reviewer comment below; see DocI_Monograph for detailed response:

Reviewer comment (DocI_Monograph, page 7): Table 6 could include the following: 1) statistical methods; 2) confounders, particularly exposure to other known neurotoxicants, and how they were measured; 3) might rename 'Assessment timing' to age of participants or just combine the information with the location/subject's column

Table 7. Studies on Other Neurodevelopmental and Cognitive Function in Children[a]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
China					
Li et al. (2004) [translated in Li et al. 2008a]	Cross-sectional Zhaozhou County, Heilongjiang Province/neonates [91]	Drinking water Range: 0.5–1.0 mg/L (control); 1.7–6.0 mg/L (high) Maternal urine during pregnancy Mean (SD): 1.74 (0.96) mg/L (control); 3.58 (1.47) mg/L (high)	Neonates (24–72 hours after delivery)	Neurodevelopmental: Neonatal behavioral neurological assessment (NBNA)	Significant differences in neurobehavioral assessment total scores between high-fluoride (36.48 ± 1.09) and control groups (38.28 ± 1.10) (subjects divided into high fluoride group and control group based on drinking water fluoride levels in place of residence); significant differences in total score of behavioral capability that includes measures of non-biological visual orientation reaction and biological visual and auditory orientation reaction between the two groups (11.34 ± 0.56 in controls compared to 10.05 ± 0.94 in high-fluoride group) No statistical adjustment for covariates
Choi et al. (2015)	Cross-sectional Mianning County/1st grade children [51]	Drinking water GM: 2.20 mg/L Children's urine GM: 1.64 mg/L Severity of fluorosis (Dean Index)	Children (ages 6–8 years)	Learning and memory: Neuropsychological tests including WRAML Visual motor ability: WRAVMA Motor ability: Finger tapping task Manual dexterity: Grooved pegboard test	Outcomes unrelated to the IQ test not significantly associated with any fluoride exposure measure Adjusted for age, sex, parity, illness before 3 years old, household income last year, and caretaker's age and education
Wang et al. (2020a)	Cross-sectional Tongxu County/school children [325]	Children's urine Mean (SD): 1.54 (0.89) mg/L	Children (ages 7–13 years)	ADHD and behavior measures: Conners' Parent Rating Scale-Revised (Chinese version) (CPRS-48)	Significant association between psychosomatic problems and urinary fluoride level (per 1-mg/L increase; β = 4.01; 95% CI: 2.74, 5.28; OR for T-score >70 = 1.97; 95% CI: 1.19, 3.27); no associations between urinary fluoride level and ADHD index or other behavioral measures Adjusted for age, sex, child's BMI, urinary creatinine, mother migrated, and father migrated

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
Mexico					
Rocha-Amador et al. (2009)	Cross-sectional Durango/elementary school children [80]	Children's urine GM (SD): 5.6 (1.7) mg/L	Children (ages 6–11 years)	Visuospatial organization and visual memory: Rey-Osterrieth Complex Figure Test, children's version	Significant correlation between urinary fluoride and visuospatial organization ($r = -0.29$) and visual memory scores ($r = -0.27$); no significant correlation with arsenic Adjusted for age
Valdez Jimenez et al. (2017)	Cohort (Prospective) Durango City and Lagos de Moreno/infants [65]	Maternal urine Range: 0.16–8.2 mg/L (all trimesters) Drinking water Range: 0.5–12.5 mg/L (all trimesters)	Infants (ages 3–15 months)	Mental development index (MDI): Bayley Scales of Infant Development II (BSDI-II) Psychomotor developmental index (PDI): Bayley Scales of Infant Development II (BSDI-II)	Significant association between log10-mg/L maternal urinary fluoride and MDI score during first trimester (adjusted $\beta = -19.05$; SE = 8.9) and second trimester (adjusted $\beta = -19.34$; SE = 7.46); no significant associations between maternal urinary fluoride and PDI score; analyses of outcomes using drinking water fluoride not performed Adjusted for age, gestational age, marginality index, and type of drinking water
Bashash et al. (2017)[c]	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [299] GCI analysis [287]	Maternal urine during pregnancy Mean (SD): 0.90 (0.35) mg/L Children's urine Mean (SD): 0.82 (0.38) mg/L	Children (age 4 years)	General cognitive index (GCI): McCarthy Scales of Children's Abilities (MSCA)	Significant association between maternal urinary fluoride and offspring GCI score (per 0.5-mg/L increase adjusted $\beta = -3.15$; 95% CI: −5.42, −0.87); associations with children's urine not significant Adjusted for gestational age; weight at birth; sex; parity (being the first child); age at outcome measurement; and maternal characteristics, including smoking history (ever smoked during the pregnancy vs. nonsmoker), marital status (married vs. not married), age at delivery, IQ, education, and cohort

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
Bashash et al. (2018)[c]	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [210]	Maternal urine during pregnancy Mean 0.85 (95% CI: 0.81, 0.90) mg/L	Children (ages 6–12 years)	ADHD: Conners' Rating Scales-Revised (CRS-R)	Significant associations between maternal urinary fluoride (per 0.5-mg/L increase) and CRS-R scores, including Cognitive Problems + Inattention Index (adjusted β = 2.54; 95% CI: 0.44, 4.63), DSM-IV Inattention Index (adjusted β = 2.84; 95% CI: 0.84, 4.84), DSM-IV ADHD Total Index (adjusted β = 2.38; 95% CI: 0.42, 4.34), and ADHD Index (adjusted β = 2.47; 95% CI: 0.43, 4.50) Adjusted for gestational age; birth weight; sex; parity; age at outcome measurement; and maternal characteristics, including smoking history (ever smoked vs. nonsmoker), marital status (married vs. not married), education, socioeconomic status, and cohort
Canada					
Barberio et al. (2017b)[d]	Cross-sectional General population/Canadian Health Measures Survey (Cycles 2 and 3) [2,221]	Children's urine Mean Cycle 2: 32.06 (95% CI: 29.65, 34.46) µmol/L Mean Cycle 3: 26.17 (95% CI: 22.57, 29.76) µmol/L	Children (ages 3–12 years)	Learning disability, ADHD (Cycle 2 only): Parent or child self-report	Significant increase in adjusted OR for learning disability (adjusted OR = 1.02; 95% CI: 1.00, 1.03) per 1-µmol/L increase in unadjusted urinary fluoride when Cycle 2 and 3 were combined; no significant associations found between urinary fluoride and ADHD (only evaluated in Cycle 2); no significant associations found when using creatinine- or specific gravity-adjusted urinary fluoride Adjusted for age and sex, household income adequacy, and highest attained education in the household

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
Riddell et al. (2019)[d]	Cross-sectional General population/Canadian Health Measures Survey (Cycles 2 and 3) [3,745]	Drinking water Mean (SD): 0.23 (0.24) mg/L [non-fluoridated water: 0.04 (0.06) mg/L; fluoridated water: 0.49 (0.22)] Community water fluoridation status (yes or no) Children's urine Mean (SD): 0.61 (0.39) mg/L [non-fluoridated water: 0.46 (0.32) mg/L; fluoridated water: 0.82 (0.54)]	Children (ages 6–17 years)	Hyperactivity/inattention: Strengths and Difficulties Questionnaire (SDQ) ADHD: parent or self-reported physician diagnosis	Significantly increased risk of ADHD with fluoride in tap water (adjusted OR = 6.10 per 1-mg/L increase; 95% CI: 1.60, 22.8) or community water fluoridation status (1.21; 95% CI: 1.03, 1.42) but not with urinary fluoride; similar results observed with attention symptoms based on the SDQ scores Adjusted for age and sex, child's BMI, ethnicity, parental education, household income, blood lead, and smoking in the home

ADHD = attention-deficit/hyperactivity disorder; BMI = body mass index; GCI = General Cognitive Index; GM = geometric mean; HOME = Home Observation Measurement of the Environment; IQ = intelligence quotient; MSCA = McCarthy Scales of Children's Abilities; SD = standard deviation; WASI = Wechsler Abbreviated Scale of Intelligence (Spanish version); WISC-IV = Wechsler Intelligence Scale for Children-Revised (as reported by Choi et al. 2015); WRAML = Wide Range Assessment of Memory and Learning; WRAVMA = Wide Range Assessment of Visual Motor Ability.

[a]Includes low risk-of-bias studies.

[b]Associations between other cognitive neurodevelopmental outcomes in children and fluoride levels were reported quantitatively, when possible. For studies with multiple analyses and results, the table summarizes key findings and is not a comprehensive summary of all findings. Results also indicated when a study reported no association, provided as a qualitative statement of no association.

[c]Bashash et al. (2017) and Bashash et al. (2018) are based on the same study population.

[d]Barberio et al. (2017b) and Riddell et al. (2019) are based on the same study population.

Summary of Results

Overall Findings

Although discussed together in this section, various health outcomes were assessed in the nine low risk-of-bias studies of other neurodevelopmental outcomes, including neurobehavioral scores in infants (two studies), cognitive tests in children other than IQ (three studies), and ADHD or learning disabilities (four studies) in children. ~~Altogether, t~~The results from eight of nine low risk-of-bias studies (three prospective cohort studies and five cross-sectional studies from seven different study populations) provide evidence of significant associations between higher fluoride exposure and cognitive neurodevelopmental outcomes in children other than decrements in IQ (see Figure A-9 through Figure A-11) (Barberio et al. 2017b; Bashash et al. 2017; Bashash et al. 2018; Li et al. 2004 [translated in Li et al. 2008a]; Riddell et al. 2019; Rocha-Amador et al. 2009; Valdez Jimenez et al. 2017; Wang et al. 2020a). Only one cross-sectional study did not find a significant association between fluoride exposure and a measure of cognitive neurodevelopment (Choi et al. 2015).

Although there is heterogeneity in the outcomes assessed and a limited number of directly comparable studies, the data provide additional evidence (beyond the consistent evidence of an association between fluoride exposure and IQ) of an association between higher fluoride exposure and cognitive or neurodevelopmental effects. The body of evidence from the nine low risk-of-bias studies is described in further detail below and is grouped into outcome categories of studies that are most comparable.

Results in Infants

Two studies evaluated neurobehavioral effects in infants either shortly after birth or at 3 to 15 months of age (Li et al. 2004 [translated in Li et al. 2008a]; Valdez Jimenez et al. 2017). Both studies observed a significant association between higher fluoride exposure and lower neurobehavioral scores. In neonates (1–3 days old), the high fluoride group (3.58 ± 1.47 mg/L fluoride based on spot maternal urine collected just prior to birth) had significantly lower total neurobehavioral assessment scores (36.48 ± 1.09 versus 38.28 ± 1.10 in controls; p-value <0.05) and total behavioral capacity scores (10.05 ± 0.94 versus 11.34 ± 0.56 in controls; p-value <0.05) compared to the control group (1.74 ± 0.96 mg/L fluoride) as measured by a standard neonatal behavioral neurological assessment (NBNA) method (Li et al. 2004 [translated in Li et al. 2008a]). In infants 3 to 15 months of age, the Mental Development Index (MDI)—which measures functions including hand-eye coordination, manipulation, understanding of object relations, imitation, and early language development—was significantly inversely associated with maternal urinary fluoride in both the first and second trimesters (adjusted βs per log10-mg/L increase = −19.05 with standard error of 8.9 for first trimester [p-value = 0.04] and −19.34 with standard error of 7.46 for second trimester [p-value = 0.013]) (Valdez Jimenez et al. 2017). Note that this study did not find an association between maternal fluoride during any trimester and the Psychomotor Developmental Index (PDI), which measures gross motor development (adjusted βs = 6.28 and 5.33 for first and second trimesters, respectively; no standard errors provided) (Valdez Jimenez et al. 2017).

Results for Cognitive Tests Other Than IQ in Children

Three studies conducted tests on cognitive function in children that were not part of an IQ test (Bashash et al. 2017; Choi et al. 2015; Rocha-Amador et al. 2009). None of the studies

Commented [A79]: Change made in response to the [redacted] Reviewer comment below; see DocC_Monograph for detailed response:

Reviewer comment (DocC_Monograph, page 17): In this section, this summary statement without further explanation is misleading. Elsewhere in this document the authors indicate that the data regarding ADHD effects contains significant heterogeneity regarding methods and outcomes and thereby precludes conclusions about ADHD and other attention-related disorders.

Commented [A80]: Change made in response to the [redacted] Reviewer comment below; see DocF_Monograph for detailed response to related reviewer comment:

Reviewer comment (DocF_Monograph, pages 5 - 11) : High?

Commented [A81]: This sentence and text throughout the monograph reflect changes to clarify that the associations are per log10-mg/L increase in fluoride exposure, in response to the [redacted] Reviewer comment below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 18): Some of the associations are really quite large, e.g. adjusted betas of -19 in the study of Valedez Jimenez et al 2017, especially for the Bayley Scale. Such associations are either suspect or are not adjusted for the concentration of fluoride appropriately (maybe it is a log unit change). This needs to be clarified.

conducted the same tests, but two of the three studies (Bashash et al. 2017; Rocha-Amador et al. 2009) observed associations between fluoride exposure and lower test scores. The General Cognitive Index (GCI) of the McCarthy Scales of Children's Abilities (MSCA) in 4-year-old children was significantly inversely associated with maternal creatinine-adjusted urinary fluoride levels during pregnancy (collected during each trimester) (adjusted β per 0.5-mg/L increase = −3.15 [95% CI: −5.42, −0.87; p-value = 0.01] in a model adjusting for main covariates including gestational age, weight at birth, sex, maternal smoking, and indicators of socioeconomic status). The association remained even after adjusting for maternal bone lead (adjusted β per 0.5-mg/L increase = −5.63 [95% CI: −8.53, −2.72; p-value <0.01]) (Bashash et al. 2017) (see Figure A-11). Choi et al. (2015), however, evaluated cognitive function endpoints in addition to IQ and found no significant associations between concurrent log-transformed water or urinary fluoride levels and Wide Range Assessment of Visual Motor Ability (WRAVMA) scores, finger tapping test scores, and grooved pegboard test scores, although there were some significant associations based on degree of fluorosis (see Figure A-11). Another study using visuoconstructional and memory scores from the Rey-Osterrieth Complex Figure Test in children 6–11 years old observed significantly lower scores with increasing concurrent child single spot urinary fluoride even after adjusting for age (partial correlation coefficients, per log-mg/L increase = −0.29 and −0.27 for copy [p-value <0.001] and immediate recall [p-value <0.001], respectively [CIs not reported]) (Rocha-Amador et al. 2009). Although these children were also exposed to arsenic, the presence of arsenic could not explain the changes because, in the area with natural contamination by fluoride and arsenic (F–As), the test scores were not significantly associated with urinary arsenic levels (partial correlation coefficients, per log-mg/L increase = −0.05 and 0.02 for copy and immediate recall, respectively [CIs not reported]). The test scores were only marginally increased from fluoride alone when both fluoride and arsenic were included simultaneously in the model (partial correlation coefficients, per log-mg/L increase = −0.32 and −0.34 for copy and immediate recall, respectively [CIs not reported]) (Rocha-Amador et al. 2009) (see Figure A-10).

Commented [A82]: This sentence reflects revisions to characterize the test more precisely as the visuoconstructional and memory score from the Rey-Osterrieth Complex Figure Test, in response to the [redacted] Reviewer comment below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 19): Please clarify what a construction task is (page 56). Do you mean a fine motor copy task?

Commented [A83]: The rest of this paragraph starting here reflects revisions to mention the results adjusted for both fluoride and arsenic, in response to the [redacted] Reviewer comment below; see DocH_Monograph for detailed response:

Reviewer comment (DocH_Monograph, page 19): Also on page 56 and highlighted in blue: this is unclear. Even though urinary arsenic is not associated with scores on these tasks, it could still very well be a confounder of the relationships between fluoride and the test scores.

Attention-related Disorders Including ADHD and Learning Disabilities in Children

Four studies evaluated attention-related disorders or learning disabilities (Barberio et al. 2017b; Bashash et al. 2018; Riddell et al. 2019; Wang et al. 2020a). All four studies found an association between increased fluoride and increased ADHD or learning disability; however, studies varied in the exposure metrics and outcomes measure. Bashash et al. (2018) evaluated behaviors associated with ADHD in children ages 6–12 years using the Conners Rating Scales-Revised (CRS-R) and observed significant associations between maternal urinary fluoride (measured during each trimester) and ADHD-like symptoms, particularly those related to inattention (an increase in 0.5 mg/L of maternal urinary fluoride was significantly associated with a 2.84-point increase [95% CI: 0.84, 4.84; p-value = 0.0054] in the DSM-IV Inattention Index and a 2.54-point increase [95% CI: 0.44, 4.63; p-value = 0.0178] in the Cognitive Problems and Inattention Index). These two scales contributed to the global ADHD Index and the DSM-IV ADHD Total Index, which were also significantly associated with higher levels of prenatal fluoride exposure (an increase of 0.5 mg/L in maternal urinary fluoride was associated with a 2.38-point increase [95% CI: 0.42, 4.34; p-value = 0.0176] in the DSM-IV ADHD Total Index and a 2.47-point increase [95% CI: 0.43, 4.50; p-value = 0.0175] in the ADHD Index) (see Figure A-11). Significant associations were not observed between maternal urinary fluoride concentrations during pregnancy and child performance on measures of hyperactivity, nor were there any significant results in children using Conners' Continuous Performance Test (CPT-II,

2nd Edition), a computerized test of sustained attention and inhibitory control (Bashash et al. 2018). Wang et al. (2020a) also used Conners' Parent Rating Scale (Chinese version) to assess behavioral outcomes in children ages 7–13 years but found only a significant association between spot urinary fluoride concentrations in children (model adjusted for creatinine) and psychosomatic problems (adjusted OR for T-score >70 per 1-mg/L increase = 1.97 [95% CI: 1.19, 3.27; p-value = 0.009] and adjusted β per 1-mg/L increase = 4.01 [95% CI: 2.74, 5.28; p-value <0.001]). No associations were found between spot urinary fluoride and the ADHD index or other behavioral measures.

Barberio et al. (2017b) evaluated learning disabilities in children 3–12 years of age, including ADHD, attention deficit disorder (ADD), and dyslexia, as part of the Canadian Health Measures Survey and found a small but significantly increased risk in self-reported (children 12 years of age) or parent- or guardian-reported (children 3–11 years of age) learning disabilities associated with higher spot urinary fluoride levels in children (adjusted OR per 1-µmol/L increase = 1.02; 95% CI: 1.00, 1.03; p-value <0.05) (see Figure A-12); however, significant associations were not observed in analyses using creatinine- or specific gravity-adjusted urinary fluoride (Barberio et al. 2017b). Barberio et al. (2017b) also reported no associations between single spot urinary fluoride and ADHD in children ages 3 to 12 years. Riddell et al. (2019) used the same Canadian Health Measured Survey but evaluated children 6–17 years old. Riddell et al. (2019) found a significantly increased risk for ADHD diagnosis with both tap water fluoride (adjusted OR per 1-mg/L increase = 6.10; 95% CI: 1.60, 22.8; p-value <0.05) and community water fluoridation status (adjusted OR per 1-mg/L increase = 1.21; 95% CI: 1.03, 1.42; p-value <0.05). A similar increase in the hyperactivity-inattention symptoms score based on the Strengths and Difficulties Questionnaire was observed with both tap water fluoride (adjusted β per 1-mg/L increase = 0.31; 95% CI: 0.04, 0.58; p-value <0.05) and community fluoridation status (adjusted β per 1-mg/L increase = 0.11; 95% CI: 0.02, 0.20; p-value <0.05). As was observed with Barberio et al. (2017b), Riddell et al. (2019) did not observe associations between specific gravity-adjusted spot urinary fluoride concentrations and either ADHD diagnosis (adjusted OR per 1-mg/L increase = 0.96; 95% CI: 0.63, 1.46) or hyperactivity-inattention symptoms (adjusted β per 1-mg/L increase = 0.31; 95% CI: −0.04, 0.66).

Summary of Key Findings for Low Risk-of-bias Studies of Other Neurodevelopmental and Cognitive Effects in Children

In summary, the high-quality studies (i.e., studies with low potential for bias) provide evidence of an association between higher fluoride exposure and neurodevelopmental and cognitive effects in children other than IQ; however, the body of evidence is limited by heterogeneity in the outcomes evaluated and few directly comparable studies. Across these outcomes, eight of nine studies reported a significant association between fluoride exposure and a measure of neurodevelopment or cognition other than IQ, which provides support for the consistency in evidence based on children's IQ studies of an association between fluoride exposure and adverse effects on cognitive neurodevelopment.

Commented [A84]: Change made in response to the [redacted] Reviewer comment below; see DocF_Monograph for detailed response to related reviewer comment:

Reviewer Comment (DocF_Monograph, pages 5 - 11) : High?

High Risk-of-bias Studies

High risk-of-bias studies (n = 6) also provide some evidence of associations between fluoride exposure and neurodevelopmental or cognitive effects in children other than effects on IQ, but the results are inconsistent and address different outcomes (Jin et al. 2016; Li et al. 1994

[translated in Li et al. 2008b]; Malin and Till 2015; Morgan et al. 1998; Mustafa et al. 2018; Shannon et al. 1986).

Risk of Bias for Neurodevelopmental or Cognitive Effect Studies in Children

The confidence in the human body of evidence was based on studies with the lowest potential for bias (i.e., studies that rated probably low or definitely low risk of bias for at least two of the three key risk-of-bias questions and did not have any other risk-of-bias concerns that would indicate serious issues with the studies). Each of the nine low risk-of-bias studies on other neurodevelopmental effects in children had little or no risk-of-bias concerns. Four of the nine studies were rated definitely low or probably low risk of bias for all risk-of-bias questions, and the remaining five studies were rated probably high risk of bias for a single question that was judged to have minimal impact on overall potential bias. None of the nine studies had a rating of definitely high risk of bias for any question. Although the nine low risk-of-bias studies had minimal or no concerns, the six studies with high overall potential for bias had several risk-of-bias concerns related to one or more of the three key risk-of-bias questions (confounding, exposure characterization, and outcome assessment). The key risk-of-bias questions are discussed below. Risk-of-bias ratings for other neurodevelopmental effect studies in children are available in Figure D-9 through Figure D-12 and Appendix E for the low and high risk-of-bias studies.

Confounding for Other Neurodevelopmental Studies in Children

Low Risk-of-bias Studies

As discussed above, there are nine studies considered to have low risk of bias when assessed across all risk-of-bias domains. Seven of nine low risk-of-bias studies were considered to have low potential for bias due to confounding because the authors addressed the three key covariates for all studies (age, sex, and socioeconomic status) and also addressed arsenic as a potential co-exposure of concern through study design or analysis. Other important covariates, including health factors, smoking, and parental characteristics, were also addressed in many of the low risk-of-bias studies. One of the studies (Bashash et al. 2018) examined several covariates in sensitivity analyses involving subsets of participants, including HOME scores, child contemporaneous fluoride exposure measured by child urinary fluoride adjusted for specific gravity, and maternal lead and mercury exposures. The authors reported that none of the sensitivity analyses indicated appreciable changes in the fluoride-related association with behaviors related to ADHD, nor was there evidence of effect modification between maternal urinary fluoride and sex.

Among the nine low risk-of-bias studies, two studies were identified that have potential for bias due to confounding (Rocha-Amador et al. 2009; Valdez Jimenez et al. 2017). Although both of these studies adjusted for several covariates through analysis or study design, Valdez Jimenez et al. (2017) did not address a potential concern for co-exposure to arsenic, and Rocha-Amador et al. (2009) does not appear to adjust for SES or address why it would not be a concern in the study population (see Appendix E for further details). Although these two studies have some potential for bias due to confounding, they are considered to have low potential for bias overall because they have low potential for bias for the other two key risk-of-bias questions (exposure characterization and outcome assessment), and no other major concerns for bias were identified.

Consistent with the IQ studies, bias due to confounding is not likely a concern for the low risk-of-bias studies.

High Risk-of-bias Studies

The six high risk-of-bias studies in the human body of evidence did not adequately address important covariates through study design or analysis. The same concerns due to potential confounding noted previously for the high risk-of-bias children's IQ studies were also present in the other neurodevelopmental high risk-of-bias studies, including not addressing the three key covariates for all studies (age, sex, SES) and/or not addressing potential co-exposures (e.g., arsenic) in areas of potential concern.

Exposure Characterization in Other Neurodevelopmental Studies in Children

Low Risk-of-bias Studies

There were no risk-of-bias concerns regarding exposure assessment in the low risk-of-bias studies. All of the low risk-of-bias studies had individual exposure data based on urine or water measures with appropriate analyses, and most of the urinary fluoride studies accounted for urinary dilution when appropriate. Although there are concerns related to the timing of urine samples (see the Risk-of-bias Considerations for Human Studies section for details), the studies that used maternal urine measured urinary fluoride multiple times throughout pregnancy (Bashash et al. 2017; Bashash et al. 2018; Valdez Jimenez et al. 2017). Another study demonstrated correlations between urinary fluoride and fluoride in the drinking water, fluorosis, or estimated dose based on water (Choi et al. 2015). Bashash et al. (2017) excluded exposure measurement outliers but found that doing so did not change the results in a meaningful way.

High Risk-of-bias Studies

A frequent critical limitation among the high risk-of-bias studies was lack of information regarding exposure or poor exposure characterization. In the high risk-of-bias studies that assessed the association between fluoride exposure and other neurodevelopmental and cognitive effects in children, fluoride exposure assessment was based on dental fluorosis, municipality-level water fluoridation prevalence data, number of years living in an area with fluorinated water, or group-level water samples. See the Exposure Characterization in IQ Studies section for further discussion on the limitations of exposure assessments in high risk-of-bias studies.

Outcome Assessment in Other Neurodevelopmental Studies in Children

Low Risk-of-bias Studies

The low risk-of-bias studies have few concerns regarding outcome assessment. Seven of the nine studies [i.e., all low risk-of-bias studies except Barberio et al. (2017b) and Riddell et al. (2019)] used appropriate methods for measuring other neurodevelopmental effects in the study population, and blinding of outcome assessors was either reported or not a concern in eight of the nine studies [i.e., all with the exception of Wang et al. (2020a)].

Among the nine low risk-of-bias studies, three were identified that have a potential for bias due to outcome assessment. One of the studies (Wang et al. 2020a) had potential concern for bias due to lack of information regarding the blinding of outcome assessors. Two of the studies (Barberio et al. 2017b; Riddell et al. 2019) were based on the same study population in Canada, where different questions were asked in Cycles 2 (2009–2011) and 3 (2012–2013) of the Canadian

Health Measures Survey (CHMS) to ascertain learning disabilities including ADHD. In Cycle 2, subjects were asked whether they had a learning disability diagnosed by a health professional and, if yes, were asked what kind. In Cycle 3, CHMS did not ask what kind of learning disability was diagnosed nor was a reason for the question omission provided. Because no reason was provided for the removal of the question, and because a question on learning disability without the specific diagnosis may be more prone to bias, this change in questioning from Cycles 2 to 3 is a potential concern. Blinding was not considered an issue in these two studies, but the methods for obtaining the information are considered to be less than ideal for measuring learning disabilities including ADHD. Although the questionnaire asked about a doctor's diagnosis of a learning disability, there was no confirmation with medical records. Moreover, these questionnaires were not validated like Conners' Rating Scales, which would have been a better method for assessing ADHD. Although the outcome assessment methods are less than ideal, there was no direct evidence that they were conducted incorrectly or that the methods would have biased the results in any specific direction. Because this was the only concern in these studies, they were considered to have low risk of bias overall.

High Risk-of-bias Studies

Among the studies on other neurodevelopmental effects with high potential for bias, there were several reasons for studies to be considered probably or definitely high risk of bias for outcome assessment. One study (Shannon et al. 1986) was considered to have probably high risk of bias based on lack of information regarding blinding of outcome assessors. One study was considered definitely high risk of bias because outcome was assessed based on a parent-completed questionnaire, and the study authors noted that the parents were informed of the study's intent and were requested to provide information on fluoride history. Other studies used outcome assessment methods that were not validated or utilized group-level measurements (i.e., school performance, working memory scores).

Confidence Assessment of Findings on Other Neurodevelopmental Effects in Children

The high-quality studies (i.e., studies with low potential for bias) provide some evidence of an association between fluoride exposure and other cognitive neurodevelopmental effects, including lower neurobehavioral scores in infants, cognitive effects other than IQ in children, and increased attention-related disorders including ADHD in children. However, due to limitations in the data set, including the heterogeneity in the outcomes assessed, a limited number of directly comparable studies, and differences in outcome assessment methods even when studies evaluated similar outcomes, there is low confidence based on this body of evidence that fluoride exposure is associated with other cognitive neurodevelopmental effects in children. Due to these limitations, the confidence assessment is not described in the same manner as the IQ in Children section or as outlined in Figure 1. Although there are limitations in the body of evidence, the low risk-of-bias studies demonstrate a relationship between higher fluoride exposure and neurodevelopmental effects, even in very young children, which supports the consistency in evidence shown in children's IQ studies of an association between fluoride exposure and adverse effects on cognitive neurodevelopment.

Commented [A85]: Change made in response to the [redacted] Reviewer comment below; see DocC_Monograph for detailed response:

Reviewer comment (DocC_Monograph, page 19): This [second] sentence directly contradicts the leading, first sentence in this paragraph. The first sentence by itself is misleading. Note: reviewer struck out "evidence" and inserted "assumption" as follows […*provide ~~evidence~~ assumption of an association between fluoride exposure and other cognitive neurodevelopmental effects"*] with red text indicating editorial change.

Commented [A86]: The following two sentences reflect revisions to further clarify why we chose not to describe the confidence assessment in the same format and level of detail as in the IQ in Children section, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 21): Sufficiently supported, but it's unclear why the same format used for the IQ studies (pages 48-49) is not used here (page 59).

Cognitive Effects in Adults

Low Risk-of-bias Studies

Overview of Studies

Two low risk-of-bias cross-sectional studies evaluated the association between fluoride exposure and cognitive effect in adults (Jacqmin et al. 1994; Li et al. 2016). These two studies used the same test for cognitive function (i.e., Mini-Mental State or MMS Examination) and used drinking water fluoride levels to assess fluoride exposure. Li et al. (2016) also measured urinary fluoride. Both studies were cross-sectional in design. One was conducted in France (Jacqmin et al. 1994) and the other in China (Li et al. 2016). Both studies were conducted in older populations (i.e., over 60 or 65 years of age).

Table 8 provides a summary of study characteristics and key findings related to fluoride exposure and cognitive effects in adults for the two low risk-of-bias studies. The purpose of the table is to summarize key findings (independent of whether an association was found) from each study and is not meant to be a comprehensive summary of all results. For each study, results are summarized for each exposure measure assessed. Results from multiple analyses using the same exposure measure may not all be presented unless conflicting results were reported.

Commented [A87]: This sentence reflects revisions to include information on counts of studies per study design, in response to the [redacted] Reviewer comment below; see DocJ_Monograph for detailed response:

Reviewer comment (DocJ_Monograph, page 12): In general, it would be helpful to the reader to describe the included study designs earlier in the respective results sections (i.e., along with the total #s of included articles, describing the # of cross-sectional, prospective cohort, etc is important).

Table 8. Studies on Cognitive Function in Adults[a]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
Jacqmin et al. (1994)	Cross-sectional France (Gironde and Dordogne)/elderly adults [3,490]	Drinking water Range: 0.03–2.03 mg	Adults (ages ≥65 years)	Cognitive function: MMS Examination	No significant increase in the prevalence of cognitive impairment with increasing fluoride quartiles No statistical adjustment for covariates for prevalence rates
Li et al. (2016)	Cross-sectional China (Inner Mongolia)/adults [511]	Drinking water daily fluoride intake Mean (SD): 2.23 (2.23) (normal group), 3.62 (6.71) (cognitive impairment group) mg Urine Mean (SD): 1.46 (1.04) (normal group), 2.47 (2.88) (cognitive impairment group) mg/L Fluorosis score Mean (SD): 0.74 (0.98) (normal group), 1.29 (1.01) (cognitive impairment group)	Adults (ages ≥60 years)	Cognitive function: MMS Examination	Subjects with cognitive impairment had a significantly higher skeletal fluorosis score and urinary fluoride concentrations; odds of increasing severity of cognitive impairment increased with urinary fluoride concentrations but were not statistically significant; no significant association with total daily water fluoride intake Adjusted for sex, age, education, marital status (married vs. not married), alcohol consumption (non-drinkers, light drinkers, moderate to heavy drinkers), smoking history (never smoker, ex-smoker, light smoker, heavy smoker), and serum homocysteine levels

GM = geometric mean; MMS = Mini-Mental State.
[a]Includes low risk-of-bias studies.
[b]Associations between cognitive effects in adults and fluoride levels were reported quantitatively, when possible. For studies with multiple analyses and results, the table summarizes key findings and is not a comprehensive summary of all findings. Results also indicate when a study reported no association, provided as a qualitative statement of no association.

Summary of Results

Results from two low risk-of-bias studies in adults did not provide enough evidence to evaluate consistency when assessing evidence for a potential association between fluoride exposure and cognitive impairment (based on the MMS Examination) (Jacqmin et al. 1994; Li et al. 2016). Jacqmin et al. (1994) did not find an association between drinking water fluoride and cognitive impairment in populations in France (n = 3,490) and found prevalence rates of cognitive impairment to be the same regardless of fluoride exposure (see Figure A-13). In contrast, Li et al. (2016) did find significantly higher urinary fluoride levels and skeletal fluorosis scores in the cognitively impaired group compared with the control group in an analysis of 38 cognitively impaired cases and 38 controls matched for several covariates, including age, sex, education, alcohol consumption, and smoking (p-value <0.05). However, the authors found no significant association between cognitive impairment and total daily water fluoride intake (adjusted ORs per 1-mg/day increase = 0.94 [95% CI: 0.85, 1.04] and 0.86 [95% CI: 0.69, 1.06] in the moderate and severe cognitive impairment groups, respectively) or urinary fluoride levels (adjusted ORs per 1-mgL increase = 1.12 [95% CI: 0.89, 1.42] and 1.25 [95% CI: 0.87, 1.81] in the moderate and severe cognitive impairment groups, respectively) in subjects from fluorosis-endemic areas of China (n = 511).

Commented [A88]: This sentence reflects revisions to more accurately make a statement regarding the consistency in results across the two studies, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 22): The analysis between the two studies may be too simplistic. The French study was done in adults with not very high exposures to fluoride. In contrast the Chinese study compared adults with skeletal fluorosis (suggesting very high exposure) with others. It may be inaccurate to suggest that these two studies were not consistent. They may (consistently) show that relatively low exposures (even if above recommended) are not associated with cognitive outcomes, but very high exposures are.
This gets at [redacted] comments before about a lack of analysis regarding doses, dose effects, or thresholds.

High Risk-of-bias Studies

The results from five out of eight high risk-of-bias studies provide evidence of cognitive impairment in adults associated with higher levels of exposure to fluoride; however, there was heterogeneity in the outcomes assessed, a limited number of directly comparable studies, and some variability in results (e.g., variation in IQ results across studies). Due to the limited number of low risk-of-bias studies identified that assess cognitive impairment in adults, the results from the high risk-of-bias studies are summarized in greater detail below than had been done in this document for bodies of evidence for IQ in children and other neurodevelopmental and cognitive effects in children.

Commented [A89]: Change made in response to the [redacted] Reviewer comment below; see DocF_Monograph for detailed response to related reviewer comment:

Reviewer comment (DocF_Monograph, page 9): High levels of

In aluminum factory workers (exposed to gaseous and particular fluoride emissions during the production of aluminum metal), significant decreases in IQ (Duan et al. 1995), diminished performance on several neurobehavioral core battery tests (NCTBs) (Guo et al. 2001 [translated in Guo et al. 2008b]), and impaired psychomotor performance and memory were observed (Yazdi et al. 2011). One study conducted on adult subjects with fluorosis (dental and skeletal) from a fluorosis-endemic area compared with healthy subjects from a non-endemic area observed significant differences for some cognitive function tests (i.e., tests of speech fluency, recognition, and working memory) but not others and generally did not observe a significant change in IQ except in the operation scores (Shao 2003). One prospective cohort study evaluated exposure to fluoride in children at 5 years of age, based on whether the children resided in areas with community water fluoridation or used fluoride toothpaste or fluoride tablets, and found no clear differences in IQ scores of the subjects at 38 years of age (Broadbent et al. 2015). One additional study suggested that populations living in areas with higher drinking water fluoride had lower levels of dementia (Still and Kelley 1980); however, the study was not focused on effects of fluoride but on whether fluoride was able to reduce the risk associated with aluminum by competing with aluminum and reducing its bioavailability. Therefore, the study was considered inadequate to evaluate the association between fluoride and dementia (Still and Kelley 1980). A more recent study in Scotland evaluated dementia rates associated with aluminum and fluoride drinking water concentrations and observed a significant increased risk of

dementia per standard deviation increase in fluoride (p-value <0.001) with the risk of dementia more than double in the highest quartile of fluoride exposure (56.3 µg/L) compared to the lowest quartile (<44.4 µg/L). The authors also found a significantly increased risk of dementia associated with increased aluminum levels at all quartiles compared with the reference group (p-values <0.05) but found no statistical interaction between aluminum and fluoride levels in relation to dementia (Russ et al. 2019). Conversely, a study in China did not find a significant association between fluoride concentrations in the drinking water and risk for dementia (Liang et al. 2003). In addition to studies that reported on cognitive impairment and exposure to fluoride, two high risk-of-bias studies were identified that reported impaired motor and sensory function (Rotton et al. 1982) and a higher prevalence of self-reported headaches, insomnia, and lethargy (Sharma et al. 2009) associated with fluoride exposure.

Risk of Bias for Cognitive Effect Studies in Adults

Due to the small number of studies with a low potential for bias (see Figure D-13 and Figure D-14), the key risk-of-bias domains (confounding, exposure characterization, outcome assessment) are not discussed separately in respective subsections, as was done for the IQ in Children and Other Neurodevelopmental and Cognitive Effects in Children bodies of evidence. The high risk-of-bias studies had concerns across several domains (see Figure D-15 and Figure D-16), but there were still relatively few studies. Therefore, the discussion for high risk-of-bias studies is also not separated into subsections by key domain.

Low Risk-of-bias Studies

Both low risk-of-bias studies on cognitive effects in adults had little or no risk-of-bias concerns. One study was rated definitely low or probably low risk of bias for all risk-of-bias questions (Li et al. 2016), and the other study was rated probably high risk of bias for a single question that was judged to have minimal impact on overall potential bias (Jacqmin et al. 1994). Jacqmin et al. (1994) had potential concern for bias due to confounding because smoking was not addressed, which has the potential to impact risk for Alzheimer's disease and rates could vary by parish (the target population consisted of men and women from 75 civil parishes in southwestern France).

High Risk-of-bias Studies

There were several issues in the eight studies in adults considered to have high potential for bias. Four of the eight studies had potential concern for bias due to lack of information on the comparison groups, or the comparison groups were considered inappropriate. All eight studies had potential concern for bias regarding covariates not being addressed, including possible co-exposures in occupational studies (e.g., aluminum) and smoking. Five of the eight studies had potential concern for bias due to lack of information regarding exposure characterization or poor exposure characterization with the most utilized exposure measure in these studies being a comparison between exposed and unexposed areas. In one case (Broadbent et al. 2015), multiple sources of fluoride exposure were assessed separately without properly controlling for the other sources of exposure, which could bias the results (see Exposure Characterization in IQ Studies for further details). Five studies also had potential for bias based on limitations in the outcome assessment, which was mainly due to lack of blinding of outcome assessors, lack of validation of the methods, or lack of sufficient details on how the outcomes were assessed.

Confidence Assessment of Findings on Cognitive Effects in Adults

The body of evidence available to examine the association between exposure to fluoride and cognitive effects in adults is limited to two low risk-of-bias cross-sectional studies. Due to the limited number of studies and a lack of evidence of an effect, there is low confidence based on this body of evidence that fluoride exposure is associated with cognitive effects in adults.

Mechanistic Data in Humans

Eight low risk-of-bias studies that evaluated fluoride exposure and mechanistic data in humans were considered potentially relevant to neurological effects. Effects on the thyroid were specifically evaluated because the NRC 2006 report identified this as a possible effect of fluoride (NRC 2006), and changes in thyroid hormones have been identified as a mechanism for neurodevelopmental effects (Haschek and Rousseaux 1991). These included effects on thyroid hormones in children (Kheradpisheh et al. 2018a; Kheradpisheh et al. 2018b; Malin et al. 2018), adults (Kheradpisheh et al. 2018a; Kheradpisheh et al. 2018b; Malin et al. 2018), or children and adults combined (Barberio et al. 2017a). In addition, some studies evaluated self-reported thyroid conditions in children and adults combined (Barberio et al. 2017a) and thyroid diseases in adults (Kheradpisheh et al. 2018b; Peckham et al. 2015) (see Figure D-17 and Figure D-18). Although the low risk-of-bias studies provide some evidence of mechanistic effects (primarily changes in thyroid stimulating hormone [TSH] levels in children), the studies were too heterogeneous or limited in number to make any determination on mechanism (see Figure 7).

Among the seven low risk-of-bias studies that reported on changes in thyroid hormones, three studies were conducted in children (Kumar et al. 2018; Singh et al. 2014; Zhang et al. 2015b) and reported increases in TSH levels. Zhang et al. (2015b) reported significant increases in TSH in children from a fluorosis-endemic area (median fluoride drinking water concentration = 1.40 mg/L; interquartile range = 1.23–1.57 mg/L) compared with a non-fluorosis-endemic area (median fluoride drinking water concentration = 0.63 mg/L; interquartile range = 0.58–0.68 mg/L), whereas 3,5,3'-triiodothyronine (T_3) or thyroxine (T_4) were not significantly different between the two groups. Similarly, Singh et al. (2014) observed significantly higher TSH levels in children without dental fluorosis who lived in a fluorosis-endemic area (fluoride drinking water concentrations of 1.6–5.5 mg/L) compared with children without dental fluorosis who lived in a non-fluorosis-endemic area (fluoride drinking water concentrations of 0.98–1.00 mg/L). When all children (with and without dental fluorosis) in the endemic area were compared with children from the non-endemic area, the TSH levels were higher in children from the fluorosis-endemic area, although results did not reach statistical significance ($p = 0.057$). Significant differences in T_4 or T_3 were not observed between groups (Singh et al. 2014). Kumar et al. (2018) also observed a significant increase in TSH levels in children from a fluorosis-endemic area (1.5–5.8 mg/L fluoride) compared with a control area (0.94–1.08 mg/L fluoride). There were also decreases in T_3 and T_4, but results were not statistically significant.

Barberio et al. (2017a) evaluated associations between fluoride and TSH levels in children and adults combined and found no relationship between fluoride exposure (measures in urine and tap water) and TSH levels. In the one study that evaluated thyroid hormone levels in adults but not children, Kheradpisheh et al. (2018b) found a significant increase in TSH associated with higher fluoride concentrations in drinking water in both adults with and without thyroid diseases such as hypothyroidism, hyperthyroidism, thyroid nodules, or thyroid cancer. Significant increases in T_3

were associated with higher fluoride in drinking water in adults without thyroid diseases, but increases in T_3 were not significant in adults with thyroid diseases. A significant association between T_4 and higher fluoride in drinking water was not observed in adults with or without thyroid diseases (Kheradpisheh et al. 2018b).

Other than changes in hormone levels, there is limited evidence of fluoride-related mechanistic effects in the three low risk-of-bias studies that evaluated thyroid-related effects. Barberio et al. (2017a) found no relationship between fluoride exposure and self-reported thyroid conditions in children and adults (children were older than 12). Kheradpisheh et al. (2018b) also found no association between fluoride exposure and hypothyroidism in an adult population in Iran. One study found a significantly higher prevalence of hypothyroidism in areas with higher fluoride concentrations in drinking water (>0.7 mg/L) compared with areas with lower fluoride drinking water concentrations (≤0.7 mg/L) (Peckham et al. 2015).

Sixteen high risk-of-bias studies were available that evaluated mechanistic data in humans associated with fluoride exposure, including effects on thyroid hormones in children (n = 9 studies), thyroid hormones in adults (Michael et al. 1996; Yasmin et al. 2013), catecholamines in adults (Michael et al. 1996) or in subjects of unknown ages (Chinoy and Narayana 1992), acetylcholinesterase (AChE) or serotonin levels in children (Lu et al. 2019; Pratap et al. 2013), brain histopathology or biochemistry in aborted fetuses (Du et al. 1992 [translated in Du et al. 2008]; Yu et al. 1996 [translated in Yu et al. 2008]), and mitochondrial fission/fusion molecules in children (Zhao et al. 2019). Similar to the low risk-of-bias studies, the high risk-of-bias studies provide some evidence of mechanistic effects (primarily changes in TSH levels in children); however, the data are insufficient to identify a clear mechanism by which fluoride causes neurodevelopmental or cognitive effects in humans.

Among high risk-of-bias studies (see Figure D-19 and Figure D-20), varying results were reported in 11 studies that evaluated associations between fluoride exposure and thyroid hormones, and a few of these studies (Lin et al. 1991; Wang et al. 2001; Yang et al. 1994 [translated in Yang et al. 2008]) were complicated by high or low iodine in the high fluoride area. When considering fluoride effects on each of the hormones individually, similar to results from low risk-of-bias studies, the most consistent evidence of fluoride-associated effects on a thyroid hormone was reported as changes in TSH levels in children, although there was some variation in the direction of association. Six of the nine high risk-of-bias studies that evaluated changes in TSH levels in children reported increases in TSH levels with higher fluoride (Lin et al. 1991; Susheela et al. 2005; Wang et al. 2001; Yang et al. 1994 [translated in Yang et al. 2008]; Yao et al. 1996; Yasmin et al. 2013). Two of the nine high risk-of-bias studies reported decreases in TSH levels in children with higher fluoride (Khandare et al. 2017; Khandare et al. 2018). One of the nine studies found no significant alterations in TSH levels in children from fluorosis-endemic areas (Hosur et al. 2012) (see Figure 8).

When considering associations between fluoride exposure and TSH, T_3, and T_4 levels together, studies that evaluated changes in all three thyroid hormones reported varying combinations of increases, decreases, or no changes in levels across the three hormones, although among the eight low and high risk-of-bias studies that evaluated associations between fluoride exposure and TSH, T_3, and T_4 levels and reported increases in TSH levels in children, seven of the eight studies found no alterations in T_3 levels (one study found an increase in T_3), and six of the eight studies found no alterations in T_4 levels (two studies found an increase in T_4). Studies also

Commented [A90]: Editorial change made during formatting

displayed variation by age in the associations between fluoride and TSH, T_3, and T_4. Due to the dynamic relationship between the thyroid gland, the pituitary gland, and the production and clearance of TSH, T_3, and T_4, the variations in results are not unexpected and do not eliminate the possibility of a mechanistic link between thyroid effects and neurodevelopmental or cognitive effects; however, the data do not support a clear indication that thyroid effects are a mechanism by which fluoride causes these effects in humans.

Endpoint	Direction of Effect ↑	NS	Grand Total
serum T3	1	3	**4**
serum T4		4	**4**
serum TSH	5	2	**7**
self-reported thyroid condition		1	**1**
thyroid disease		1	**1**
hypothyroidism prevalence	1		**1**

Figure 7. Number of Low Risk-of-bias Studies that Evaluated Thyroid Hormones in Children and Adults by Endpoint and Direction of Association

Interactive figure and additional study details are available at https://public.tableau.com/app/profile/ntp.visuals/viz/FluorideTableauDashboards/ReadMe. This figure displays study counts for low risk-of-bias studies in both children and adults, as these counts are most relevant to the summary of fluoride-related mechanistic effects in low risk-of-bias studies. Counts for high risk-of bias studies and studies by age (i.e., children, adults, or children/adults combined) can also be accessed in the interactive figure. Study counts are tabulated by significance (unless study footnotes in the interactive figure indicate that statistical significance was not tested)—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with significantly increased results.

Endpoint	Direction of Effect ↑	↓	NS	Grand Total
serum T3	1	1	6	**8**
serum T4	3		5	**8**
serum TSH	6	2	1	**9**

Figure 8. Number of High Risk-of-bias Studies that Evaluated Thyroid Hormones in Children by Endpoint and Direction of Association

Interactive figure and additional study details are available at https://public.tableau.com/app/profile/ntp.visuals/viz/FluorideTableauDashboards/ReadMe. This figure displays study counts for high risk-of-bias studies in children, as these counts are most relevant to the summary of associations between fluoride and thyroid hormones in high risk-of-bias studies. Counts for low risk-of bias studies, studies in adults, or all studies combined can also be accessed in the interactive figure. Study counts are tabulated by significance (unless study footnotes in the interactive figure indicate that statistical significance was not tested)—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with significantly increased results.

In addition to evaluating thyroid hormone levels, a few high risk-of-bias studies evaluated other mechanistic data associated with fluoride exposure; however, the data are insufficient to identify a clear mechanism by which fluoride might cause neurodevelopmental or cognitive effects in humans. Serum epinephrine and norepinephrine were significantly increased in a fluoride-

endemic region (it was not reported whether subjects were children or adults) compared with a non-endemic region (Chinoy and Narayana 1992). A separate study reported that ~~S~~serum epinephrine and norepinephrine (referred to as adrenaline and noradrenaline in the study) were significantly increased in adults in a fluoride-endemic area (fluoride in the drinking water ranged from 1.0–6.53 ppm) compared with a control area (fluoride in the drinking water ranged from 0.56–0.72 ppm) (Michael et al. 1996). Serum AChE was significantly reduced in children from a high fluoride region compared with a lower fluoride region (Pratap et al. 2013). Serum serotonin was significantly increased in children from Turkey who were drinking water containing 2.5 mg/L of fluoride compared with children drinking bottled water or water containing <0.5 mg/L of fluoride (Lu et al. 2019). Aborted fetuses from high fluoride areas in China were found to have histological changes in the brain and significant changes in neurotransmitter levels compared with a control area (Du et al. 1992 [translated in Du et al. 2008]; Yu et al. 1996 [translated in Yu et al. 2008]).

Commented [A91]: Change made in response to the [redacted] comment below; see DocC_Monograph for detailed response:

Reviewer comment (DocC_Monograph, page 19): For the non-health professional reader, the use of different nomenclature for the same neurotransmitter is confusing.

There are also two more recent low risk-of-bias studies that evaluated polymorphisms in dopamine-related genes; however, a determination on mechanism cannot be made at this time due to the limited number of studies. For children (10–12 years old) with a Val158Met polymorphism in the *COMT* gene (i.e., catechol-O-methyltransferase), which results in slower degradation and greater availability of dopamine within the brain, a stronger association between increasing urinary fluoride levels and decreasing IQ was reported (Zhang et al. 2015b). For children (7–12 years old) with a dopamine receptor-2 (DRD2) Taq 1A polymorphism (which is involved in reduced D2 receptor density and availability) and the TT (variant) genotype, a significant inverse association between log urinary fluoride and IQ was observed; however, this significant relationship was not observed in children with the CC (wild-type) or CT (hybrid) genotypes (Cui et al. 2018).

Animal Learning and Memory Data

NTP provided a review of the experimental animal evidence in the earlier draft monographs (NTP 2020) and agrees with the NASEM committee's comments (NASEM 2020; 2021) that the experimental animal database is of poor quality, with many studies suffering from major reporting deficiencies (*Sup01_Monograph_NASEM_Feb_2021.pdf*). NTP acknowledges that further efforts to disentangle the potential for motor activity deficits to influence tests of learning and memory in the fluoride literature are warranted. Overall, these general issues and deficiencies with the experimental animal database led to NTP's conclusion that the animal studies are currently *inadequate* to inform the question of an association between fluoride exposure and neurodevelopmental and cognitive effects in humans. Therefore, this systematic review does not include an experimental animal section.

Mechanistic Data in Animals

There are a wide variety of studies in animals that evaluate mechanistic effects potentially related to neurological changes following oral fluoride exposure (see Appendix F); however, the mechanisms underlying fluoride-associated cognitive neurodevelopmental effects are not well characterized, and review of the data did not identify a mode of action for fluoride effects on IQ in children. Categories of mechanistic endpoints with the largest amount of available data include changes in biochemical components of the brain or neurons, neurotransmitters, oxidative stress, histopathology, and thyroid function. Limiting the data to studies with at least one

exposure at or below 20 ppm fluoride drinking water equivalents (gavage and dietary exposures were backcalculated into equivalent drinking water concentrations for comparison) still provided a sufficient number of studies for evaluation of these mechanistic endpoints. This evaluation is provided in Appendix F. Neurotransmitter and biochemical changes in the brain and neurons were considered the mechanistic areas with the greatest potential to demonstrate effects of fluoride on the brain of animals in the lower dose range and provide evidence of changes in the brain that may relate to lower IQ in children (see Appendix F). Histological data can be useful in determining whether effects are occurring in the brain at lower fluoride concentrations; however, author descriptions of these effects may be limited, thereby making it difficult to directly link histological changes in the brain to learning and memory effects. Oxidative stress is considered a general mechanistic endpoint that cannot be specifically linked to neurodevelopmental or cognitive effects in humans; however, like histopathology, it may help in identifying changes in the brain occurring at lower concentrations of fluoride. Although any effects in the brain or neurological tissue at lower concentrations of fluoride may support reduced IQ in humans, it may be difficult to distinguish the potential effects of fluoride on learning and memory functions from other neurological or general health outcomes.

In Vitro Data on Neurodevelopmental or Cognitive Effects

Although in vitro studies were identified as part of the systematic review process, NTP determined that the information on neurological effects from these studies is too general, and results cannot necessarily be attributed to effects on learning and memory or other cognitive functions at this time. The in vitro data may help support specific mechanisms identified from in vivo mechanistic data; however, as described above, no specific mechanism has been determined for fluoride effects on learning and memory or other neurodevelopmental or cognitive outcomes.

Discussion

This systematic review evaluated the available animal and human literature concerning the association between fluoride exposure and cognitive neurodevelopment. The available data on potential mechanisms to evaluate biological plausibility were also assessed. The ~~potential~~ health benefits of fluoride with respect to oral health are acknowledged but are not the focus of this review.

This review extended NTP's previous evaluation of the experimental animal data (NTP 2016). Although the animal data provide some evidence of effects of fluoride on neurodevelopment, they give little insight into the question of whether fluoride influences IQ. This is due to deficiencies identified in the animal body of evidence. Mechanistic studies in humans provide some evidence of adverse neurological effects of fluoride. However, these studies were too heterogenous and limited in number to make any determination on biological plausibility.

The literature on adults is also limited; therefore, it was determined that there is low confidence in the body of evidence from studies that evaluate fluoride exposure and adult cognition. Compared to the literature in adults, there is a much more extensive literature in children.

The literature in children was separated into studies assessing IQ and studies assessing other cognitive or neurodevelopmental outcomes. There is low confidence in the body of evidence from studies that evaluate fluoride exposure and other cognitive or neurodevelopmental outcomes in children. This body of evidence is made up of~~Altogether, the results from eight of~~ nine high-quality studies (three prospective cohort and ~~five~~ six cross-sectional studies from seven different study populations) and six low-quality studies. Eight of the nine high-quality studies provide some evidence that fluoride is associated with other cognitive or neurodevelopmental outcomes in children. The data also suggest that neurodevelopmental effects occur in very young children. However, the confidence in this body of evidence is low because the number of studies is limited, and there is too much heterogeneity in the outcomes measured, ages assessed, and methods used~~,~~ to directly compare studies of any one outcome. Additional studies on outcomes such as attention-deficit hyperactivity disorder (ADHD) and other attention-related disorders, where there is some evidence of an effect of fluoride exposure, would be necessary to critically assess the data.

Most of the epidemiological studies (n = 72) assessed the association between fluoride exposure and IQ in children. Although all studies, both high- and low-quality, were considered, this evaluation focuses on the high-quality, low risk-of-bias studies in children for two reasons. First, there are fewer limitations and greater confidence in the results of the high-quality studies. Second, there is a relatively large number of high-quality studies (n = 19), such that the body of evidence from these studies could be used to evaluate confidence in the association between fluoride exposure and changes in children's IQ.

This review finds, with moderate confidence, that higher fluoride exposure is associated with lower IQ in children. The association between higher fluoride exposure and lower IQ in children was consistent across different study populations, study locations, study quality/risk-of-bias determinations, study designs, exposure measures, and types of exposure data (group-level and individual-level). There were 19 low risk-of-bias studies that were conducted in 15 study populations, across 5 countries, and evaluating more than 7,000 children. Of these 19 studies, 18

Commented [A92]: Change made in response to the [redacted] Reviewer comment below; see DocF_Monograph for detailed response:

Reviewer comment (DocF_Monograph, page 2): Please confirm with [redacted] for references, but [redacted] believe there is a significant amount of data out there to suggest that it's more than just "potential health benefits." Recommend expanding upon this a bit more to describe some of the health benefits that have been shown.

Commented [A93]: Change made in response to the [redacted] Reviewer comment below; see DocC_Monograph for detailed response:

Reviewer comment (DocC_Monograph, page 20): This appears contrary to the preceeding sentence and is not valuable to present description of the details of low-confidence studies.

Commented [A94]: Change made in response to the [redacted] Reviewer comment below; see DocF_Monograph for detailed response to several related reviewer comments:

Reviewer comment (DocF_Monograph, pages 5 - 11) : High?

reported an association between higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximated or exceeded the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] and lower IQ. These include 3 prospective cohort studies and 15 cross-sectional studies (12 of which indicated that exposure likely preceded the outcome). Forty-six of 53 low-quality studies in children also reported an association between higher fluoride exposure and lower IQ.

Many studies in this assessment relied on drinking-water fluoride levels (both group-level measures and individual-level measures), rather than measures of total fluoride exposure, to establish exposed versus "unexposed" or reference groups. Although fluoride in water is a major source of exposure [comprising 40% to 70% of total exposure (US EPA 2010)], other sources of fluoride provide variable amounts that depend on personal preferences and habits. The use of dental products containing fluoride and consuming foods and beverages prepared with fluoridated water can also result in measurable exposures (US EPA 2010). Green et al. (2019) suggested that significant exposures occur from black tea consumption. Thus, drinking water fluoride levels may, but usually do not, reflect total fluoride exposure. This could be a potential limitation in studies that rely on water fluoride data to assess fluoride exposure (in particular, earlier studies). However, because water is only part of a person's total exposure to fluoride, this limitation would likely result in an underestimate of exposure to fluoride. In addition, this limitation is less of a concern in areas where fluoride in the drinking water is high because drinking water likely contributes a large proportion of the total fluoride intake in those areas as compared with areas where fluoride in the drinking water is lower.

This review found that the quality of exposure assessment has improved over the years. More recent studies by Valdez Jimenez et al. (2017), Bashash et al. (2017), and Green et al. (2019) used individual measures of urinary fluoride, either maternal urine collected prenatally or children's urine, which ~~confirmed~~ support the association between higher total fluoride exposure and lower children's IQ and other cognitive neurodevelopmental effects. Studies using different types of exposure measures reported similar findings of an association (NRC 2006), which strengthens confidence in earlier studies that reported IQ deficits with ~~increasing~~ high group-level fluoride exposure. However, there is less certainty in the quantitative estimates of the magnitude of IQ deficits from earlier studies that used group-level exposure measures than the estimates from more recent studies that used individual-level exposure measures.

It is worth noting that there are circumstances wherein typical children's water consumption considered with water fluoride levels may substantially underestimate total fluoride exposure. One example is bottle-fed infants wherein nutrition is provided by powdered formula that is rehydrated with fluoridated water (Till et al. 2020). To decrease an exclusively formula-fed infant's exposure to fluoride, for the purpose of reducing risk of dental fluorosis, the Centers for Disease Control and Prevention recommends using low-fluoride bottled water to mix with infant formula (CDC 2015). A few studies also support the possibility of heightened sensitivities to the detrimental cognitive effects of fluoride exposure in individuals with certain genetic polymorphisms in dopamine receptor D2 or catechol-O-methyltransferase (Cui et al. 2018; Zhang et al. 2015b), potentially impacting dopamine catabolism and receptor sensitivity. Differential exposures to fluoride and genetic susceptibilities of children to fluoride may represent special situations that would appear to warrant further research.

Commented [A95]: Change made in response to the [redacted] Reviewer comment below; see DocC_Monograph for detailed response:

Review comment (DocC_Monograph, page22): The association is not causal and therefore, suggest changing the word [confirmed] to "support"... the association.

Commented [A96]: Change made in response to the [redacted] Reviewer comment below; see DocC_Monograph for detailed response:

Reviewer comment (DocC_Monograph, page 22): This statement appears to imply that there are no studies with negative findings and the sentence asserts a dose-response relationship for which data are incomplete.

The following section briefly recaps the strength of the epidemiological evidence for an association between fluoride exposure and cognitive neurodevelopmental deficits. This is followed by a more detailed listing of limitations of the evidence base and limitations of the systematic review, with some suggestions of areas where further research may be most beneficial.

Limitations of the Evidence Base

Limitations in the epidemiological studies with low risk of bias include:

- Few studies are available that assessed the association between fluoride exposure and cognitive function (particularly IQ) in adults and attention-related disorders including ADHD in children and adults.
- Heterogeneity in outcomes was assessed for other neurobehavioral outcomes, limiting the assessment of other possible effects in children.
- Studies rarely separated the results by sex or provided information to indicate that sex was not a modifying factor.
- Associations between lower total fluoride exposure [e.g., represented by populations whose total fluoride exposure was lower than the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] and children's IQ remain unclear. More studies at lower exposure levels are needed to fully understand potential associations in ranges typically found in the United States (i.e., <1.5 mg/L in water). However, it should be noted that, as of April 2020, CWS supplying water with ≥1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population (~1.9 million people) (CDC Division of Oral Health 2020). This indicates that the moderate confidence in the association between higher fluoride exposure and lower IQ is relevant, at a minimum to children living in these areas of the United States where fluoride in drinking water is known to be at or above 1.5 mg/L. This is only compounded by additional exposures to fluoride from other sources.
- No studies investigating the association between fluoride exposure and neurodevelopmental or cognitive effects in adults or children have been conducted in the United States.
- No studies are available to evaluate lifelong exposure in adults, or fluoride exposure over a child's lifetime and neurodevelopmental or cognitive changes over time.
- The database does not allow for comparison of ages and possible changes at different developmental stages in children to assess if there is a delay in development or if associations persist.
- The database does not allow for establishing clear correlations between prenatal and postnatal exposures.

Limitations in the epidemiological studies with high risk of bias include:

- Many of the original publications were in a non-English language and provided limited details on methodology.

Commented [A97]: Change made in response to the [redacted] Reviewer comment below; see DocF_Monograph for detailed response:

Reviewer comment (DocF_Monograph, page 16): Elaborate a little further.

Commented [A98]: The next three limitations (starting with this one) were added in response to [redacted] comments from the [redacted] Reviewer listed below; see DocH_Monograph for detailed response. *Note a* [redacted] *Reviewer also suggested an issue that was addressed in one of these limitations and is described in the subsequent comment (see DocJ_Monograph for detailed response)*:

Reviewer comment (DocH_Monograph, page 7): Cumulative exposure: The authors should make clear that exposure during gestation likely implies that there is continuing exposure in the post natal period. Further, these two exposure periods are likely highly correlated, making conclusions regarding a critical period of exposure difficult. The converse is also true – i.e. if exposure is measured in the post natal or childhood period, and especially if it is from drinking water, then there was likely exposure in the prenatal period as well.

Reviewer comment (DocH_Monograph, page 14): The results also need to be interpreted based on age of test administration. Some higher order functions do not develop until later ages and thus cannot be tested well in younger children. Also, as with other neurotoxins, deficits can occur at a variety of ages, and either persist or not. So the age at assessment becomes an important variable in the interpretation of findings and should be accounted for in the discussion.

Commented [A99]: This limitation was added in response to the [redacted] Reviewer comment below; see DocJ_Monograph for detailed response:

Reviewer comment (DocJ_Monograph, page 22): Yes, the approach used to assess risk of bias was clearly described and generally appropriate.
Though, it is unclear whether these studies adequately captured a critical fluoride exposure window likely to impact neurocognitive health (i.e., does fluoride exposure in older adulthood impact neurocognitive health?) For example, lifelong fluoride exposure, and/or fluoride exposure at different lifestages that may be more critical to neurocognitive development, were not captured in these cross-sectional studies. Thus, it raises questions as to whether these cross-sectional studies are truly "low risk of bias," or are "lower" risk of bias than others.

- Studies lacked information regarding exposure and/or had serious limitations in the exposure assessment. Exposure assessment concerns include limited individual exposure information, a lack of information on fluoride sampling methods and timing of the exposure measurements, a lack of quantitation of levels of fluoride in drinking water in a few studies, and a lack of individual-level information on fluorosis in areas reported to be endemic for fluorosis.
- The comparison groups in studies conducted in areas endemic for fluorosis still may have been exposed to high levels of fluoride or levels similar to those used in water fluoridation in the United States. This factor may have limited the ability to detect true effects.
- Studies did not provide sufficient direct information (e.g., participation rates or methods for selection) to evaluate selection bias.
- Failure to address important covariates was an issue for many studies. Some studies conducted simple statistical analyses without accounting for any covariates in the analysis, although many noted similarities between the study populations. In cases where adjustments in analyses were made, often these studies did not account for covariates considered critical for that study population and outcome including co-exposures.
- Studies conducted in areas with high, naturally occurring fluoride levels in drinking water often did not account for potential exposures to arsenic or iodine deficiencies in study subjects in areas where these substances were likely to occur.
- Studies lacked information on whether the outcome assessors were blind to the exposure group, including studies that examined children in their schools and subjects from high-fluoride communities.

Limitations in the animal and mechanistic evidence base include:

- The overall quality of the experimental animal studies is poor, and there are relatively few well-designed and well-performed studies at lower fluoride exposure levels (i.e., <20 ppm, which is roughly equivalent to human exposure of <4 ppm).
- The understanding of the specific molecular events responsible for fluoride's adverse effects on neurobehavioral function is poor.

A key data gap in the human and animal bodies of evidence includes the need for mechanistic insight into fluoride-related neurodevelopmental or cognitive changes.

Strengths of the Evidence Base

Strengths in the epidemiological evidence base include:

- There are 72 studies directly addressing the relationship between fluoride exposure and children's IQ.
- There are 12 high-quality cross-sectional studies with low risk of bias providing evidence that exposure occurred prior to outcome assessment in those studies.
- Studies are from diverse geographic locations that included data for more than 7,000 children.

- There are 19 high-quality studies evaluating the same outcome (i.e., IQ) and 9 evaluating other neurodevelopmental outcomes.
- Reported ~~responses~~ associations between ~~to~~higher fluoride exposure and lower children's IQ are consistent in the vast majority of studies of both low and high quality.
- Reported ~~responses~~ associations between ~~to~~higher fluoride exposure and lower children's IQ are consistent across different study populations, study designs, and exposure measures.
- Findings of studies with group- and individual-level information on exposure and outcomes are similar.
- A wide variety of important covariates are either addressed by study design or captured across the evidence base, with no consistent patterns that would suggest an alternative explanation.

Limitations of the Systematic Review

This systematic review has few limitations. The human body of evidence included a large database of observational studies. Most of the observational studies were cross-sectional; however, 12 of these were considered to provide sufficient evidence that exposure occurred prior to the outcome. In addition, the systematic review covered a wide range of study designs, populations, and measures of fluoride exposure. The systematic review was designed to cover reports on all potential mechanistic data including effects on the thyroid. After review of the studies evaluating thyroid effects, studies that only evaluated goiters and other effects on thyroid size were not considered in this review. This is not considered a limitation because these studies did not include specific information on thyroid hormones that could indicate a mechanism for thyroid involvement in neurodevelopment. In addition, review of the mechanistic data was limited to in vivo studies with at least one concentration below 20 ppm. This is not considered a limitation for the systematic review because the mechanistic body of evidence was used to evaluate biological plausibility for the effects observed in humans; therefore, data were limited to concentrations that would be more reflective of human exposures. The decision to not more closely evaluate the in vitro data is not considered a limitation because there were sufficient in vivo data, and no key events were identified where in vitro data would provide additional insight.

The supplemental literature search for non-English-language studies not indexed in traditional databases supports the comprehensive nature of the literature search strategy for this systematic review. In the absence of guidance on the most complete non-English-language databases that may contain health studies of fluoride, a standard systematic review approach for database selection was followed whereby a set of exemplar documents, called 'seed studies' were used. Databases were selected that identified non-English-language studies of fluoride that we were aware of and were not captured in searches of databases from the main literature search. This informed approach influenced the selection process; however, this is not considered a limitation because it provided an objective measure by which to compare databases. Following the recommendation of the NASEM committee in its review of the September 16, 2020, draft monograph, the experimental animal section has been removed and is not included in this monograph. Although the deficiencies identified in the animal body of evidence support this removal (see Animal Learning and Memory Data for further explanation), NTP acknowledges

Commented [A100]: Change made in response to the [redacted] Reviewer comment below; see DocC_Monograph for detailed response:

Reviewer comment (DocC_Monograph, page 23): Of the 72 studies, there are some with equivocal or contrary results.

Commented [A101]: Change made in response to the [redacted] Reviewer comment below; see DocF_Monograph for detailed response to related reviewer comment:

Reviewer comment (DocF_Monograph, pages 5 - 11) : High?

Commented [A102]: Change made in response to the [redacted] Reviewer comment below; see DocC_Monograph for detailed response:

Reviewer comment (DocC_Monograph, page 23): This statement is contrary to the earlier declarations of significant heterogeneity across studies and thus, high risk of bias.

Commented [A103]: The following four sentences were added in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 4): Potential bias as there was no systematic selection of Chinese databases to be searched. Two databases were selected because they contained studies of which the authors were aware.

Recommendation: As NASEM noted, this introduces potential bias. [redacted] suggests this be added as a limitation.

that the absence of the experimental animal data is a limitation of this systematic review. For the purpose of this review, NTP considers the experimental animal data to be *inadequate* to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans.

Summary

This systematic review evaluated the available animal and human literature concerning the association between fluoride exposure and cognitive neurodevelopment. The available data on potential mechanisms to evaluate biological plausibility were also assessed. Existing animal studies provide little insight into the question of whether fluoride exposure affects IQ. Human mechanistic studies were too heterogenous and limited in number to make any determination on biological plausibility. The body of evidence from studies on adults is also limited and provides low confidence that fluoride exposure is associated with adverse effects on adult cognition. There is, however, a large body of evidence on IQ effects in children. There is also some evidence that higher fluoride exposure is associated with other neurodevelopmental and cognitive effects; although, because of the heterogeneity of the outcomes, there is low confidence in the literature for these other effects. This review finds, with moderate confidence, that higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] is consistently associated with lower IQ in children. More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ.

Commented [A104]: Change made in response to the [redacted] Reviewer comment below; see DocF_Monograph for detailed response to several related reviewer comments:

Reviewer comment (DocF_Monograph, pages 5 - 11) : High

Commented [A105]: The following two sentences reflect revisions in response to related comments from [redacted] Reviewers listed below to focus the emphasis on the data on which we base our confidence statement and to acknowledge the need for further studies at lower exposure levels; see DocG_Monograph and DocH_Monograph, respectively for detailed response:

Reviewer comment (DocG_Monograph, page 18): An extension of a prior comment, on page 48 (Confidence Assessment of Findings on IQ in Children), the review does not provide evidence to support the statement that the high fluoride exposure should be interpreted as "mainly greater than the WHO Drinking Water Quality Guideline [≥ 1.5 mg/L]".

Reviewer comment (DocH_Monograph, page15): In general, the confidence rating in the body of evidence for this outcome is supported. However, several concerns necessitate a refinement of this confidence rating.
In agree with the prior review in that conclusions can only be made above the WHO drinking water limit for fluoride. It seems as though there is a lack of dose response curve estimation for lower levels of exposure, so an inference cannot be made over the entire range of exposure. Indeed, it is this lower dose range that is of interest for the US population.

Commented [A106]: This sentence and similar sentences throughout the monograph were revised to add "higher" as a qualifier when appropriate to describe fluoride exposures and to provide a benchmark (the 1.5 mg F/L WHO Guidelines for Drinking-Water Quality) to aid in describing total exposures above which moderate confidence was determined for children's IQ studies, in response to the [redacted] Reviewer comment below; see DocG_Monograph for detailed response:

Reviewer comment (DocG_Monograph, page 14): Throughout, it is important to talk about *higher* exposure to fluoride, not just exposure to fluoride. Everyone is exposed to fluoride so describing the at-risk group as being exposed to fluoride is meaningless (and confusing).

References

Adedara IA, Abolaji AO, Idris UF, Olabiyi BF, Onibiyo EM, Ojuade TD, Farombi EO. 2017a. Neuroprotective influence of taurine on fluoride-induced biochemical and behavioral deficits in rats. Chem Biol Interact. 261:1-10.

Adedara IA, Olabiyi BF, Ojuade TD, Idris UF, Onibiyo EM, Farombi EO. 2017b. Taurine reverses sodium fluoride-mediated increase in inflammation, caspase-3 activity, and oxidative damage along the brain-pituitary-gonadal axis in male rats. Can J Phys Pharmacol. 95:1019-1029. https://doi.org/10.1139/cjpp-2016-0641

Akinrinade ID, Memudu AE, Ogundele OM. 2015a. Fluoride and aluminium disturb neuronal morphology, transport functions, cholinesterase, lysosomal and cell cycle activities. Pathophysiology. 22:105-115.

Akinrinade ID, Memudu AE, Ogundele OM, Ajetunmobi OI. 2015b. Interplay of glia activation and oxidative stress formation in fluoride and aluminium exposure. Pathophysiology. 22:39-48.

An J, Mei S, Liu A, Fu Y, Wang C. 1992. [Effect of high level of fluoride on children's intelligence]. Chin J Control Endem Dis. 7(2):93-94.

Arbuckle TE, Fraser WD, Fisher M, Davis K, Liang CL, Lupien N, Bastien S, Velez MP, von Dadelszen P, Hemmings DG et al. 2013. Cohort profile: The maternal-infant research on environmental chemicals research platform. Paediatr Perinat Epidemiol. 27(4):415-425. https://doi.org/10.1111/ppe.12061

Baba NA, Raina R, Verma PK, Sultana M. 2014. Alterations in plasma and tissue acetylcholinesterase activity following repeated oral exposure of chlorpyrifos alone and in conjunction with fluoride in Wistar rats. Proc Indian Natl Sci Acad B Biol Sci. 84:969-972.

Banala R, Nagapuri K, Mohd K, Reddy M, Karnati P. 2018. Carica Papaya leaf extract as a neuroprotective agent against behavioral and neurotransmitter changes in brain of the rat treated with sodium fluoride in pre- and post-natal periods. Pharmacogn Mag. 14(55):123-131. https://doi.org/10.4103/pm.pm_378_17

Barberio AM, Hosein FS, Quinonez C, McLaren L. 2017a. Fluoride exposure and indicators of thyroid functioning in the Canadian population: Implications for community water fluoridation. J Epidemiol Community Health. 71:1019-1025. https://doi.org/10.1136/jech-2017-209129

Barberio AM, Quinonez C, Hosein FS, McLaren L. 2017b. Fluoride exposure and reported learning disability diagnosis among Canadian children: Implications for community water fluoridation. Can J Public Health. 108:229-239. https://doi.org/10.17269/cjph.108.5951

Bartos M, Gumilar F, Gallegos CE, Bras C, Dominguez S, Monaco N, Esandi MDC, Bouzat C, Cancela LM, Minetti A. 2018. Alterations in the memory of rat offspring exposed to low levels of fluoride during gestation and lactation: Involvement of the alpha7 nicotinic receptor and oxidative stress. Reprod Toxicol. 81:108-114. https://doi.org/10.1016/j.reprotox.2018.07.078

Bashash M, Thomas D, Hu H, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Ettinger AS, Wright R, Zhang Z et al. 2017. Prenatal fluoride exposure and cognitive outcomes in

children at 4 and 6-12 years of age in Mexico. Environ Health Perspect. 125(9):1-12. https://doi.org/10.1289/ehp655

Bashash M, Marchand M, Hu H, Till C, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Green R, Schnaas L et al. 2018. Prenatal fluoride exposure and attention deficit hyperactivity disorder (ADHD) symptoms in children at 6-12 years of age in Mexico City. Environ Int. 121(Pt 1):658-666. https://doi.org/10.1016/j.envint.2018.09.017

Bhatnagar M, Rao P, Sushma J, Bhatnagar R. 2002. Neurotoxicity of fluoride: Neurodegeneration in hippocampus of female mice. Indian J Exp Biol. 40:546-554.

Bhatnagar M, Sukhwal P, Suhalka P, Jain A, Joshi C, Sharma D. 2011. Effects of fluoride in drinking water on NADPH-diaphorase neurons in the forebrain of mice: A possible mechanism of fluoride neurotoxicity. Fluoride. 44:195-209.

Broadbent JM, Thomson WM, Moffitt TE, Poulton R. 2015. Community water fluoridation and intelligence response. Am J Public Health. 105:3-4.

California Office of Environmental Health Hazard Assessment (OEHHA). 2011. Meeting synopsis and slide presentations: carcinogen identification committee meeting held on October 12, 2011. California Office of Environmental Health Hazard Assessment. http://oehha.ca.gov/prop65/public_meetings/cic101211synop.html. [19 August 2019]

CDC Division of Oral Health. 2020. Personal communication. September 3, 2020.

Centers for Disease Control and Prevention (CDC). 2013. Community water fluoridation: Fluoridation statistics. Atlanta, GA: Centers for Disease Control and Prevention. https://www.cdc.gov/fluoridation/statistics/2012stats.htm. [19 August 2019]

Centers for Disease Control and Prevention (CDC). 2015. Community water fluoridation FAQs: Infant formula Atlanta, GA: Centers for Disease Control and Prevention. https://www.cdc.gov/fluoridation/faqs/infant-formula.html. [22 September 2021]

Chen Y. 2012. Organophosphate-induced brain damage: Mechanisms, neuropsychiatric and neurological consequences, and potential therapeutic strategies. Neurotox. 33:391-400.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 1991. [Research on the intellectual development of children in high fluoride areas]. Chin J Control Endem Dis. 6(Suppl):99-100.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 2008. Research on the intellectual development of children in high fluoride areas. Fluoride. 41:120-124.

Chinoy NJ, Narayana MV. 1992. Studies on fluorosis in Mehsana District of North Gujarat. Proc Zool Soc. 45:157-161.

Choi AL, Sun G, Zhang Y, Grandjean P. 2012. Developmental fluoride neurotoxicity: A systematic review and meta-analysis. Environ Health Perspect. 120:1362-1368.

Choi AL, Zhang Y, Sun G, Bellinger DC, Wang K, Yang XJ, Li JS, Zheng Q, Fu Y, Grandjean P. 2015. Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study. Neurotoxicol Teratol. 47:96-101.

Chouhan S, Flora SJS. 2008. Effects of fluoride on the tissue oxidative stress and apoptosis in rats: Biochemical assays supported by IR spectroscopy data. Toxicology. 254:61-67.

Chouhan S, Lomash V, Flora SJ. 2010. Fluoride-induced changes in haem biosynthesis pathway, neurological variables and tissue histopathology of rats. J Appl Toxicol. 30:63-73.

Cui Y, Zhang B, Ma J, Wang Y, Zhao L, Hou C, Yu J, Zhao Y, Zhang Z, Nie J et al. 2018. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. Ecotoxicol Environ Saf. 165:270-277. https://doi.org/10.1016/j.ecoenv.2018.09.018

Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. 2020. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. Neurosci Lett. 729:134981. https://doi.org/10.1016/j.neulet.2020.134981

Ding Y, Sun H, Han H, Wang W, Ji X, Liu X, Sun D. 2011. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. J Hazard Mater. 186:1942-1946.

Du L, Wan C, Cao X, Liu J. 1992. [The effect of fluorine on the developing human brain]. Chin J Pathol. 21(4):218-220.

Du L, Wan C, Cao X, Liu J. 2008. The effect of fluorine on the developing human brain. Fluoride. 41:327-330.

Duan J, Zhao M, Wang L, Fang D, Wang Y, Wang W. 1995. A comparative analysis of the results of multiple tests in patients with chronic industrial fluorosis. Guizhou Med J. 18(3):179-180.

Duan Q, Jiao J, Chen X, Wang X. 2018. Association between water fluoride and the level of children's intelligence: A dose-response meta-analysis. Public Health. 154:87-97. https://doi.org/10.1016/j.puhe.2017.08.013

Gais S, Schonauer M. 2017. Untangling a cholinergic pathway from wakefulness to memory. Neuron. 94(4):696-698.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008a. [Effects of fluoride on learning and memory and cholinesterase activity in rat brains]. Chin J Endemiol. 27:128-130.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008b. [Level of oxidative stress in rat brains and learning and memory function of rats with chronic fluorosis]. Chin J Endemiol. 27:371-373.

Gao Q, Liu YJ, Guan ZZ. 2009. Decreased learning and memory ability in rats with fluorosis: Increased oxidative stress and reduced cholinesterase activity in the brain. Fluoride. 42:277-285.

Garman RH, Li AA, Kaufmann W, Auer RN, Bolon B. 2016. Recommended methods for brain processing and quantitative analysis in rodent developmental neurotoxicity studies. Toxicol Pathol. 44(1):14-42. https://doi.org/10.1177/0192623315596858

Green R, Lanphear B, Hornung R, Flora D, Martinez-Mier EA, Neufeld R, Ayotte P, Muckle G, Till C. 2019. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. JAMA Pediatr.E1-E9.

Green R, Rubenstein J, Popoli R, Capulong R, Till C. 2020. Sex-specific neurotoxic effects of early-life exposure to fluoride: A review of the epidemiologic and animal literature. Current Epidemiology Reports. 7(4):263-273. https://doi.org/10.1007/s40471-020-00246-1

Guner S, Uyar-Bozkurt S, Haznedaroglu E, Mentes A. 2016. Dental fluorosis and catalase immunoreactivity of the brain tissues in rats exposed to high fluoride pre- and postnatally. Biol Trace Elem Res. 174:150-157.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 1991. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. Chin J Epidemiol. 10(2):98-100.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 2008a. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. Fluoride. 41:125-128.

Guo ZY, He YH, Zhu QX. 2001. [Research on the neurobehavioral function of workers occupationally exposed to fluoride]. Ind Health & Occup Dis. 27(6):346-348.

Guo ZY, He YH, Zhu QX. 2008b. Research on the neurobehavioral function of workers occupationally exposed to fluoride. Fluoride. 41:152-155.

Guyatt GH, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, Debeer H et al. 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 64(4):383-394. https://doi.org/10.1016/j.jclinepi.2010.04.026

Haschek W, Rousseaux C. 1991. Handbook of toxicologic pathology. 1st ed.: Academic Press.

Health Canada. 2015. Third report on human biomonitoring of environmental chemicals in Canada - Results of the Canadian Health Measures Survey Cycle 3 (2012–2013). Ottawa, Ontario. https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/ewh-semt/alt_formats/pdf/pubs/contaminants/chms-ecms-cycle3/chms-ecms-cycle3-eng.pdf.

Higgins JP, Green S. 2011. Cochrane handbook for systematic reviews of interventions. New York, NY: John Wiley & Sons.

Hosur MB, Puranik RS, Vanaki S, Puranik SR. 2012. Study of thyroid hormones free triiodothyronine (FT3), free thyroxine (FT4) and thyroid stimulating hormone (TSH) in subjects with dental fluorosis. Eur J Dent. 6:184-190.

Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, Sedykh A, Thayer K, Merrick BA, Walker V et al. 2020. SWIFT-Active Screener: Accelerated document screening

through active learning and integrated recall estimation. Environ Int. 138:105623. https://doi.org/10.1016/j.envint.2020.105623

Ibarluzea J, Gallastegi M, Santa-Marina L, Jiménez Zabala A, Arranz E, Molinuevo A, Lopez-Espinosa MJ, Ballester F, Villanueva CM, Riano I et al. 2021. Prenatal exposure to fluoride and neuropsychological development in early childhood: 1-to 4 years old children. Environ Res. 207:112181. https://doi.org/10.1016/j.envres.2021.112181

International Programme on Chemical Safety (IPCS). 2002. Fluorides. Geneva: World Health Organization, International Programme on Chemical Safety. Environmental Health Criteria 227. https://inchem.org/documents/ehc/ehc/ehc227.htm.

Jacqmin H, Commenges D, Letenneur L, Barberger-Gateau P, Dartigues JF. 1994. Components of drinking water and risk of cognitive impairment in the elderly. Am J Epidemiol. 139:48-57.

Jia B, Zong L, Lee JY, Lei J, Zhu Y, Xie H, Clemens JL, Feller MC, Na Q, Dong J et al. 2019. Maternal supplementation of low dose fluoride alleviates adverse perinatal outcomes following exposure to intrauterine inflammation. Scientific Reports. 9(1):2575. https://doi.org/10.1038/s41598-018-38241-8

Jiang C, Zhang S, Liu H, Guan Z, Zeng Q, Zhang C, Lei R, Xia T, Wang Z, Yang L et al. 2014. Low glucose utilization and neurodegenerative changes caused by sodium fluoride exposure in rat's developmental brain. Neuromolecular Med. 16:94-105.

Jin T, Han T, Wei Y, Wu Y, Wang Z, Zhang H. 2016. Investigation on working memory level of children aged 8-12 years in coal-burning fluorosis area. Journal of environment and health.409-411.

Jones S, Burt BA, Petersen PE, Lennon MA. 2005. The effective use of fluorides in public health. Bull World Health Organ. 83:670-676.

Khan AM, Raina R, Dubey N, Verma PK. 2017. Effect of deltamethrin and fluoride co-exposure on the brain antioxidant status and cholinesterase activity in Wistar rats. Drug Chem Toxicol. 41(2):1-5. https://doi.org/10.1080/01480545.2017.1321009

Khandare AL, Gourineni SR, Validandi V. 2017. Dental fluorosis, nutritional status, kidney damage, and thyroid function along with bone metabolic indicators in school-going children living in fluoride-affected hilly areas of Doda District, Jammu and Kashmir, India. Environ Monit Assess. 189:579. https://doi.org/10.1007/s10661-017-6288-5

Khandare AL, Validandi V, Gourineni SR, Gopalan V, Nagalla B. 2018. Dose-dependent effect of fluoride on clinical and subclinical indices of fluorosis in school going children and its mitigation by supply of safe drinking water for 5 years: An Indian study. Environ Monit Assess. 190:110. https://doi.org/10.1007/s10661-018-6501-1

Kheradpisheh Z, Mahvi AH, Mirzaei M, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018a. Correlation between drinking water fluoride and TSH hormone by ANNs and ANFIS. J Environ Health Sci Eng. 16(1):11-18. https://doi.org/10.1007/s40201-018-0290-x

Kheradpisheh Z, Mirzaei M, Mahvi AH, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018b. Impact of drinking water fluoride on human thyroid hormones: A case-control study. Sci Rep. 8:2674. https://doi.org/10.1038/s41598-018-20696-4

Kumar V, Chahar P, Kajjari S, Rahman F, Bansal DK, Kapadia JM. 2018. Fluoride, thyroid hormone derangements and its correlation with tooth eruption pattern among the pediatric population from endemic and non-endemic fluorosis areas. J Contemp Dent Pract. 19(12):1512-1516.

Li F, Chen X, Huang R, Xie Y. 2009. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. J Environ Health. 26(4):838-840.

Li J, Yao L, Q.L. S, Wu CY. 2004. [Effects of high fluoride level on neonatal neurobehavioral development]. Chin J Endemiol. 23(5):463-465.

Li J, Yao L, Shao QL, Wu CY. 2008a. Effects of high fluoride level on neonatal neurobehavioral development. Fluoride. 41:165-170.

Li M, Gao Y, Cui J, Li Y, Li B, Liu Y, Sun J, Liu X, Liu H, Zhao L et al. 2016. Cognitive impairment and risk factors in elderly people living in fluorosis areas in China. Biol Trace Elem Res. 172:53-60.

Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. 2010. [Investigation and analysis of children's IQ and dental fluorosis in high fluoride area]. Chin J Pest Control. 26(3):230-231.

Li Y, Li X, Wei S. 1994. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. J West Chin University Med Sci. 25(2):188-191.

Li Y, Li X, Wei S. 2008b. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. Fluoride. 41:331-335.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2003. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. Chin J Public Health Manag. 19(4):337-338.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2008c. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. Fluoride. 41:161-164.

Liang C, Ji R, Cao J, Jiang Y, Yu B, Ma F, Wu Y, Ying B, Zhang Y, Sun S et al. 2003. Study on the relationship between drinking water trace elements and cognitive ability of the elderly. Health Res.436-440.

Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. 1991. [High fluoride and low iodine environment and subclinical cretinism in Xinjiang]. Endem Dis Bull. 6(2):62-67.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. Alterations of nAChRs and ERK1/2 in the brains of rats with chronic fluorosis and their connections with the decreased capacity of learning and memory. Toxicol Lett. 192:324-329.

Lou DD, Guan ZZ, Liu YJ, Liu YF, Zhang KL, Pan JG, Pei JJ. 2013. The influence of chronic fluorosis on mitochondrial dynamics morphology and distribution in cortical neurons of the rat brain. Arch Toxicol. 87:449-457.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S et al. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. Physiol Behav. 206:76-83. https://doi.org/10.1016/j.physbeh.2019.02.017

Ma Q, Huang H, Sun L, Zhou T, Zhu J, Cheng X, Duan L, Li Z, Cui L, Ba Y. 2017. Gene-environment interaction: Does fluoride influence the reproductive hormones in male farmers modified by ERα gene polymorphisms? Chemosphere. 188:525-531. https://doi.org/10.1016/j.chemosphere.2017.08.166

Malin AJ, Till C. 2015. Exposure to fluoridated water and attention deficit hyperactivity disorder prevalence among children and adolescents in the United States: An ecological association. Environ Health. 14:17.

Malin AJ, Riddell J, McCague H, Till C. 2018. Fluoride exposure and thyroid function among adults living in Canada: Effect modification by iodine status. Environ Int. 121(Pt 1):667-674. https://doi.org/10.1016/j.envint.2018.09.026

McPherson CA, Zhang G, Gilliam R, Brar SS, Wilson R, Brix A, Picut C, Harry GJ. 2018. An evaluation of neurotoxicity following fluoride exposure from gestational through adult ages in Long-Evans hooded rats. Neurotoxicol Res.1-18. https://doi.org/10.1007/s12640-018-9870-x

Mesram N, Nagapuri K, Banala RR, Nalagoni CR, Karnati PR. 2016. Quercetin treatment against NaF induced oxidative stress related neuronal and learning changes in developing rats. J King Saud Univ Sci. 29:221-229.

Michael M, Barot VV, Chinoy NJ. 1996. Investigations of soft tissue functions in fluorotic individuals of North Gujarat. Fluoride. 29:63-71.

Miller K, Howard B, Phillips J, Shah M, Mav D, Thayer K, Shah R. 2016. SWIFT-Active screener: Reducing literature screening effort through machine learning for systematic reviews. Cochrane Colloquium Seoul; October 25 2016; Seoul, Korea.

~~Moher D, Liberati A, Tetzlaff J, Altman D. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med. 6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097~~

Morgan L, Allred E, Tavares M, Bellinger D, Needleman H. 1998. Investigation of the possible associations between fluorosis, fluoride exposure, and childhood behavior problems. Pediatr Dent. 20:244-252.

Mustafa DE, Younis UM, Elhag SAA. 2018. The relationship between the fluoride levels in drinking water and the schooling performance of children in rural areas of Khartoum State, Sudan. Fluoride. 51(2):102-113.

Nageshwar M, Sudhakar K, Reddy KP. 2018. Quercetin ameliorates oxidative stress, neural damage of brain and behavioral impairment of rat with fluoride exposure. Int J Pharm Sci Res. 9(8):3247-3256. https://doi.org/10.13040/IJPSR.0975-8232.9(8).3247-56

National Academies of Sciences Engineering and Medicine (NASEM). 2020. Review of the draft NTP monograph: Systematic review of fluoride exposure and neurodevelopmental and cognitive health effects. Washington, DC: National Academies of Sciences, Engineering and Medicine. https://doi.org/10.17226/25715.

National Academies of Sciences Engineering and Medicine (NASEM). 2021. Review of the revised NTP Monograph on the systematic review of fluoride exposure and neurodevelopmental and cognitive health effects: A letter report. Washington, DC: National Academies of Sciences, Engineering and Medicine. https://doi.org/10.17226/26030.

National Institute for Occupational Safety and Health (NIOSH). 1984. Fluoride in urine. Washington, DC: National Institute for Occupational Safety and Health. Method 8308.

National Research Council (NRC). 2006. Committee on fluoride in drinking water, board on environmental studies and toxicology. Fluoride in drinking water: A scientific review of EPA's standards.: National Research Council. http://www.nap.edu/catalog/11571/fluoride-in-drinking-water-a-scientific-review-of-epas-standards. [19 August 2019]

National Toxicology Program (NTP). 2016. Systematic literature review on the effects of fluoride on learning and memory in animal studies. Research Triangle Park, NC: National Toxicology Program. NTP Research Report 1. https://ntp.niehs.nih.gov/ntp/results/pubs/rr/reports/01fluoride_508.pdf. [19 August 2019]

National Toxicology Program (NTP). 2019. Health Assessment Workspace Collaborative (HAWC) Page: Flouride (2019). Research Triangle Park, NC: U.S. Department of Health and Human Services, National Institutes of Health, National Toxicology Program. https://hawcproject.org/assessment/405/.

National Toxicology Program (NTP). 2020. Revised draft NTP Monograph on the systematic review of fluoride exposure and neurodevelopmental and cognitive health effects. Research Triangle Park, NC: National Toxicology Program. https://www.nationalacademies.org/event/10-19-2020/docs/DDA97C9170D1A255D69C004CEB77B698E8D005011EFB.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. Environ Pollut. 233:889-899. https://doi.org/10.1016/j.envpol.2017.09.015

Nkpaa KW, Onyeso GI. 2018. Rutin attenuates neurobehavioral deficits, oxidative stress, neuro-inflammation and apoptosis in fluoride treated rats. Neurosci Lett. 682:92-99. https://doi.org/10.1016/j.neulet.2018.06.023

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE , Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. PLoS Med. 18(3):e1003583. 10.1371/journal.pmed.1003583

Peckham S, Lowery D, Spencer S. 2015. Are fluoride levels in drinking water associated with hypothyroidism prevalence in England? A large observational study of GP practice data and fluoride levels in drinking water. J Epidemiol Community Health. 69:619-624.

Podgorski J, Berg M. 2020. Global threat of arsenic in groundwater. Science. 368(6493):845-850. https://doi.org/10.1126/science.aba1510

Pulungan ZS, Sofro ZM, Partadiredja G. 2016. Sodium fluoride does not affect the working memory and number of pyramidal cells in rat medial prefrontal cortex. Anat Sci Int. 93(1):128-138.

Reddy YP, Tiwari SK, Shaik AP, Alsaeed A, Sultana A, Reddy PK. 2014. Effect of sodium fluoride on neuroimmunological parameters, oxidative stress and antioxidative defenses. Toxicol Mech Methods. 24:31-36.

Ren D, Li K, Liu D. 1989. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. Chin J Control Endem Dis. 4(4):251.

Ren D, Li K, Liu D. 2008. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. Fluoride. 41:319-320.

Riddell JK, Malin AJ, Flora D, McCague H, Till C. 2019. Association of water fluoride and urinary fluoride concentrations with attention deficit hyperactivity disorder in Canadian youth. Environ Int. 133:105190. https://doi.org/10.1016/j.envint.2019.105190

Rocha-Amador D, Navarro ME, Carrizales L, Morales R, Calderon J. 2007. Decreased intelligence in children and exposure to fluoride and arsenic in drinking water. Cad Saude Publica. 23(Suppl 4):S579-587.

Rocha-Amador D, Navarro M, Trejo-Acevedo A, Carrizales L, Perez-Maldonado I, Diaz-Barriga F, Calderon J. 2009. Use of the Rey-Osterrieth Complex Figure Test for neurotoxicity evaluation of mixtures in children. Neurotox. 30:1149-1154.

Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environ Health Perspect. 122(7):711-718.

Rotton J, Tikofsky RS, Feldman HT. 1982. Behavioral effects of chemicals in drinking water. J Appl Psychol. 67:230-238.

Russ TC, Killin LOJ, Hannah J, Batty GD, Deary IJ, Starr JM. 2019. Aluminium and fluoride in drinking water in relation to later dementia risk. Brit J Psychol. 14:1-6. https://doi.org/10.1192/bjp.2018.287

Santa-Marina L, Jimenez-Zabala A, Molinuevo A, Lopez-Espinosa M, Villanueva C, Riano I, Ballester F, Sunyer J, Tardon A, Ibarluzea J. 2019. Fluorinated water consumption in pregnancy and neuropsychological development of children at 14 months and 4 years of age. Environ Epidemiol. 3:386-387. https://dx.doi.org/10.1097/01.EE9.0000610304.33479.18

Saxena S, Sahay A, Goel P. 2012. Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India. J Neurosci Rural Pract. 3:144-149.

Scientific Committee on Health and Environmental Risks (SCHER). 2011. Critical review of any new evidence on the hazard profile, health effects, and human exposure to fluoride and the fluoridating agents of drinking water. Scientific Committee on Health and Environmental Risks. http://ec.europa.eu/health/scientific_committees/environmental_risks/docs/scher_o_139.pdf. [19 August 2019]

Seraj B, Shahrabi M, Shadfar M, Ahmadi R, Fallahzadeh M, Eslamlu HF, Kharazifard MJ. 2012. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. J Dent. 9:221-229.

Shan KR, Qi XL, Long YG, Nordberg A, Guan ZZ. 2004. Decreased nicotinic receptors in PC12 cells and rat brains influenced by fluoride toxicity: A mechanism relating to a damage at the level in post-transcription of the receptor genes. Toxicology. 200:169-177.

Shannon FT, Fergusson DM, Horwood LJ. 1986. Exposure to fluoridated public water supplies and child health and behaviour. New Zeal Med J. 99:416-418.

Shao Q. 2003. [Study of cognitive function impairment caused by chronic fluorosis]. Chin J Endemiol. 22(4):336-338.

Sharma JD, Sohu D, Jain P. 2009. Prevalence of neurological manifestations in a human population exposed to fluoride in drinking water. Fluoride. 42:127-132.

Singh N, Verma KG, Verma P, Sidhu GK, Sachdeva S. 2014. A comparative study of fluoride ingestion levels, serum thyroid hormone & TSH level derangements, dental fluorosis status among school children from endemic and non-endemic fluorosis areas. Springerplus. 3:7.

Soto-Barreras U, Escalante-Villalobos KY, Holguin-Loya B, Perez-Aguirre B, Nevarez-Rascon A, Martinez-Martinez RE, Loyola-Rodriguez JP. 2019. Effect of fluoride in drinking water on dental caries and IQ in children. Fluoride. 52:474-482.

Still CN, Kelley P. 1980. The incidence of primary degenerative dementia vs. water fluoride content in South Carolina, USA. Neurotox. 1:125-132.

Sudhakar K, Reddy KP. 2018. Protective effects of Abelmoschus moschatus seed extract on neurotransmitter system of developing brain of Wistar rats with gestational and post-natal exposure of sodium fluoride. Int J Green Pharm. 12:S131-S139.

Sudhir KM, Chandu GN, Prashant GM, Subba Reddy VV. 2009. Effect of fluoride exposure on intelligence quotient (IQ) among 13-15 year old school children of known endemic area of fluorosis, Nalgonda District, Andhra Pradesh. J Indian Assoc Public Health Dent. 2009(13):88-94.

Sun M, Li S, Wang Y, Li F. 1991. [Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis]. J Guiyang Med Coll. 16(3):204-206.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2000. [Effects of high fluoride drinking water on the cerebral function of mice]. Chin J Epidemiol. 19:262-263.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2008. Effects of high fluoride drinking water on the cerebral function of mice. Fluoride. 41:148-151.

Susheela AK, Bhatnagar M, Vig K, Mondal NK. 2005. Excess fluoride ingestion and thyroid hormone derangements in children living in Delhi, India. Fluoride. 38:98-108.

Till C, Green R, Grundy JG, Hornung R, Neufeld R, Martinez-Mier EA, Ayotte P, Muckle G, Lanphear B. 2018. Community water fluoridation and urinary fluoride concentrations in a national sample of pregnant women in Canada. Environ Health Perspect. 126(10):107001. https://doi.org/10.1289/ehp3546

Till C, Green R, Flora D, Hornung R, Martinez-Mier EA, Blazer M, Farmus L, Ayotte P, Muckle G, Lanphear B. 2020. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. Environ Int. 134:105315. https://doi.org/10.1016/j.envint.2019.105315

Trivedi M, Sangai N, Patel R, Payak M, Vyas S. 2012. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. Fluoride. 45(4):377-383.

Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. 2007. Effect of high fluoride water on intelligence of school children in India. Fluoride. 40:178-183.

Tsunoda M, Aizawa Y, Nakano K, Liu Y, Horiuchi T, Itai K, Tsunoda H. 2005. Changes in fluoride levels in the liver, kidney, and brain and in neurotransmitters of mice after subacute administration of fluoride. Fluoride. 38:284-292.

U.S. Department of Health and Human Services (US DHHS). 2015. U.S. Public Health Service recommendation for fluoride concentration in drinking water for the prevention of dental caries. U.S. Department of Health and Human Services. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4547570/. [19 August 2019]

U.S. Environmental Protection Agency (US EPA). 2010. Fluoride: Exposure and relative source contribution analysis. Washington, DC: U.S. Environmental Protection Agency. 820-R-10-015. http://www.epa.gov/dwstandardsregulations/fluoride-risk-assessment-and-relative-source-contribution. [19 August 2019]

Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, Costilla-Salazar R, Calderon Hernandez J, Alcaraz Contreras Y, Rocha-Amador DO. 2017. In utero exposure to fluoride and cognitive development delay in infants. Neurotox. 59:65-70. https://doi.org/10.1016/j.neuro.2016.12.011

Villa A, Anabalon M, Zohouri V, Maguire A, Franco AM, Rugg-Gunn A. 2010. Relationships between fluoride intake, urinary fluoride excretion and fluoride retention in children and adults: An analysis of available data. Caries Res. 44(1):60-68. https://doi.org/10.1159/000279325

Wang A, Duan L, Huang H, Ma J, Zhang Y, Ma Q, Guo Y, Li Z, Cheng X, Zhu J et al. 2020a. Association between fluoride exposure and behavioural outcomes of school-age children: A pilot study in China. Int J Environ Health Res.1-10. https://doi.org/10.1080/09603123.2020.1747601

Wang G, Gao M, Zhang M, Yang M, Xiang Q. 2012. Correlation between total fluoride intake and children's IQ. J Southeast Univ Med Ed.743-746.

Wang M, Liu L, Li H, Li Y, Liu H, Hou C, Zeng Q, Li P, Zhao Q, Dong L et al. 2020b. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. Environ Int. 134:105229. https://doi.org/10.1016/j.envint.2019.105229

Wang X, Wang L, Hu P, Guo X, Luo X. 2001. [Effects of high iodine and high fluorine on children's intelligence and thyroid function]. Chin J Endemiol. 20(4):288-290.

Watanabe M, Kono K, Orita Y, Dote T, Usuda K, Takahashi Y, Yoshida Y. 1995. Influence of dietary fluoride intake on urinary fluoride concentration and evaluation of corrected levels in spot urine. Fluoride. 28(2):61-70.

Waugh DT. 2019. Fluoride exposure induces inhibition of sodium/iodide symporter (NIS) contributing to impaired iodine absorption and iodine deficiency: Molecular mechanisms of inhibition and implications for public health. Int J Environ Res Public Health. 16(6). https://doi.org/10.3390/ijerph16061086

World Health Organization (WHO). 2008. Guidelines for drinking-water quality [electronic resource]: Incorporating 1st and 2nd addenda. Geneva, Switzerland: World Health Organization. Third Edition. Vol. 1. https://apps.who.int/iris/bitstream/handle/10665/204411/9789241547611_eng.pdf?sequence=1&isAllowed=y.

World Health Organization (WHO). 2017. Guidelines for drinking-water quality. World Health Organization. 4th ed. + 1st add. https://apps.who.int/iris/handle/10665/254637.

Xiang Q, Liang Y, Chen L, Wang C, Chen B, Chen X, Zhou M. 2003a. Effect of fluoride in drinking water on children's intelligence. Fluoride. 36:84-94.

Xiang Q, Liang Y, Chen B, Chen L. 2011. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. Fluoride. 44:191-194.

Xiang Q, Wang Y, Yang M, Zhang M, Xu Y. 2013. Level of fluoride and arsenic in household shallow well water in Wamiao and Xinhuai villages in Jiangsu province, China. Fluoride. 46:192-197.

Xiang QY, Liang YX, Zhou MS, Zang HB. 2003b. Blood lead of children in Wamiao-Xinhuai intelligence study. Fluoride. 36:198-199.

Yang Y, Wang X, X. G, Hu P. 1994. [The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine]. Chin J Epidemiol. 15(4):296-298.

Yang Y, Wang X, Guo X, Hu P. 2008. The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine. Fluoride. 41:336-339.

Yao L, Zhou J, Wang S, Cui K, Lin F. 1996. Analysis on TSH and intelligence level of children with dental Fluorosis in a high fluoride area. Lit Inf Prev Med. 2(1):26-27.

Yasmin S, Ranjan S, D'Souza D. 2013. Effect of excess fluoride ingestion on human thyroid function in Gaya Region, Bihar, India. Toxicol Environ Chem. 95:1235-1243.

Yazdi SM, Sharifian A, Dehghani-Beshne M, Momeni VR, Aminian O. 2011. Effects of fluoride on psychomotor performance and memory of aluminum potroom workers. Fluoride. 44:158-162.

Yu X, Chen J, Li Y, Liu H, Hou C, Zeng Q, Cui Y, Zhao L, Li P, Zhou Z et al. 2018. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. Environ Int. 118:116-124. https://doi.org/10.1016/j.envint.2018.05.042

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 1996. [Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis]. Chin J Endemiol. 15(5):257-259.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 2008. Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis. Fluoride. 41:134-138.

Zhang KL, Lou DD, Guan ZZ. 2015a. Activation of the AGE/RAGE system in the brains of rats and in SH-SY5Y cells exposed to high level of fluoride might connect to oxidative stress. Neurotoxicol Teratol. 48:49-55.

Zhang S, Zhang X, Liu H, Qu W, Guan Z, Zeng Q, Jiang C, Gao H, Zhang C, Lei R et al. 2015b. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. Toxicol Sci. 144:238-245.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C et al. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. Arch Toxicol. 93(3):709-726. https://doi.org/10.1007/s00204-019-02390-0

Zhao Q, Tian Z, Zhou G, Niu Q, Chen J, Li P, Dong L, Xia T, Zhang S, Wang A. 2020. SIRT1-dependent mitochondrial biogenesis supports therapeutic effects of resveratrol against neurodevelopment damage by fluoride. Theranostics. 10:4822-4838. https://doi.org/10.7150/thno.42387

Zhou G, Tang S, Yang L, Niu Q, Chen J, Xia T, Wang S, Wang M, Zhao Q, Liu L et al. 2019. Effects of long-term fluoride exposure on cognitive ability and the underlying mechanisms: Role of autophagy and its association with apoptosis. Toxicol Appl Pharmacol. 378:114608. https://doi.org/10.1016/j.taap.2019.114608

Zhou T, Duan L-J, Ding Z, Yang R-P, Li S-H, Xi Y, Cheng X-M, Hou J-X, Wen S-B, Chen J et al. 2012. Environmental fluoride exposure and reproductive hormones in male living in endemic fluorosis villages in China. Life Sci J. 9(4):1-7.

Zohouri FV, Swinbank CM, Maguire A, Moynihan PJ. 2006. Is the fluoride/creatinine ratio of a spot urine sample indicative of 24-h urinary fluoride? Community Dent Oral Epidemiol. 34(2):130-138. https://doi.org/10.1111/j.1600-0528.2006.00269.x

Appendix A. Data Figures: Neurodevelopmental or Cognitive Effects and Outcomes

Figures

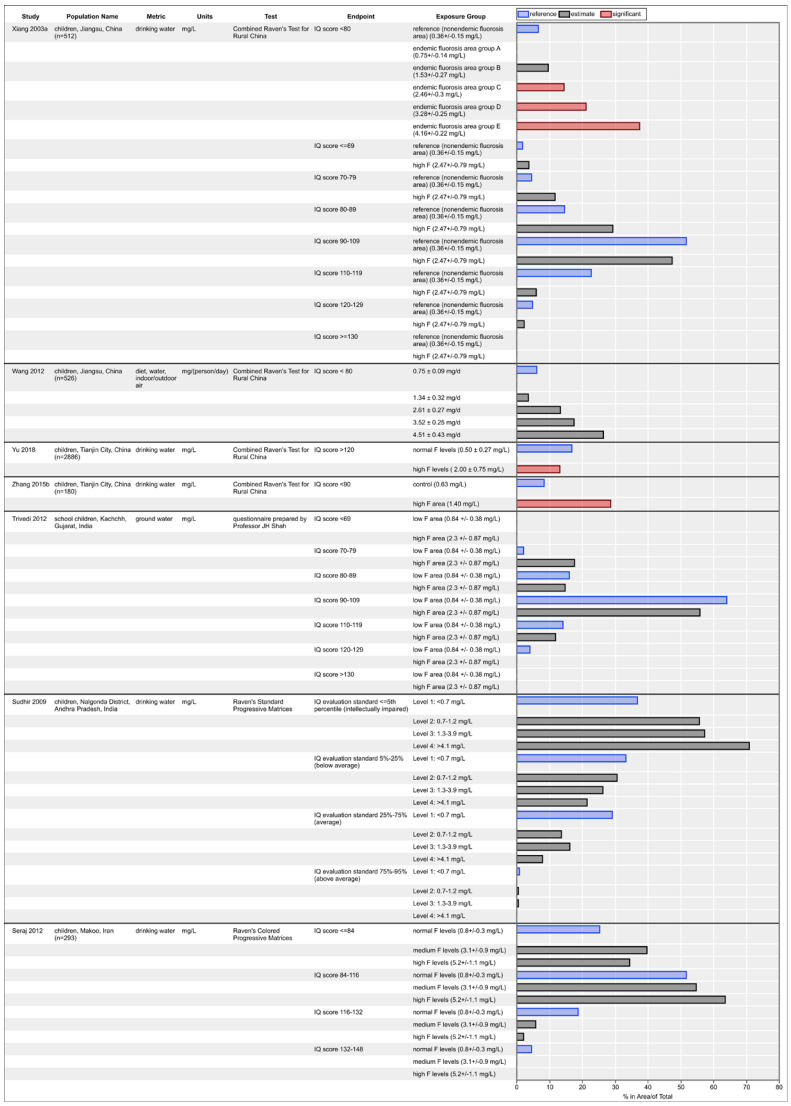


Figure A-1. Distribution of IQ in Children by Fluoride Exposure (Low Risk-of-bias Studies; Presented as % in Area or % of Total Group)

Reference group indicated by blue bars; other bars represent response estimates with red indicating statistical significance compared with the reference group.
An interactive version of Figure A-1 and additional study details in HAWC here (NTP 2019). "F" represents fluoride. For IQ distribution results by drinking water fluoride level provided in Xiang et al. (2003a), Trivedi et al. (2012), Sudhir et al. (2009), and Seraj et al. (2012) and rate of low IQ scores by fluoride intake provided in Wang et al. (2012), statistical significance was not evaluated.

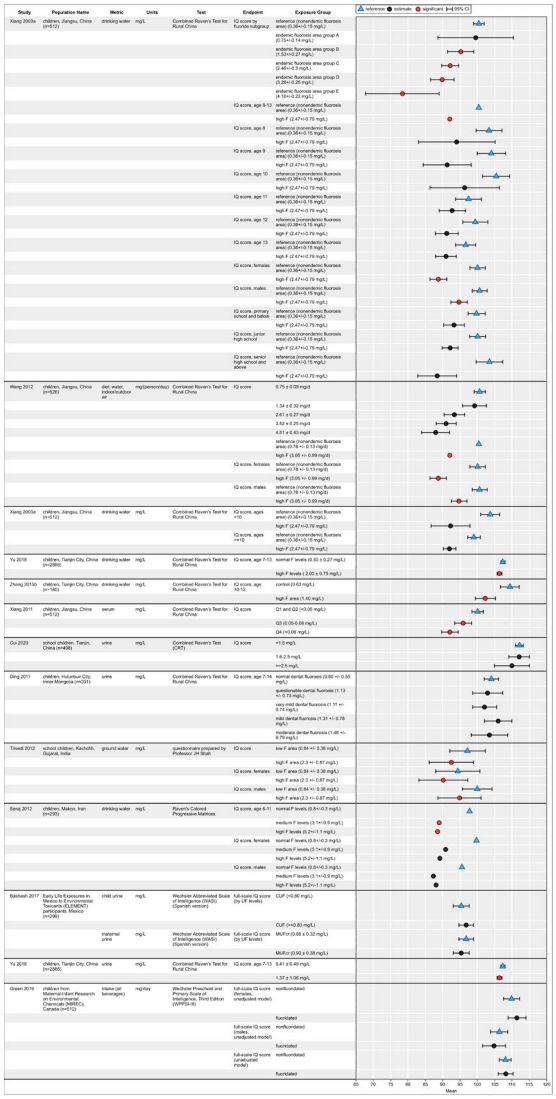


Figure A-2. Mean IQ in Children by Fluoride Exposure (Low Risk-of-bias Studies)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. An interactive version of Figure A-2 and additional study details in HAWC here (NTP 2019) "F" represents fluoride. Three additional publications based on subsample of the larger Yu et al. (2018) cohort were identified (Zhao et al. 2019; Zhao et al. 2020; Zhou et al. 2019); however, results from these studies are not presented here. The main study by Yu et al. (2018) is considered a better representation of the IQ results. For all studies, SDs are available and can be viewed in HAWC (NTP 2019) by clicking the data points within the plot area; however, 95% CIs could not be calculated for Seraj et al. (2012) because Ns are not available for exposure groups.

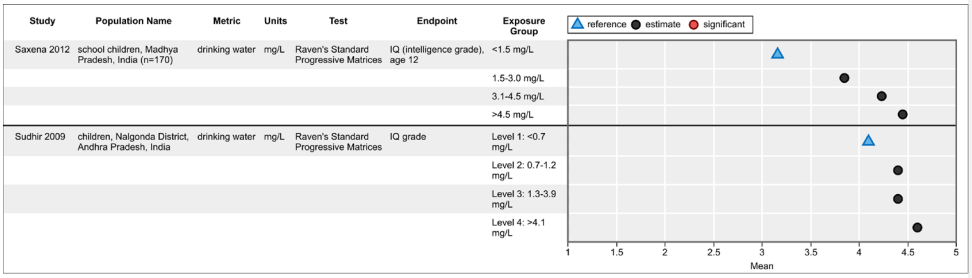


Figure A-3. Intelligence Grade in Children by Fluoride Exposure (Low Risk-of-bias Studies; Presented as Mean)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. An interactive version of Figure A-3 and additional study details in HAWC here (NTP 2019). For Saxena et al. (2012), children's intelligence was measured using Raven's Standard Progressive Matrices. Children's scores were converted to percentile, and specific grades were allotted based on the percentiles. Grades ranged from intellectually superior (Grade I) to intellectually impaired (Grade V). Results for Soto-Barreras et al. (2019) are not presented here. Outcomes in the study were presented as levels of fluoride exposure associated with each intelligence grade. Results reported were not significant.

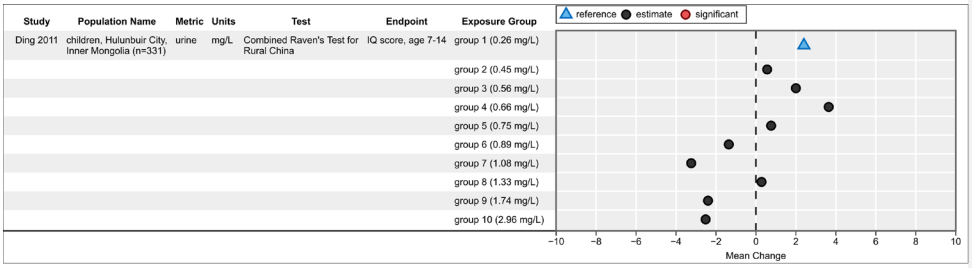
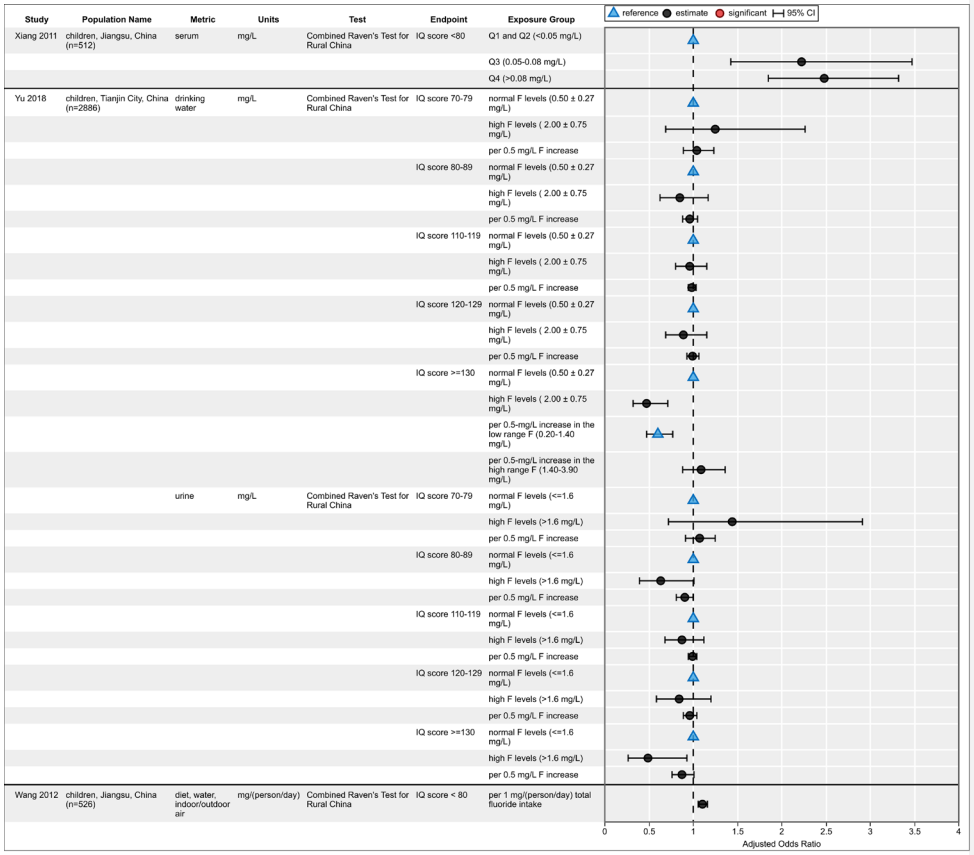


Figure A-4. Mean Change in IQ in Children by Fluoride Exposure (Low Risk-of-bias Studies)

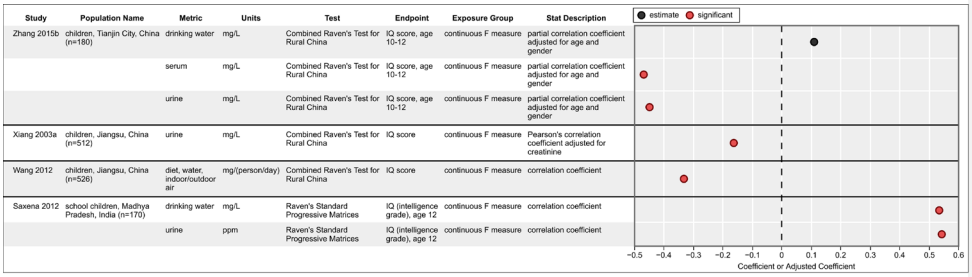
Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. An interactive version of Figure A-4 and additional study details in HAWC here (NTP 2019). For Ding et al. (2011), SDs are available and can be viewed in HAWC (NTP 2019) by clicking the data points within the plot area; however, 95% CIs could not be calculated because Ns for each exposure group are not available.



Study	Population Name	Metric	Units	Test	Endpoint	Exposure Group
Xiang 2011	children, Jiangsu, China (n=512)	serum	mg/L	Combined Raven's Test for Rural China	IQ score <80	Q1 and Q2 (<0.05 mg/L)
						Q3 (0.05-0.08 mg/L)
						Q4 (>0.08 mg/L)
Yu 2018	children, Tianjin City, China (n=2886)	drinking water	mg/L	Combined Raven's Test for Rural China	IQ score 70-79	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score 80-89	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score 110-119	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score 120-129	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score >=130	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5-mg/L increase in the low range F (0.20-1.40 mg/L)
						per 0.5-mg/L increase in the high range F (1.40-3.90 mg/L)
		urine	mg/L	Combined Raven's Test for Rural China	IQ score 70-79	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score 80-89	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score 110-119	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score 120-129	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score >=130	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
Wang 2012	children, Jiangsu, China (n=526)	diet, water, indoor/outdoor air	mg/(person/day)	Combined Raven's Test for Rural China	IQ score < 80	per 1 mg/(person/day) total fluoride intake

Figure A-5. Associations between Fluoride Exposure and IQ Scores in Children (Low Risk-of-bias Studies; Presented as Adjusted OR)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. Cutoffs for the dichotomous outcome are listed in the Endpoint column.
An interactive version of Figure A-5 and additional study details in HAWC here (NTP 2019). For Xiang et al. (2011), there was a significant linear trend across different levels of serum fluoride for IQ score <80 ($p < 0.001$). For Yu et al. (2018), significance levels by IQ score were not reported.



Study	Population Name	Metric	Units	Test	Endpoint	Exposure Group	Stat Description
Zhang 2015b	children, Tianjin City, China (n=180)	drinking water	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
		serum	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
		urine	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
Xiang 2003a	children, Jiangsu, China (n=512)	urine	mg/L	Combined Raven's Test for Rural China	IQ score	continuous F measure	Pearson's correlation coefficient adjusted for creatinine
Wang 2012	children, Jiangsu, China (n=526)	diet, water, indoor/outdoor air	mg/(person/day)	Combined Raven's Test for Rural China	IQ score	continuous F measure	correlation coefficient
Saxena 2012	school children, Madhya Pradesh, India (n=170)	drinking water	mg/L	Raven's Standard Progressive Matrices	IQ (intelligence grade), age 12	continuous F measure	correlation coefficient
		urine	ppm	Raven's Standard Progressive Matrices	IQ (intelligence grade), age 12	continuous F measure	correlation coefficient

Figure A-6. Correlations between Fluoride Exposure and IQ Score in Children (Low Risk-of-bias Studies; Presented as Correlation Coefficient)

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-6 and additional study details in HAWC here (NTP 2019). "F" represents fluoride. For Saxena et al. (2012), a significant relationship between water fluoride level and intelligence grade was observed. Increasing intelligence grades reflected increasing levels of impairment (reduced intelligence) in children.

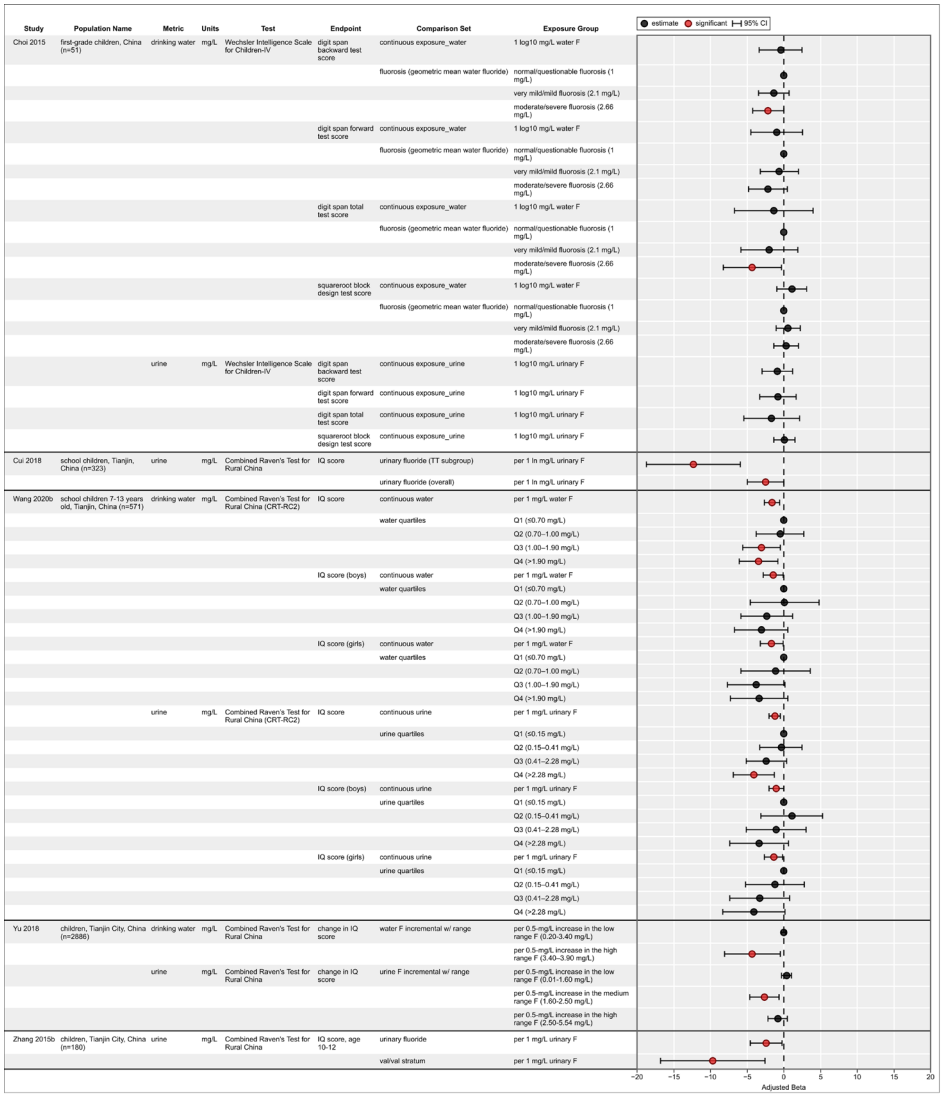


Figure A-7. Associations between Fluoride Exposure and IQ Score in Children (Low Risk-of-bias Studies; Presented as Adjusted Beta)—China

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-7 and additional study details in HAWC here (NTP 2019). "F" represents fluoride. For Yu et al. (2018), authors note an obvious decrease in the IQ score at water fluoride exposure levels between 3.40 mg/L and 3.90 mg/L and a similar adverse effect on IQ scores at urinary fluoride exposure levels from 1.60 mg/L to 2.50 mg/L, and so the changes in IQ score are indicated as significant; however, significance levels for change in IQ score were not reported.

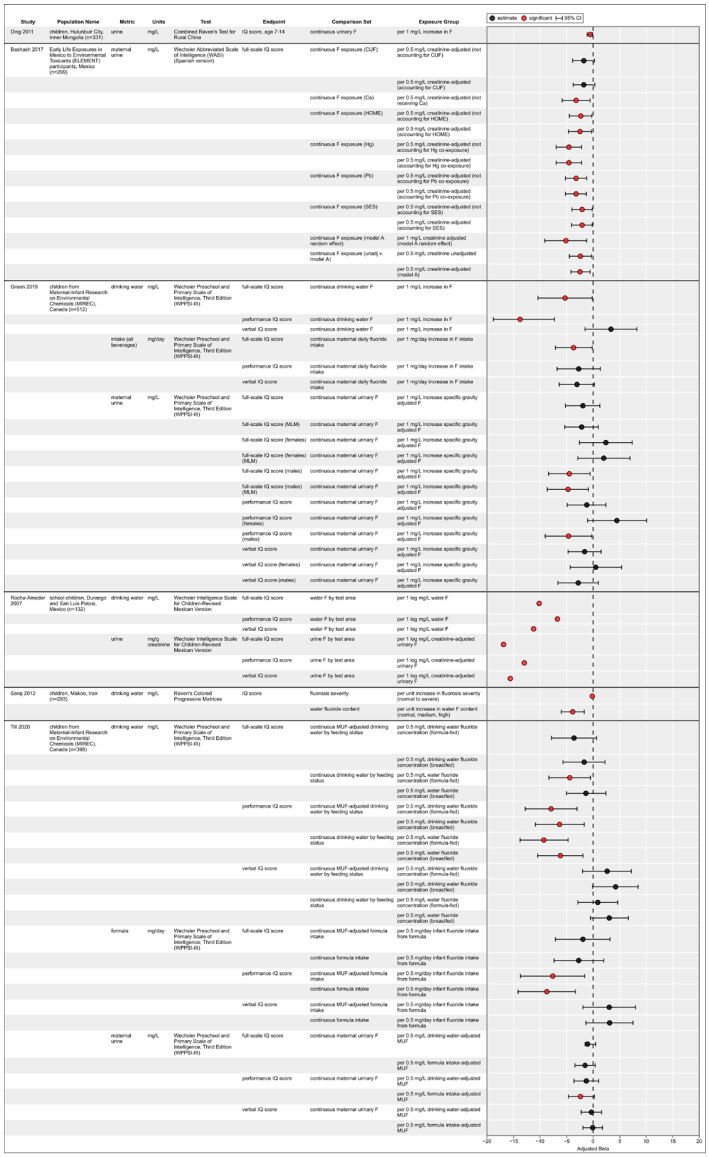


Figure A-8. Associations between Fluoride Exposure and IQ Score in Children (Low Risk-of-bias Studies; Presented as Adjusted Beta)—Areas Other Than China

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-8 and additional study details in HAWC here (NTP 2019). "F" represents fluoride.

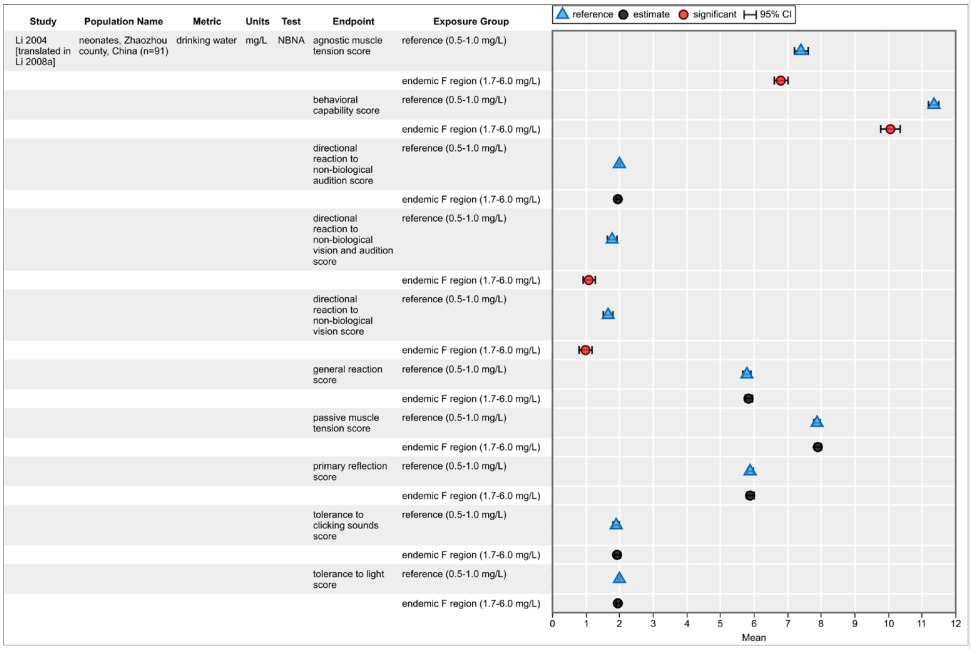


Figure A-9. Mean Motor/Sensory Scores in Children by Fluoride Exposure (Low Risk-of-bias Studies)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-9 and additional study details in HAWC here (NTP 2019). "F" represents fluoride. 95% CIs are small and are within figure symbols and may be difficult to see. Values for SDs and 95% CIs can be viewed in HAWC (NTP 2019) by clicking the data points within the plot area. Total neonatal behavioral neurological assessment (NBNA) score was also significantly reduced in the endemic F region versus reference region (not shown).

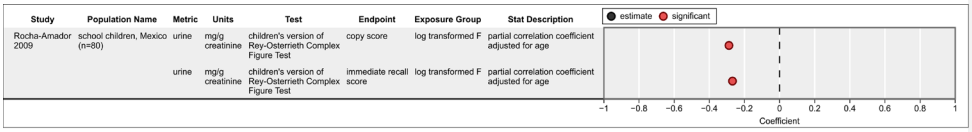


Figure A-10. Correlations between Fluoride Exposure and Other Cognitive Effects in Children (Low Risk-of-bias Studies; Presented as Correlation Coefficient)

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-10 and additional study details in HAWC here (NTP 2019). "F" represents fluoride.

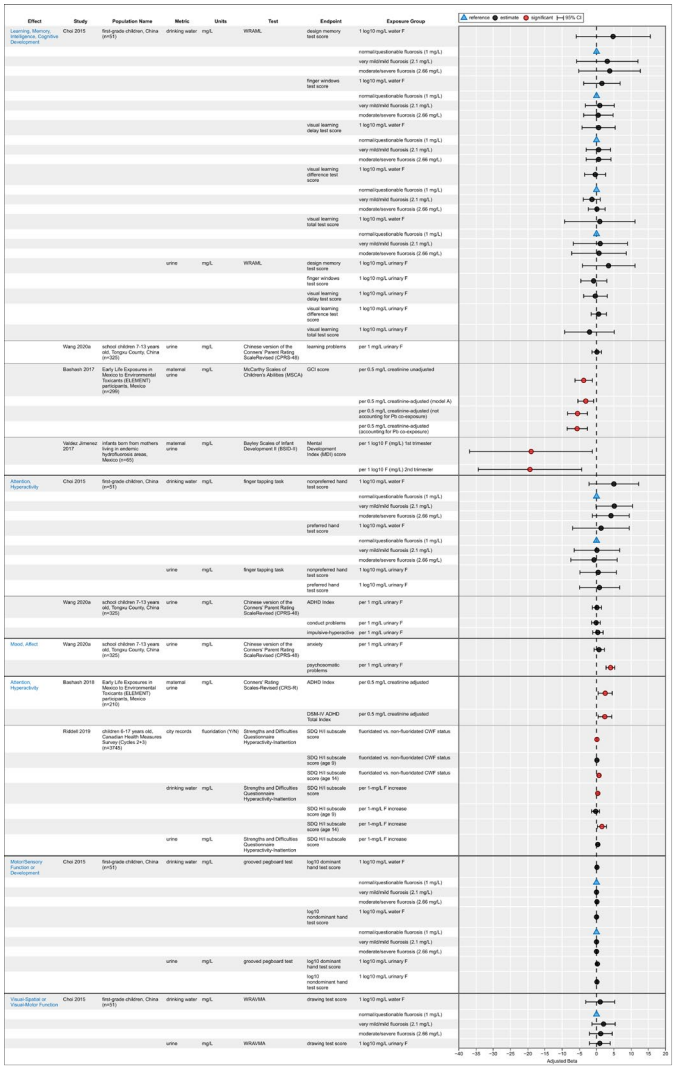


Figure A-11. Associations between Fluoride Exposure and Other Neurodevelopmental Effects in Children (Low Risk-of-bias Studies; Presented as Adjusted Beta)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. An interactive version of Figure A-11 and additional study details in HAWC here (NTP 2019). "F" represents fluoride. Bashash et al. (2018) observed significant associations between maternal urinary fluoride and ADHD-like symptoms related to inattention (an increase in 0.5 mg/L of maternal urinary fluoride was associated with a 2.84-point increase in the DSM-IV Inattention Index and a 2.54-point increase in Cognitive Problems and Inattention Index). These two scales contributed to the global ADHD Index and the DSM-IV ADHD Total Index shown here.

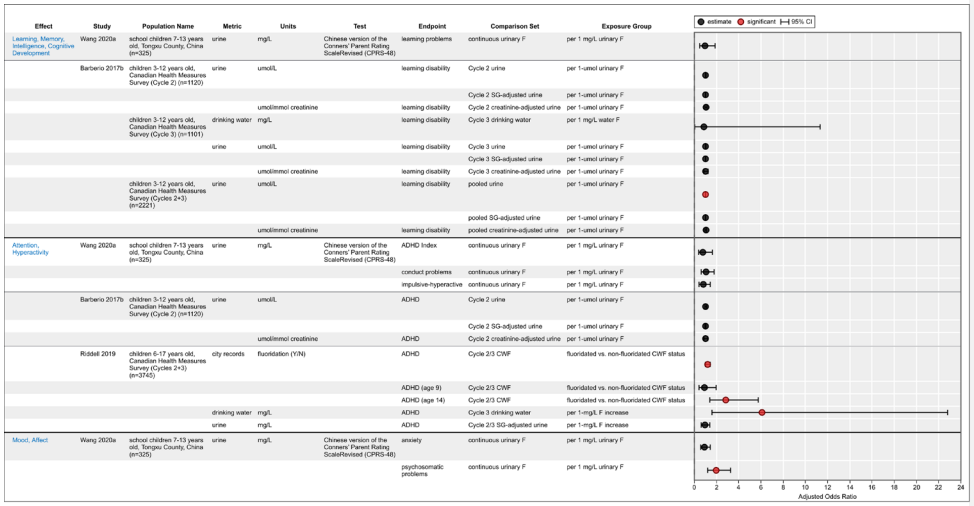


Figure A-12. Associations between Fluoride Exposure and Other Neurodevelopmental Effects in Children (Low Risk-of-bias Studies; Presented as Adjusted OR)

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-12 and additional study details in HAWC here (NTP 2019). "F" represents fluoride. Drinking water results for Barberio et al. (2017b) have a large confidence interval and are not completely visible in the figure. 95% CIs are 0.068–11.33 and can be viewed in HAWC (NTP 2019) by clicking the OR within the plot area.

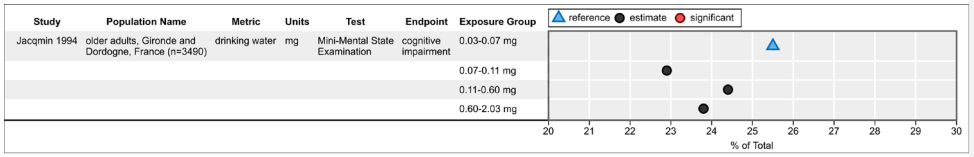


Figure A-13. Cognitive Impairment in Adults by Fluoride Exposure (Low Risk-of-bias Studies; Presented as % of Total Group)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-13 and additional study details in HAWC here (NTP 2019). Results from Li et al. (2016) suggested that fluoride exposure may be a risk factor for cognitive impairment in elderly subjects; however, results from the study were not conducive to presentation in this visualization.

Appendix B. Literature Search and Document Review Details

Table of Contents

Tables

B.1. Introduction

NTP initially published a systematic review of the experimental animal literature in 2016 that was subsequently expanded to include human epidemiological studies, mechanistic studies, and newer experimental animal literature. Table B-1 provides a timeline of key activities contributing to the 2022 NTP monograph including the multiple literature searches, draft monographs, and document review activities that have occurred since 2016.

Table B-2 is a summary of the specific search terms used for the PubMed database. In order to ensure inclusion of relevant papers, the strategy for this search was broad for the consideration of neurodevelopmental or cognitive endpoints and comprehensive for fluoride as an exposure or treatment. The specific search strategies for other databases are available in the protocol (https://ntp.niehs.nih.gov/go/785076).

Table B-1. Literature Search and Document Review Timeline

Date	Action
July 2016	Published 2016 NTP monograph of the systematic literature review on the effects of fluoride on learning and memory in animals only
June 2017	Published protocol for a new NTP monograph on systematic review on effects of fluoride on neurodevelopment and cognition from evidence in human, experimental animal, and mechanistic data
April 2019	Completed final literature search for 2019 draft NTP monograph on human, experimental animal, and mechanistic data (i.e., updated through April 2019)
May 2019	Published 2019 revised protocol for 2019 draft NTP monograph
September 2019	Sent 2019 draft NTP monograph for review by NASEM committee
February 2020	Received NASEM committee's review report of 2019 draft NTP monograph; began the following key changes in response to NASEM report: • Expanded literature search to non-English-language databases • Conducted meta-analysis on children's IQ and fluoride exposure • Revised protocol for monograph to include additional information.
May 2020	Completed final literature search for 2020 draft NTP monograph on human experimental animal and mechanistic data (i.e., updated through May 2020 and expanded to include non-English-language databases)
September 2020	Published 2020 revised protocol for 2020 draft NTP monograph
September 2020	Sent 2020 draft NTP monograph for second review by NASEM committee
February 2021	Received NASEM committee's review report of revised 2020 draft NTP monograph; made the following key changes in response to NASEM report: • Removed hazard step and hazard conclusions • Removed meta-analysis to publish separately.
December 2021	Sent 2021 draft NTP monograph on the state of the science for external peer review
April 2022	Published final 2022 NTP monograph on the state of the science

Commented [A107]: This paragraph and Table B-1 were added to provide further clarity on the progression of this multiyear assessment, in response to related comments from [redacted]Reviewers listed below; see DocI_Monograph and DocJ_Monograph, respectively, for detailed response. Note, one of the reviewers provided four separate comments on this issue.

Reviewer comment (DocI_Monograph, page 8): The rationale for date selection needs to be more clearly articulated. The specific dates are included in the appendix, perhaps they could be included in the main text for clarity in the methods.

Reviewer comment (DocJ_Monograph, page 4): Because this review has an extensive history that could be difficult for a reader to follow (i.e., the original 2016 review, and drafts from 2019, 2020, and the current draft), it would be helpful to develop a table or flowchart that documents that history. For example, you may consider noting the purpose/research question, findings, and noteworthy differences from previous/subsequent versions.
See comments below, but the literature search section, in particular, was a little difficult to follow - and having the "big picture" of the review in a table or flowchart to refer to, would better allow the reader to follow all of the searches conducted, and how the differ, yet fit together to contribute to the present document.

Reviewer comment (DocJ_Monograph, page 6): The literature search section was somewhat confusing to follow, though, given the complexity of updating reviews, etc it is understandable why multiple searches were conducted. See previous comment regarding the various iterations of this review, historically, and how a table or flowchart may help the reader understand the progression of this review, and thus, better follow the searches that were carried out.
For example, you may consider adding sub-headings within this section to distinguish which searches were run to capture which types of studies.

Reviewer comment (DocJ_Monograph, page 4): It is not clear why the "hazard assessment step" was removed from the methodology. Is it because the authors deemed the step not possible based on available evidence? Or is it because the hazard assessment step will occur separately, taking into consideration both the review and the results of meta-analysis?

Reviewer comment (DocJ_Monograph, page 10): It is not clear why the meta-analysis portion of this review is being prepared as a separate report.

Table B-2. PubMed Search Terms

Database	Search Terms
PUBMED	((Fluorides[mh:noexp] OR fluorides, topical[mh] OR sodium fluoride[mh] OR Fluorosis, Dental[mh] OR fluorosis[tiab] OR fluorid*[tiab] OR flurid*[tiab] OR fluorin*[tiab] OR florin*[tiab]) NOT (18F[tiab] OR f-18[tiab] OR 19F[tiab] OR f-19[tiab] OR f-labeled[tiab] OR "fluorine-18"[tiab] OR "fluorine-19"[tiab] OR pet-scan[tiab] OR radioligand*[tiab])) AND ((Aryl Hydrocarbon Hydroxylases[mh] OR Aryl Hydrocarbon Receptor Nuclear Translocator[mh] OR Behavior and Behavior Mechanisms[mh] OR Gene Expression Regulation[mh] OR Glucuronosyltransferase[mh] OR Intelligence tests[mh] OR Malate Dehydrogenase[mh] OR Mediator Complex Subunit 1[mh] OR Mental disorders[mh] OR Mental processes[mh] OR Monocarboxylic Acid Transporters[mh] OR Myelin Basic Protein[mh] OR nervous system[mh] OR nervous system diseases[mh] OR nervous system physiological phenomena[mh] OR Neurogranin[mh] OR Oligodendroglia[mh] OR Peroxisome Proliferator-Activated Receptors[mh] OR Psychological Phenomena and Processes[mh] OR Receptors, thyroid hormone[mh] OR Receptors, thyrotropin[mh] OR Retinoid X Receptors[mh] OR thyroid diseases[mh] OR thyroid hormones[mh] OR Thyrotropin-releasing hormone[mh] OR Thyroxine-Binding Proteins[mh] OR Pregnane X Receptor[supplementary concept] OR thyroid-hormone-receptor interacting protein[supplementary concept] OR Constitutive androstane receptor[supplementary concept] OR Academic performance[tiab] OR auditory[tiab] OR cortical[tiab] OR delayed development[tiab] OR developmental impairment[tiab] OR developmental-delay*[tiab] OR developmental-disorder*[tiab] OR euthyroid[tiab] OR gait[tiab] OR glia*[tiab] OR gliogenesis[tiab] OR hyperactiv*[tiab] OR impulse-control[tiab] OR iodide peroxidase[tiab] OR IQ[tiab] OR ischemi*[tiab] OR locomotor[tiab] OR mental deficiency[tiab] OR mental development[tiab] OR mental illness[tiab] OR mental-deficit[tiab] OR mobility[tiab] OR mood[tiab] OR morris-maze[tiab] OR morris-water[tiab] OR motor abilit*[tiab] OR Motor activities[tiab] OR motor performance[tiab] OR nerve[tiab] OR neural[tiab] OR neurobehav*[tiab] OR Neurocognitive impairment[tiab] OR neurodegenerat*[tiab] OR Neurodevelopment*[tiab] OR neurodisease*[tiab] OR neurologic*[tiab] OR neuromuscular[tiab] OR neuron*[tiab] OR neuropath*[tiab] OR obsessive compulsive[tiab] OR OCD[tiab] OR olfaction[tiab] OR olfactory[tiab] OR open-field-test[tiab] OR passive avoidance[tiab] OR plasticity[tiab] OR senil*[tiab] OR sociab*[tiab] OR speech*[tiab] OR spelling[tiab] OR stereotypic-movement*[tiab] OR synap*[tiab] OR tauopath*[tiab] OR Thyroglobulin[tiab] OR Thyroid disease*[tiab] OR thyroid gland[tiab] OR thyroid hormone*[tiab] OR thyronine*[tiab] OR visual motor[tiab] OR Visuospatial processing[tiab] OR water maze[tiab]) OR ((active-avoidance[tiab] OR ADHD[tiab] OR alzheimer*[tiab] OR amygdala[tiab] OR antisocial[tiab] OR anxiety[tiab] OR anxious[tiab] OR asperger*[tiab] OR attention deficit[tiab] OR autism[tiab] OR autistic[tiab] OR behavioral[tiab] OR behaviors[tiab] OR behavioural[tiab] OR behaviours[tiab] OR bipolar[tiab] OR cerebellum[tiab] OR cognition[tiab] OR cognitive[tiab] OR communication-disorder*[tiab] OR comprehension[tiab] OR cranial[tiab] OR dementia[tiab] OR dendrit*[tiab] OR dentate-gyrus[tiab] OR depression[tiab] OR dextrothyroxine[tiab] OR diiodothyronine*[tiab] OR diiodotyrosine[tiab] OR down syndrome[tiab] OR dyslexia[tiab] OR entorhinal cortex[tiab] OR epilep*[tiab] OR gangli*[tiab] OR goiter[tiab] OR graves-disease[tiab] OR hearing[tiab] OR hippocamp*[tiab] OR human development[tiab] OR hyperthyroid*[tiab] OR hypothalam*[tiab] OR hypothyroid*[tiab] OR impulsiv*[tiab] OR Intellectual disability[tiab] OR intelligence[tiab] OR language[tiab] OR learning[tiab] OR lewy bod*[tiab] OR long-term potentiation[tiab] OR long-term synaptic depression[tiab] OR memory[tiab] OR mental disorder*[tiab] OR mental recall[tiab] OR monoiodotyrosine[tiab] OR Motor activity[tiab] OR motor skill*[tiab] OR multiple sclerosis[tiab] OR myxedema[tiab] OR Nervous system[tiab] OR nervous-system[tiab] OR neurit*[tiab] OR optic[tiab] OR palsy[tiab] OR panic[tiab] OR parahippocamp*[tiab] OR paranoia[tiab] OR paranoid[tiab] OR parkinson*[tiab] OR perception[tiab] OR perforant*[tiab] OR personality[tiab] OR phobia[tiab] OR problem solving[tiab] OR proprioception[tiab] OR psychomotor[tiab] OR reflex[tiab] OR risk taking[tiab] OR schizophrenia[tiab] OR seizure*[tiab] OR sensation*[tiab] OR sleep[tiab] OR smell[tiab] OR spatial behavior[tiab] OR stroke[tiab] OR substantia-nigra[tiab] OR taste[tiab] OR thyroiditis[tiab] OR thyrotoxicosis[tiab] OR Thyrotropin[tiab] OR thyroxine[tiab] OR triiodothyronine[tiab] OR vision[tiab]) NOT medline[sb]))

Appendix C. Detailed Literature Search Results and List of Included Studies

Table of Contents

C.1. Detailed Literature Search Results

C.1.1. Literature Search Results Counts and Title and Abstract Screening

The electronic database searches retrieved 25,450 unique references in total (20,883 references during the initial search conducted in December 2016, 3,657 references during the literature search updates [including the final updated search conducted for the primary epidemiological studies on May 1, 2020], and 910 references from the supplemental Chinese database searches); 11 additional references were identified by technical advisors or from reviewing reference lists in published reviews and included studies. As a result of title and abstract screening, 1,036 references were moved to full-text review, and 24,425 references were excluded (11,402 by manual screening for not satisfying the PECO criteria and 13,023 based on the SWIFT-Active Screener algorithm).

C.1.2. Full-text Review

Among the 1,036 references that underwent full-text review, 489 were excluded at that stage with reasons for exclusion documented; 333 references were excluded for not satisfying the PECO criteria; and 156 references from the May 2020 searches (main literature search update and supplemental Chinese database searches) were excluded for not including information that would materially advance the human, animal in vivo, or mechanistic findings (see the Main Literature Search section for a description of the methodology). These screening results are outlined in a study selection diagram that reports numbers of studies excluded for each reason at the full-text review stage (see Figure 2) [using reporting practices outlined in ~~Moher et al. (2009)~~Page et al. (2021)]. After full-text review, 547 studies were considered relevant with primary neurodevelopmental or cognitive outcomes, secondary neurobehavioral outcomes, and/or outcomes related to thyroid function. A few studies assessed data for more than one evidence stream (human, non-human mammal, and/or in vitro), and several human and animal studies assessed more than one type of outcome (e.g., primary and secondary outcomes). The number of included studies is summarized below:

- 167 human studies (84 primary only; 13 secondary only; 5 primary and secondary; 8 primary and thyroid; 2 secondary and thyroid; and 55 thyroid only);
- 339 non-human mammal studies (7 primary only; 186 secondary only; 67 primary and secondary; 6 primary, secondary, and thyroid; 4 secondary and thyroid; and 69 thyroid only); and,
- 60 in vitro/mechanistic studies (48 neurological and 12 thyroid).

One publication contained human, experimental non-human mammal, and in vitro data. Three publications contained both human and experimental non-human mammal data. Fourteen publications contained data relevant to both experimental non-human mammal studies and in vitro studies.

Commented [A108]: Change made in response to the [redacted] Reviewer comment below; see DocB1_Monograph for detailed response.

Reviewer comment (DocB1_Monograph, page 5): The authors are using an old version of the PRISMA flow diagram - The 2020 PRISMA flow diagram can be found here: https://prisma-statement.org/prismastatement/flowdiagram.aspx

C.2. List of Included Studies

C.2.1. Studies in Humans

As described in Figure 2, 167 human studies were included; however, full data extraction was conducted only on studies with neurological outcomes or thyroid hormone data. Data extraction was completed using HAWC (NTP 2019). Data were extracted from a subset of included studies in humans (n = 124) and are available in HAWC based on outcome. The following lists of references are organized as studies that are available in HAWC followed by studies that are not available in HAWC. Specifically, data for primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence) and secondary neurobehavioral outcomes (anxiety, aggression, motor activity, or biochemical changes), as well as thyroid hormone level data, were extracted from included human studies and are available in HAWC. Data for included studies identified through the 2020 literature search update were extracted only for primary neurodevelopmental or cognitive outcomes; a subset of these studies (n = 7) also included secondary neurobehavioral outcomes and/or thyroid hormone level data that were not extracted because those data would not materially advance the human or mechanistic findings. Included human studies that evaluated only other thyroid-related effects such as goiters or thyroid size (n = 43) were not extracted and are not available in HAWC. The list below presents the 167 human studies that were included in the review. An overview of the screening results is outlined in the study selection diagram (Figure 2) that reports numbers of included studies as well as numbers of studies excluded for each reason at the full-text review stage.

C.2.1.1. Studies Available in HAWC

An J, Mei S, Liu A, Fu Y, Wang C. 1992. [Effect of high level of fluoride on children's intelligence]. *Chin J Control Endem Dis* 7(2): 93-94.

Aravind A, Dhanya RS, Narayan A, Sam G, Adarsh VJ, Kiran M. 2016. Effect of fluoridated water on intelligence in 10-12-year-old school children. *J Int Soc Prev Community Dent* 6(Suppl 3): S237-S242.

Bai A, Li Y, Fan Z, Li X, Li P. 2014. [Intelligence and growth development of children in coal-burning-borne arsenism and fluorosis areas: An investigation study]. *Chin J Endemiol* 33(2): 160-163.

Barberio AM, Hosein FS, Quinonez C, McLaren L. 2017. Fluoride exposure and indicators of thyroid functioning in the Canadian population: Implications for community water fluoridation. *J Epidemiol Community Health* 71: 1019-1025.

Barberio AM, Quinonez C, Hosein FS, McLaren L. 2017. Fluoride exposure and reported learning disability diagnosis among Canadian children: Implications for community water fluoridation. *Can J Public Health* 108: 229-239.

Bashash M, Thomas D, Hu H, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Ettinger AS, Wright R, Zhang Z, Liu Y, Schnaas L, Mercado-Garcia A, Tellez-Rojo MM, Hernandez-Avila M. 2017. Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6-12 years of age in Mexico. *Environ Health Perspect* 125(9): 1-12.

Bashash M, Marchand M, Hu H, Till C, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Green R, Schnaas L, Mercado-Garcia A, Hernandez-Avila M, Tellez-Rojo MM. 2018. Prenatal fluoride exposure and attention deficit hyperactivity disorder (ADHD) symptoms in children at 6-12 years of age in Mexico City. *Environ Int* 121(Pt 1): 658-666.

Broadbent JM, Thomson WM, Moffitt TE, Poulton R. 2015. Community water fluoridation and intelligence response. *Am J Public Health* 105: 3-4.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 1991. [Research on the intellectual development of children in high fluoride areas]. *Chin J Control Endem Dis* 6(Suppl): 99-100.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 2008. Research on the intellectual development of children in high fluoride areas. *Fluoride* 41: 120-124.

Chinoy NJ, Narayana MV. 1992. Studies on fluorosis in Mehsana District of North Gujarat. *Proc Zool Soc* 45: 157-161.

Choi AL, Zhang Y, Sun G, Bellinger DC, Wang K, Yang XJ, Li JS, Zheng Q, Fu Y, Grandjean P. 2015. Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study. *Neurotoxicol Teratol* 47: 96-101.

Cui Y, Zhang B, Ma J, Wang Y, Zhao L, Hou C, Yu J, Zhao Y, Zhang Z, Nie J, Gao T, Zhou G, Liu H. 2018. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. *Ecotoxicol Environ Saf* 165: 270-277.

Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. 2020. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. *Neurosci Lett* 729: 134981.

Das K, Mondal NK. 2016. Dental fluorosis and urinary fluoride concentration as a reflection of fluoride exposure and its impact on IQ level and BMI of children of Laxmisagar, Simlapal Block of Bankura District, W.B., India. *Environ Monit Assess* 188: 218.

Ding Y, Sun H, Han H, Wang W, Ji X, Liu X, Sun D. 2011. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. *J Hazard Mater* 186: 1942-1946.

Du L, Wan C, Cao X, Liu J. 1992. [The effect of fluorine on the developing human brain]. *Chin J Pathol* 21(4): 218-220.

Du L, Wan C, Cao X, Liu J. 2008. The effect of fluorine on the developing human brain. *Fluoride* 41: 327-330.

Duan J, Zhao M, Wang L, Fang D, Wang Y, Wang W. 1995. A comparative analysis of the results of multiple tests in patients with chronic industrial fluorosis. Guizhou Med J 18(3): 179-180. Erickson JD, Hay S. 1976. Water fluoridation and congenital malformations: No association. *J Am Dent Assoc* 93: 981-984.

Erickson JD, Hay S. 1976. Water fluoridation and congenital malformations: No association. *J Am Dent Assoc* 93: 981-984.

Erickson JD. 1980. Down syndrome, water fluoridation, and maternal age. *Teratology* 21: 177-180.

Eswar P, Nagesh L, Devaraj CG. 2011. Intelligent quotients of 12-14 year old school children in a high and low fluoride village in India. *Fluoride* 44: 168-172.

Fan Z, Dai H, Bai A, Li P, Li T, Li G. 2007. Effect of high fluoride exposure in children's intelligence. *J Environ Health* 24(10): 802-803.

Green R, Lanphear B, Hornung R, Flora D, Martinez-Mier EA, Neufeld R, Ayotte P, Muckle G, Till C. 2019. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. *JAMA Pediatr*: E1-E9.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 1991. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. *Chin J Epidemiol* 10(2): 98-100.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 2008. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. *Fluoride* 41: 125-128.

Guo ZY, He YH, Zhu QX. 2001. [Research on the neurobehavioral function of workers occupationally exposed to fluoride]. *Ind Hlth & Occup Dis* 27(6): 346-348.

Guo ZY, He YH, Zhu QX. 2008. Research on the neurobehavioral function of workers occupationally exposed to fluoride. *Fluoride* 41: 152-155.

He H, Cheng ZS, Liu WQ. 1989. [Effects of fluorine on the human fetus]. *J Control Endem Dis* 4(3): 136-138.

He H, Cheng ZS, Liu WQ. 2008. Effects of fluorine on the human fetus. *Fluoride* 41: 321-326.

He MX, Zhang CN. 2010. [Investigation of children's intelligence quotient and dental fluorosis in drinking water-type of endemic fluorosis area in Pucheng County, Shaanxi Province before and after drinking water change]. *Chin J Endemiol* 29: 547-548.

Hong F, Wang H, Yang D, Zhang Z. 2001. [Investigation on the intelligence and metabolism of iodine and fluoride in children with high iodine and fluoride]. *Chin J Control Endem Dis* 12-14.

Hong FG, Cao YX, Yang D, Wang H. 2001. [Research on the effects of fluoride on child intellectual development under different environmental conditions]. *Chin Prim Health Care* 15(3): 56-57.

Hong FG, Cao YX, Yang D, Wang H. 2008. Research on the effects of fluoride on child intellectual development under different environmental conditions. *Fluoride* 41: 156-160.

Hosur MB, Puranik RS, Vanaki S, Puranik SR. 2012. Study of thyroid hormones free triiodothyronine (FT3), free thyroxine (FT4) and thyroid stimulating hormone (TSH) in subjects with dental fluorosis. *Eur J Dent* 6: 184-190.

Jacqmin H, Commenges D, Letenneur L, Barberger-Gateau P, Dartigues JF. 1994. Components of drinking water and risk of cognitive impairment in the elderly. *Am J Epidemiol* 139: 48-57.

Jin T, Han T, Wei Y, Wu Y, Wang Z, Zhang, H. 2016. [Investigation on working memory level of children aged 8-12 years in coal-burning fluorosis area]. *J Environ Health* 33(5): 409-411.

Jin T, Wang Z, Wei Y, Wu Y, Han T, Zhang H. 2017. [Investigation on intelligence level of children aged 8-12 years old in coal-burning fluorosis area]. *J Environ Health* 34(3): 229-231.

Kang J, Cheng Y, Wu K, Lin S, He G, Jin Y. 2011. Effect of exposure to fluoride and arsenic in drinking water of Hangjinhouqi on children's intelligence. *Chinese School Health*: 679-681.

Karimzade S, Aghaei M, Mahvi AH. 2014. Investigation of intelligence quotient in 9-12 year-old children exposed to high- and low-drinking water fluoride in West Azerbaijan Province, Iran. *Fluoride* 47: 9-14.

Khan SA, Singh RK, Navit S, Chadha D, Johri N, Navit P, Sharma A, Bahuguna R. 2015. Relationship between dental fluorosis and intelligence quotient of school going children in and around Lucknow District: A cross-sectional study. *J Clin Diagn Res* 9(11): 10-15.

Khandare AL, Gourineni SR, Validandi V. 2017. Dental fluorosis, nutritional status, kidney damage, and thyroid function along with bone metabolic indicators in school-going children living in fluoride-affected hilly areas of Doda District, Jammu and Kashmir, India. *Environ Monit Assess* 189: 579.

Khandare AL, Validandi V, Gourineni SR, Gopalan V, Nagalla B. 2018. Dose-dependent effect of fluoride on clinical and subclinical indices of fluorosis in school going children and its mitigation by supply of safe drinking water for 5 years: An Indian study. *Environ Monit Assess* 190: 110.

Kheradpisheh Z, Mahvi AH, Mirzaei M, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018. Correlation between drinking water fluoride and TSH hormone by ANNs and ANFIS. *J Environ Health Sci Eng* 16(1): 11-18.

Kheradpisheh Z, Mirzaei M, Mahvi AH, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018. Impact of drinking water fluoride on human thyroid hormones: A case-control study. *Sci Rep* 8: 2674.

Kumar V, Chahar P, Kajjari S, Rahman F, Bansal DK, Kapadia JM. 2018. Fluoride, thyroid hormone derangements and its correlation with tooth eruption pattern among the pediatric population from endemic and non-endemic fluorosis areas. *J Contemp Dent Pract* 19(12): 1512-1516.

Kundu H, Basavaraj P, Singla A, Gupta R, Singh K, Jain S. 2015. Effect of fluoride in drinking water on children's intelligence in high and low fluoride areas of Delhi. *J Indian Assoc Public Health Dent* 13(2): 116-121.

Lamberg M, Hausen H, Vartiainen T. 1997. Symptoms experienced during periods of actual and supposed water fluoridation. *Community Dent Oral Epidemiol* 25: 291-295.

Li F, Chen X, Huang R, Xie Y. 2009. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. *J Environ Health* 26(4): 838-840.

Li J, Yao L, Shao QL, Wu CY. 2004. [Effects of high fluoride level on neonatal neurobehavioral development]. *Chin J Endemiol* 23(5): 463-465.

Li J, Yao L, Shao QL, Wu CY. 2008. Effects of high fluoride level on neonatal neurobehavioral development. *Fluoride* 41: 165-170.

Li M, Gao Y, Cui J, Li Y, Li B, Liu Y, Sun J, Liu X, Liu H, Zhao L, Sun D. 2016. Cognitive impairment and risk factors in elderly people living in fluorosis areas in China. *Biol Trace Elem Res* 172: 53-60.

Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. 2010. Investigation and analysis of children's intelligence and dental fluorosis in high fluoride area. *J Med Pest Control* 26(3): 230-231.

Li XS, Zhi JL, Gao RO. 1995. Effect of fluoride exposure on the intelligence of children. *Fluoride* 28: 189-192.

Li Y, Li X, Wei S. 1994. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. *J West Chin University Med Sci* 25(2): 188-191.

Li Y, Li X, Wei S. 2008. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. *Fluoride* 41: 331-335.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2003. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. *Chin J Public Health Manag* 19(4): 337-338.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2008. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. *Fluoride* 41: 161-164.

Liang C, Ji R, Cao J, Jiang Y, Yu B, Ma F, Wu Y, Ying B, Zhang Y, Sun S, Li Y, Emsley CL, Gao S, Hall KS, Hendrie HC. 2003. [Study on the relationship between drinking water trace elements and cognitive ability of the elderly]. *Health Res* 436-440.

Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. 1991. High fluoride and low iodine environment and subclinical cretinism in Xinjiang. *Endem Dis Bull* 6(2): 62-67.

Liu S, Lu Y, Sun Z, Wu L, Wang X, Yan S. 2000. [Report on the intellectual ability of children living in high-fluoride water areas]. *Chin J Control Endem Dis* 15(4): 231-232.

Liu SL, Lu Y, Sun ZR, Wu L, Lu WL, Wang XW, Yan S. 2008. Report on the intellectual ability of children living in high-fluoride water areas. *Fluoride* 41: 144-147.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S, Mentes A, DenBesten PK. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. *Physiol Behav* 206: 76-83.

Lu Y, Sun ZR, Wu LN, Wang X, Lu W, Liu SS. 2000. Effect of high-fluoride water on intelligence in children. *Fluoride* 33: 74-78.

Luo Y, Ma R, Liu Z, Guan Z, Lou D, Zheng D. 2018. [Intelligence investigation and forensic significance of children in coal-burning fluorosis area]. *Chin J Forensic Medicine* 33(6): 590-593.

Malin AJ, Till C. 2015. Exposure to fluoridated water and attention deficit hyperactivity disorder prevalence among children and adolescents in the United States: An ecological association. *Environ Health* 14: 17.

Malin AJ, Riddell J, McCague H, Till C. 2018. Fluoride exposure and thyroid function among adults living in Canada: Effect modification by iodine status. *Environ Int* 121(Pt 1): 667-674.

Michael M, Barot VV, Chinoy NJ. 1996. Investigations of soft tissue functions in fluorotic individuals of North Gujarat. *Fluoride* 29: 63-71.

Mondal D, Dutta G, Gupta S. 2016. Inferring the fluoride hydrogeochemistry and effect of consuming fluoride-contaminated drinking water on human health in some endemic areas of Birbhum District, West Bengal. *Environ Geochem Health* 38: 557-576.

Morgan L, Allred E, Tavares M, Bellinger D, Needleman H. 1998. Investigation of the possible associations between fluorosis, fluoride exposure, and childhood behavior problems. *Pediatr Dent* 20: 244-252.

Mustafa DE, Younis UM, Elhag SAA. 2018. The relationship between the fluoride levels in drinking water and the schooling performance of children in rural areas of Khartoum State, Sudan. *Fluoride* 51(2): 102-113.

Nagarajappa R, Pujara P, Sharda AJ, Asawa K, Tak M, Aapaliya P, Bhanushali N. 2013. Comparative assessment of intelligence quotient among children living in high and low fluoride areas of Kutch, India: A pilot study. *Iran J Public Health* 42: 813-818.

Peckham S, Lowery D, Spencer S. 2015. Are fluoride levels in drinking water associated with hypothyroidism prevalence in England? A large observational study of GP practice data and fluoride levels in drinking water. *J Epidemiol Community Health* 69: 619-624.

Poureslami HR, Horri A, Garrusi B. 2011. A comparative study of the IQ of children age 7-9 in a high and a low fluoride water city in Iran. *Fluoride* 44: 163-167.

Pratap SV, Singh CD, Sandeep T, Sandeep K, Vikas G, Mukesh T, Anurag T. 2013. A correlation between serum vitamin, acetylcholinesterase activity and IQ in children with excessive endemic fluoride exposure in Rajasthan, India. *Int Res J Medical Sci* 1(3): 12-16.

Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. 1990. [Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children]. *J Control Endem Dis* 5(4): 203-204.

Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. 2008. Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children. *Fluoride* 41: 115-119.

Ratan SK, Rattan KN, Pandey RM, Singhal S, Kharab S, Bala M, Singh V, Jhanwar A. 2008. Evaluation of the levels of folate, vitamin B12, homocysteine and fluoride in the parents and the

affected neonates with neural tube defect and their matched controls. *Pediatr Surg Int* 24: 803-808.

Razdan P, Patthi B, Kumar JK, Agnihotri N, Chaudhari P, Prasad M. 2017. Effect of fluoride concentration in drinking water on intelligence quotient of 12-14-year-old children in Mathura District: A cross-sectional study. *J Int Soc Prev Community Dent* 7: 252-258.

Ren D, Li K, Liu D. 1989. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. *Chin J Control Endem Dis* 4(4): 251.

Ren D, Li K, Liu D. 2008. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. *Fluoride* 41: 319-320.

Riddell JK, Malin AJ, Flora D, McCague H, Till C. 2019. Association of water fluoride and urinary fluoride concentrations with attention deficit hyperactivity disorder in Canadian youth. *Environ Int* 133: 105190.

Rocha-Amador D, Navarro ME, Carrizales L, Morales R, Calderon J. 2007. Decreased intelligence in children and exposure to fluoride and arsenic in drinking water. *Cad Saude Publica* 23 Suppl 4: S579-587.

Rocha-Amador D, Navarro M, Trejo-Acevedo A, Carrizales L, Perez-Maldonado I, Diaz-Barriga F, Calderon J. 2009. Use of the Rey-Osterrieth Complex Figure Test for neurotoxicity evaluation of mixtures in children. *Neurotox* 30: 1149-1154.

Rotton J, Tikofsky RS, Feldman HT. 1982. Behavioral effects of chemicals in drinking water. *J Appl Psychol* 67: 230-238.

Russ TC, Killin LOJ, Hannah J, Batty GD, Deary IJ, Starr JM. 2019. Aluminium and fluoride in drinking water in relation to later dementia risk. *Brit J Psychol* 14: 1-6.

Saxena S, Sahay A, Goel P. 2012. Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India. *J Neurosci Rural Pract* 3: 144-149.

Sebastian ST, Sunitha S. 2015. A cross-sectional study to assess the intelligence quotient (IQ) of school going children aged 10-12 years in villages of Mysore District, India with different fluoride levels. *J Indian Soc Pedod Prev Dent* 33: 307-311.

Seraj B, Shahrabi M, Falahzade M, Falahzade F, Akhondi N. 2006. Effect of high fluoride concentration in drinking water on children's intelligence. *J Dent Med* 19(2): 80-86.

Seraj B, Shahrabi M, Shadfar M, Ahmadi R, Fallahzadeh M, Eslamlu HF, Kharazifard MJ. 2012. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. *J Dent* 9: 221-229.

Shannon FT, Fergusson DM, Horwood LJ. 1986. Exposure to fluoridated public water supplies and child health and behaviour. *New Zeal Med J* 99: 416-418.

Shao Q. 2003. Study of cognitive function impairment caused by chronic fluorosis. *Chin J Endemiol* 22(4): 336-338.

Sharma JD, Sohu D, Jain P. 2009. Prevalence of neurological manifestations in a human population exposed to fluoride in drinking water. *Fluoride* 42: 127-132.

Shivaprakash PK, Ohri K, Noorani H. 2011. Relation between dental fluorosis and intelligence quotient in school children of Bagalkot District. *J Indian Soc Pedod Prev Dent* 29: 117-120.

Singh A, Jolly SS, Devi P, Bansal BC, Singh SS. 1962. Endemic fluorosis: An epidemiological, biochemical and clinical study in the Bhatinda District of Panjab. *Indian J Med Res* 50: 387-398.

Singh N, Verma KG, Verma P, Sidhu GK, Sachdeva S. 2014. A comparative study of fluoride ingestion levels, serum thyroid hormone & TSH level derangements, dental fluorosis status among school children from endemic and non-endemic fluorosis areas. *Springerplus* 3: 7.

Soto-Barreras U, Escalante-Villalobos KY, Holguin-Loya B, Perez-Aguirre B, Nevarez-Rascon A, Martinez-Martinez RE, Loyola-Rodriguez JP. 2019. Effect of fluoride in drinking water on dental caries and IQ in children. *Fluoride* 52: 474-482.

Still CN, Kelley P. 1980. The incidence of primary degenerative dementia vs. water fluoride content in South Carolina, USA. *Neurotox* 1: 125-132.

Sudhir KM, Chandu GN, Prashant GM, Subba Reddy VV. 2009. Effect of fluoride exposure on intelligence quotient (IQ) among 13-15 year old school children of known endemic area of fluorosis, Nalgonda District, Andhra Pradesh. *J Indian Assoc Public Health Dent* 2009(13): 88-94.

Sun M, Li S, Wang Y, Li F. 1991. Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis. *J Guiyang Med Coll* 16(3): 204-206.

Susheela AK, Bhatnagar M, Vig K, Mondal NK. 2005. Excess fluoride ingestion and thyroid hormone derangements in children living in Delhi, India. *Fluoride* 38: 98-108.

Tamboli BL, Mathur GM, Mathur AP, Lalla SK, Goyal OP. 1980. Prevalence of fluorosis in Pratabpura and Surajpura villages, District Ajmer (Rajasthan). *Indian J Med Res* 71: 57-67.

Till C, Green R, Flora D, Hornung R, Martinez-Mier EA, Blazer M, Farmus L, Ayotte P, Muckle G, Lanphear B. 2020. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. *Environ Int* 134: 105315.

Tripathi P, Sultana N. 2007. Fluoride content of groundwater and prevalence of dental, skeletal and neurological stage of fluorosis in Tehsil Purwa of Unnao. *Indian J Environ Prot* 27: 737-739.

Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. 2007. Effect of high fluoride water on intelligence of school children in India. *Fluoride* 40: 178-183.

Trivedi M, Sangai N, Patel R, Payak M, Vyas S. 2012. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. *Fluoride* 45(4): 377-383.

Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, Costilla-Salazar R, Calderon Hernandez J, Alcaraz Contreras Y, Rocha-Amador DO. 2017. In utero exposure to fluoride and cognitive development delay in infants. *Neurotox* 59: 65-70.

Wang A, Duan L, Huang H, Ma J, Zhang Y, Ma Q, Guo Y, Li Z, Cheng X, Zhu J, Zhou G, Ba Y. 2020. Association between fluoride exposure and behavioural outcomes of school-age children: A pilot study in China. *Int J Environ Health Res*: 1-10.

Wang G, Yang D, Jia F, Wang H. 1996. [A study of the IQ levels of four- to seven-year-old children in high fluoride areas]. *Endem Dis Bull* 11(1): 60-62.

Wang G, Yang D, Jia F, Wang H. 2008. A study of the IQ levels of four- to seven-year-old children in high fluoride areas. *Fluoride* 41: 340-343.

Wang G, Gao M, Zhang M, Yang M, Xiang Q. 2012. Correlation between total fluoride intake and children's IQ. *J Southeast Univ Med Ed*: 743-746.

Wang G, Zhang M, Wang Q, Han A, Gao M, Lin P, Xiang Q. 2017. [Investigation on the relationship between serum fluoride content and IQ of children before and after reducing fluoride to water]. *Capital Public Health* 11(6): 274-277.

Wang M, Liu L, Li H, Li Y, Liu H, Hou C, Zeng Q, Li P, Zhao Q, Dong L, Zhou G, Yu X, Liu L, Guan Q, Zhang S, Wang A. 2020. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. *Environ Int* 134: 105229.

Wang S, Wang L, Hu P, Guo S, Law S. 2001. [Effects of high iodine and high fluorine on children's intelligence and thyroid function]. *Chin J Endemiol* 20(4): 288-290.

Wang SX, Wang ZH, Cheng XT, Li J, Sang ZP, Zhang XD, Han LL, Qiao XY, Wu ZM, Wang ZQ. 2005. [Investigation and evaluation on intelligence and growth of children in endemic fluorosis and arsenism areas]. *Chin J Endemiol* 24: 179-182.

Wang SX, Wang ZH, Cheng XT, Li J, Sang ZP, Zhang XD, Han LL, Qiao XY, Wu ZM, Wang ZQ. 2007. Arsenic and fluoride exposure in drinking water: Children's IQ and growth in Shanyin County, Shanxi Province, China. *Environ Health Perspect* 115: 643-647.

Wang SY, Zhang HX, Fan W, Fang SJ, Kang PP, Chen XG, Yu MJ. 2005. [The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children]. *J Appl Clin Ped* 20(9): 897-899.

Wang SY, Zhang HX, Fan W, Fang SJ, Kang PP, Chen XG, Yu MJ. 2008. The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children. *Fluoride* 41: 344-348.

Wang Z, Wang S, Zhang X, Li J, Zheng X, Hu C. 2006. Investigation of children's growth and development under long-term fluoride exposure. *Chin J Control Endem Dis* 21(4): 239-241.

Wei N, Li Y, Deng J, Xu S, Guan Z. 2014. [The effects of comprehensive control measures on intelligence of school-age children in coal-burning-borne endemic fluorosis areas]. *Chin J Endemiol* 33(2): 320-322.

Xiang Q, Liang Y, Chen L, Wang C, Chen B, Chen X, Zhou M. 2003. Effect of fluoride in drinking water on children's intelligence. *Fluoride* 36: 84-94.

Xiang Q, Liang Y, Chen B, Chen L. 2011. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. *Fluoride* 44: 191-194.

Xu Y, Lu C, Zhang X. 1994. Effect of fluoride on children's intelligence. *Endem Dis Bull* 2: 83-84.

Yang Y, Wang X, Guo X, Hu P. 1994. [The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine]. *Chin J Epidemiol* 15(4): 296-298.

Yang Y, Wang X, Guo X, Hu P. 2008. The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine. *Fluoride* 41: 336-339.

Yao L, Zhou J, Wang S, Cui K, Lin F. 1996. Analysis of TSH levels and intelligence of children residing in high fluorosis areas. *Lit Inf Prev Med* 2(1): 26-27.

Yao Y. 1997. Comparative assessment of the physical and mental development of children in endemic fluorosis area with water improvement and without water improvement. *Lit Inf Prev Med* 3(1): 42-43.

Yasmin S, Ranjan S, D'Souza D. 2013. Effect of excess fluoride ingestion on human thyroid function in Gaya Region, Bihar, India. *Toxicol Environ Chem* 95: 1235-1243.

Yazdi SM, Sharifian A, Dehghani-Beshne M, Momeni VR, Aminian O. 2011. Effects of fluoride on psychomotor performance and memory of aluminum potroom workers. *Fluoride* 44: 158-162.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 1996. [Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis]. *Chin J Endemiol* 15(5): 257-259.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 2008. Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis. *Fluoride* 41: 134-138.

Yu X, Chen J, Li Y, Liu H, Hou C, Zeng Q, Cui Y, Zhao L, Li P, Zhou Z, Pang S, Tang S, Tian K, Zhao Q, Dong L, Xu C, Zhang X, Zhang S, Liu L, Wang A. 2018. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. *Environ Int* 118: 116-124.

Zhang J, Yao H, Chen Y. 1998. [Effect of high level of fluoride and arsenic on children's intelligence]. *Chin J Public Health* 17(2): 57.

Zhang P, Cheng L. 2015. [Effect of coal-burning endemic fluorosis on children's physical development and intellectual level]. *Chin J Control Endem Dis* 30(6): 458-459.

Zhang S, Zhang X, Liu H, Qu W, Guan Z, Zeng Q, Jiang C, Gao H, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Cui Y, Yu L, Wang A. 2015. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. *Toxicol Sci* 144: 238-245.

Zhao LB, Liang GH, Zhang DN, Wu XR. 1996. Effect of a high fluoride water supply on children's intelligence. *Fluoride* 29: 190-192.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao Q, Tian Z, Zhou G, Niu Q, Chen J, Li P, Dong L, Xia T, Zhang S, Wang A. 2020. SIRT1-dependent mitochondrial biogenesis supports therapeutic effects of resveratrol against neurodevelopment damage by fluoride. *Theranostics* 10: 4822-4838.

Zhao Y, Cui Y, Yu J, Zhang B, Nie J, Zhao L, Zhang Z, Liu H. 2018. [Study on the relationship between water-borne high iodine and thyroid hormone and children's intelligence level]. *J Environ Health* 35(1): 6-9.

Zhou G, Tang S, Yang L, Niu Q, Chen J, Xia T, Wang S, Wang M, Zhao Q, Liu L, Li P, Dong L, Yang K, Zhang S, Wang A. 2019. Effects of long-term fluoride exposure on cognitive ability and the underlying mechanisms: Role of autophagy and its association with apoptosis. *Toxicol Appl Pharmacol* 378: 114608.

C.2.1.2. Studies Not Available in HAWC

Bachinskii PP, Gutsalenko OA, Naryzhniuk ND, Sidora VD, Shliakhta AI. 1985. [Action of the body fluorine of healthy persons and thyroidopathy patients on the function of hypophyseal-thyroid the system]. *Probl Endokrinol* 31: 25-29.

Balabolkin MI, Mikhailets ND, Lobovskaia RN, Chernousova NV. 1995. [The interrelationship of the thyroid and immune statuses of workers with long-term fluorine exposure]. *Ter Arkh* 67: 41-42.

Baum K, Boerner W, Reiners C, Moll E. 1981. [Bone density and thyroid function in adolescents in relation to fluoride content of drinking water]. *Fortschr Med* 99: 1470-1472.

Berry WTC, Whittles JH. 1963. Absence of effect of fluoride upon the incidence of thyroid enlargements in Wiltshire schoolgirls. *Mon Bull Minist Health Public Health Lab Serv* 22: 50-52.

Cherkinskii SN, Zaslavskaia RM. 1956. [Significance of fluorides in potable water in the development of endemic goiter]. *Probl Endokrinol Gormonoter* 2: 70-75.

Choubisa SL. 2001. Endemic fluorosis in southern Rajasthan, India. *Fluoride* 34: 61-70.

Chuka A, Zhukovskil V, Mirku I, Postel'Niku D. 1964. Prezhdevremennoe starenie naseleniya v zone rasprostraneniya endemicheskogo zoba. *Vestnik Akad Med Nauk Sssr* 19: 23-27.

Dai HX, Zeng P, Wang KY, Zhang XG, Ma ZJ, Zhou YG, Fan ZX, Guo SH. 2013. [Analysis of a survey results of patients with suspected high iodine goiter in Liuji Town Fuping County of Shaanxi Province]. *Chin J Endemiol* 32: 408-411.

Day T, Powell-Jackson P. 1972. Fluoride, water hardness, and endemic goitre. *Lancet* 299(7761): 1135-1138.

Desai VK, Solanki DM, Bansal RK. 1993. Epidemiological study of goitre in endemic fluorosis district of Gujarat. *Fluoride* 26: 187-190.

Díaz-Cadórniga FJ, Delgado E, Tartón T, Valdés MM, Méndez A, Fernández MT, Rojo C. 2003. Endemic goiter associated with high iodine intake in primary school children in the Saharawi Arab Democratic Republic. *Endocrinol Nutr* 50: 357-362.

Eichner R, Borner W, Henschler D, Kohler W, Moll E. 1981. [Osteoporosis therapy and thyroid function. Influence of 6 months of sodium fluoride treatment on thyroid function and bone density]. *Fortschr Med* 99: 342-348.

Fiorentini S, Galeazzi M, Visintin B. 1947. II fluoro in natura come agente morbigeno II. La fluorosi die Campagnano di Roma. III. Un focolaio di fluorosi umana a Campagnano di Roma. IV. Osservazioni radiologiche sui processi alveolari, sulle ossa mascellari, e sul paradenzio degli abitanti die Campagnano. V. Zona fluorotica intorno a Campagnano di Roma. VI. Frequenza e caratteri clinici della carie dentale in soggetti fluorotici. *Rend Ist Superiore Sanita* 10: 721-804.

Gabovich RD, Verzhikovskaia NV. 1958. [Effect of fluoride compounds on absorption of radioactive iodine by the thyroid gland in humans and in experimental conditions]. *Probl Endokrinol Gormonoter* 4: 49-54.

Galletti P, Held HR, Korrodi H, Wegmann T. 1956. [Investigations on the possible damaging effect of fluorine on the thyroid gland]. *Helv Med Acta* 23: 601-605.

Galletti PM, Joyet G, Jallut O. 1957. [Effect of sodium fluoride on thyroid function in Basedow's Disease]. *Helv Med Acta* 24: 209-215.

Galletti PM, Joyet G. 1958. Effect of fluorine on thyroidal iodine metabolism in hyperthyroidism. *J Clin Endocrinol Metab* 18: 1102-1110.

Gas'kov AI, Savchenkov MF, Iushkov NN. 2005. [The specific features of the development of iodine deficiencies in children living under environmental pollution with fluorine compounds]. *Gig Sanit*: 53-55.

Gedalia I, Brand N. 1963. The relationship of fluoride and iodine in drinking water in the occurrence of goiter. *Arch Int Pharmacodyn Ther* 142: 312-315.

Grimm H. 1973. [The physical development of schoolchildren under the influence of drinking water fluoridation in Karl Marx Stadt]. *Dtsch Gesundheitsw* 28: 2363-2369.

Hasling C, Nielsen HE, Melsen F, Mosekilde L. 1987. Safety of osteoporosis treatment with sodium fluoride, calcium phosphate and vitamin D. *Miner Electrolyte Metab* 13: 96-103.

Hidehiko T. 1958. On the relation between the distribution of endemic goiter and the fluorine content of natural water in Hidaka Province, Hokkaido. *Folia Pharmacol Jpn* 54: 225-229.

Hoffmann-Axthelm W. 1953. [Observations on the influence of fluorine on dental enamel and thyroid gland]. *Dtsch Zahnarztl Z* 8: 757-765.

Jentzer A. 1956. [Effect of fluorine on the iodine content of the human thyroid gland]. *Bull Schweiz Akad Med Wiss* 12: 539-543.

Jooste PL, Weight MJ, Kriek JA, Louw AJ. 1999. Endemic goitre in the absence of iodine deficiency in schoolchildren of the Northern Cape Province of South Africa. *Eur J Clin Nutr* 53: 8-12.

Kolomiitseva MG. 1961. [The content of fluorine in the external environment of the Upper Altai autonomous region and its role in the etiology of endemic goiter]. *Gig Sanit* 26: 101-103.

Korrodi H, Wegmann T, Galleti P, Held HR. 1955. [Caries prophylaxis and the untoward effects of fluor on the thyroid gland]. *Schweiz Med Wochenschr* 85: 1016-1019.

Kutlucan A, Kale Koroglu B, Numan Tamer M, Aydin Y, Baltaci D, Akdogan M, Ozturk M, Vural H, Ermis F. 2013. The investigation of effects of fluorosis on thyroid volume in school-age children. *Med Glas* 10: 93-98.

Latham MC, Grech P. 1967. The effects of excessive fluoride intake. *Am J Public Health* 57: 651-660.

Leone NC, Leatherwood EC, Petrie IM, Lieberman L. 1964. Effect of fluoride on thyroid gland: Clinical study. *J Am Dent Assoc* 69: 179-180.

Levi JE, Silberstein HE. 1955. Lack of effect of fluorine ingestion on uptake of iodine 131 by the thyroid gland. *J Lab Clin Med* 45: 348-351.

McGlashan N, Chelkowska E, Sasananan S. 2010. A survey of goiter morbidity in Ban Mae Toen, northwest Thailand. *Southeast Asian J Trop Med Public Health* 41: 1200-1208.

Rathore S, Meena C, Gonmei Z, Dwivedi S, Toteja GS, Bala K. 2018. Study of excess fluoride ingestion and thyroid hormone derangement in relation with different fluoride levels in drinking water among children of Jodhpur District, Rajasthan, India. *Asian J Microbiol Biotechnol Environ Sci* 20: 327-331.

Reisenauer R, Rezler D, Křemenová J, Preininger Q. 1961. [Fluorization of the waters in Czechoslovakia. IV. Endocrinological control of results of two years' fluorization of drinking-water in school children]. *Cesk Stomatol* 61: 91-97.

Romer TEZ, Kowalczyk B, Kacprzak M, Wiktorowski M. 1976. [Incidence of goiter in pubertal girls of the Piotrkow Region and iodide content in drinking water]. *Endokrynol Pol* 27: 373-380.

Savchenkov MF, Efimova NV, Manueva RS, Nikolaeva LA, Shin NS. 2016. [Thyroid gland pathology in children population exposed to the combination of iodine deficiency and fluoride pollution of environment]. *Gig Sanit* 95: 1201-1205.

Shtifanova AK. 1962. [The fluorine content in water, soil and vegetal products of the Alma-Atinsk District areas and its role in the etiology of dental caries and endemic goiter]. *Zdravookhranenie Kazakhstana*: 60-63.

Siddiqui AH. 1969. Incidence of simple goiter in areas of endemic fluorosis in Nalgonda District, Andhra Pradesh, India. *Fluoride* 2: 200-205.

Sidora VD, Shliakhta AI, Iugov VK, Kas'ianenko AS, Piatenko VG. 1983. [Indices of the pituitary-thyroid system in residents of cities with various fluorine concentrations in drinking water]. *Probl Endokrinol* 29: 32-35.

Sung FC, Chen KP, Chen CY, Tai PW, Yang CF. 1973. Studies of the effect of salt iodization on endemic goiter in Taiwan. IV. A survey of drinking water in relation to endemic goiter. *J Fomosan Med Assoc* 72: 96-103.

Tokar VI, Voroshnin VV, Sherbakov SV. 1989. [Chronic effects of fluorides on the pituitary-thyroid system in industrial workers]. *Gig Tr Prof Zabol*: 19-22.

Wespi HJ. 1954. [Iodized-fluoridized salt for the prevention of goiter and caries]. *Schweiz Med Wochenschr* 84: 885-890.

Yu YN. 1985. [Effects of chronic fluorosis on the thyroid gland]. *Chin Med J* 65: 747-7479.

C.2.2. Studies in Non-human Animals

As described in Figure 2, 339 non-human mammal studies were included; however, full data extraction was conducted only on studies with primary neurological outcomes and/or secondary functional neurological outcomes (e.g., motor activity). Data extraction was completed using HAWC (NTP 2019). Data were extracted from a subset of included studies in animals (n = 123) and are available in HAWC based on outcome. The following lists of references are organized as studies that are available in HAWC followed by studies that are not available in HAWC. Specifically, all primary outcomes and functional neurological secondary outcomes (e.g., motor activity) were extracted from animal studies and are available in HAWC, including studies from the NTP (2016) assessment. Studies are also available in HAWC that evaluated mechanistic effects related to oral fluoride exposure at or below 20 ppm fluoride drinking water equivalents for categories of mechanistic endpoints with the largest amount of available data (i.e., biochemistry of the brain or neurons, neurotransmission, oxidative stress, and histopathology [n = 70]); however, these mechanistic data were generally not extracted. Several animal studies assessed primary neurological outcomes and/or functional neurological secondary outcomes and mechanistic effects in the four mechanistic categories listed above (n = 56). In total, 140 animal studies are available in HAWC (70 with primary neurological outcomes and/or secondary functional neurological outcomes without relevant mechanistic data; 15 with relevant mechanistic data only; and 55 with primary and/or secondary functional neurological outcomes with relevant mechanistic data). Studies that evaluated other mechanistic endpoints, as well as studies that assessed only mechanistic effects at fluoride levels above 20 ppm fluoride drinking water equivalents, are not available in HAWC (n = 199). The list below presents the 339 non-human animal studies that were included in the review. An overview of the screening results is outlined in the study selection diagram (Figure 2) that reports numbers of included studies as well as numbers of studies excluded for each reason at the full-text review stage.

C.2.2.1. Studies Available in HAWC

Adedara IA, Abolaji AO, Idris UF, Olabiyi BF, Onibiyo EM, Ojuade TD, Farombi EO. 2017. Neuroprotective influence of taurine on fluoride-induced biochemical and behavioral deficits in rats. *Chem Biol Interact* 261: 1-10.

Adedara IA, Olabiyi BF, Ojuade TD, Idris UF, Onibiyo EM, Farombi EO. 2017. Taurine reverses sodium fluoride-mediated increase in inflammation, caspase-3 activity, and oxidative damage along the brain-pituitary-gonadal axis in male rats. *Can J Phys Pharmacol* 95: 1019-1029.

Agustina F, Sofro ZM, Partadiredja G. 2018. Subchronic administration of high-dose sodium fluoride causes deficits in cerebellar purkinje cells but not motor coordination of rats. *Biol Trace Elem Res* 188(2): 424-433.

Ahmad KR, Noor S, Jabeen S, Nauroze T, Kanwal MA, Raees K, Abbas T. 2017. Amelioration by jambul fruit extract of fluoride-induced hepato-nephronal histopathologies and impaired neuromotor capacity in mice. *Fluoride* 50: 2-14.

Akinrinade ID, Memudu AE, Ogundele OM. 2015. Fluoride and aluminium disturb neuronal morphology, transport functions, cholinesterase, lysosomal and cell cycle activities. *Pathophysiology* 22: 105-115.

Akinrinade ID, Memudu AE, Ogundele OM, Ajetunmobi OI. 2015. Interplay of glia activation and oxidative stress formation in fluoride and aluminium exposure. *Pathophysiology* 22: 39-48.

Baba NA, Raina R, Verma PK, Sultana M. 2014. Alterations in plasma and tissue acetylcholinesterase activity following repeated oral exposure of chlorpyrifos alone and in conjunction with fluoride in Wistar rats. *Proc Indian Natl Sci Acad B Biol Sci* 84: 969-972.

Bagmut I, Kolisnyk I, Titkova A, Babiy L, Filipchenko S. 2018. The antioxidant system enzymes' activity in rats' brain, intoxicated with sodium fluoride in subtoxic doses. *Arch Balkan Med Union* 53(4): 506-511.

Balaji B, Kumar EP, Kumar A. 2015. Evaluation of standardized bacopa monniera extract in sodium fluoride-induced behavioural, biochemical, and histopathological alterations in mice. *Toxicol Ind Health* 31: 18-30.

Balayssac D, Richard D, Authier N, Nicolay A, Jourdan D, Eschalier A, Coudore F. 2002. Absence of painful neuropathy after chronic oral fluoride intake in Sprague-Dawley and Lou/C rats. *Neurosci Lett* 327: 169-172.

Banala RR, Karnati PR. 2015. Vitamin A deficiency: An oxidative stress marker in sodium fluoride (NaF) induced oxidative damage in developing rat brain. *Int J Dev Neurosci* 47: 298-303.

Banala R, Nagapuri K, Mohd K, Reddy M, Karnati P. 2018. Carica Papaya leaf extract as a neuroprotective agent against behavioral and neurotransmitter changes in brain of the rat treated with sodium fluoride in pre- and post-natal periods. *Pharmacogn Mag* 14(55): 123-131.

Banji D, Banji OJ, Pratusha NG, Annamalai AR. 2013. Investigation on the role of spirulina platensis in ameliorating behavioural changes, thyroid dysfunction and oxidative stress in offspring of pregnant rats exposed to fluoride. *Food Chem* 140: 321-331.

Baran-Poesina V, Negres S, Dobrescu D, Dimcevici-Poesina N, Dimcevici-Poesina A, Feghiu A, Soare T, Militaru M. 2013. Experimental pharmacological researches regarding the influence of sodium fluoride in allopathic and homeopathic doses on central nervous system's performances:

A correlation between behavioral response in classic maze test and morphological aspects of cerebral cortex. *Farmacia* 61: 781-799.

Bartos M, Gumilar F, Bras C, Gallegos CE, Giannuzzi L, Cancela LM, Minetti A. 2015. Neurobehavioural effects of exposure to fluoride in the earliest stages of rat development. *Physiol Behav* 147: 205-212.

Bartos M, Gumilar F, Gallegos CE, Bras C, Dominguez S, Monaco N, Esandi MDC, Bouzat C, Cancela LM, Minetti A. 2018. Alterations in the memory of rat offspring exposed to low levels of fluoride during gestation and lactation: Involvement of the alpha7 nicotinic receptor and oxidative stress. *Reprod Toxicol* 81: 108-114.

Basha PM, Rai P, Begum S. 2011. Fluoride toxicity and status of serum thyroid hormones, brain histopathology, and learning memory in rats: A multigenerational assessment. *Biol Trace Elem Res* 144: 1083-1094.

Basha PM, Sujitha NS. 2012. Combined impact of exercise and temperature in learning and memory performance of fluoride toxicated rats. *Biol Trace Elem Res* 150: 306-313.

Bataineh HN, Nusier MK. 2006. Impact of 12-week ingestion of sodium fluoride on aggression, sexual behavior, and fertility in adult male rats. *Fluoride* 39: 293-301.

Bera I, Sabatini R, Auteri P, Flace P, Sisto G, Montagnani M, Potenza MA, Marasciulo FL, Carratu MR, Coluccia A, Borracci P, Tarullo A, Cagiano R. 2007. Neurofunctional effects of developmental sodium fluoride exposure in rats. *Eur Rev Med Pharmacol Sci* 11: 211-224.

Bhatnagar M, Rao P, Sushma J, Bhatnagar R. 2002. Neurotoxicity of fluoride: Neurodegeneration in hippocampus of female mice. *Indian J Exp Biol* 40: 546-554.

Bhatnagar M, Sukhwal P, Suhalka P, Jain A, Joshi C, Sharma D. 2011. Effects of fluoride in drinking water on NADPH-diaphorase neurons in the forebrain of mice: A possible mechanism of fluoride neurotoxicity. *Fluoride* 44: 195-209.

Chen H, Geng D. 2011. [The change of cognition induced by chronic fluoride in rats]. *Acta Academiae Medicinae Xuzhou* 31(5): 319-322.

Chen J, Niu Q, Xia T, Zhou G, Li P, Zhao Q, Xu C, Dong L, Zhang S, Wang A. 2018. ERK1/2-mediated disruption of BDNF-TrkB signaling causes synaptic impairment contributing to fluoride-induced developmental neurotoxicity. *Toxicology* 410: 222-230.

Chinoy NJ, Shah SD. 2004. Biochemical effects of sodium fluoride and arsenic trioxide toxicity and their reversal in the brain of mice. *Fluoride* 37: 80-87.

Chioca LR, Raupp IM, Da Cunha C, Losso EM, Andreatini R. 2008. Subchronic fluoride intake induces impairment in habituation and active avoidance tasks in rats. *Eur J Pharmacol* 579: 196-201.

Chouhan S, Flora SJS. 2008. Effects of fluoride on the tissue oxidative stress and apoptosis in rats: Biochemical assays supported by IR spectroscopy data. *Toxicology* 254: 61-67.

Chouhan S, Lomash V, Flora SJ. 2010. Fluoride-induced changes in haem biosynthesis pathway, neurological variables and tissue histopathology of rats. *J Appl Toxicol* 30: 63-73.

Cui YS, Zhong Q, Li WF, Liu ZH, Wang Y, Hou CC. 2017. [Effects of fluoride exposure on thyroid hormone level and intelligence in rats]. *Chin J Ind Hyg Occup Dis* 35: 888-892.

Dabrowska E. 1997. Effect of different fluorine doses on the supraoptic nucleus of the rat. *Folia Histochem Cytobiol* 35: 115-116.

Dong Y, Wang Y, Wei N, Guan Z. 2015. [Expression levels of brain muscarinic acetylcholine receptor in offspring rats of drinking-water borne fluorosis]. *Chin J Endemiol* 34: 326-330.

Dong YT, Wang Y, Wei N, Zhang QF, Guan ZZ. 2015. Deficit in learning and memory of rats with chronic fluorosis correlates with the decreased expressions of M1 and M3 muscarinic acetylcholine receptors. *Arch Toxicol* 89: 1981-1991.

Dong YT, Wei N, Qi XL, Liu XH, Chen D, Zeng XX, Guan ZZ. 2017. Attenuating effect of vitamin E on the deficit of learning and memory of rats with chronic fluorosis: The mechanism may involve muscarinic acetylcholine receptors. *Fluoride* 50: 354-364.

Dong YW, Y. Wei, N. Guan, Z. 2015. [Expression of muscarinic acetylcholine receptors in the brain of rats with chronic fluorosis]. *Chin J Endemiol* 34(2): 84-88.

Ekambaram P, Paul V. 2001. Calcium preventing locomotor behavioral and dental toxicities of fluoride by decreasing serum fluoride level in rats. *Environ Toxicol Pharmacol* 9: 141-146.

Ekambaram P, Paul V. 2002. Modulation of fluoride toxicity in rats by calcium carbonate and by withdrawal of fluoride exposure. *Pharmacol Toxicol* 90: 53-58.

Ekambaram P, Paul V. 2003. Effect of vitamin D on chronic behavioral and dental toxicities of sodium fluoride in rats. *Fluoride* 36: 189-197.

El-lethey HS, Kamel MM, Shaheed IB. 2010. Neurobehavioral toxicity produced by sodium fluoride in drinking water of laboratory rats. *J Am Sci* 6(5): 54-63.

El-lethey HS, Kamel MM. 2011. Effects of black tea in mitigation of sodium fluoride potency to suppress motor activity and coordination in laboratory rats. *J Am Sci* 7(4): 243-254.

El-lethey HS, Shaheed IB. 2011. Potential health impact of black tea against Na-F-induced alterations in territorial aggression, sexual behaviour and fertility of male rats. *Life Sci J* 8: 828-839.

Elliott L. 1967. Lack of effect of administration of fluoride on the central nervous system of rats. *Acta Pharmacol Toxicol (Copenh)* 25: 323-328.

Flace P, Benagiano V, Vermesan D, Sabatini R, Inchingolo AM, Auteri P, Ambrosi G, Tarullo A, Cagiano R. 2010. Effects of developmental fluoride exposure on rat ultrasonic vocalization, acoustic startle reflex and pre-pulse inhibition. *Eur Rev Med Pharmacol Sci* 14: 507-512.

Gabovich RD. 1962. [On the problem of the effect of fluorine in drinking water on the functional state of the central nervous system]. *Gig Sanit* 27: 10-12.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008. [Level of oxidative stress in rat brains and learning and memory function of rats with chronic fluorosis]. *Chin J Endemiol* 27: 371-373.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008. [Effects of fluoride on learning and memory and cholinesterase activity in rat brains]. *Chin J Endemiol* 27: 128-130.

Gao Q, Liu YJ, Guan ZZ. 2009. Decreased learning and memory ability in rats with fluorosis: Increased oxidative stress and reduced cholinesterase activity in the brain. *Fluoride* 42: 277-285.

Gao Y, Liu L, Young L, Huan L, Jin H. 2009. Effects of learning and memory of fluoride and the antagonism of selenium in rats. *Studies of Trace Elements and Health* 26(2): 1-3.

Ge QD, Tan Y, Luo Y, Wang WJ, Zhang H, Xie C. 2018. MiR-132, miR-204 and BDNF-TrkB signaling pathway may be involved in spatial learning and memory impairment of the offspring rats caused by fluorine and aluminum exposure during the embryonic stage and into adulthood. *Environ Toxicol Pharmacol* 63: 60-68.

Ge Y, Chen L, Yin Z, Song X, Ruan T, Hua L, Liu J, Wang J, Ning H. 2018. Fluoride-induced alterations of synapse-related proteins in the cerebral cortex of ICR offspring mouse brain. *Chemosphere* 201: 874-883.

Gopal K, Saxena R, Gupta GSD, Rana MD, Agrawal D. 2006. Fluoride induced alterations in neurobehavioural and cardiovascular responses in rats. *J Adv Zool* 27: 1-7.

Gui CZ, Ran LY, Wu CX, Long YG, He J, Zhang H, Guan ZZ. 2009. [Changes in learning and memory ability and brain cholinesterase activity in the rats with coal burning fluorosis]. *Chin J Endemiol* 28: 497-500.

Gui CZ, Ran LY, Li JP, Guan ZZ. 2010. Changes of learning and memory ability and brain nicotinic receptors of rat offspring with coal burning fluorosis. *Neurotoxicol Teratol* 32: 536-541.

Gui CZ, Ran LY, Guan ZZ. 2011. [Expression levels of brain nicotinic acetylcholine receptor mRNA and protein in coal-burning type of fluorosis rats]. *Chin J Endemiol* 30: 239-242.

Guner S, Uyar-Bozkurt S, Haznedaroglu E, Mentes A. 2016. Dental fluorosis and catalase immunoreactivity of the brain tissues in rats exposed to high fluoride pre- and postnatally. *Biol Trace Elem Res* 174: 150-157.

Han H, Du W, Zhou B, Zhang W, Xu G, Niu R, Sun Z. 2014. Effects of chronic fluoride exposure on object recognition memory and mRNA expression of SNARE complex in hippocampus of male mice. *Biol Trace Elem Res* 158: 58-64.

Hong JH, Ge YM, Ning HM, Wang JD. 2005. [Effects of High Fluoride and Low Iodine on Learning-Memory and TchE of Brain in Offspring Rats]. *Chin Prev Med* 6: 489-491.

Inkielewicz I, Krechniak J. 2004. Fluoride effects on glutathione peroxidase and lipid peroxidation in rats. *Fluoride* 37: 7-12.

Jain A, Mehta VK, Chittora RA, Mahdi A, Bhatnagar M. 2015. Melatonin ameliorates fluoride induced neurotoxicity in young rats: An in vivo evidence. *Asian J Pharm Clin Res* 8: 164-167.

Jetti R, Raghuveer CV, Mallikarjuna RC. 2016. Protective effect of ascorbic acid and Ginkgo biloba against learning and memory deficits caused by fluoride. *Toxicol Ind Health* 32: 183-187.

Jia B, Zong L, Lee JY, Lei J, Zhu Y, Xie H, Clemens JL, Feller MC, Na Q, Dong J, McLane MW, Jones-Beatty K, Burd I. 2019. Maternal supplementation of low dose fluoride alleviates adverse perinatal outcomes following exposure to intrauterine inflammation. *Sci Rep* 9(1): 2575.

Jiang C, Zhang S, Liu H, Guan Z, Zeng Q, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Zhang X, Cui Y, Yu L, Wang A. 2014. Low glucose utilization and neurodegenerative changes caused by sodium fluoride exposure in rat's developmental brain. *Neuromolecular Med* 16: 94-105.

Jiang S, Su J, Yao S, Zhang Y, Cao F, Wang F, Wang H, Li J, Xi S. 2014. Fluoride and arsenic exposure impairs learning and memory and decreases mGluR5 expression in the hippocampus and cortex in rats. *PLoS One* 9(4): e96041.

Khan AM, Raina R, Dubey N, Verma PK. 2017. Effect of deltamethrin and fluoride co-exposure on the brain antioxidant status and cholinesterase activity in Wistar rats. *Drug Chem Toxicol* 41(2): 1-5.

Kinawy AA, Al-Eidan AA. 2018. Impact of prenatal and postnatal treatment of sodium fluoride and aluminum chloride on some hormonal and sensorimotor aspects in rats. *Biol Trace Elem Res*: 1-8.

Kivrak Y. 2012. Effects of fluoride on anxiety and depression in mice. *Fluoride* 45: 302-306.

Li M, Cui J, Gao YH, Zhang W, Sun LY, Liu XN, Liu Y, Sun DJ. 2015. Pathological changes and effect on the learning and memory ability in rats exposed to fluoride and aluminum. *Toxicol Res* 4: 1366-1373.

Li X, Zhang J, Niu R, Manthari RK, Yang K, Wang J. 2019. Effect of fluoride exposure on anxiety- and depression-like behavior in mouse. *Chemosphere* 215: 454-460.

Liu F, Ma J, Zhang H, Liu P, Liu YP, Xing B, Dang YH. 2014. Fluoride exposure during development affects both cognition and emotion in mice. *Physiol Behav* 124: 1-7.

Liu WX. 1989. [Experimental study of behavior and cerebral morphology of rat pups generated by fluorotic female rat]. *Chin J Pathol* 18: 290-292.

Liu YJ, Gao Q, Wu CX, Long YG, Guan ZZ. 2009. [Modified expression of extracellular signal-regulated protein kinase signal transduction in rat brains and changed capacity of learning and memory of rats with chronic fluorosis]. *Chin J Endemiol* 28: 32-35.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. Alterations of nAChRs and ERK1/2 in the brains of rats with chronic fluorosis and their connections with the decreased capacity of learning and memory. *Toxicol Lett* 192: 324-329.

Liu YJ, Gao Q, Long YG, Yu YN, Guan ZZ. 2011. [Influence of chronic fluorosis on expression of phospho-Elk-1 in rat brains]. *Chin J Endemiol* 30: 251-255.

Lou DD, Guan ZZ, Liu YJ, Liu YF, Zhang KL, Pan JG, Pei JJ. 2013. The influence of chronic fluorosis on mitochondrial dynamics morphology and distribution in cortical neurons of the rat brain. *Arch Toxicol* 87: 449-457.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S, Mentes A, DenBesten PK. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. *Physiol Behav* 206: 76-83.

Ma J, Liu F, Liu P, Dong YY, Chu Z, Hou TZ, Dang YH. 2015. Impact of early developmental fluoride exposure on the peripheral pain sensitivity in mice. *Int J Dev Neurosci* 47: 165-171.

Manusha S, Sudhakar K, Reddy KP. 2019. Protective effects of allium sativum extract against sodium fluoride induced neurotoxicity. *Int J Pharm Sci Res* 10(2): 625-633.

McPherson CA, Zhang G, Gilliam R, Brar SS, Wilson R, Brix A, Picut C, Harry GJ. 2018. An evaluation of neurotoxicity following fluoride exposure from gestational through adult ages in Long-Evans hooded rats. *Neurotoxicol Res*: 1-18.

Mesram N, Nagapuri K, Banala RR, Nalagoni CR, Karnati PR. 2016. Quercetin treatment against NaF induced oxidative stress related neuronal and learning changes in developing rats. *J King Saud Univ Sci* 29: 221-229.

Mullenix PJ, Denbesten PK, Schunior A, Kernan WJ. 1995. Neurotoxicity of sodium fluoride in rats. *Neurotoxicol Teratol* 17: 169-177.

Nageshwar M, Sudhakar K, Reddy NCC, Reddy KP. 2017. Neuroprotective effects of curcumin on sodium fluoride induced behavioural and enzymatic changes in brain and muscles of rat. *J Environ Biol* 38: 675-681.

Nageshwar M, Sudhakar K, Reddy KP. 2018. Quercetin ameliorates oxidative stress, neural damage of brain and behavioral impairment of rat with fluoride exposure. *Int J Pharm Sci Res* 9(8): 3247-3256.

Nian W, Wang X, Shao D, Yu Q, Ouyang W, Zhang Z, Ruan Q. 2018. Effects of subchronic exposure to fluorine on hippocampal injury in mice and its molecular mechanism. *Acta Sci Circumst* 38(11): 4512-4519.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. *Environ Pollut* 233: 889-899.

Niu R, Sun Z, Wang J, Cheng Z. 2008. Effects of fluoride and lead on locomotor behavior and expression of nissl body in brain of adult rats. *Fluoride* 41: 276-282.

Niu R, Sun Z, Cheng Z, Li Z, Wang J. 2009. Decreased learning ability and low hippocampus glutamate in offspring rats exposed to fluoride and lead. *Environ Toxicol Pharmacol* 28: 254-258.

Niu R, Liu S, Wang J, Zhang J, Sun Z. 2014. Proteomic analysis of hippocampus in offspring male mice exposed to fluoride and lead. *Biol Trace Elem Res* 162: 227-233.

Niu R, Xue X, Zhao Y, Sun Z, Yan X, Li X, Feng C, Wang J. 2015. Effects of fluoride on microtubule ultrastructure and expression of Tubalpha1a and Tubbeta2a in mouse hippocampus. *Chemosphere* 139: 422-427.

Niu R, Chen H, Manthari RK, Sun Z, Wang J, Zhang J, Wang J. 2018. Effects of fluoride on synapse morphology and myelin damage in mouse hippocampus. *Chemosphere* 194: 628-633.

Nkpaa KW, Onyeso GI. 2018. Rutin attenuates neurobehavioral deficits, oxidative stress, neuro-inflammation and apoptosis in fluoride treated rats. *Neurosci Lett* 682: 92-99.

Paul V, Ekambaram P, Jayakumar AR. 1998. Effects of sodium fluoride on locomotor behavior and a few biochemical parameters in rats. *Environ Toxicol Pharmacol* 6: 187-191.

Pereira M, Dombrowski PA, Losso EM, Chioca LR, Da Cunha C, Andreatini R. 2011. Memory impairment induced by sodium fluoride is associated with changes in brain monoamine levels. *Neurotoxicol Res* 19: 55-62.

Pulungan ZS, Sofro ZM, Partadiredja G. 2016. Sodium fluoride does not affect the working memory and number of pyramidal cells in rat medial prefrontal cortex. *Anat Sci Int* 93(1): 128-138.

Raghu J, Raghuveer VC, Rao MC, Somayaji NS, Babu PB. 2013. The ameliorative effect of ascorbic acid and Ginkgo biloba on learning and memory deficits associated with fluoride exposure. *Interdiscip Toxicol* 6: 217-221.

Raju S, Sivanesan SK, Gudemalla K. 2019. Cognitive enhancement effect of Ginkgo biloba extract on memory and learning impairments induced by fluoride neurotoxicity. *Int J Res Pharm Sci* 10(1): 129-134.

Reddy MM, Karnati PR. 2015. Protective effects of aqueous extract of fruit pulp of tamarindus indica on motor activity and metabolism of the gastrocnemius muscle of rats treated with fluoride. *Int J Toxicol Pharmacol Res* 7: 241-246.

Reddy YP, Tiwari SK, Shaik AP, Alsaeed A, Sultana A, Reddy PK. 2014. Effect of sodium fluoride on neuroimmunological parameters, oxidative stress and antioxidative defenses. *Toxicol Mech Methods* 24: 31-36.

Rumiantsev GI, Novikov SM, Mel'nikova NN, Levchenko NI, Kozeeva EE. 1988. [Experimental study of the biological effect of salts of hydrofluosilicic acid]. *Gig Sanit*: 80-82.

Sarkozi K, Horvath E, Vezer T, Papp A, Paulik E. 2015. Behavioral and general effects of subacute oral arsenic exposure in rats with and without fluoride. *Int J Environ Health Res* 25: 418-431.

Shah SD, Chinoy NJ. 2004. Adverse effects of fluoride and/or arsenic on the cerebral hemisphere of mice and recovery by some antidotes. *Fluoride* 37: 162-171.

Shalini B, Sharma JD. 2015. Beneficial effects of emblica officinalis on fluoride-induced toxicity on brain biochemical indexes and learning-memory in rats. *Toxicol Int* 22: 35-39.

Shan KR, Qi XL, Long YG, Nordberg A, Guan ZZ. 2004. Decreased nicotinic receptors in PC12 cells and rat brains influenced by fluoride toxicity: A mechanism relating to a damage at the level in post-transcription of the receptor genes. *Toxicology* 200: 169-177.

Sharma C, Suhalka P, Bhatnagar M. 2018. Curcumin and resveratrol rescue cortical-hippocampal system from chronic fluoride-induced neurodegeneration and enhance memory retrieval. *Int J Neurosci*: 1-15.

Shen X, Zhang Z, Xu X. 2004. [Effect of iodine and selenium on learning memory impairment induced by fluorosis and blood biochemical criterion of rats]. *Occupation and Health* 20(1): 6-8.

Sudhakar K, Nageshwar M, Pratap Reddy K. 2017. Seed extract of Abelmoschus moschatus medik reverses NAF-induced behavioral changes through neurodegeneration and oxidative stress in brain of rat. *Asian J Pharm Clin Res* 10: 165-171.

Sudhakar K, Nageshwar M, Reddy KP. 2018. Protective effect of okra, Abelmoschus moschatus seed extract on developing brain of rats during pre- and post-natal fluoride exposure. *Int J Pharm Sci Res* 9: 1519-1528.

Sudhakar K, Nageshwar M, Reddy KP. 2018. Abelmoschus moschatus extract reverses altered pain and neurohistology of a rat with developmental exposure of fluoride. *J Appl Pharm Sci* 8(6): 94-104.

Sudhakar K, Reddy KP. 2018. Protective effects of Abelmoschus moschatus seed extract on neurotransmitter system of developing brain of Wistar rats with gestational and post-natal exposure of sodium fluoride. *Int J Green Pharm* 12: S131-S139.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2000. [Effects of high fluoride drinking water on the cerebral function of mice]. *Chin J Epidemiol* 19: 262-263.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2008. Effects of high fluoride drinking water on the cerebral function of mice. *Fluoride* 41: 148-151.

Sun Z, Zhang Y, Xue X, Niu R, Wang J. 2018. Maternal fluoride exposure during gestation and lactation decreased learning and memory ability, and glutamate receptor mRNA expressions of mouse pups. *Hum Exp Toxicol* 37: 87-93.

Trivedi MH, Verma RJ, Chinoy NJ. 2007. Amelioration by black tea of sodium fluoride-induced changes in protein content of cerebral hemisphere, cerebellum and medulla oblongata in brain region of mice. *Acta Poloniae Pharm* 64: 221-225.

Trivedi MH, Verma RJ, Chinoy NJ. 2009. Mitigation of sodium fluoride induced toxicity in mice brain by black tea infusion. *Fluoride* 42: 29-33.

Trivedi MH, Verma RJ, Sangai NP, Chinoy NJ. 2011. Black tea extract mitigation of NaF-induced lipid peroxidation in different regions of mice brains. *Fluoride* 44: 243-254.

Trivedi MH, Verma RJ, Sangai NP, Chinoy NJ. 2012. Mitigation by black tea extract of sodium fluoride induced histopathological changes in brain of mice. *Fluoride* 45: 13-26.

Tsunoda M, Aizawa Y, Nakano K, Liu Y, Horiuchi T, Itai K, Tsunoda H. 2005. Changes in fluoride levels in the liver, kidney, and brain and in neurotransmitters of mice after subacute administration of fluoride. *Fluoride* 38: 284-292.

Varner JA, Jensen KF, Horvath W, Isaacson RL. 1998. Chronic administration of aluminum-fluoride or sodium-fluoride to rats in drinking water: Alterations in neuronal and cerebrovascular integrity. *Brain Res* 784(1-2): 284-298.

Verma RJ, Trivedi MH, Chinoy NJ. 2007. Black tea amelioration of sodium fluoride-induced alterations of DNA, RNA, and protein contents in the cerebral hemisphere, cerebellum, and medulla oblongata regions of mouse brain. *Fluoride* 40: 7-12.

Wang G, Li J, Zhu H, Zhu J. 2006. Effect of different doses of chronic exposure of fluoride on rat learning and memory behavior. *Studies of Trace Elements and Health* 23(2): 1-2.

Wang JD, Ge YM, Ning HM, Wang SL. 2004. Effects of high fluoride and low iodine on biochemical indexes of the brain and learning-memory of offspring rats. *Fluoride* 37: 201-208.

Wang J, Zhang Y, Guo Z, Li R, Xue X, Sun Z, Niu R. 2018. Effects of perinatal fluoride exposure on the expressions of miR-124 and miR-132 in hippocampus of mouse pups. *Chemosphere* 197: 117-122.

Wei N, Dong Y, Wang Y, Guan Z. 2014. [Effects of chronic fluorosis on neurobehavioral development in offspring of rats and antagonistic effect of vitamin E]. *Chin J Endemiol* 33: 125-128.

Whitford GM, Whitford JL, Hobbs SH. 2009. Appetitive-based learning in rats: Lack of effect of chronic exposure to fluoride. *Neurotoxicol Teratol* 31: 210-215.

Wu CX, Gu XL, Ge YM, Zhang JH, Wang JD. 2006. Effects of high fluoride and arsenic on brain biochemical indexes and learning-memory in rats. *Fluoride* 39: 274-279.

Wu NP, Zhao ZL, Gao WH, Li XQ. 1995. [Behavioral teratology in rats exposed to fluoride.] *Chin J Endemiol* 12(5): 271-273.

Wu NP, Zhao ZL, Gao WH, Li XQ. 2008. Behavioral teratology in rats exposed to fluoride. *Fluoride* 41: 129-133.

Xu X, Shen X, Zhang Z. 2001. Effect of fluorosis on mice learning and memory behaviors and brain SOD activity and MDA content *China Public Health* 17(1): 8-10.

Yang L, Jin P, Wang X, Zhou Q, Lin X, Xi S. 2018. Fluoride activates microglia, secretes inflammatory factors and influences synaptic neuron plasticity in the hippocampus of rats. *Neurotox* 69: 108-120.

Yu Q, Shao D, Zhang R, Ouyang W, Zhang Z. 2019. Effects of drinking water fluorosis on L-type calcium channel of hippocampal neurons in mice. *Chemosphere* 220: 169-175.

Yuan J, Li Q, Niu R, Wang J. 2019. Fluoride exposure decreased learning ability and the expressions of the insulin receptor in male mouse hippocampus and olfactory bulb. *Chemosphere* 224: 71-76.

Zhang C, Ren C, Chen H, Geng R, Fan H, Zhao H, Guo K, Geng D. 2013. The analog of Ginkgo biloba extract 761 is a protective factor of cognitive impairment induced by chronic fluorosis. *Biol Trace Elem Res* 153: 229-236.

Zhang KL, Lou DD, Guan ZZ. 2015. Activation of the AGE/RAGE system in the brains of rats and in SH-SY5Y cells exposed to high level of fluoride might connect to oxidative stress. *Neurotoxicol Teratol* 48: 49-55.

Zhang J, Zhu W, Zhang Z. 2009. [The effect of fluorine exposure of pregnant rats on the learning and memory capabilities of baby rats]. *Chinese Journal of Public Health* 25(11): 1347-1348.

Zhang J, Zhu WJ, Xu XH, Zhang ZG. 2011. Effect of fluoride on calcium ion concentration and expression of nuclear transcription factor kappa-B rho65 in rat hippocampus. *Exp Toxicol Pathol* 63: 407-411.

Zhang J, Zhang Z. 2013. Effects of chronic fluorosis on camkiiα, c-FOS, BAX, and BCL-2 channel signaling in the hippocampus of rats. *Fluoride* 46: 135-141.

Zhang Z, Shen X, Xu X. 2001. [Effects of selenium on the damage of learning-memory ability of mice induced by fluoride]. *J Hyg Res* 30: 144-146.

Zhang Z, Xu X, Shen X, Xua XH. 1999. [Effect of fluoride exposure on synaptic structure of brain areas related to learning-memory in mice]. *J Hyg Res* 28(4): 210-212.

Zhang Z, Xu X, Shen X, Xua XH. 2008. Effect of fluoride exposure on synaptic structure of brain areas related to learning-memory in mice. *Fluoride* 41: 139-143.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao XL, Wu JH. 1998. Actions of sodium fluoride on acetylcholinesterase activities in rats. *Biomed Environ Sci* 11: 1-6.

Zheng X, Sun Y, Ke L, Ouyang W, Zhang Z. 2016. Molecular mechanism of brain impairment caused by drinking-acquired fluorosis and selenium intervention. *Environ Toxicol Pharmacol* 43: 134-139.

Zhu W, Zhang J, Zhang Z. 2011. Effects of fluoride on synaptic membrane fluidity and PSD-95 expression level in rat hippocampus. *Biol Trace Elem Res* 139: 197-203.

Zhu YL, Zheng YJ, LV XM, Ma Y, Zhang J. 2012. Effects of fluoride exposure on performance in water labyrinth and monoamine neurotransmitters of rats. *Journal of Xinjiang Medical University* 3: 014.

Zhu YP, Xi SH, Li MY, Ding TT, Liu N, Cao FY, Zeng Y, Liu XJ, Tong JW, Jiang SF. 2017. Fluoride and arsenic exposure affects spatial memory and activates the ERK/CREB signaling pathway in offspring rats. *Neurotox* 59: 56-64.

C.2.2.2. Studies Not Available in HAWC

Abdelaleem MM, El-Tahawy NFG, Abozaid SMM, Abdel-Hakim SA. 2018. Possible protective effect of curcumin on the thyroid gland changes induced by sodium fluoride in albino rats: Light and electron microscopic study. *Endocr Regul* 52: 59-68.

Abd-Elhakim YM, Mohammed AT, Ali HA. 2018. Impact of subchronic exposure to triclosan and/or fluoride on estrogenic activity in immature female rats: The expression pattern of calbindin-D9k and estrogen receptor alpha genes. *J Biochem Mol Toxicol* 32(2): 22027.

Abdumajidov OR. 2004. [Sex differences in lipid peroxidation and antioxidant defense of the brain tissue in intoxication with low doses of inorganic compounds]. *Uzbekiston Tibbiet Zhurnali*: 58-60.

Adebayo OL, Shallie PD, Salau BA, Ajani EO, Adenuga GA. 2013. Comparative study on the influence of fluoride on lipid peroxidation and antioxidants levels in the different brain regions of well-fed and protein undernourished rats. *J Trace Elem Med Biol* 27: 370-374.

Adedara IA, Ojuade TJD, Olabiyi BF, Idris UF, Onibiyo EM, Ajeigbe OF, Farombi EO. 2016. Taurine ameliorates renal oxidative damage and thyroid dysfunction in rats chronically exposed to fluoride. *Biol Trace Elem Res*: 1-8.

Ahmed SK, Kalleny NK, Attia AAEM, Elkateb LA. 2015. The possible protective role of chromium chloride against sodium fluoride-induced changes in the structure of the cerebellar cortex of the adult male albino rat. *Egypt J Histol* 38: 402-414.

Al Badawi MH, Mahmoud OM, Salem NA. 2016. Therapeutic potential of omega-3 against sodium fluoride toxicity on the cerebellar cortex of adult male albino rats: Histological and immunohistochemical study. *Egypt J Histol* 39: 170-178.

Alhayani A, Elshal EB, Aal IHA, Al-Shammeri E, Kabra H. 2013. Does vitamin E protect against sodium fluoride toxicity on the cerebellar cortex of albino rats? *Middle East J Sci Res* 16: 1019-1026.

Ameeramja J, Raghunath A, Perumal E. 2018. Tamarind seed coat extract restores fluoride-induced hematological and biochemical alterations in rats. *Environ Sci Pollut Res Int* 25(26): 26157-26166.

Antonyan OA. 1980. [Lipid per oxidation in fluorosis and the protective role of dietary factors]. *Zh Eksp Klin Med* 20: 381-388.

Ardelean I, Racoveanu N, Manescu S, Lupulescu A, Diaconescu M, Ghelerter L. 1964. Experimental investigations concerning the effect of fluorine on the thyroid gland. *Rum Med Rev* 8: 17-20.

Atmaca N, Atmaca HT, Kanici A, Anteplioglu T. 2014. Protective effect of resveratrol on sodium fluoride-induced oxidative stress, hepatotoxicity and neurotoxicity in rats. *Food Chem Toxicol* 70: 191-197.

Auskaps AM, Shaw JH. 1955. Hemoglobin concentration, thyroid weight and growth rate in rats during minimum fluoride ingestion. *J Nutr* 55: 611-621.

Bagmut I, Kolisnyk I, Titkova A, Petrenko T, Filipchenko S. 2018. Content of catecholamines in blood serum of rats under fluoride intoxication. *Georgian Med News* (280-281): 125-129.

Bakalyan PH, Antonyan OA. 1981. [Effect of fluorosis on glutathione peroxidase and glutathione reductase activities and sulfhydryl groups]. *Zh Eksp Klin Med* 21: 10-14.

Basha PM, Madhusudhan N. 2010. Pre and post natal exposure of fluoride induced oxidative macromolecular alterations in developing central nervous system of rat and amelioration by antioxidants. *Neurochem Res* 35: 1017-1028.

Basha PM, Madhusudhan N. 2011. Effect of maternal exposure of fluoride on oxidative stress markers and amelioration by selected antioxidants in developing central nervous system of rats. *Biologia* 66: 187-193.

Basha PM, Rai P, Begum S. 2011. Evaluation of fluoride-induced oxidative stress in rat brain: A multigeneration study. *Biol Trace Elem Res* 142: 623-637.

Basha PM, Sujitha NS. 2012. Combined influence of intermittent exercise and temperature stress on the modulation of fluoride toxicity. *Biol Trace Elem Res* 148: 69-75.

Basha PM, Saumya SM. 2013. Suppression of mitochondrial oxidative phosphorylation and TCA enzymes in discrete brain regions of mice exposed to high fluoride: Amelioration by panax ginseng (ginseng) and lagerstroemia speciosa (banaba) extracts. *Cell Mol Neurobiol* 33: 453-464.

Basha MP, Begum S, Madhusudhan N. 2014. Antioxidants in the management of fluoride induced neural oxidative stress in developing rats. *Int J Pharm Sci Res* 5: 201-206.

Benetato G, Giuran AM, Cirmaciu R, Cirje M, Petrescu A, Vacariu A. 1970. [Effect of fluorine in drinking water on the metabolism of Ca and Mg and on neuromuscular excitability: Experimental studies and clinical observations]. *Rev Roum Physiol* 7: 335-352.

Bharti VK, Srivastava RS. 2009. Fluoride-induced oxidative stress in rat's brain and its amelioration by buffalo (Bubalus bubalis) pineal proteins and melatonin. *Biol Trace Elem Res* 130: 131-140.

Bhatnagar M, Rao P, Saxena A, Bhatnagar R, Meena P, Barbar S, Chouhan A, Vimal S. 2006. Biochemical changes in brain and other tissues of young adult female mice from fluoride in their drinking water. *Fluoride* 39: 280-284.

Bilgili A, Akdogan M, Yildiz M, Eraslan G, Cetin N. 2004. The effects of fluoride on thyroid hormones in rabbits. *Indian Vet J* 81: 986-988.

Bobek S, Kahl S, Ewy Z. 1976. Effect of long-term fluoride administration on thyroid hormones level blood in rats. *Endocrinol Exp* 10: 289-295.

Bouaziz H, Ammar E, Ghorbel H, Ketata S, Jamoussi K, Ayadi F, Guermazi F, Zeghal N. 2004. Effect of fluoride ingested by lactating mice on the thyroid function and bone maturation of their suckling pups. *Fluoride* 37: 133-142.

Bouaziz H, Soussia L, Guermazi F, Zeghal N. 2005. Fluoride-induced thyroid proliferative changes and their reversal in female mice and their pups. *Fluoride* 38: 185-192.

Bouaziz HB, Amara I, Essefi M, Croute F, Zeghal N. 2010. Fluoride-induced brain damages in suckling mice. *Pestic Biochem Physiol* 96: 24-29.

Chauhan SS, Ojha S, Mahmood A. 2013. Effects of fluoride and ethanol administration on lipid peroxidation systems in rat brain. *Indian J Exp Biol* 51: 249-255.

Chen J, Chen X, Yang K, Xia T, Xie H. 2002. [Studies on DNA damage and apoptosis in rat brain induced by fluoride]. *Chin J Prev Med* 36: 222-224.

Chirumari K, Reddy PK. 2007. Dose-dependent effects of fluoride on neurochemical milieu in the hippocampus and neocortex of rat brain. *Fluoride* 40: 101-110.

Chouhan S, Yadav A, Kushwah P, Kaul RK, Flora SJS. 2011. Silymarin and quercetin abrogates fluoride induced oxidative stress and toxic effects in rats. *Mol Cell Toxicol* 7: 25-32.

Clay AB, Suttie JW. 1987. Effect of dietary fluoride on dairy cattle: Growth of young heifers. *J Dairy Sci* 70: 1241-1251.

Czechowicz K, Osada A, Slesak B. 1974. Histochemical studies on the effect of sodium fluoride on metabolism in Purkinje's cells. *Folia Histochem Cytochem* 12: 37-44.

Demole V, Lerch P. 1956. [Normality of fixation of radioactive iodine in the thyroid of rats during experimental fluorosis]. *Helv Physiol Pharmacol Acta* 14(4): 62-63.

Dhurvey V, Patil V, Thakare M. 2017. Effect of sodium fluoride on the structure and function of the thyroid and ovary in albino rats (rattus norvegicus). *Fluoride* 50: 235-246.

Domzalska E. 1966. [Influence of sodium fluoride on hypophysis, thyroid gland, parathyroid, and adrenal gland in the white rat]. *Czas Stomatol* 19: 839-844.

El-lethey HS, Kamel MM, Shaheed IB. 2011. Perinatal exposure to sodium fluoride with emphasis on territorial aggression, sexual behaviour and fertility in male rats. *Life Sci J* 8: 686-694.

Flora SJS, Mittal M, Mishra D. 2009. Co-exposure to arsenic and fluoride on oxidative stress, glutathione linked enzymes, biogenic amines and DNA damage in mouse brain. *J Neurol Sci* 285: 198-205.

Flora SJS, Mittal M, Pachauri V, Dwivedi N. 2012. A possible mechanism for combined arsenic and fluoride induced cellular and DNA damage in mice. *Metallomics* 4: 78-90.

Gabovich RD, Verzhikovskaia NV. 1958. [Effect of fluoride compounds on absorption of radioactive iodine by the thyroid gland in humans and in experimental conditions]. *Probl Endokrinol Gormonoter* 4: 49-54.

Galamini-Ligori M, Di Blasi F. 1961. [Action of sodium fluoride on the thyroid of hypophysectomized rats]. *Boll Soc Ital Biol Sper* 37: 1503-1506.

Galletti P, Held HR, Korrodi H, Wegmann T. 1956. [Investigations on the possible damaging effect of fluorine on the thyroid gland]. *Helv Med Acta* 23: 601-605.

Ge Y, Ning H, Feng C, Wang H, Yan X, Wang S, Wang J. 2006. Apoptosis in brain cells of offspring rats exposed to high fluoride and low iodine. *Fluoride* 39: 173-178.

Ge Y, Niu R, Zhang J, Wang J. 2011. Proteomic analysis of brain proteins of rats exposed to high fluoride and low iodine. *Arch Toxicol* 85: 27-33.

Ge YM, Ning HM, Wang SL, Wang JD. 2005. DNA damage in thyroid gland cells of rats exposed to long-term intake of high fluoride and low iodine. *Fluoride* 38: 318-323.

Ge YM, Ning HM, Wang SL, Wang JD. 2005. Comet assay of DNA damage in brain cells of adult rats exposed to high fluoride and low iodine. *Fluoride* 38: 209-214.

Ge YM, Ning HM, Wang SL, Wang JD. 2005. Effects of high fluoride and low iodine on brain histopathology in offspring rats. *Fluoride* 38: 127-132.

Ge YM, Ning HM, Gu XL, Yin M, Yang XF, Qi YH, Wang JD. 2013. Effects of high fluoride and low iodine on thyroid function in offspring rats. *J Integr Agric* 12: 502-508.

Guan ZZ. 1986. [Morphology of the brain of the offspring of rats with chronic fluorosis]. *Chin J Pathol* 15: 297-299.

Guan Z, Wang Y, Xiao K. 1997. [Influence of experimental fluorosis on phospholipid content and fatty acid composition in rat brain]. *Chin Med J* 77: 592-596.

Guan Z-Z, Wang Y-N, Xiao K-Q, Dai D-Y, Chen Y-H, Liu J-L, Sindelar P, Dallner G. 1998. Influence of chronic fluorosis on membrane lipids in rat brain. *Neurotoxicol Teratol* 20: 537-542.

Guan ZZ, Shan KR, Wang YN, Dallner G. 2006. [Changes of lipids and nicotinic receptors in rat brains and pheochromocytoma with fluorosis]. *Chin J Endemiol* 25: 121-124.

Gushchin SK. 1951. [Effect of sodium fluoride on iodine metabolism in rabbit tissue organs; on the etiology of endemic goiter]. *Gig Sanit* 2: 45-48.

Hamza RZ, Al-Harbi MS. 2014. Sodium fluoride induced neurotoxicity and possible antioxidant role of selenium and curcumin in male mice. *Biosci Biotechnol Res Asia* 11: 81-87.

Hamza RZ, El-Shenawy NS, Ismail HAA. 2015. Protective effects of blackberry and quercetin on sodium fluoride-induced oxidative stress and histological changes in the hepatic, renal, testis and brain tissue of male rat. *J Basic Clin Physiol Pharmacol* 26: 237-251.

Haojun Z, Yaoling W, Ke Z, Jin L, Junling W. 2012. Effects of NaF on the expression of intracellular Ca2+ fluxes and apoptosis and the antagonism of taurine in murine neuron. *Toxicol Mech Methods* 22: 305-308.

Hara K. 1980. Studies on fluorosis especially effects of fluoride on thyroid metabolism. *J Dent Health* 30: 42-57.

Harris NO, Hayes RL. 1955. A tracer study of the effect of acute and chronic exposure to sodium fluoride on the thyroid iodine metabolism of rats. *J Dent Res* 34: 470-477.

Hassan HA, Abdel-Aziz AF. 2010. Evaluation of free radical-scavenging and anti-oxidant properties of black berry against fluoride toxicity in rats. *Food Chem Toxicol* 48: 1999-2004.

Hoogstratten B, Leone NCLG, Shupe J, Greenwood DA, Lieberman J. 1965. Effect of fluorides on hematopoietic system, liver, and thyroid gland in cattle. *J Amer Med Assoc* 192: 26-32.

Inkielewicz I, Rogowska M, Krechniak J. 2006. Lipid peroxidation and antioxidant enzyme activity in rats exposed to fluoride and ethanol. *Fluoride* 39: 53-59.

Inkielewicz I, Czarnowski W. 2008. Oxidative stress parameters in rats exposed to fluoride and aspirin. *Fluoride* 41: 76-82.

Inkielewicz-Stepniak I, Czarnowski W. 2010. Oxidative stress parameters in rats exposed to fluoride and caffeine. *Food Chem Toxicol* 48: 1607-1611.

Jiang P, Li G, Zhou X, Wang C, Qiao Y, Liao D, Shi D. 2018. Chronic fluoride exposure induces neuronal apoptosis and impairs neurogenesis and synaptic plasticity: Role of GSK-3beta/beta-catenin pathway. *Chemosphere* 214: 430-435.

Jiang SF, Xi SH, Yao SQ, Tong JW, Zhang YS, Wang Q, Su J, Li MY. 2013. [Effects of fluoride, arsenic and co-exposure on expression of Bcl-2 and Bax in hippocampus and cerebral cortex of rats]. *Chin J Endemiol* 32: 365-369.

Jiang Y, Guo X, Sun Q, Shan Z, Teng W. 2016. Effects of excess fluoride and iodide on thyroid function and morphology. *Biol Trace Elem Res* 170: 382-389.

Jin TX, Guan ZZ, Zhang H. 2011. [The effect of fluoride on α subunit of calcium/calmodulin-dependent protein kinase-II mRNA and protein expression in central nervous system]. *Chin J Endemiol* 30: 247-250.

Jonderko G, Kita K, Pietrzak J, Primus-Slowinska B, Ruranska B, Zylka-Wloszczyk M, Straszecka J. 1983. [Effect of subchronic sodium fluoride poisoning on the thyroid gland of rabbits with normal and increased supply of iodine]. *Endokrynol Pol* 34: 195-203.

Kahl S, Bobek S. 1975. [Effect of fluoride administration on radiothyroxine turnower in rats]. *Endokrynol Pol* 26: 391-396.

Kahl S, Ewy Z. 1975. Effect of single and long term sodium fluoride administration on biosynthesis of the thyroid hormone in rats. *Fluoride* 8: 191-198.

Kapoor V, Prasad T, Paliwal VK. 2001. Blood biochemical constituents in calves following subclinical levels of fluoride toxicosis. *Fluoride* 34: 126-131.

Karawya FS, Zahran NM, Azzam EZ. 2015. Is water fluoridation a hidden cause of obesity? Histological study on thyroid follicular cells of albino rats. *Egypt J Histol* 38: 547-557.

Kaur T, Bijarnia RK, Nehru B. 2009. Effect of concurrent chronic exposure of fluoride and aluminum on rat brain. *Drug Chem Toxicol* 32: 215-221.

Kelimu A, Liu KT, Lian J, Hu HH, Zheng YJ, Wang TM. 2008. [Effects of vitamin C and E on the ultrastructure in liver, kidney and brain of fluorosis rats]. *Chin J Endemiol* 27: 378-381.

Kinawy AA. 2019. Synergistic oxidative impact of aluminum chloride and sodium fluoride exposure during early stages of brain development in the rat. *Environ Sci Pollut Res Int* 26(11): 10951-10960.

Knizhnikov VA. 1959. [Effect of potable water with high fluoride concentration on thyroid function]. *Gig Sanit* 24: 20-25.

Knizhnikov VA, Tsypin AB, Shcherbova EN, Bugryshev PF. 1963. [The effect of drinking water with an increased fluorine content on the bioelectrical activity of the brain and heart under experimental conditions]. *Gig Sanit* 28: 16-19.

Kondo T, Yoshida M, Kasahara K. 1976. [Acute fluorosis in female rats: Time of inhibition and recovery of cholinesterase in serum and salivary glands]. *Jpn J Dent Health* 26: 187-192.

Kowalewska M. 1974. [Biopotentials of the organ of hearing in chronic sodium fluoride poisoning]. *J Pol Otolaryngol* 28: 417-424.

Krechniak J, Inkielewicz I. 2005. Correlations between fluoride concentrations and free radical parameters in soft tissues of rats. *Fluoride* 38: 293-296.

Leonard BE. 1972. Effect of phentolamine on the increase in brain glycolysis following the intraventricular administration of dibutyryl-3,5-cyclic adenosine monophosphate and sodium fluoride to mice. *Biochem Pharmacol* 21: 115-117.

Liu G, Zhang W, Jiang P, Li X, Liu C, Chai C. 2012. Role of nitric oxide and vascular endothelial growth factor in fluoride-induced goitrogenesis in rats. *Environ Toxicol Pharmacol* 34: 209-217.

Li H, Cai Q, Wang D. 2012. [Effect of fluoride on the expression of rat thyroid peroxidase mRNA]. *Chin J Endemiol* 31: 515-517.

Li H, Cai Q, Wang D. 2012. [Effects of fluoride on rat thyroid morphology, thyroid peroxidase activity and the expression of thyroid peroxidase protein]. *Chin J Endemiol* 31: 271-274.

Liu H, Hou C, Zeng Q, Zhao L, Cui Y, Yu L, Wang L, Zhao Y, Nie J, Zhang B, Wang A. 2016. Role of endoplasmic reticulum stress-induced apoptosis in rat thyroid toxicity caused by excess fluoride and/or iodide. *Environ Toxicol Pharmacol* 46: 277-285.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. [Changes of the c-Jun N-terminal kinase in the brains of rats with chronic fluorosis]. *Chin J Endemiol* 29: 608-612.

Liu YJ, Guan ZZ, Gao Q, Pei JJ. 2011. Increased level of apoptosis in rat brains and SH-SY5Y cells exposed to excessive fluoride-A mechanism connected with activating JNK phosphorylation. *Toxicol Lett* 204: 183-189.

Lohakare J, Pattanaik AK. 2013. Effects of addition of fluorine in diets differing in protein content on urinary fluoride excretion, clinical chemistry and thyroid hormones in calves. *Brazilian J Anim Sci* 42: 751-758.

Long YG, Wang YN, Chen J, Jiang SF, Nordberg A, Guan ZZ. 2002. Chronic fluoride toxicity decreases the number of nicotinic acetylcholine receptors in rat brain. *Neurotoxicol Teratol* 24: 751-757.

Lou DD, Liu YF, Zhang KL, Yu YN, Guan ZZ. 2011. [Changes of reactive oxygen species level and mitochondria fission-fusion in cortical neurons of rats with chronic fluorosis]. *Chin J Endemiol* 30: 256-260.

Lou DD, Liu YF, Qin SL, Zhang KL, Yu YN, Guan ZZ. 2012. [Changed transcription level of mitochondrial fission and fusion gene loci in cortical neurons of rats with chronic fluorosis]. *Chin J Endemiol* 31: 125-129.

Lou DD, Zhang KL, Qin SL, Liu YF, Yu YN, Guan ZZ. 2012. [Alteration of mitochondrial distribution and gene expression of fission 1 protein in cortical neurons of rats with chronic fluorosis]. *Chin J Pathol* 41: 243-247.

Lou DD, Pan JG, Zhang KL, Qin SL, Liu YF, Yu YN, Guan ZZ. 2013. [Changed expression of mito-fusion 1 and mitochondrial fragmentation in the cortical neurons of rats with chronic fluorosis]. *Chin J Prev Med* 47: 170-174.

Lou DD, Zhang KL, Pan JG, Qin SL, Liu YF, Yu YN, Guan ZZ. 2013. [Influence of chronic fluorosis on the expression of mitochondrial fission protein dynamin-related 1 in the cortical neurons of rats]. *Chin J Prev Med* 47: 561-564.

Lou DD, Zhang KL, Qin SL, Liu YF, Liu YJ, Guan ZZ. 2013. [Effects of chronic fluorosis on 4.8 kb mitochondrial DNA in liver, kidney and brain of rats]. *Chin J Endemiol* 32: 121-124.

Lou DD, Guan ZZ, Pei JJ. 2014. Alterations of apoptosis and expressions of Bax and Bcl-2 in the cerebral cortices of rats with chronic fluorosis. *Fluoride* 47: 199-207.

Luo GY, Niu RY, Sun ZL, Zhang JH, Wang JM, Wang C, Wang JD. 2011. Reduction of CaMKII expression in the hippocampus of rats from ingestion of fluoride and/or lead. *Fluoride* 44: 63-69.

Ma T, Liu D, Song K. 1999. Cytochemical study of neuron enzyme at anterior horn of spinal cord in rats with experimental fluorosis. *J Chin Med Univ* 28: 81-82.

Ma TX, Yu HT, Song KQ. 2008. [Expression of c-fos and Caspase 8 in cerebral cortex of rats with experimental fluorosis]. *Chin J Endemiol* 27: 131-133.

Mach Z, Zygulska-Machowa H. 1959. O wplywie fluoru na przemiane J131 [Russian and English summ.]. *Endokrynol Pol* 10: 157-162.

Machida H. 1989. [A study on the rabbit thermoregulatory system effects of high dose sodium fluoride]. *Dent Sci Rep* 89: 607-626.

Madan J, Puri JP, Singh JK. 2009. Growth, feed efficiency and blood profile of buffalo calves consuming high levels of fluoride. *Trop Anim Health Prod* 41: 295-298.

Madhusudhan N, Basha PM, Begum S, Ahmed F. 2009. Fluoride-induced neuronal oxidative stress and its amelioration by antioxidants in developing rats. *Fluoride* 42: 179-187.

Madhusudhan N, Basha PM, Rai P, Ahmed F, Prasad GR. 2010. Effect of maternal fluoride exposure on developing CNS of rats: Protective role of Aloe vera, Curcuma longa and Ocimum sanctum. *Indian J Exp Biol* 48: 830-836.

Manocha SL, Warner H, Olkowski ZL. 1975. Cytochemical response of kidney, liver and nervous system of fluoride ions in drinking water. *Histochem J* 7: 343-355.

Mansour HH, Tawfik SS. 2012. Efficacy of lycopene against fluoride toxicity in rats. *Pharm Biol* 50: 707-711.

Mietkiewski K, Walczak M, Trojanowicz R. 1966. [Effect of sodium fluoride on the neurosecretory system in guinea pigs]. *Endokrynol Pol* 17: 121-131.

Mohamed NE. 2016. The role of calcium in ameliorating the oxidative stress of fluoride in rats. *Biol Trace Elem Res* 170: 128-144.

Muhlemann HR, Schneider R. 1956. [Mitotic activity of rat thyroid epithelium after administration of fluoridated drinking water]. *Schweiz Med Wochenschr* 86: 625-627.

Nabavi SF, Eslami S, Moghaddam AH, Nabavi SM. 2011. Protective effects of curcumin against fluoride-induced oxidative stress in the rat brain. *Neurophysiology* 43: 287-291.

Nabavi SF, Moghaddam AH, Nabavi SM, Eslami S. 2011. Protective effect of curcumin and quercetin on thyroid function in sodium fluoride intoxicated rats. *Fluoride* 44: 147-152.

Nabavi SF, Habtemariam S, Jafari M, Sureda A, Nabavi SM. 2012. Protective role of gallic acid on sodium fluoride induced oxidative stress in rat brain. *Bull Environ Contam Toxicol* 89: 73-77.

Nabavi SF, Nabavi SM, Latifi AM, Mirzaei M, Habtemariam S, Moghaddam AH. 2012. Mitigating role of quercetin against sodium fluoride-induced oxidative stress in the rat brain. *Pharm Biol* 50: 1380-1383.

Nabavi SF, Nabavi SM, Habtemariam S, Moghaddam AH, Sureda A, Mirzaei M. 2013. Neuroprotective effects of methyl-3-O-methyl gallate against sodium fluoride-induced oxidative stress in the brain of rats. *Cell Mol Neurobiol* 33: 261-267.

Nabavi SM, Sureda A, Nabavi SF, Latifi AM, Moghaddam AH, Hellio C. 2012. Neuroprotective effects of silymarin on sodium fluoride-induced oxidative stress. *J Fluor Chem* 142: 79-82.

Narayanaswamy M, Piler MB. 2010. Effect of maternal exposure of fluoride on biometals and oxidative stress parameters in developing CNS of rat. *Biol Trace Elem Res* 133: 71-82.

Narbutt B, Romer TE, Grabski J, Szymik N. 1971. [Influence of natrium fluoride on the structure of the rat thyroid]. *Endokrynol Pol* 22: 445-451.

Narbutt B, Romer TE, Grabski J, Szymik N. 1971. [The influence of sodium fluoride on the morphology of the thyroid gland in rats]. *Endokrynol Pol* 22: 361-365.

Niu RY, Sun ZL, Cheng ZT, Liu HT, Chen HC, Wang JD. 2008. Effects of fluoride and lead on N-methyl-D-aspartate receptor 1 expression in the hippocampus of offspring rat pups. *Fluoride* 41: 101-110.

Niu R, Wang J, Sun Z, Xue X, Yan X, Zhang J. 2015. Transcriptional regulatory dynamics of the hypothalamic-pituitary-testicular axis in male mice exposed to fluoride. *Environ Toxicol Pharmacol* 40: 557-562.

Niu R, Zhang Y, Liu S, Liu F, Sun Z, Wang J. 2015. Proteome alterations in cortex of mice exposed to fluoride and lead. *Biol Trace Elem Res* 164: 99-105.

Ogilvie AL. 1952. Histological findings in the kidney, liver, pancreas, adrenal and thyroid gland of the rat following sodium fluoride administration. *J Dent Res* 31: 598-598.

Okayasu I, Tsuchida M, Yanagisawa F. 1985. Hyperplastic nodules of thyroid parafollicular cells (C cells) in rats induced by prolonged low dose ingestion of NaF. *Fluoride* 18: 111-117.

Pal S, Sarkar C. 2014. Protective effect of resveratrol on fluoride induced alteration in protein and nucleic acid metabolism, DNA damage and biogenic amines in rat brain. *Environ Toxicol Pharmacol* 38: 684-699.

Pan Y, Lu P, Yin L, Chen K, He Y. 2015. Effect of fluoride on the proteomic profile of the hippocampus in rats. *Z Naturforsch C* 70: 151-157.

Phillips PH, Lamb AR. 1934. Histology of certain organs and teeth in chronic toxicosis due to fluorin. *Arch Path* 17: 169-176.

Portela ML. 1972. [Biochemical effects in the prolonged ingestion of fluorides in rats]. *Arch Latinoam Nutr* 22: 291-308.

Prestes DS, Filappi A, Schossler DR, Duarte FA, Dressler VL, Flores EMM, Cecim M. 2009. Functional and histological evaluations of thyroid of sheep submitted to sodium fluoride administration. *Arq Bras Med Vet Zootec* 61: 293-298.

Puentes F, Cremer HD. 1966. Experiments on fluoride-iodine antagonism in the thyroid gland. *Adv Fluorine Res* 4: 213-220.

Qian W, Miao K, Li T, Zhang Z. 2013. Effect of selenium on fluoride-induced changes in synaptic plasticity in rat hippocampus. *Biol Trace Elem Res* 155: 253-260.

Qing-Feng S, Ying-Peng X, Tian-Tong X. 2019. Matrix metalloproteinase-9 and p53 involved in chronic fluorosis induced blood-brain barrier damage and neurocyte changes. *Arch Med Sci* 15(2): 457-466.

Qiu YH, Kong DM, Yang Q, Zhao N. 2010. [Influence of high-fluoride on thyroid function and brain damage in rats]. *Chin J Endemiol* 29: 146-149.

Raghavendra M, Ravindra RK, Raghuveer YP, Narasimha JK, Uma MRV, Navakishor P. 2016. Alleviatory effects of hydroalcoholic extract of cauliflower (brassica oleracea var. botrytis) on thyroid function in fluoride intoxicated rats. *Fluoride* 49: 84-90.

Rakhov GM. 1964. [Effect of calcium and fluorine in drinking water on the iodine metabolism and status of the thyroid gland in iodine insufficiency in food]. *Gig Sanit* 29: 12-17.

Ranpariya VL, Parmar SK, Sheth NR, Chandrashekhar VM. 2011. Neuroprotective activity of matricaria recutita against fluoride-induced stress in rats. *Pharm Biol* 49: 696-701.

Reddy KP, Sailaja G, Krishnaiah C. 2009. Protective effects of selenium on fluoride induced alterations in certain enzymes in brain of mice. *J Environ Biol* 30: 859-864.

Rogalska A, Kuter K, Zelazko A, Glogowska-Gruszka A, Swietochowska E, Nowak P. 2017. Fluoride alteration of [3H]glucose uptake in Wistar rat brain and peripheral tissues. *Neurotoxicol Res* 31: 436-443.

Saka O, Hallac P, Urgancioğlu I. 1965. The effect of fluoride on the thyroid of the rat. *New Istanbul Contrib Clin Sci* 8: 87-90.

Samanta A, Chanda S, Bandyopadhyay B, Das N. 2016. Establishment of drug delivery system nanocapsulated with an antioxidant (+)-catechin hydrate and sodium meta borate chelator against sodium fluoride induced oxidative stress in rats. *J Trace Elem Med Biol* 33: 54-67.

Sarkar C, Das N, Pal S, Dinda B. 2014. Oxidative stress induced alteration of protein and nucleic acid metabolism in fluoride-intoxicated rat brain: Protection by 3α-hydroxy olean-12-en-27-oic acid isolated from neanotis wightiana. *Int J Pharm Sci Res* 5: 3047-3066.

Sarkar C, Pal S. 2014. Ameliorative effect of resveratrol against fluoride-induced alteration of thyroid function in male Wistar rats. *Biol Trace Elem Res* 162: 278-287.

Sarkar C, Pal S, Das N, Dinda B. 2014. Ameliorative effects of oleanolic acid on fluoride induced metabolic and oxidative dysfunctions in rat brain: Experimental and biochemical studies. *Food Chem Toxicol* 66: 224-236.

Seffner W, Teubener W, Runde H, Wiedner H, Vogt J, Otto G, Zschau E, Geinitz D, Franke J. 1990. Boron as an antidote to fluorosis? II. Studies on various organs of pigs. *Fluoride* 23: 68-79.

Selim AOA, El-Haleem MR, Ibrahim IH. 2012. Effect of sodium fluoride on the thyroid gland of growing male albino rats: Histological and biochemical study. *Egypt J Histol* 35: 470-482.

Shao Q, Wang Yn, Guan Z. 2000. [Influence of free radical inducer on the level of oxidative stress in brain of rats with fluorosis]. *Chin J Prev Med* 34: 330-332.

Sharma C, Suhalka P, Sukhwal P, Jaiswal N, Bhatnagar M. 2014. Curcumin attenuates neurotoxicity induced by fluoride: An in vivo evidence. *Pharmacogn Mag* 10: 61-65.

Shashi A. 1992. Studies on alterations in brain lipid metabolism following experimental fluorosis. *Fluoride* 25: 77-84.

Shashi A. 1993. Nucleic acid levels in thyroid gland in acute and chronic fluoride intoxication. *Fluoride* 26: 191-196.

Shashi A, Singh JP, Thapar SP. 1994. Effect of long-term administration of fluoride on levels of protein, free amino acids and RNA in rabbit brain. *Fluoride* 27: 155-159.

Shashi A. 2003. Histopathological investigation of fluoride-induced neurotoxicity in rabbits. *Fluoride* 36: 95-105.

Shashi A, Neetika S, Bhardwaj M. 2009. Neuronal DNA damage and apoptosis in brain of rat exposed to fluoride. *Asian J Microbiol Biotechnol Environ Sci* 11: 629-632.

Shayiq RM, Raza H, Kidwai AM. 1986. Fluoride and lipid peroxidation a comparative study in different rat tissues. *Bull Environ Contam Toxicol* 37: 70-76.

Shen QF, Li HN, Xu TT, Xia YP. 2012. [Damage of blood brain barrier of spinal cord in rats with chronic fluorosis]. *Chin Med J* 92: 2357-2361.

Shen Q, Tian R, Li H, Xu T, Xia Y. 2014. [White matter injury of spinal cord in rats with chronic fluorosis and recovery after defluoridation]. *Chin Med J* 94: 1189-1192.

Shen X, Zhang Z, Xu X. 2004. [Influence of combined iodine and fluoride on phospholipid and fatty acid composition in brain cells of rats]. *J Hyg Res* 33: 158-161.

Shivarajashankara YM, Shivashankara AR, Bhat PG, Rao SH. 2001. Effect of fluoride intoxication on lipid peroxidation and antioxidant systems in rats. *Fluoride* 34: 108-113.

Shivarajashankara YM, Shivashankara AR, Bhat PG, Rao SH. 2002. Brain lipid peroxidation and antioxidant systems of young rats in chronic fluoride intoxication. *Fluoride* 35: 197-203.

Shivarajashankara YM, Shivashankara AR, Bhat PG, Rao SM, Rao SH. 2002. Histological changes in the brain of young fluoride-intoxicated rats. *Fluoride* 35: 12-21.

Siebenhuner L, Miloni E, Burgi H. 1984. [Effects of fluoride on thyroid hormone biosynthesis: Studies in a highly sensitive test system]. *Klin Wochenschr* 62: 859-861.

Singh R, Srivastava AK, Gangwar NK. 2017. Clinico-pathological studies on the co-exposure of cypermethrin and fluoride in experimental rats with ameliorative action of Vitamin E. *Vet Pract* 18(2): 207-210.

Soni KK, Shrivastava VK. 2007. Sodium fluoride induced histopathological changes in thyroid gland of male mus musculus. *Biochem Cell Arch* 7: 317-320.

Stee EW. 1968. *Effect of sodium fluoride and AMOX (NF3O) on growth and thyroid function in the rat*. No. AMRL-TR-67-189. Wright-Patterson Air Force Base, OH: pp. 67.

Štolc V, Podoba J. 1960. Effect of fluoride on the biogenesis of thyroid hormones. *Nature* 188: 855-856.

Sugivama Y. 1967. [The effect of sodium fluoride administration on the parathyroid glands]. *Hirosaki Med J* 19: 520-529.

Sun Y, Ke L, Zheng X, Li T, Ouyang W, Zhang Z. 2016. Effects of different levels of calcium intake on brain cell apoptosis in fluorosis rat offspring and its molecular mechanism. *Biol Trace Elem Res*: 1-12.

Takata H. 1958. The effect of fluorine upon the uptake of I131 by the thyroid glands. *Folia Pharmacol Jpn* 54: 230-236.

Teng Y, Zhang J, Zhang Z, Feng J. 2017. The effect of chronic fluorosis on calcium ions and CaMKIIα, and c-fos expression in the rat hippocampus. *Biol Trace Elem Res*: 295-302.

Trabelsi M, Guermazi F, Zeghal N. 2001. Effect of fluoride on thyroid function and cerebellar development in mice. *Fluoride* 34: 165-173.

Tsuchida M, Okayasu I, Kohyama Y, Kurihara H, Tanaka H, Yanagisawa F, Date C, Hayashi M, Mui K, Asada M. 1986. Effects of long term, low dose ingestion of fluoride on the thyroid gland in rats. *Stud Environ Sci* 27: 307-312.

Vani ML, Reddy KP. 2000. Effects of fluoride accumulation on some enzymes of brain and gastrocnemius muscle of mice. *Fluoride* 33: 17-26.

Wang C, Liang C, Ma J, Manthari RK, Niu R, Wang J, Wang J, Zhang J. 2018. Co-exposure to fluoride and sulfur dioxide on histological alteration and DNA damage in rat brain. *J Biochem Mol Toxicol* 32.

Wang H, Yang Z, Zhou B, Gao H, Yan X, Wang J. 2009. Fluoride-induced thyroid dysfunction in rats: Roles of dietary protein and calcium level. *Toxicol Ind Health* 25: 49-57.

Wang J, Niu R, Sun Z, Lv L, Smith GW. 2008. Effects of protein and calcium supplementation on bone metabolism and thyroid function in protein and calcium deficient rabbits exposed to fluoride. *Fluoride* 41: 283-291.

Wang JD, Ge YM, Ning HM, Wang SL. 2004. Effects of high fluoride and low iodine on oxidative stress and antioxidant defense of the brain in offspring rats. *Fluoride* 37: 264-270.

Wang JL. 2007. [Effect of fluoride on the intracellular Ca2+ in neurons of mice]. *Chin J Endemiol* 26: 505-507.

Wang Y, Guan Z, Xiao K. 1997. [Changes of coenzyme Q content in brain tissues of rats with fluorosis]. *Chin J Prev Med* 31: 330-333.

Wang Y, Dong Y, Wei N, Guan Z. 2015. [Influence of chronic fluorosis on expression of quinone oxidoreductase-1 and heme oxygenase-1 in rat brains]. *Chin J Endemiol* 34: 250-253.

Wedzisz A, Cieciura J. 1988. Effect of small sodium fluoride feed supplements on the serum thyroid hormone content of rats. *Bromatol Chem Toksykol* 21: 174-175.

Wei N, Dong YT, Deng J, Wang Y, Qi XL, Yu WF, Xiao Y, Zhou JJ, Guan ZZ. 2018. Changed expressions of N-methyl-d-aspartate receptors in the brains of rats and primary neurons exposed to high level of fluoride. *J Trace Elem Med Biol* 45: 31-40.

Yan N, Liu Y, Liu S, Cao S, Wang F, Wang Z, Xi S. 2016. Fluoride-induced neuron apoptosis and expressions of inflammatory factors by activating microglia in rat brain. *Mol Neurobiol* 53: 4449-4460.

Yang H, Xing R, Liu S, Yu H, Li P. 2016. Gamma-Aminobutyric acid ameliorates fluoride-induced hypothyroidism in male Kunming mice. *Life Sci* 146: 1-7.

Yang H, Xing R, Liu S, Yu H, Li P. 2019. Analysis of the protective effects of gamma-aminobutyric acid during fluoride-induced hypothyroidism in male Kunming mice. *Pharm Biol* 57(1): 29-37.

Yang M, Ren Z, Zhou B, Guan Z, Yu W. 2017. [Expression of endonuclease G in the brain tissue of rats with chronic fluorosis]. *Chin J Endemiol* 36: 327-332.

Yuan SD, Xie QW, Lu FY. 1993. Changes of serotonin content and turnover rate in hypothalamus of female rat during fluorosis. *Fluoride* 26: 57-60.

Zhai JX, Guo ZY, Hu CL, Wang QN, Zhu QX. 2003. [Studies on fluoride concentration and cholinesterase activity in rat hippocampus]. *Chin J Ind Hyg Occup Dis* 21: 102-104.

Zhan CW, Huo DJ. 1988. Ultrastructural findings in liver, kidneys, thyroid-gland and cardiac-muscle of rabbits following sodium-fluoride administration. *Fluoride* 21: 32-38.

Zhan XA, Xu ZR, Li JX, Wang M. 2005. Effects of fluorosis on lipid peroxidation and antioxidant systems in young pigs. *Fluoride* 38: 157-161.

Zhan XA, Li JX, Wang M, Xu ZR. 2006. Effects of fluoride on growth and thyroid function in young pigs. *Fluoride* 39: 95-100.

Zhang KL, Lou DD, Liu YF, Qin SL, Guan ZZ. 2012. [Changes of P-glycoprotein and nuclear factor κB in the cerebral cortex of rat with chronic fluorosis]. *Chin J Endemiol* 31: 613-616.

Zhang KL, Lou DD, Guan ZZ. 2013. [Expression of receptor for advanced glycation endproducts and nuclear factor κB in brain hippocampus of rat with chronic fluorosis]. *Chin J Endemiol* 32: 625-628.

Zhang WD, Zhang Y, Liu GY, Jiang P, Chai CY. 2008. [Effects of fluoride on ultrastructure of thyroids in rats]. *Chin J Endemiol* 27: 622-624.

Zhang ZG, Wang XY, Nian WW, Liao QX, Zhang R, Ouyang W. 2017. Effects of calcium on drinking fluorosis-induced hippocampal syntaptic plasticity impairment in the offspring of rats. *Transl Neurosci* 8: 191-200.

Zhao W, Zhu H, Yu Z, Aoki K, Misumi J, Zhang X. 1998. Long-term effects of various iodine and fluorine doses on the thyroid and fluorosis in mice. *Endocr Regul* 32: 63-70.

Zhao WY. 1988. [A preliminary study of the interaction of iodine and fluoride in experimental iodine goiter and fluorosis]. *Chin J Prev Med* 22: 146-148.

Zhao XL, Gao WH, Zhao ZL. 1994. [Effects of sodium fluoride on the activity of Ca2+Mg(2+)-ATPase in synaptic membrane in rat brain]. *Chin J Prev Med* 28: 264-266.

Zhavoronkov AA, Polyakova GA. 1973. Morphological and functional state of the hypothalamo-hypophyseal neurosecretory system in experimental fluorosis. *Bull Exp Biol Med* 75: 194-196.

Zhou B, Luo G, Wang C, Niu R, Wang J. 2014. Effects of fluoride on expression of cytokines in the hippocampus of adult rats. *Fluoride* 47: 191-198.

C.2.3. In Vitro Experimental Studies

As described in Figure 2, 60 in vitro experimental studies were included; however, data extraction was not conducted on in vitro studies. Therefore, in vitro experimental studies are not available in HAWC (NTP 2019) with the exception of in vitro studies that also reported in vivo non-human animal data that met the relevant criteria for being made available in HAWC. The following lists of references are organized as studies that are available in HAWC (n = 6) followed by studies that are not available in HAWC (n = 54).

C.2.3.1. Studies Available in HAWC

Chen J, Niu Q, Xia T, Zhou G, Li P, Zhao Q, Xu C, Dong L, Zhang S, Wang A. 2018. ERK1/2-mediated disruption of BDNF-TrkB signaling causes synaptic impairment contributing to fluoride-induced developmental neurotoxicity. *Toxicology* 410: 222-230.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. *Environ Pollut* 233: 889-899.

Shan KR, Qi XL, Long YG, Nordberg A, Guan ZZ. 2004. Decreased nicotinic receptors in PC12 cells and rat brains influenced by fluoride toxicity: A mechanism relating to a damage at the level in post-transcription of the receptor genes. *Toxicology* 200: 169-177.

Zhang KL, Lou DD, Guan ZZ. 2015. Activation of the AGE/RAGE system in the brains of rats and in SH-SY5Y cells exposed to high level of fluoride might connect to oxidative stress. *Neurotoxicol Teratol* 48: 49-55.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao XL, Wu JH. 1998. Actions of sodium fluoride on acetylcholinesterase activities in rats. *Biomed Environ Sci* 11: 1-6.

C.2.3.2. Studies Not Available in HAWC

Ardelean I, Racoveanu N, Manescu S, Lupulescu A, Diaconescu M, Ghelerter L. 1964. Experimental investigations concerning the effect of fluorine on the thyroid gland. *Rum Med Rev* 8: 17-20.

Chen J, Chen X, Yang K. 2000. [Effects of selenium and zinc on the DNA damage caused by fluoride in pallium neural cells of rats]. *J Hyg Res* 29: 216-217.

Chen L, Ning H, Yin Z, Song X, Feng Y, Qin H, Li Y, Wang J, Ge Y, Wang W. 2017. The effects of fluoride on neuronal function occurs via cytoskeleton damage and decreased signal transmission. *Chemosphere* 185: 589-594.

Chen R, Zhao LD, Liu H, Li HH, Ren C, Zhang P, Guo KT, Zhang HX, Geng DQ, Zhang CY. 2017. Fluoride induces neuroinflammation and alters Wnt signaling pathway in BV2 microglial cells. *Inflammation* 40: 1123-1130.

Cheng TJ, Chen TM, Chen CH, Lai YK. 1998. Induction of stress response and differential expression of 70 kDa stress proteins by sodium fluoride in HeLa and rat brain tumor 9L cells. *J Cell Biochem* 69: 221-231.

Deng MF, Zhu D, Liu YP, He WW, Gui CZ, Guan ZZ. 2018. Attenuation by 7-nitroindazole of fluoride-induced toxicity in SH-SY5Y cells exposed to high fluoride: Effects on nitric oxide, nitric oxide synthetase activity, nNOS, and apoptosis. *Fluoride* 51(4): 328-339.

Flores-Mendez M, Ramirez D, Alamillo N, Hernandez-Kelly LC, Del Razo LM, Ortega A. 2014. Fluoride exposure regulates the elongation phase of protein synthesis in cultured Bergmann glia cells. *Toxicol Lett* 229: 126-133.

Gao Q, Liu YH, Guan ZZ. 2008. Oxidative stress might be a mechanism connected with the decreased alpha 7 nicotinic receptor influenced by high-concentration of fluoride in SH-SY5Y neuroblastoma cells. *Toxicol In Vitro* 22: 837-843.

Goschorska M, Gutowska I, Baranowska-Bosiacka I, Piotrowska K, Metryka E, Safranow K, Chlubek D. 2018. Influence of acetylcholinesterase inhibitors used in Alzheimer's Disease treatment on the activity of antioxidant enzymes and the concentration of glutathione in THP-1 macrophages under fluoride-induced oxidative stress. *Int J Environ Res Pub Health* 16(1).

Guan ZZ, Shan KR, Xiu J, Long YG. 2005. [Fluorosis on expression of nicotinic acetylcholine receptors in protein and gene levels in human SH-SY5Y neuroblastoma cells]. *Chin J Prev Med* 39: 26-29.

Guan ZZ, Shan KR, Wang YN, Dallner G. 2006. [Changes of lipids and nicotinic receptors in rat brains and pheochromocytoma with fluorosis]. *Chin J Endemiol* 25: 121-124.

Haojun Z, Yaoling W, Ke Z, Jin L, Junling W. 2012. Effects of NaF on the expression of intracellular Ca2+ fluxes and apoptosis and the antagonism of taurine in murine neuron. *Toxicol Mech Methods* 22: 305-308.

Hong-Liang L, Qiang Z, Yu-Shan C, Lei Z, Gang F, Chang-Chun H, Liang Z, Aiguo W. 2014. Fluoride-induced thyroid cell apoptosis. *Fluoride* 47: 161-169.

Inkielewicz-Stepniak I, Radomski MW, Wozniak M. 2012. Fisetin prevents fluoride- and dexamethasone-induced oxidative damage in osteoblast and hippocampal cells. *Food Chem Toxicol* 50: 583-589.

Jin TX, Guan ZZ, Zhang H. 2011. [The effect of fluoride on α subunit of calcium/calmodulin-dependent protein kinase-II mRNA and protein expression in central nervous system]. *Chin J Endemiol* 30: 247-250.

Kariya T, Kotani M, Field JB. 1974. Effects of sodium fluoride and other metabolic inhibitors on basal and TSH stimulated cyclic AMP and thyroid metabolism. *Metab Clin Exper* 23: 967-973.

Ke L, Zheng X, Sun Y, Ouyang W, Zhang Z. 2016. Effects of sodium fluoride on lipid peroxidation and PARP, XBP-1 expression in PC12 cell. *Biol Trace Elem Res* 173: 161-167.

Lee J, Han YE, Favorov O, Tommerdahl M, Whitsel B, Lee CJ. 2016. Fluoride induces a volume reduction in CA1 hippocampal slices via MAP kinase pathway through volume regulated anion channels. *Exp Neurobiol* 25: 72-78.

Levesque L, Mizzen CA, McLachlan DR, Fraser PE. 2000. Ligand specific effects on aluminum incorporation and toxicity in neurons and astrocytes. *Brain Res* 877: 191-202.

Li H, Gao MT, Xu KY, Wang CY. 2007. Effect of sodium fluoride on the primary porcine thyroid cells and thyroid peroxidase activity. *J Clin Rehabil Tissue Eng Res* 11: 7425-7428.

Li H, Gao MT, Xu KY, Cui MY, Dai X. 2008. [Effect of fluoride on thyroid functioning in primary porcine thyrocyte]. *Chin J Endemiol* 27: 38-40.

Li H, Huang H, Xu Y, Gao Y, Liu Z. 2010. [Toxic effects of fluoride on rat cerebral cortex astrocytes in vitro]. *J Hyg Res* 39: 86-88.

Liu H, Zeng Q, Cui Y, Yu L, Zhao L, Hou C, Zhang S, Zhang L, Fu G, Liu Y, Jiang C, Chen X, Wang A. 2014. The effects and underlying mechanism of excessive iodide on excessive fluoride-induced thyroid cytotoxicity. *Environ Toxicol Pharmacol* 38: 332-340.

Liu HL, Zeng Q, Cui YS, Zhao L, Zhang L, Fu G, Hou CC, Zhang S, Yu LY, Jiang CY, Wang ZL, Chen XM, Wang AG. 2014. The role of the IRE1 pathway in excessive iodide- and/or fluoride-induced apoptosis in Nthy-ori 3-1 cells in vitro. *Toxicol Lett* 224: 341-348.

Liu YJ, Guan ZZ, Gao Q, Pei JJ. 2011. Increased level of apoptosis in rat brains and SH-SY5Y cells exposed to excessive fluoride-A mechanism connected with activating JNK phosphorylation. *Toxicol Lett* 204: 183-189.

Liu Y, Gao Q, Tang Z, Zhang X, Guan Z. 2015. [The expression and correlation between neural nicotinic acetylcholine receptor subunit α3 and mitogen-activated protein kinase cell signaling transduction pathway in human neuroblastoma cell line SH-SY5Y overexposed to fluoride]. *Chin J Endemiol* 34: 553-558.

Madaoui S, Rappaport L, Nunez J. 1974. Prostaglandins and in vitro TSH-dependent iodide binding by rat thyroid glands. *Biochimie* 56: 109-113.

Nakagawa-Yagi Y, Saito Y, Kitoh N, Ogane N, Fujisawa E, Nakamura H. 1993. Fluoride causes suppression of neurite outgrowth in human neuroblastoma via an influx of extracellular calcium. *Biochem Biophys Res Commun* 191: 727-736.

Ong J, Kerr DIB. 1995. Interactions of N-ethylmaleimide and aluminium fluoride with GABA(B) receptor function in rat neocortical slices. *Eur J Pharmacol* 287: 197-200.

Pastan I, Macchia V, Katzen R. 1968. Effect of fluoride on the metabolic activity of thyroid slices. *Endocrinology* 83: 157-160.

Rubakhova VM. 1977. [Effect of serotonin and sodium fluoride on visceral nerve conductors]. *Vyestsi Akademii Navuk BSSR Syeryya Biyalahichnykh Navuk* 1: 117-119.

Shayiq RM, Raza H, Kidwai AM. 1986. Fluoride and lipid peroxidation a comparative study in different rat tissues. *Bull Environ Contam Toxicol* 37: 70-76.

Shuhua X, Ziyou L, Ling Y, Fei W, Sun G. 2012. A role of fluoride on free radical generation and oxidative stress in BV-2 microglia cells. *Mediators Inflamm* 2012: 1-8.

Singh P, Das TK. 2019. Ultrastructural localization of 4-hydroxynonenal adducts in fluoride-exposed cells: Protective role of dietary antioxidants. *Fluoride* 52(1): 49-58.

Taylor P. 1972. Comparison of the effects of various agents on thyroidal adenyl cyclase activity with their effects on thyroid hormone release. *J Endocrinol* 54: 137-145.

Tu W, Zhang Q, Liu Y, Han LY, Wang Q, Chen PP, Zhang S, Wang AG, Zhou X. 2018. Fluoride induces apoptosis via inhibiting SIRT1 activity to activate mitochondrial p53 pathway in human neuroblastoma SH-SY5Y cells. *Toxicol Appl Pharmacol* 347: 60-69.

van der Voet GB, Schijns O, de Wolff FA. 1999. Fluoride enhances the effect of aluminium chloride on interconnections between aggregates of hippocampal neurons. *Arch Physiol Biochem* 107: 15-21.

Wang JL. 2007. [Effect of fluoride on the intracellular Ca2+ in neurons of mice]. *Chin J Endemiol* 26: 505-507.

Wang J, Gao Y, Cheng X, Yang J, Zhao Y, Xu H, Zhu Y, Yan Z, Manthari RK, Mehdi OM, Wang J. 2019. GSTO1 acts as a mediator in sodium fluoride-induced alterations of learning and memory related factors expressions in the hippocampus cell line. *Chemosphere* 226: 201-209.

Wei N, Dong YT, Deng J, Wang Y, Qi XL, Yu WF, Xiao Y, Zhou JJ, Guan ZZ. 2018. Changed expressions of N-methyl-d-aspartate receptors in the brains of rats and primary neurons exposed to high level of fluoride. *J Trace Elem Med Biol* 45: 31-40.

Willems CB-V, Sande J, Dumont JE. 1972. Inhibition of thyroid secretion by sodium fluoride in vitro. *Biochim Biophys Acta* 264: 197-204.

Woodward JJ, Harms J. 1992. Potentiation of N-methyl-D-aspartate-stimulated dopamine release from rat brain slices by aluminum fluoride and carbachol. *J Neurochem* 58: 1547-1554.

Wu J, Cheng M, Liu Q, Yang J, Wu S, Lu X, Jin C, Ma H, Cai Y. 2015. Protective role of tert-butylhydroquinone against sodium fluoride-induced oxidative stress and apoptosis in PC12 cells. *Cell Mol Neurobiol* 35: 1017-1025.

Xia T, Zhang M, He WH, He P, Wang AG. 2007. [Effects of fluoride on neural cell adhesion molecules mRNA and protein expression levels in primary rat hippocampal neurons]. *Chin J Prev Med* 41: 475-478.

Xu B, Xu Z, Xia T, He P, Gao P, He W, Zhang M, Guo L, Niu Q, Wang A. 2011. Effects of the Fas/Fas-L pathway on fluoride-induced apoptosis in SH-SY5Y cells. *Environ Toxicol* 26: 86-92.

Xu Z, Xu B, Xia T, He W, Gao P, Guo L, Wang Z, Niu Q, Wang A. 2013. Relationship between intracellular Ca2+ and ROS during fluoride-induced injury in SH-SY5Y cells. *Environ Toxicol* 28: 307-312.

Yamashita K, Field JB. 1972. Elevation of cyclic guanosine 3,5; monophosphate levels in dog thyroid slices caused by acetylcholine and sodium fluoride. *J Biol Chem* 247: 7062-7066.

Yan L, Liu S, Wang C, Wang F, Song Y, Yan N, Xi S, Liu Z, Sun G. 2013. JNK and NADPH oxidase involved in fluoride-induced oxidative stress in BV-2 microglia cells. *Mediators Inflamm* 2013: 895-975.

Zhang CY, Chen R, Wang F, Ren C, Zhang P, Li Q, Li HH, Guo KT, Geng DQ, Liu CF. 2016. EGb-761 attenuates the anti-proliferative activity of fluoride via DDK1 in PC-12 cells. *Neurochem Res* 42(2): 606-614.

Zhang M, Wang A, He W, He P, Xu B, Xia T, Chen X, Yang K. 2007. Effects of fluoride on the expression of NCAM, oxidative stress, and apoptosis in primary cultured hippocampal neurons. *Toxicology* 236: 208-216.

Zhang M, Wang A, Xia T, He P. 2008. Effects of fluoride on DNA damage, S-phase cell-cycle arrest and the expression of NF-kappaB in primary cultured rat hippocampal neurons. *Toxicol Lett* 179: 1-5.

Zhang S, Zheng X, Sun Y, Wang Y, Zhang Z. 2015. Alterations in oxidative stress and apoptosis in cultured PC12 cells exposed to fluoride. *Fluoride* 48: 213-222.

Zhao L, Xiao Y, Deng CM, Tan LC, Guan ZZ. 2016. Protective effect of lovastatin on neurotoxicity of excessive fluoride in primary hippocampal neurons. *Fluoride* 49: 36-46.

Zhao XL, Gao WH, Zhao ZL. 1994. [Effects of sodium fluoride on the activity of Ca2+Mg(2+)-ATPase in synaptic membrane in rat brain]. *Chin J Prev Med* 28: 264-266.

Appendix D. Risk-of-bias Figures

Figures

D.1. Studies in Humans

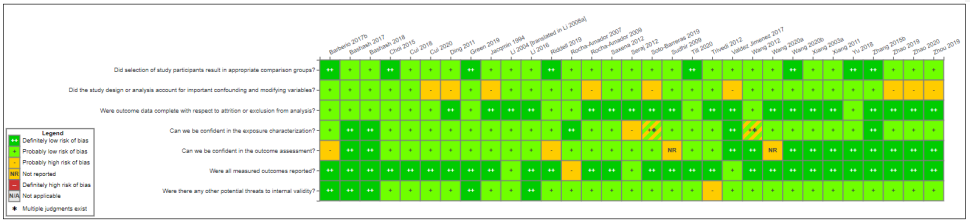


Figure D-1. Risk-of-bias Heatmap for Low Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-1 and additional study details in HAWC here (NTP 2019).

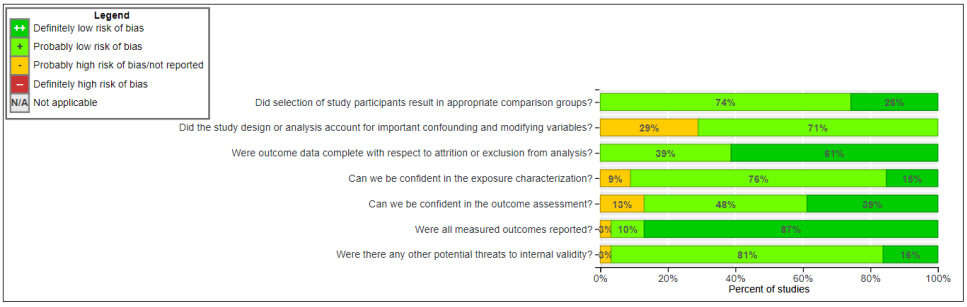


Figure D-2. Risk-of-bias Bar Chart for Low Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-2 and additional study details in HAWC here (NTP 2019).

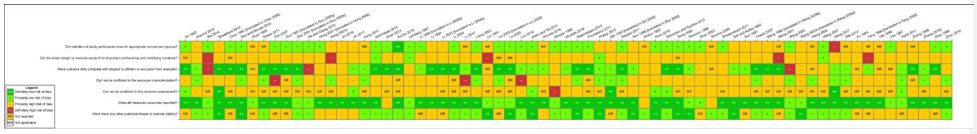


Figure D-3. Risk-of-bias Heatmap for High Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-3 and additional study details in HAWC here (NTP 2019).

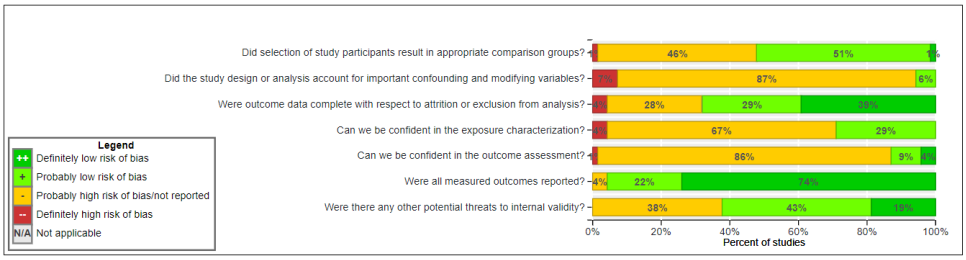


Figure D-4. Risk-of-bias Bar Chart for High Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-4 and additional study details in HAWC here (NTP 2019).

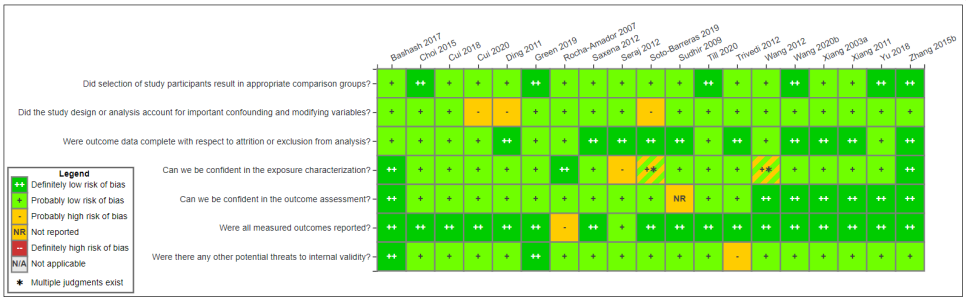


Figure D-5. Risk-of-bias Heatmap for Low Risk-of-bias Children's IQ Studies Following Fluoride Exposure

An interactive version of Figure D-5 and additional study details in HAWC here (NTP 2019).

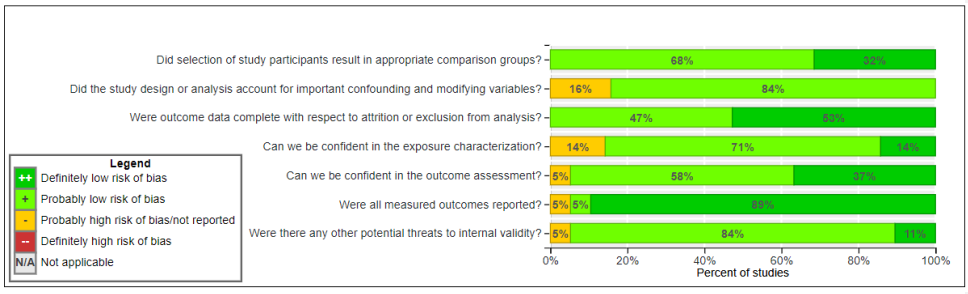


Figure D-6. Risk-of-bias Bar Chart for Low Risk-of-bias Children's IQ Studies Following Fluoride Exposure

An interactive version of Figure D-6 and additional study details in HAWC here (NTP 2019).

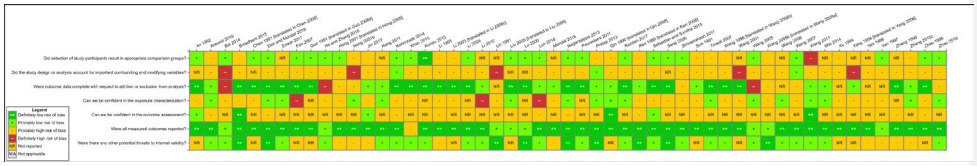


Figure D-7. Risk-of-bias Heatmap for High Risk-of-bias Children's IQ Studies Following Fluoride Exposure

An interactive version of Figure D-7 and additional study details in HAWC here (NTP 2019).

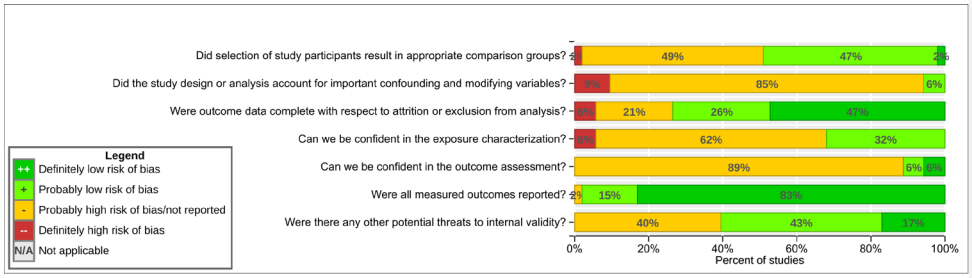


Figure D-8. Risk-of-bias Bar Chart for High Risk-of-bias Children's IQ Studies Following Fluoride Exposure

An interactive version of Figure D-8 and additional study details in HAWC here (NTP 2019).

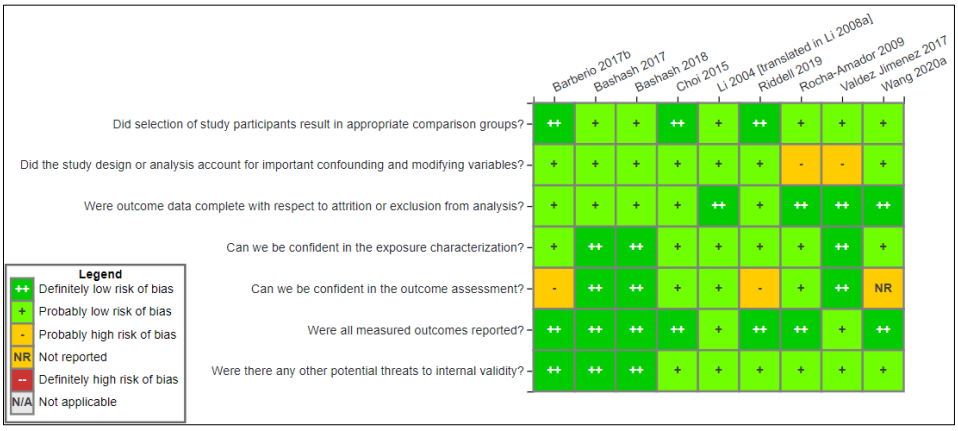


Figure D-9. Risk-of-bias Heatmap for Low Risk-of-bias Children's Other Neurodevelopmental Effect Studies Following Fluoride Exposure

An interactive version of Figure D-9 and additional study details in HAWC here (NTP 2019).

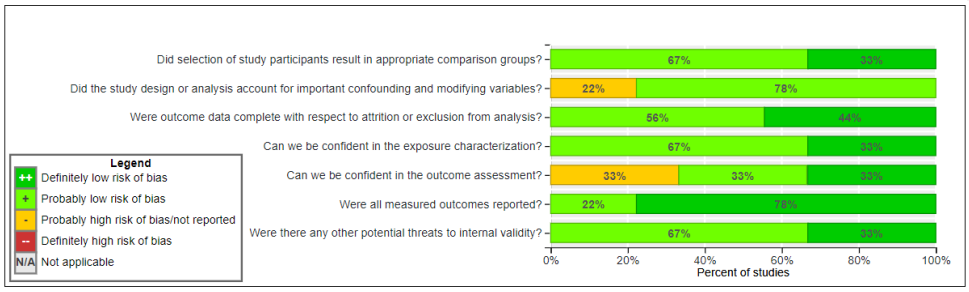


Figure D-10. Risk-of-bias Bar Chart for Low Risk-of-bias Children's Other Neurodevelopmental Effect Studies Following Fluoride Exposure

An interactive version of Figure D-10 and additional study details in HAWC here (NTP 2019).

Figure D-11. Risk-of-bias Heatmap for High Risk-of-bias Children's Other Neurodevelopmental Effect Studies Following Fluoride Exposure

An interactive version of Figure D-11 and additional study details in HAWC here (NTP 2019).

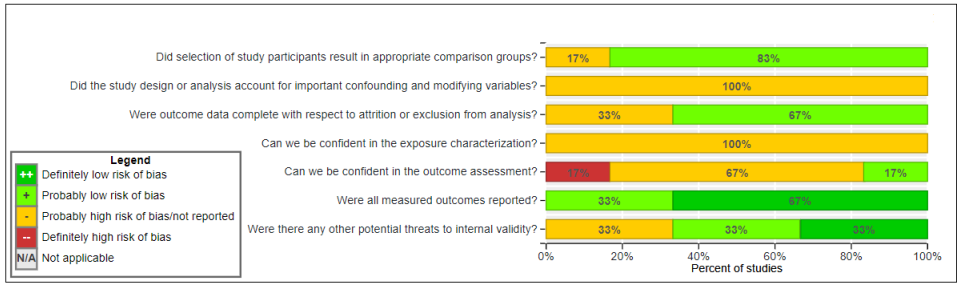


Figure D-12. Risk-of-bias Bar Chart for High Risk-of-bias Children's Other Neurodevelopmental Effect Studies Following Fluoride Exposure

An interactive version of Figure D-12 and additional study details in HAWC here (NTP 2019).

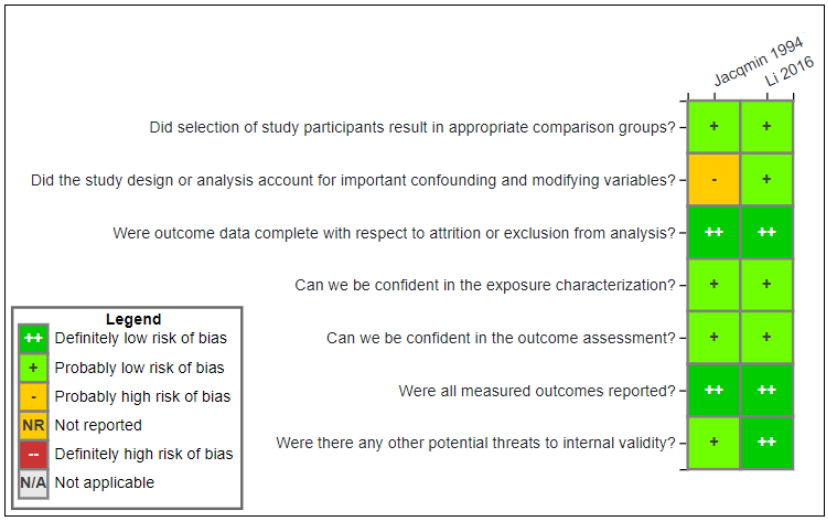


Figure D-13. Risk-of-bias Heatmap for Low Risk-of-bias Adult Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-13 and additional study details in HAWC here (NTP 2019).

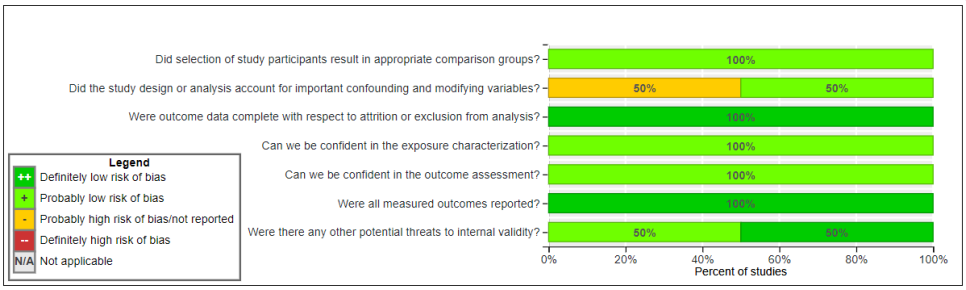


Figure D-14. Risk-of-bias Bar Chart for Low Risk-of-bias Adult Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-14 and additional study details in HAWC here (NTP 2019).

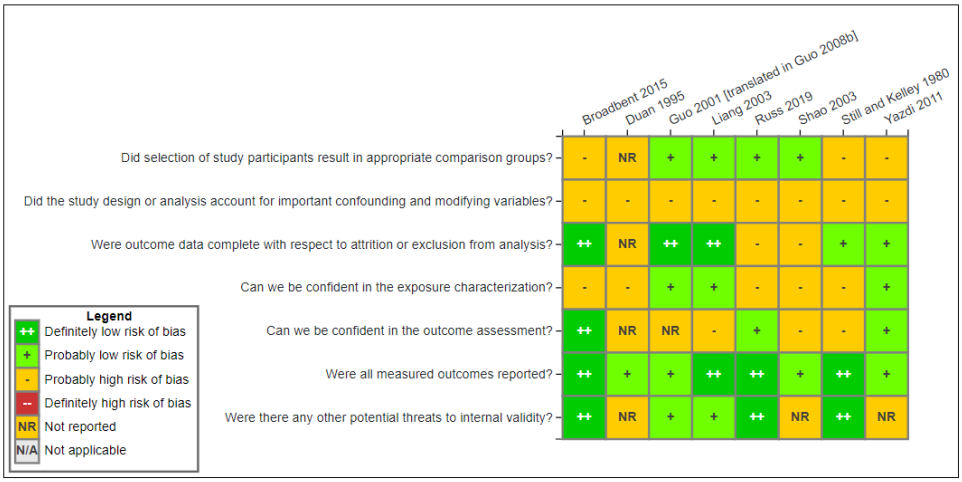


Figure D-15. Risk-of-bias Heatmap for High Risk-of-bias Adult Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-15 and additional study details in HAWC here (NTP 2019).

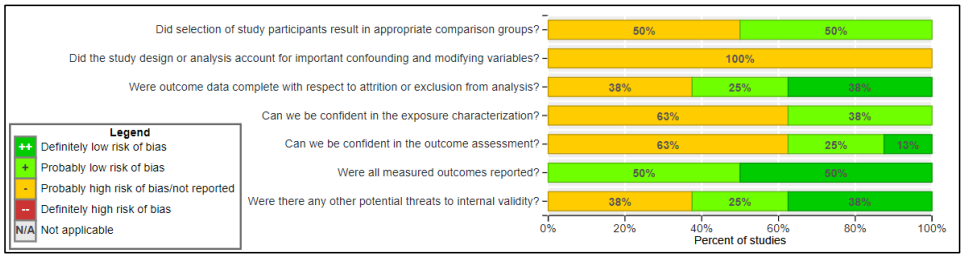


Figure D-16. Risk-of-bias Bar Chart for High Risk-of-bias Adult Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-16 and additional study details in HAWC here (NTP 2019).

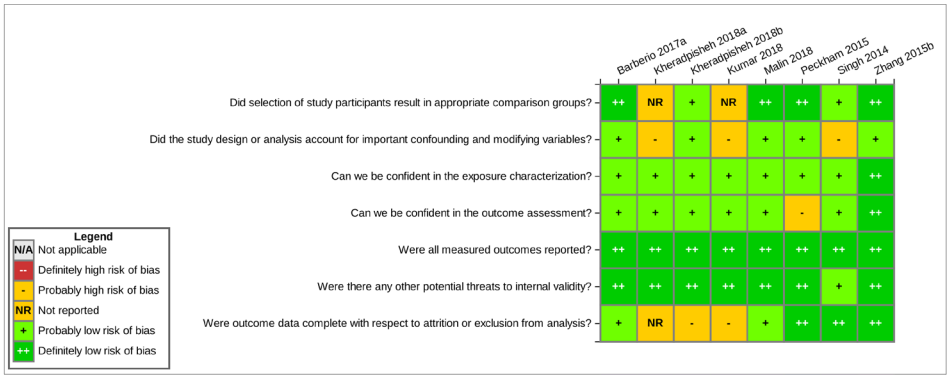


Figure D-17. Risk-of-bias Heatmap for Low Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure

An interactive version of Figure D-17 and additional study details in HAWC here (NTP 2019).

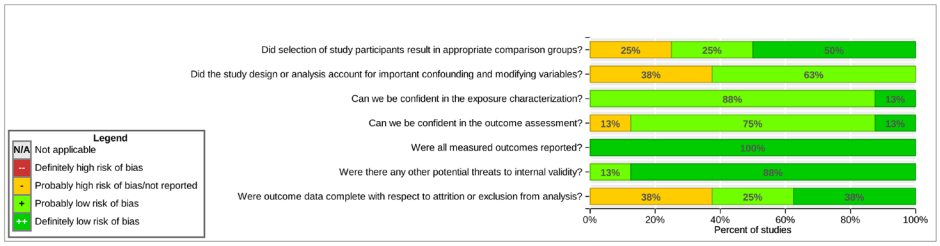


Figure D-18. Risk-of-bias Bar Chart for Low Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure

An interactive version of Figure D-18 and additional study details in HAWC here (NTP 2019).

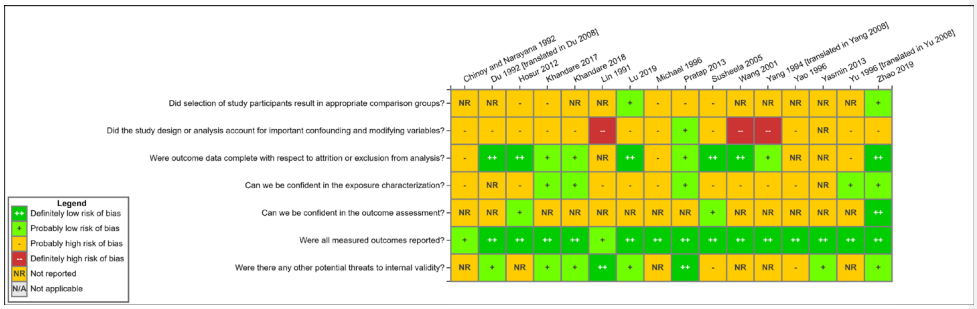


Figure D-19. Risk-of-bias Heatmap for High Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure

An interactive version of Figure D-19 and additional study details in HAWC here (NTP 2019).

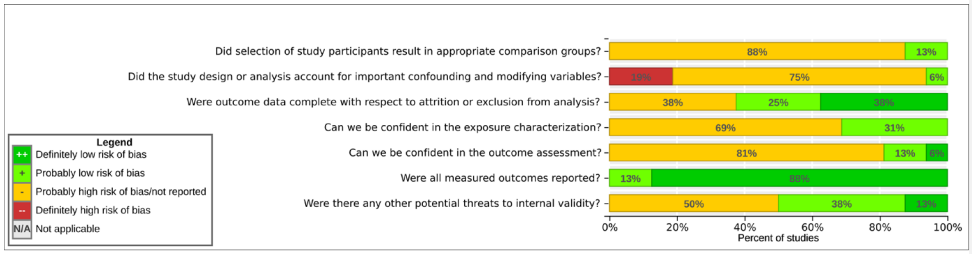


Figure D-20. Risk-of-bias Bar Chart for High Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure

An interactive version of Figure D-20 and additional study details in HAWC here (NTP 2019).

D.2. Studies in Non-human Animals

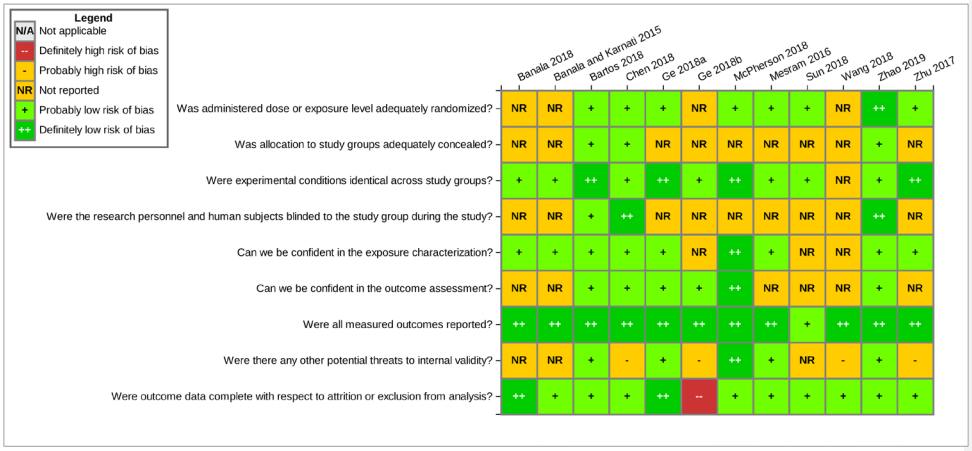


Figure D-21. Risk-of-bias Heatmap for New Developmental Animal Learning and Memory Studies Following Fluoride Exposure

An interactive version of Figure D-21 and additional study details in HAWC here (NTP 2019).

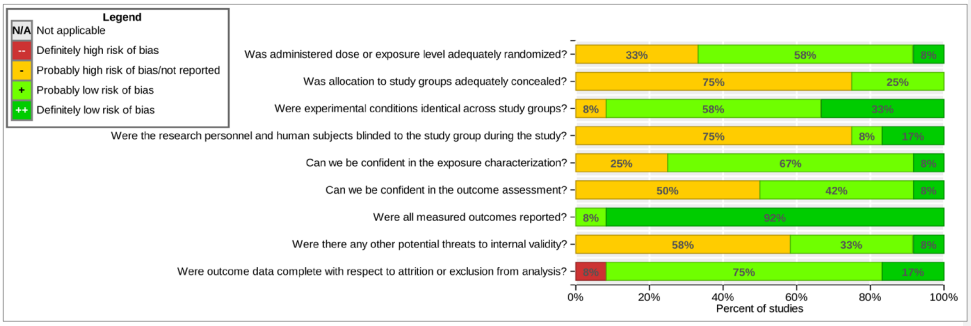


Figure D-22. Risk-of-bias Bar Chart for New Developmental Animal Learning and Memory Studies Following Fluoride Exposure

An interactive version of Figure D-22 and additional study details in HAWC here (NTP 2019).

Figure D-23. Risk-of-bias Heatmap for New Adult Animal Learning and Memory Studies Following Fluoride Exposure

An interactive version of Figure D-23 and additional study details in HAWC here (NTP 2019).

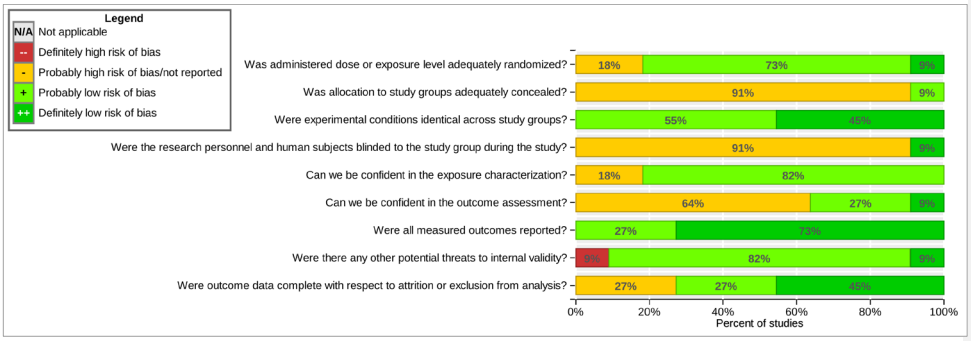


Figure D-24. Risk-of-bias Bar Chart for New Adult Animal Learning and Memory Studies Following Fluoride Exposure

An interactive version of Figure D-24 and additional study details in HAWC here (NTP 2019).

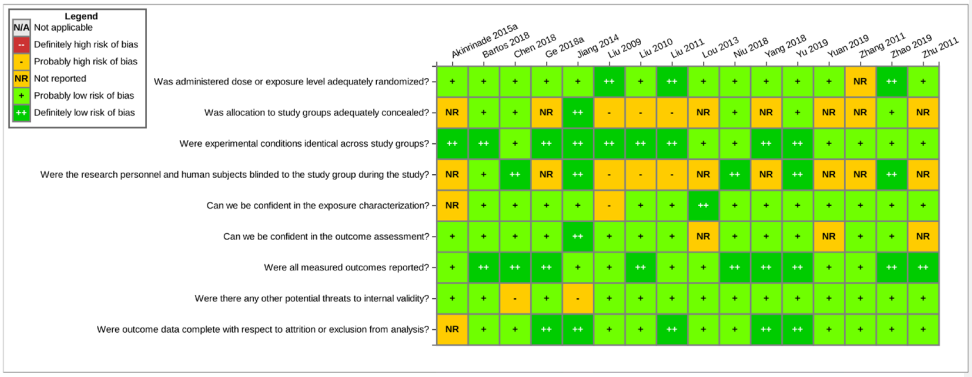


Figure D-25. Risk-of-bias Heatmap for Low Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure

An interactive version of Figure D-25 and additional study details in HAWC here (NTP 2019).

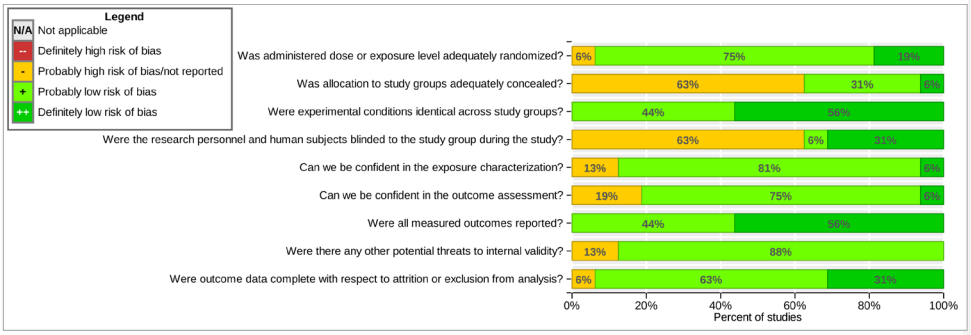


Figure D-26. Risk-of-bias Bar Chart for Low Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure

An interactive version of Figure D-26 and additional study details in HAWC here (NTP 2019).

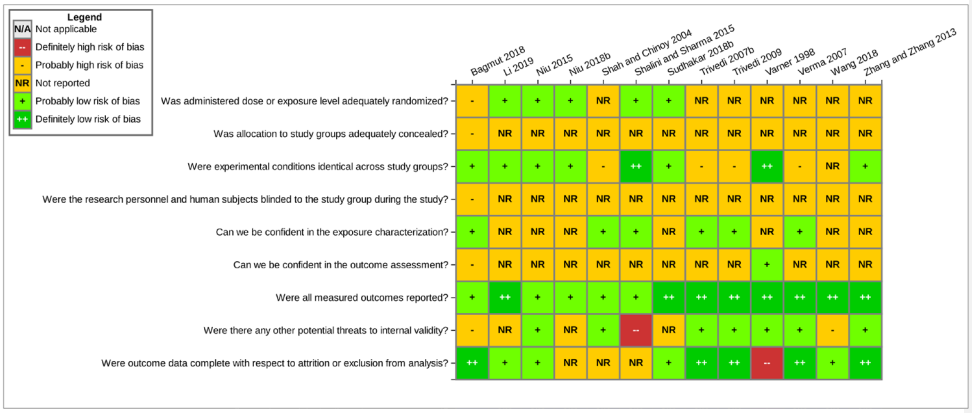


Figure D-27. Risk-of-bias Heatmap for High Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure

An interactive version of Figure D-27 and additional study details in HAWC here (NTP 2019).

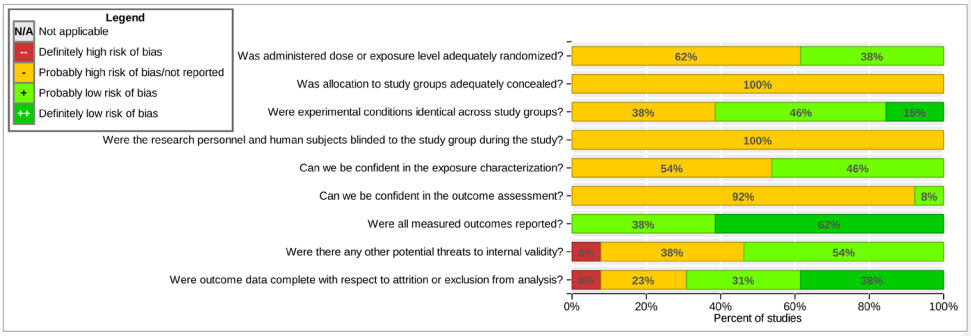


Figure D-28. Risk-of-bias Bar Chart for High Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure

An interactive version of Figure D-28 and additional study details in HAWC here (NTP 2019).

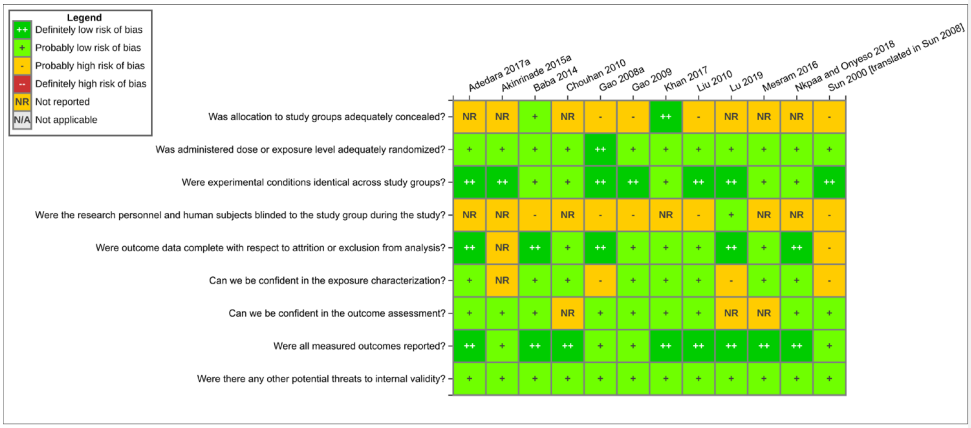


Figure D-29. Risk-of-bias Heatmap for Low Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure

An interactive version of Figure D-29 and additional study details in HAWC here (NTP 2019).

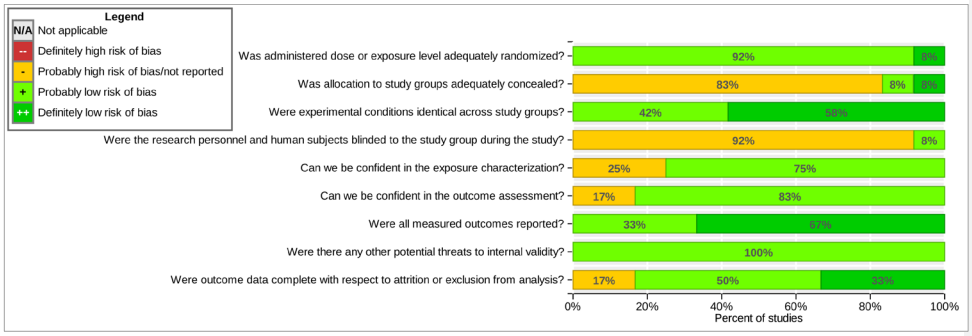


Figure D-30. Risk-of-bias Bar Chart for Low Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure

An interactive version of Figure D-30 and additional study details in HAWC here (NTP 2019).

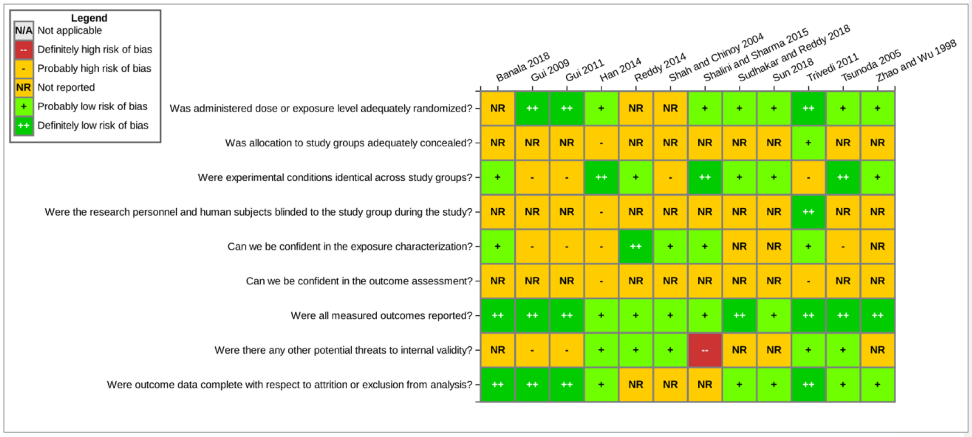


Figure D-31. Risk-of-bias Heatmap for High Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure

An interactive version of Figure D-31 and additional study details in HAWC here (NTP 2019).

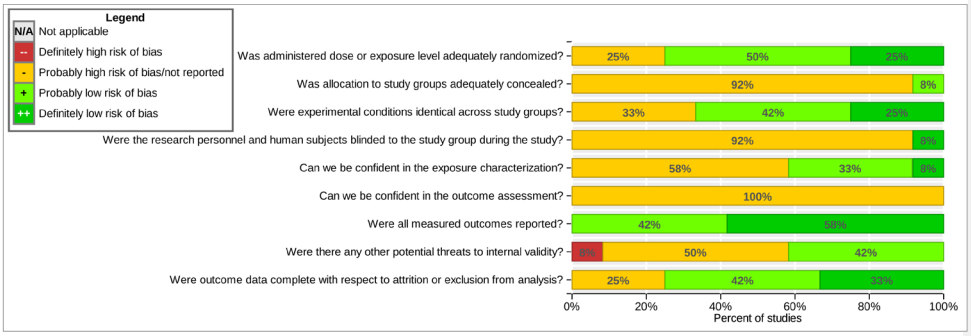


Figure D-32. Risk-of-bias Bar Chart for High Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure

An interactive version of Figure D-32 and additional study details in HAWC here (NTP 2019).

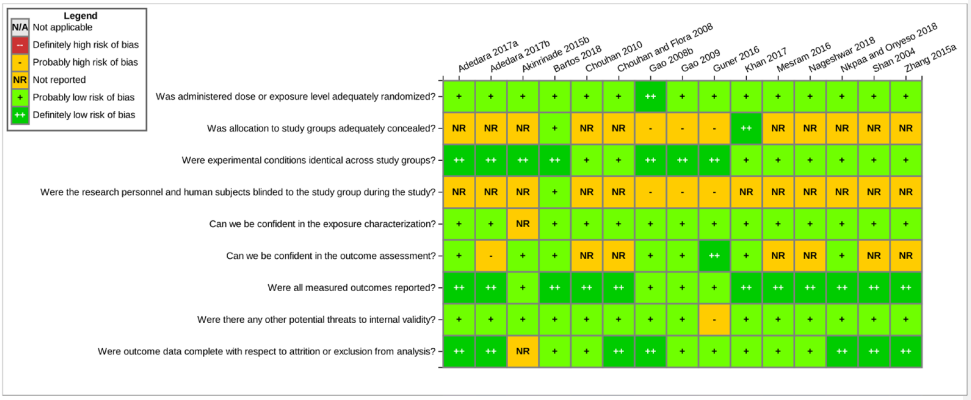


Figure D-33. Risk-of-bias Heatmap for Low Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure

An interactive version of Figure D-33 and additional study details in HAWC here (NTP 2019).

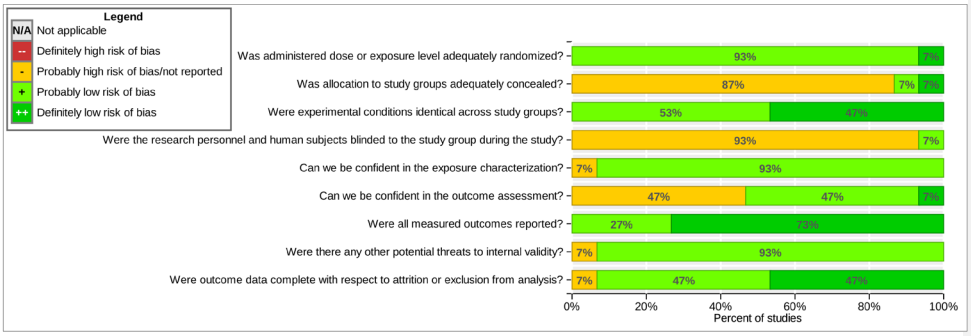


Figure D-34. Risk-of-bias Bar Chart for Low Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure

An interactive version of Figure D-34 and additional study details in HAWC here (NTP 2019).

Figure D-35. Risk-of-bias Heatmap for High Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure

An interactive version of Figure D-35 and additional study details in HAWC here (NTP 2019).

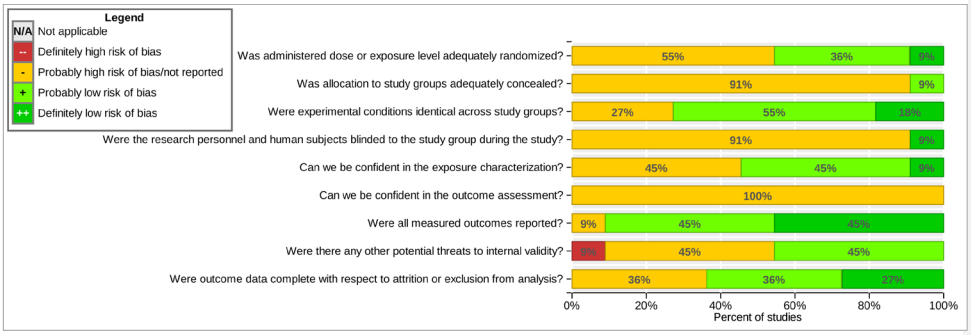


Figure D-36. Risk-of-bias Bar Chart for High Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure

An interactive version of Figure D-36 and additional study details in HAWC here (NTP 2019).

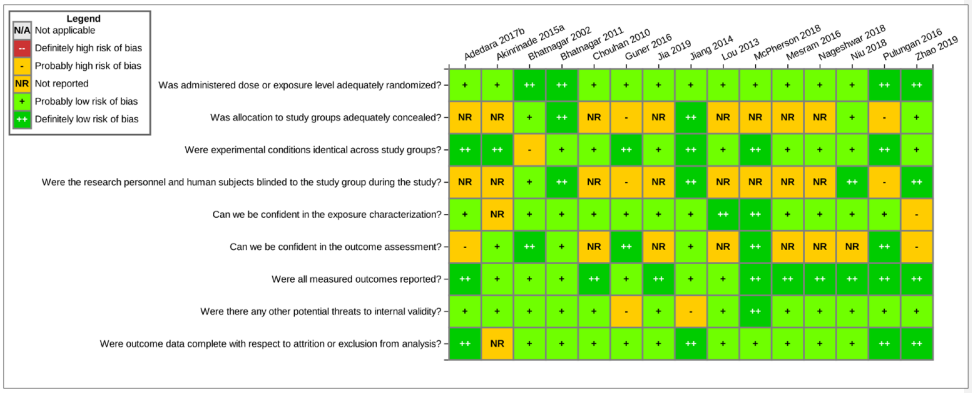


Figure D-37. Risk-of-bias Heatmap for Low Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure

An interactive version of Figure D-37 and additional study details in HAWC here (NTP 2019).

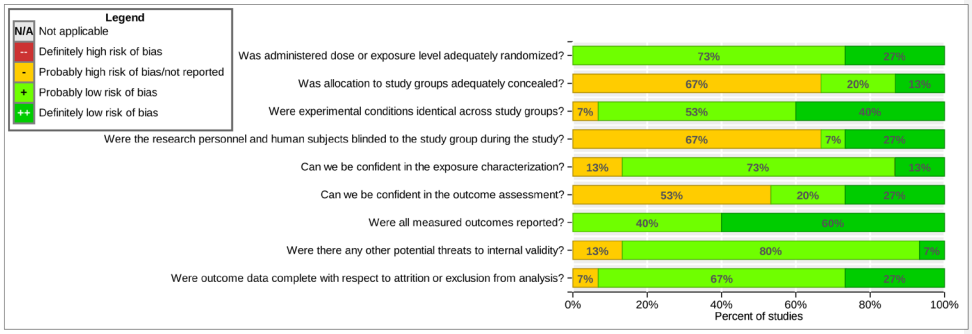


Figure D-38. Risk-of-bias Bar Chart for Low Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure

An interactive version of Figure D-38 and additional study details in HAWC here (NTP 2019).

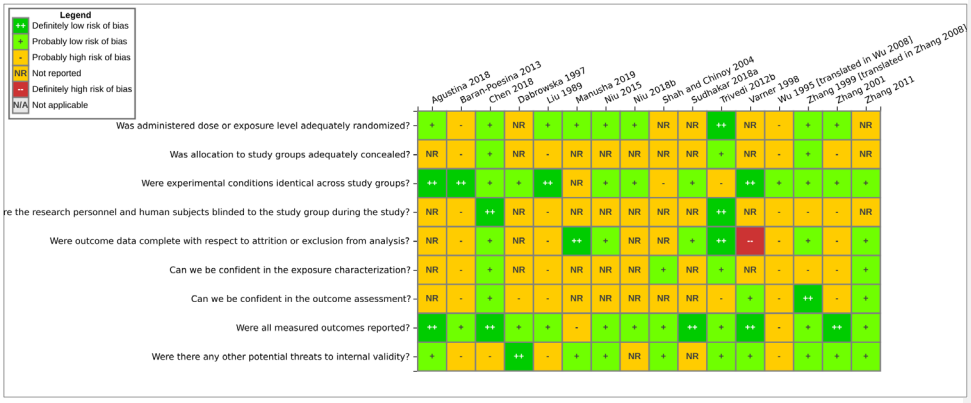


Figure D-39. Risk-of-bias Heatmap for High Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure

An interactive version of Figure D-39 and additional study details in HAWC here (NTP 2019).

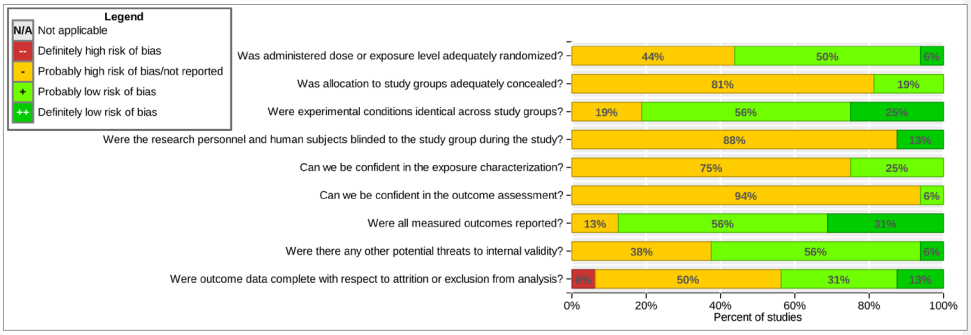


Figure D-40. Risk-of-bias Bar Chart for High Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure

An interactive version of Figure D-40 and additional study details in HAWC here (NTP 2019).

Appendix E. Details for Low Risk-of-bias Studies

Table of Contents

Commented [A109]: Text in *Other potential threats* throughout Appendix E reflects revisions made to consider study-specific failures to account for sampling strategy or clustering in determining potential for bias, in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 5): Clustering: NASEM identified that in some population studies, participants living in the same communities were assigned the same measure of fluoride exposure without considering the effect in the data analysis. These correlations may artificially increase the statistical power.

Recommendation: Limitations should note the studies where clustering was a potential threat and specifically whether the investigators addressed this.

E.1. IQ Studies

E.1.1. Bashash et al. (2017)

E.1.1.1. Study Details

- ***Study design:*** Prospective cohort
- ***Population:*** Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants (pregnant mothers and their children aged 4 or 6–12 years).
- ***Study area:*** Mexico City, Mexico
- ***Sample size:*** 299 mother-child pairs, of whom 211 had data for the IQ analyses.
- ***Data relevant to the review:*** Adjusted and unadjusted associations between IQ scores and maternal or child's urinary fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant association between maternal urinary fluoride and IQ score (adjusted β = −2.50 per 0.5 mg/L increase; 95% CI: −4.12, −0.59). No significant associations with children's urinary fluoride.

E.1.1.2. Risk of Bias

- Author contacts:
  - Authors were contacted for additional information on whether clustering was addressed. The authors provided results from additional models with cohort as a random effect, which informed the rating decision for the following risk-of-bias domains: Other.
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Study participants were selected from two different cohorts from three hospitals in Mexico City that serve low-to-moderate income populations. One cohort was from an observational study of prenatal lead exposure and neurodevelopment outcomes, and the other was from a randomized trial of the effect of calcium on maternal blood lead levels. The authors state that participants had no history of psychiatric disorders, high-risk pregnancies, gestational diabetes, illegal drug use, or continuous prescription drugs, but no information on smoking habits was considered. Study populations appear to be similar, but there may be some differences because subjects were selected from two different cohorts that were recruited from slightly different time periods.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar despite the subjects coming from different original study populations wherein different methods were used for recruitment.
- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Data were collected via questionnaire on maternal age, education, marital status at first prenatal visit, birth order, birth weight, gestational age at delivery, maternal smoking, maternal IQ, and HOME scores. All models were

Commented [A110]: Change made in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 5): The NTP response to NASEM's comments indicate they contacted investigators on specific issues, and that some study authors responded while others did not. Recommendation: NTP should provide an assessment of how non-response would affect the grading of these studies. This may be tied to the NASEM concern regarding transparency in the risk-of-bias assessment. [redacted] requests an appendix and language in the Discussion detailing which studies received a change to their risk-of-bias determination.

adjusted for gestational age at birth, sex, birth weight, birth order, age at testing, maternal marital status, smoking history, age at delivery, maternal IQ, education, and cohort, with additional testing for children's urinary fluoride, mercury, lead, and calcium. Sensitivity analyses additionally adjusted for HOME score. Important covariates not considered included BMI, iodine deficiency, arsenic, and maternal mental health and nutrition. Arsenic is assumed not to be a potential co-exposure in this population because the study authors did not discuss it as an issue, but did consider other co-exposures. Arsenic is included in the water quality control program in Mexico City and is not considered a concern in this population.

- *Potentially important study-specific covariates:* All key covariates were addressed.
  - *Direction/magnitude of effect size:* Not applicable.
- *Basis for rating:* Probably low risk of bias based on direct evidence that key covariates, including other potential co-exposures, were addressed and indirect evidence that the methods used to collect the information were valid and reliable and that arsenic is not likely to be an issue in this study population.

- Attrition:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although there was a large amount of attrition, the study authors clearly describe all reasons for attrition and also provide characteristics to compare those participants included to those excluded. There were some slight differences between those included and those excluded, but there is nothing to indicate that the attrition would potentially bias the results.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Urinary fluoride concentrations were determined in spot urine samples (2nd morning void) collected from mothers (during at least one trimester) and children ages 6–12 years. Fluoride content was measured using ion-selective electrode-based assays. QC methods were described including between laboratory correlations. All samples were measured in duplicate. Extreme outliers were excluded. Urinary dilution was addressed by using creatinine-adjusted levels.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- Outcome:
  - *Rating:* Definitely low risk of bias (++)

  - *Summary:* Outcome was assessed using the McCarthy Scales of Children's Abilities (MSCA) in 4-year-old children (translated into Spanish) and the Wechsler Abbreviated Scale of Intelligence (WASI) in 6–12-year-olds. The WASI is a well-established test, and the validity of both tests is well documented by the authors. Inter-examiner reliability was evaluated and reported with a correlation of 0.99 (++ for methods). The study report stated that psychologists were blind to the children's fluoride exposure (++ for blinding). Overall rating for methods and blinding = ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study. Statistical tests of bivariate associations (using Chi-square tests for categorical variables and analysis of variance [ANOVA]) were used to compare the means of the outcomes or exposure within groups based on the distribution of each covariate. Generalized additive models (GAMs) were used to estimate the adjusted association between fluoride exposure and measures of children's intelligence. Residual diagnostics were used to examine model assumptions and identify any potentially influential observations. Results are reported as adjusted effects and 95% CIs. In sensitivity analyses, regression models accounted for clustering at the cohort level by using cohort as a fixed effect in the models. Although using cohort as a random effect would be more appropriate, using individual-level exposure data and accounting for numerous important covariates in the models likely captured the cohort effect. Additional models with cohort as a random effect were also subsequently made available via personal communication with the study authors and showed similar results to the main model.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk if bias is based on direct evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include

individual exposure measurements, blinding of outcome assessor to participants' fluoride exposure, and the prospective cohort study design.

E.1.2. Choi et al. (2015)

E.1.2.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** First-grade children (ages 6–8 years)
- ***Study area:*** Mianning County in southern Sichuan, China
- ***Sample size:*** 51 first-grade children
- ***Data relevant to the review:*** Associations between IQ (digit span for auditory span and working memory and block design for visual organization and reasoning components of WISC-IV only) with continuous urine or drinking water fluoride levels. Study also had information based on dental fluorosis score.
- ***Reported association with fluoride exposure:*** Yes: Compared to the normal/questionable dental fluorosis, the moderate/severe dental fluorosis group was associated with significantly lower total (adjusted β = −4.28; 95% CI: −8.22, −0.33) and backward (adjusted β = −2.13; 95% CI: −4.24, −0.02) digit span scores. Linear associations between total digit span and log-transformed fluoride in urine (adjusted β = −1.67; 95% CI: −5.46, 2.12) and in drinking water (adjusted β = −1.39; 95% CI: −6.76, 3.98) were observed but not significant. Other outcomes not significantly associated with fluoride exposure.

E.1.2.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information because it was not necessary.
- Population selection:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were selected during the same time frame using the same methods. Fifty-one first-grade children residing in Mianning County in southern Sichuan, China were included in this pilot study. It is not specified whether the 51 children represented all the first-grade children from this area or whether some refused to participate. Children who did not speak Chinese, were not students at the Primary School of Sunshui Village in Mianning County, or those with chronic or acute disease that might affect neurobehavioral function tests were excluded. Demographic characteristics are presented in Table 1 of the study, which indicates that subjects were similar. Important covariates are adjusted for in the statistical analyses.
  - *Basis for Rating:* Definitely low risk of bias based on direct evidence that the exposure groups were similar and were recruited within the same time frame using the same methods with no evidence of differences in participation/response rates.

- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The parents or guardians completed a questionnaire on demographic and personal characteristics of the children (sex, age at testing, parity, illnesses before age 3, and past medical history) and caretakers (age, parity, education and occupational histories, residential history, and household income). A 20-µL capillary blood sample was collected at the school by a Mianning County Center for Disease Control (CDC) health practitioner and tested for possible iron deficiency, which could have been used as a covariate of neurodevelopmental performance. Important covariates that were not assessed include maternal BMI, parental mental health, maternal smoking status, maternal reproductive factors, parental IQ, and HOME score. However, the study authors noted that confounding bias appeared to be limited due to the minimal diversity in the social characteristics of the subjects. The study authors indicated that CDC records documented that levels of other contaminants, including arsenic and lead, were very low in the area. Iodine differences were not specifically addressed, but there is no indication from the information provided that this might have been a concern.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Probably low risk of bias because there is direct evidence that the key covariates were considered and indirect evidence that co-exposure to arsenic was likely not an issue in this area and that methods used for collecting the information were valid and reliable.
- Attrition:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The majority of results were reported for the 51 children stated to be included in the pilot study. In Table 5 of the study, the N for each dental fluorosis category totals only 43, but the text indicates 8 children did not have a Dean Index because permanent teeth had not erupted.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study used three different measurements of fluoride exposure: well water fluoride concentrations from the residence during pregnancy and onwards, fluoride concentrations from children's first morning urine samples, and degree of children's dental fluorosis. Fluoride concentrations in community well water were measured and recorded by Mianning County CDC; specific analytic methods were not reported, but it is likely that standard methods were used because the

analyses were conducted by the CDC and were likely the same as those used to measure the fluoride in urine. Migration of subjects was noted to be limited. Well water fluoride concentrations of the mother's residence during pregnancy and onward were used to characterize a child's lifetime exposure. To provide a measure of the accumulated body burden, each child was given a 330-mL (11.2-oz) bottle of Robust© distilled water (free from fluoride and other contaminants) to drink the night before the clinical examinations, after emptying the bladder and before bedtime. The first urine sample the following morning was collected at home, and the fluoride concentration was determined on a 5-mL sample using an ion-specific electrode at the Mianning CDC. There is no indication that urinary fluoride levels accounted for dilution, nor was it clear that the method of administering water to the children and collection methods sufficiently controlled for differences in dilution. One of the investigators, a dentist, performed a blinded dental examination on each child's permanent teeth to rate the degree of dental fluorosis using the Dean Index. The Dean Index is a commonly used index in epidemiological studies and remains the gold standard in the dentistry armamentarium. The Index has the following classifications: normal, questionable, very mild, mild, moderate, and severe. Quality control (QC) procedures are not reported but were likely appropriate.

        - *Direction/magnitude of effect size:* Current levels were used to assess lifetime exposure. This is likely to be a non-differential exposure misclassification, and direction of bias is unknown. Because subject migration appears to be limited, it is likely that the current fluoride levels are adequate reflections of past exposure. Dental fluorosis would be an indicator that exposure occurred in the past, and there was a fair correlation between degree of dental fluorosis and current urine and water fluoride levels, with both increasing with increasing levels of dental fluorosis.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.
- Outcome:
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* The study authors adopted culture-independent tests considered feasible for children aged 6 to 8 years. The Wide Range Assessment of Memory and Learning (WRAML) was used for the assessment of memory and learning. Three subtests were also used. The Finger Windows subtest assesses sequential visual memory. The Design Memory subtest assesses the ability to reproduce designs from memory following a brief exposure. The Visual Learning subtest assesses the ability to learn the locations of pictured objects over repeated exposures. The Wechsler Intelligence Scale for Children-Revised (WISC-IV) includes digit span for auditory span and working memory and block design for visual organization and reasoning. The grooved pegboard test assesses manual dexterity. The tests used have been validated on a Western population. Although there is no information provided to indicate that the tests were validated on the study population, the study authors indicated that the tests were culture-

independent (+ for methods). Blinding of the outcome assessors to participants' fluoride exposure, or steps to minimize potential bias were not reported. However, it is unlikely that the assessors had knowledge of the individual exposure as children all came from the same area, and water and urine levels were tested at the CDC. (+ for blinding). Overall = +.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that all outcomes were assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.

- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses are appropriate. Multiple regression models evaluate the associations between exposure indicators and test scores after adjusting for covariates. Specific regression models are not described or refenced, just stated to be "standard regression analysis with confounder adjustment." The distributions of fluoride concentrations in urine and water are skewed and log10-transformed to approximate a Gaussian distribution (test not specified). Results are reported as adjusted effects and 95% CIs. There is no evidence that residual diagnostics were used to examine model assumptions; however, the impact on the effect estimates is expected to be minimal.
    - *Other potential concerns:* It should be noted that this study was a pilot study and, therefore, had a relatively small sample size (i.e., 51 children).
  - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in the confounding, exposure, and outcome domains. Study strengths include individual fluoride measurements with blinding at outcome assessment likely. All key covariates and many other important covariates were considered in the study design or analysis.

E.1.3. Cui et al. (2018)

E.1.3.1. Study Details

- ***Study design:*** Cross-sectional

- ***Population:*** School children aged 7–12 years from four schools in two districts in China with different fluoride levels
- ***Study area:*** Jinghai and Dagang in Tianjin City, China
- ***Sample size:*** 323 school children
- ***Data relevant to the review:*** IQ scores by urine fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant association between IQ score and log-transformed urinary fluoride (adjusted β = −2.47; 95% CI: −4.93, −0.01).

E.1.3.2. Risk of Bias

- Author contacts:
  - Authors were contacted in June 2019 to obtain additional information for risk-of-bias evaluation. Additional information provided by the authors informed the rating decision for the following risk-of-bias domains: Detection (outcome assessment).
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Four schools were selected from the same district in China. The schools were selected based on levels of fluoride in the local drinking water and the degree of school cooperation. No details were provided on the number of schools in given areas or the difficulty in getting school cooperation. It was noted that the residents in the four areas had similar living habits, economic situations, and educational standards. Although authors do not provide the specific data to support this, fluoride levels and IQ scores were provided by different subject characteristics. The areas were classified as historically endemic fluorosis and non-fluorosis. Cluster sampling was used to select the grades in each school according to previously set child ages, and classroom was randomly selected with all students within a selected classroom included. Reasons for exclusion do not appear to be related to exposure or outcome.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and recruited within the same time frame using the same methods, with no evidence of differences in participation/response rates.
- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The measurements of all covariates were obtained by structured questionnaires that were completed by children with the help of their parents. Covariates that were assessed include: sex, age, child's ethnicity, child's BMI, birth (normal vs. abnormal), mother's age at delivery, mother's education, income per family member, mother's smoking/alcohol during pregnancy, family member smoking, environmental noise, iodine region (non-endemic vs. iodine-excess-endemic area), factory within 30 m of residence, iodine salt, diet supplements, seafood/pickled food/tea consumption, surface water consumption, physical activity, stress, anger, anxiety/depression, trauma, having a cold 5 times a year,

Commented [A111]: Change made in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 5): The NTP response to NASEM's comments indicate they contacted investigators on specific issues, and that some study authors responded while others did not. Recommendation: NTP should provide an assessment of how non-response would affect the grading of these studies. This may be tied to the NASEM concern regarding transparency in the risk-of-bias assessment. [redacted] requests an appendix and language in the Discussion detailing which studies received a change to their risk-of-bias determination.

thyroid disease in relatives, mental retardation in relatives, and cancer in relatives. Covariates not considered include parity, maternal and paternal IQ, and quantity and quality of caregiving environment (e.g., HOME score). The authors report that there were no other environmentally toxic substances that might have affected intelligence, such as high arsenic or iodine deficiency according to the Tianjin Centers for Disease Prevention and Control.

- *Potentially important study-specific covariates:* All key covariates were considered in this study.
  - *Direction/magnitude of effect size:* Not applicable.
- *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key covariates were considered, methods for collecting the information were valid and reliable, and co-exposure to arsenic was likely not an issue in this area.

- Attrition:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of the 400 children enrolled, 35 were excluded because they did not have informed consent signed by a guardian or they moved out of the area. Forty-two children were excluded because they did not have a DRD2 genotyping measurement. It is unclear whether these children were from the same schools or whether they were evenly distributed throughout the study area. It is also unclear whether the excluded subjects were similar to those included in the study. In the study, some analyses had fewer than the 323 subjects, but this seems reasonable given the subgroups that were being evaluated.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although children were selected based on area fluoride levels, fluoride in the urine was used in the analysis. Urine was collected from each child during the morning of enrollment and analyzed within a week. Fluoride levels were measured using an ion-selective electrode according to the China standard. A brief description of the method was provided, but no QC methods were reported. The study authors did not account for urinary dilution in the spot samples.
    - *Direction/magnitude of effect size:* Not accounting for dilution could cause some exposure misclassification. The direction and magnitude would depend on where the differences occurred.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.
- Outcome:
  - *Rating:* Probably low risk of bias (+)

- *Summary:* IQ was measured by professionals using the Combined Raven's Test–The Rural in China method, which is the appropriate test for the study population (++ for methods). Blinding or other methods to reduce bias were not reported. Although it was unlikely that the outcome assessor would have knowledge of the child's urine fluoride levels, there was potential that they would know whether the child was from an endemic or non-endemic area if the IQ tests were conducted at the child's school, and there was no information provided on how the IQ tests were administered. Correspondence with the study author noted the cross-sectional nature of the study with outcome and exposure assessed at the same time, making the outcome assessors blind to the exposure status of participants. However, there was still potential for knowledge of the area (+ for blinding).
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.

- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses were appropriate. Multiple linear regression models were applied to evaluate the relationship between urine fluoride levels and IQ scores, accounting for numerous important covariates. The urinary fluoride levels were log-transformed due to a skewed distribution. Residual diagnostics were used to examine model assumptions. Model robustness was tested through bootstrap, sensitivity analysis after excluding potential outliers, and cross-validation techniques. Results are reported as adjusted effects and 95% CIs. The analysis did not account for clustering at the school level or at the grade level (students were from four schools in grades selected via a clustered sampling method). There is no evidence that the sampling strategy was otherwise accounted for via sampling weights. The impact of these factors on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for several important covariates.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate, and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key covariates were considered in the study design or analysis.

E.1.4. Cui et al. (2020)

E.1.4.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** School children aged 7–12 years
- ***Study area:*** Tianjin City, China (one randomly selected school from each district based on iodine levels in the water), presumably was an expansion of the Cui et al. (2018) study
- ***Sample size:*** 498 school children
- ***Data relevant to the review:*** IQ scores by urine fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: A 2-point decrease in IQ was observed in the highest urinary fluoride group compared to the lowest urinary fluoride group (i.e., 110.00 in ≥2.5-mg/L group versus 112.16 in <1.6-mg/L group); however, the results did not achieve statistical significance based on a one-way ANOVA comparing the three different urinary fluoride categories only.

E.1.4.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for the 2020 publication. Authors were contacted in June 2019 for additional information on the Cui et al. (2018) publication. Additional information provided by the authors regarding Cui et al. (2018) informed the rating decision for the following risk-of-bias domains: Detection (outcome assessment). ~~Information obtained from that correspondence may have been used for additional information in the 2020 publication.~~
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were recruited from 2014 to 2018. One school was selected from each district where water concentrations of water iodine were <10, 10–100, 100–150, 150–300 and >300 µg/L. In each school, classes were randomly sampled for the appropriate age group of 7–12 years old. A table of subject characteristics was provided by IQ. A total of 620 children were recruited, and 122 children who did not have complete information or enough blood sample were excluded. Reasons for exclusion do not appear to be related to exposure or outcome. The characteristics of the 498 included children are presented.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited within the same time frame

Commented [A112]: Change made in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 5): The NTP response to NASEM's comments indicate they contacted investigators on specific issues, and that some study authors responded while others did not. Recommendation: NTP should provide an assessment of how non-response would affect the grading of these studies. This may be tied to the NASEM concern regarding transparency in the risk-of-bias assessment. [redacted] requests an appendix and language in the Discussion detailing which studies received a change to their risk-of-bias determination.

using the same methods, with no evidence of differences in participation/response rates.

- Confounding:
  - *Rating:* Probably high risk of bias (−)
  - *Summary:* It was noted by the study authors that there were no other environmental poisons except water fluoride. Other studies also conducted in this area of China noted specifically that arsenic was not a concern. Iodine was addressed as that was one of the main points of the study. Twenty-one factors (provided in Table 1 of the study) were selected as covariates, and a homemade questionnaire of unspecified validity was used for obtaining the information. It was noted that child age, stress, and anger were significantly associated with IQ although it is unclear whether these varied by fluoride level. However, Cui et al. (2018) indicate that stress and anger were not significantly associated with fluoride, and it was assumed that results would be similar for this study even though more children were included.
  - *Potentially important study-specific covariates:* Age (children 7–12 years old)
    - *Direction/magnitude of effect size:* Age is a key covariate for IQ, even in the narrow age range evaluated in this study. The direction of the association may depend on the number of children in each age group within the different urinary fluoride categories; however, these data were not provided. In general, there were fewer subjects ≤9 years of age (i.e., 111) compared to >9 years of age (i.e., 387) with a significantly higher IQ in the ≤9-year-old age group. Therefore, if exposure were higher in the older subjects, this could likely bias the association away from the null.
  - *Basis for rating:* Probably high risk of bias because there is indirect evidence that age was not addressed as a key covariate and it may be related to both IQ and exposure.
- Attrition:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of the 620 children recruited, 122 (20%) were excluded due to incomplete information or inadequate blood sample. No information was provided to indicate whether there were similarities or differences in the children included versus the children excluded, but exclusion is unlikely to be related to either outcome or exposure.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children's morning urine was collected with a clean polyethylene tube, and fluoride was measured using a fluoride ion-selective electrode following Chinese standard WS/T 89-2015. A brief description was provided, but no QC

methods were reported. The study authors do not account for urinary dilution in the spot samples.

- *Direction/magnitude of effect size:* Not accounting for dilution could cause some exposure misclassification. The direction and magnitude would depend on where the differences occurred.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.

- Outcome:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* IQ was measured using the Combined Raven's Test, which is an appropriate test for the study population (++ for methods). Blinding was not mentioned; however, the outcome assessors would not likely have had knowledge of the child's urinary fluoride. Subjects appear to have been recruited based on iodine levels; therefore, it is unlikely that there would have been any knowledge of potential fluoride exposure. Correspondence with the study authors for the Cui et al. (2018) study also indicated that the outcome assessors would have been blind.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* One-way ANOVA was used to make comparisons between mean IQ by urinary fluoride levels. Consideration of heterogeneity of variances was not reported. There is no adjustment for covariates or for clustering of children at the school level. There is no evidence that the sampling strategy was otherwise accounted for (i.e., via sampling weights). The impact of these factors on the effect estimates is expected to be minimal given the use of individual-level data. The primary focus of the study was to evaluate associations between IQ and thyroid hormone or dopamine levels (not between IQ and fluoride levels). It should also be noted that more advanced analyses used for thyroid hormone- and dopamine-IQ associations

still lacked adjustment for school and accounting for clustering of children from the same school.

- *Other potential concerns:* None identified.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate, and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements, but the study is limited by the cross-sectional study design, lack of accounting for urine dilution, and lack of addressing age as a key covariate.

E.1.5. Ding et al. (2011)

E.1.5.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Elementary school children aged 7–14 years old
- ***Study area:*** Hulunbuir City, Inner Mongolia, China
- ***Sample size:*** 331 school children
- ***Data relevant to the review:*** IQ mean difference based on 10 categories of urine fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant association between urinary fluoride and IQ score (each 1-mg/L increase in urinary fluoride was associated with a decrease in IQ score of 0.59 points; 95% CI: −1.09, −0.08).

E.1.5.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information because it was not necessary.
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study randomly selected 340 7–14-year-olds from four nearby elementary schools in Hulunbuir. Authors stated that the four elementary schools appeared to be very similar in teaching quality. The study authors noted that they followed the principles of matching social and natural factors like economic situation, educational standards, and geological environments as much as possible; however, how this was done is unclear and no table of study subject characteristics by group was provided.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited within the same time frame using the same methods, with no evidence of differences in participation/response rates.
- Confounding:

  - *Rating:* Probably high risk of bias (−)
  - *Summary:* It was noted that none of the four sites had other potential neurotoxins, including arsenic, in their drinking water. Details were not provided, except for a reference supporting the statement. In addition, iodine deficiency was noted as not being issue in any of the four areas. Age was the only key covariate adjusted for in the regression model. Although dental fluorosis severity by % female was reported, not enough data were provided to determine whether sex should have been considered in the regression model. The study authors note that future studies will include covariates such as parents' educational attainment, mother's age at delivery, and household income.
  - *Potentially important study-specific covariates:* Sex
    - *Direction/magnitude of effect size:* There is not enough information to determine whether there was an effect from sex. There were some differences in dental fluorosis level by sex, but it is unclear how this might impact the results or whether the distribution of sex differed by age.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that there were differences in sex that were not considered in the study design or analyses.
- Attrition:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data were relatively complete (i.e., <5% loss). Of the 340 subjects selected for inclusion, 5 were excluded because they lived in the area for less than a year with an additional 4 not consenting to participate.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analysis was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Spot urine samples were collected and measured using China CDC standards. All samples were analyzed twice using a fluoride ion-selective electrode. Recovery rates were specified as 95%–105% with an LOD of 0.05 mg/L. Water samples were collected from small-scale central water supply systems and tube wells with handy pumps and were processed using standard methods, similar to the urine samples. Quality assurance validation was reported. A blind professional examiner evaluated the children for dental fluorosis using Dean's Index. All urine and water samples were above the LOD. Urine levels were the primary exposure measure used in the analysis. The study authors did not account for urinary dilution in the spot samples. The mean urine fluoride concentration was correlated with the dental fluorosis levels.
    - *Direction/magnitude of effect size:* Spot urine samples that did not account for dilution could have exposure misclassification. The misclassification is likely non-differential, and the potential direction of bias is unknown.

  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.
- Outcome:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* IQ was determined using the Combined Raven's Test–The Rural in China (CRT-RC3) (++ for methods). Although blinding was not reported, it is unlikely that the IQ assessors had knowledge of the children's urine levels or even of the water levels from the four sites, as these were sent to a separate lab for testing (+ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses were reasonable (ANOVA and multiple linear regression), but consideration of homogeneity of variance was not reported. The NASEM committee's review (NASEM 2021) pointed out a potential concern regarding the lack of accounting for clustering at the school level because children were selected from four elementary schools. However, as outlined in the ***Selection*** domain, the authors stated that they followed the principles of matching social and natural factors like economic situation, educational standards, and geological environments to the extent possible and that the four elementary schools appeared to be very similar in teaching quality. There is no evidence that the sampling strategy was otherwise accounted for (i.e., via sampling weights). The impact of these factors on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for age as a key covariate.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and that there were no other potential threats of risk of bias.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure

measurements, but the study is limited by the cross-sectional study design, lack of accounting for urine dilution, and lack of consideration of sex as a key covariate.

E.1.6. Green et al. (2019)

E.1.6.1. Study Details

- ***Study design:*** Prospective cohort
- ***Population:*** Maternal-Infant Research on Environmental Chemicals (MIREC) participants (pregnant mothers and their children aged 3–4 years)
- ***Study area:*** 10 cities, Canada
- ***Sample size:*** 512 mother-child pairs (238 from non-fluoridated areas, 162 from fluoridated areas; 264 females, 248 males)
- ***Data relevant to the review:*** Adjusted linear regression models evaluating associations between IQ in both sexes together and separately, with maternal urinary fluoride across all three trimesters or with estimated maternal fluoride intake.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower full-scale IQ per 1-mg/L increase in maternal urinary fluoride in boys (adjusted β = −4.49) but not girls (adjusted β = 2.40) and not in both sexes combined (adjusted β = −1.95); significantly lower full-scale IQ per 1-mg increase in maternal intake in both sexes combined (adjusted β = −3.66 [no sex interaction]); significantly lower full-scale IQ per 1-mg/L increase in drinking water fluoride in both sexes combined (adjusted β = −5.29 [no sex interaction]).

E.1.6.2. Risk of Bias

- Author contacts:
  - Authors were contacted in June 2019 for additional information for the risk-of-bias evaluation. Additional information provided by the authors informed the rating decision for the following risk-of-bias domains: Other.
- Population selection:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Pregnant women were recruited from the same population during the same time frame and using the same methods as the MIREC program. Methods were reported in detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were similar and were recruited with the same methods during the same time frame.
- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study considered several possible covariates, including maternal age, pre-pregnancy BMI, marriage status, birth country, race, maternal education, employment, income, HOME score, smoking during pregnancy, secondhand smoke in the home, alcohol consumption during pregnancy, parity, sex, age at

Commented [A113]: Change made in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 5): The NTP response to NASEM's comments indicate they contacted investigators on specific issues, and that some study authors responded while others did not. Recommendation: NTP should provide an assessment of how non-response would affect the grading of these studies. This may be tied to the NASEM concern regarding transparency in the risk-of-bias assessment. [redacted] requests an appendix and language in the Discussion detailing which studies received a change to their risk-of-bias determination.

testing, gestational age, birth weight, time of void, and time since last void. The study also conducted secondary analyses to test for lead, mercury, arsenic, and PFOA. There is no indication of any other potential co-exposures in this study population. Iodine deficiency or excess could not be assessed but is not expected to differentially occur. The study was not able to assess parental IQ or mental health disorders. Methods used to obtain the information included questionnaires and laboratory tests.

- *Potentially important study-specific covariates:* All key covariates were addressed.
  - *Direction/magnitude of effect size:* Not applicable.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and direct evidence that key covariates, including potential co-exposures, were addressed.

- Attrition:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of the 610 recruited children, 601 (98.5%) completed testing. Of the 601 mother-child pairs, 512 (85.2%) had all three maternal urine samples and complete covariate data, and 400 (66.6%) had data available to estimate fluoride intake.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Spot urine samples from all three trimesters of pregnancy were evaluated using appropriate methods, and results were adjusted for creatinine and specific gravity. Fluoride intake was estimated based on fluoride water levels, and information on consumption of tap water and other water-based beverages (e.g., tea, coffee) was obtained via questionnaire.
    - *Direction/magnitude of effect size:* There is not any specific direction or magnitude of bias expected. Urinary fluoride levels are reflective of a recent exposure. Having measurements from all three trimesters of pregnancy provides a better representation of actual exposure than a single measurement, although the potential for missed high exposure is possible. However, the possibility of the occurrence of missed high exposure would be similar in all females and would be non-differential. For the fluoride intake, exposure was based on the fluoride levels in the water at the residence. If women worked outside the home and the majority of intake occurred from areas outside the home (and were different from levels in the home), there is potential to bias toward the null.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- Outcome:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The Wechsler Preschool and Primary Scale of Intelligence was normalized for ages 2.5–<4.0 and sex using the U.S population-based norms. Blinding was not reported, but it is unlikely that the outcome assessors had knowledge of the maternal fluoride level or were aware of whether the city had fluoridated water.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes were reported.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Multivariate linear regression analyses were used to evaluate the associations between maternal urinary fluoride and fluoride intake and children's IQ scores. Regression diagnostics were used to test assumptions for linearity, normality, and homogeneity. There were no potential influential observations (based on Cook's distance). Sensitivity analyses showed that the effects of maternal urinary fluoride (MUF), fluoride intake, and water fluoride were robust to the exclusion of two very low IQ scores in males (<70). City was accounted for as a covariate in the regression models published. Additional models with city as a random effect were also subsequently made publicly available and showed similar results to the main model.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, prospective cohort design, and the consideration of key covariates.

E.1.7. Rocha-Amador et al. (2007)

E.1.7.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 6–10 years
- ***Study area:*** Moctezuma (low fluoride, low arsenic) and Salitral (high fluoride, high arsenic) of San Luis Potosí State and 5 de Febrero (high fluoride, high arsenic) of Durango State, Mexico
- ***Sample size:*** 132 children
- ***Data relevant to the review:*** Associations between full-scale IQ, performance IQ, verbal IQ, and child's urine or water fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant associations between log-transformed fluoride and IQ scores (full-scale IQ adjusted βs of −10.2 [water] and −16.9 [urine]; CIs not reported); arsenic also present, but the effect from arsenic was smaller (full-scale IQ adjusted βs of −6.15 [water] and −5.72 [urine]; CIs not reported).

E.1.7.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information to inform the risk-of-bias evaluation because it was not necessary.
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* All children in 1st through 3rd grades in three rural areas in Mexico (n = 480) were screened for study eligibility, including age, time at residence, and address. Authors report that the three selected communities were similar in population and general demographic characteristics. Children who had lived in the area since birth and were 6–10 years old were eligible to participate (n = 308). Of the 308 children, 155 were randomly selected and the response rate was 85%, but participation was not reported by area. It was noted, however, that no significant differences in age, sex, or time of residence were observed between participants and non-participants. Time frame for selection was not mentioned but appears to be similar. Sociodemographic characteristics of subjects was provided in Table 1 of the study. There was a significant difference in SES and transferrin saturation, but these were considered in the analysis.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the populations were similar, and differences were noted and addressed in the analysis.
- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study design or analysis accounted for age, sex, SES, transferrin saturation, weight, height, blood lead levels, and mother's education. Arsenic

Commented [A114]: Change made in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response.

Reviewer comment (DocA1_Monograph, page 3): The NTP re5ponse to NASEM's comments indicate they contacted investigators on specific issues, and that some study authors responded while others did not. Recommendation: NTP should provide an assessment of how non-response would affect the grading of these studies. This may be tied to the NASEM concern regarding transparency in the risk-of-bias assessment. [redacted] requests an appendix and language in the Discussion detailing which studies received a change to their risk-of-bias determination.

levels were highly correlated with fluoride levels; however, arsenic and fluoride were evaluated alone, and arsenic was found to have less of an effect on IQ than fluoride. This provides evidence that arsenic had been addressed as a co-exposure and cannot explain the association between fluoride exposure and decreased IQ. Smoking was not addressed and methods for measuring many of the covariates were not reported.

- *Potentially important study-specific covariates:* Arsenic
  - *Direction/magnitude of effect size:* The presence of arsenic in this study, which also demonstrated an association, would likely bias the association away from the null. Although arsenic may contribute to some of the magnitude of the observed effect of fluoride (the exact impact of arsenic on the magnitude cannot be assessed), the presence of arsenic does not fully explain the observed association between fluoride exposure and IQ. The presence of arsenic may affect the magnitude of the association between fluoride and IQ, but it has no impact on the direction of the association.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and direct evidence that key covariates were addressed.

- Attrition:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of 155 children randomly selected for study participation, 85% responded to enroll. According to the authors, there were no significant differences in age, sex, or time of residence between responders and non-responders. However, no data were provided to support this, and no breakdown of responders/non-responders by region was provided. Data were provided for the 132 children agreeing to participate.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Urine samples were collected on the same day as psychological evaluations and were analyzed for fluoride according to NIOSH Method 8308 (Fluoride in Urine). For QC, a reference standard was also used (NIST SRM 2671a). Urine samples were also analyzed for arsenic by using the Atomic Absorption Spectrophotometer with hydride system and a reference standard for QC. Levels were adjusted for urinary creatinine levels to account for dilution in the spot samples. Tap water samples were collected from each child's home on the day of biological monitoring. Fluoride was measured with a sensitive, specific ion electrode. Detailed methods are provided including internal quality controls. It was noted that in the high fluoride group, it was common to drink bottled water low in fluoride and to use the tap water only for cooking; therefore, urine was

considered the most appropriate measure of exposure. Only children who had lived at the same residence since birth were included.

- *Direction/magnitude of effect size:* Not applicable.

- *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- Outcome:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Neuropsychological profiles were assessed through the WISC-RM (revised for Mexico). This is a well-established test appropriately adjusted for the study population. However, no additional validation was provided (+ for methods). The study report stated that the test assessors were masked to both arsenic and fluoride water levels (++ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Probably high risk of bias (−)
  - *Summary:* It was reported that an interaction between fluoride and arsenic was measured, but it was noted only in the discussion that the study design precluded testing statistical interaction between fluoride and arsenic. This provides indirect evidence of selective reporting.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that there was selective reporting.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study. Multivariate linear analyses were used to evaluate the associations between fluoride in water and urine and children's IQ scores. Exposures were natural log-transformed, but the rationale was not provided. Regression diagnostics were not used to test model assumptions for linearity, normality, and homogeneity. The analyses did not account for clustering at the community level. The three selected communities were similar in population and general demographic characteristics. Although the analysis used individual-level exposures rather than area-level exposures, if the exposure levels within a certain area were highly correlated (which might be expected), then the results might still be biased. However, the overall impact on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for multiple important covariates.

- *Other potential concerns:* None identified.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and blinding of outcome assessors to participants' fluoride exposure, but it is limited by the cross-sectional study design and the inability to completely rule out the influence of arsenic in the results.

E.1.8. Saxena et al. (2012)

E.1.8.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 12 years
- ***Study area:*** Madhya Pradesh, India
- ***Sample size:*** 170 children
- ***Data relevant to the review:*** Mean IQ grade (not standard scores; higher IQ grades are associated with lower intelligence) by water fluoride quartiles, continuous water fluoride, or continuous urinary fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant correlations between IQ score and water (r = 0.534) and urinary (r = 0.542) fluoride levels. Significant increase in mean IQ grade (i.e., increase in proportion of children with intellectual impairment) with increasing urinary fluoride in adjusted analyses.

E.1.8.2. Risk of Bias

- Author contacts:
  - Authors were contacted in August ~~of~~ 2017 to obtain additional information for risk-of-bias evaluation. Additional information provided by the authors informed the rating decision for the following risk-of-bias domains: Detection (outcome assessment).
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There was indirect evidence that subjects were similar and were recruited using the same methods during the same time frame. The study participants were selected from a stratified cluster of geographic areas based on fluoride concentration in groundwater. According to the authors, the selected villages were similar in population and demographic characteristics. Data are provided to show the breakdown in SES, parental education, height/age, and weight/height, and no significant differences were noted. Participation was stated to be voluntary, but participation rates were not provided. It is unclear whether the 170 subjects were selected with 100% participation or whether the 170 subjects were all who were asked to participate, but it appears that all subjects participated.

Commented [A115]: Change made in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 5): The NTP response to NASEM's comments indicate they contacted investigators on specific issues, and that some study authors responded while others did not. Recommendation: NTP should provide an assessment of how non-response would affect the grading of these studies. This may be tied to the NASEM concern regarding transparency in the risk-of-bias assessment. [redacted] requests an appendix and language in the Discussion detailing which studies received a change to their risk-of-bias determination.

Timing of the recruitment was not provided but is assumed to occur during the same time frame.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and recruited using the same methods during the same time frame.

- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There was indirect evidence that key covariates, including potential co-exposures, were addressed using reasonable methods. A questionnaire, completed with the assistance of parents, was used to collect information on child characteristics (age, sex, height, weight), residential history, medical history (including illness affecting the nervous system and head trauma), educational level of the head of the family (in years), and SES of the family. The SES was recorded according to the Pareek and Trivedi classification. The nutritional status of the children was calculated using Waterlow's classification, which defines two groups for malnutrition using height-for-age ratio (chronic condition) and weight for height ratio (acute condition). Within both groups, it categorizes the malnutrition as normal, mildly impaired, moderately impaired, or severely impaired. Urinary lead and arsenic were analyzed using the atomic absorption spectrophotometer. Urinary iodine was measured using the Dunn method. Authors do not report which covariates were included in the multivariate regression models; however, there was no difference in reported demographic characteristics. All subjects were the same age, and there was no difference in iodine, lead, or arsenic between the groups. Mean urinary arsenic levels increased with increasing fluoride even though there was no significant difference by group.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and that key covariates, including potential co-exposures, were addressed.
- Attrition:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results were provided for all 170 children stated to be included in the study.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence of no attrition.
- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* A sample of 200 mL of drinking water was collected at each child's home. The fluoride levels were analyzed by a fluoride ion-selective electrode. Each subject was also asked to collect a sample of his/her first morning urine. The

fluoride content in the urine was determined using a fluoride ion-selective electrode. QA/QC and LOD were not reported, and urinary dilution was not assessed. Although only current levels were measured, children who had changed their water source since birth were excluded.

- *Direction/magnitude of effect size:* Spot urine samples that did not account for dilution could have exposure misclassification. The misclassification is likely non-differential and not likely to bias in any specific direction. Children who had changed water source since birth were excluded, but it was not specifically noted that the fluoride in the water source was stable over the years.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- Outcome:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Intelligence was assessed using Raven's Standard Progressive Matrices and categorized into five grade levels. Although it was not noted that the test was validated to the study population, the test is visual and would be applicable to most populations (+ for methods). There is no mention of blinding by test administrators or evaluators, and the exposure groups come from different geographic areas. It was also not reported who measured the levels of fluoride from the home or urine samples. Correspondence with the study authors indicated that the outcome assessors were blind to the children's fluoride status (++ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* One-way analysis of variance (ANOVA), simple linear regression, and multiple linear regression were used to compare mean intelligence grades by water fluoride levels and to assess the association between grades and urinary fluoride. Consideration of heterogeneity of variance (for ANOVA) was not reported. Regression diagnostics were not

used to test model assumptions for linearity, normality, and homogeneity. Given the ordinal nature of the intelligence grade variable (score from 1 to 5), ordinal logistic regression would have been a more appropriate method. There was no adjustment for area-level clustering in multivariate analyses (although subjects were selected via stratified cluster sampling from two areas). Although the analysis used individual-level exposures rather than area-level exposures, if the exposure levels within a certain area were highly correlated (which might be expected), then the results might still be biased. However, the overall impact on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for important covariates.

        - *Other potential concerns:* None identified.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate, and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and the consideration of key covariates, but it was limited by the cross-sectional study design and lack of addressing dilution in the urine samples.

E.1.9. Seraj et al. (2012)

E.1.9.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 6–11 years
- ***Study area:*** five villages, Makoo, Iran
- ***Sample size:*** 293 children
- ***Data relevant to the review:*** IQ (mean and distribution) assessed by Raven's Colored Progressive Matrices and presented by fluoride area; beta was also provided for water fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant association between water fluoride and IQ score (adjusted β per 1-mg/L increase in water fluoride = −3.865; CIs not reported); significantly higher IQ score in normal area (97.77 ± 18.91) compared with medium (89.03 ± 12.99) and high (88.58 ± 16.01) areas.

E.1.9.2. Risk of Bias

- Author contacts:
    - Authors were not contacted for additional information because it was not necessary.
- Population selection:
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Subjects were selected from five villages in Makoo. The villages were stated to all be rural with similar general demographic and geographic

characteristics and were comparable in terms of SES and parental occupations. Children were 6–11 years old. Age, sex, and education were taken into account in the analysis. No other characteristics were provided or discussed. Participation rates were not reported. There is indirect evidence that the populations were similar, and some possible differences were addressed.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and recruited using the same methods during the same time frame.

- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Age, sex, dental fluorosis intensity, and educational levels (child's and parents') were evaluated as important covariates. Other covariates such as smoking were not discussed. Information was obtained from a detailed questionnaire. Lead was measured but found only in low levels in the drinking water throughout the study regions. Iodine in the water was also stated to be measured, and residents were receiving iodine-enriched salt. Arsenic was not addressed, but there is no evidence that arsenic levels would vary across villages in this area. Based on water quality maps, co-exposure to arsenic is likely not a major concern in this area.
  - *Potentially important study-specific covariates:* Arsenic.
    - *Direction/magnitude of effect size:* Conceptually, if there were differential amounts of arsenic in the different villages, co-exposure to arsenic could bias the association, with the direction of the bias dependent on where the arsenic was present; however, arsenic was not expected to be a major concern in this study area based on water quality maps.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and that key covariates, including potential co-exposures, were addressed or were not likely to be an issue in the study area.
- Attrition:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Attrition was low if it occurred. It was noted that 293 out of 314 children living in the villages were recruited. It is not clear whether 21 children were excluded based on exclusion criteria or whether they refused to participate; however, this accounts for less than 10% of the population, and results were available for all 293 subjects.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was minimal, adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
  - *Rating:* Probably high risk of bias (−)

- *Summary:* Exposure was primarily based on area of residence. Fluoride in the groundwater was analyzed by the SPADNS (Sulfophenylazo dihydroxynaphthalene-disulfonate) method, utilizing the 4000 UV-Vis spectrophotometer in the environmental health engineering laboratory of the Public Health School of the Tehran University of Medical Sciences. Specific details were not provided on methods of collection or sample locations or whether these locations represented the primary sources of drinking water for the subjects. Villages were categorized into normal (0.5–1 ppm), moderate (3.1 ± 0.9 ppm), and high (5.2 ± 1.1 ppm) fluoride based on the mean fluoride content of all seasons presumably for the stated 12-year time period. Subjects were stated to be long-life residents of the village. Dental fluorosis was also measured and increased in severity with fluoride levels; however, all areas had some degree of dental fluorosis. Although authors used an average fluoride level in varying seasons over presumably 12 years, they used a less-established method without reporting reliability or validity, and they did not provide data to indicate that the mean was truly representative of the fluoride levels over time and throughout the village. Although dental fluorosis severity increased with increasing fluoride levels, the data could also indicate potential exposure misclassification.
  - *Direction/magnitude of effect size:* The presence of dental fluorosis in all groups indicates that there may have been different exposures in some children at a younger age. Although there were only about 20 children in the "normal" fluoride group with very mild to mild dental fluorosis, this could bias the results toward the null because those children may have experienced a higher level of fluoride at some point. The other two fluoride groups were exposed to fluoride levels that likely exceeded those in the "normal" fluoride group.
- *Basis for rating:* Probably high risk of bias based on indirect evidence that exposure was assessed using insensitive methods.

- Outcome:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Intelligence was evaluated using Raven's Color Progressive Matrices. This is a well-established method. Although the study authors did not provide data to indicate that the methods were valid in this study population, the test is designed to be culturally diverse (+ for methods). The study report stated that test administrators were blinded to subjects' exposure status (++ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that outcomes were assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported. However, because the study author did not report the method for

obtaining the betas in Table 4 of the study, it is not clear whether these were adjusted or unadjusted regression coefficients.

- *Basis for rating:* Probably low risk of bias based on direct evidence that all the study's measured outcomes were reported, but the results were not sufficiently reported.

- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical methods for comparisons of IQ level by exposure groups were reasonable (ANOVA, post hoc test, and Kruskal-Wallis test), but consideration of heterogeneity of variance was not reported. Clustering at the village levels was not accounted for in multivariate analyses, which used area-level water fluoride levels. Because the exposure levels within a certain area are highly correlated (which might be expected), the results are likely to be biased. There was adjustment for some individual-level important covariates, and the children were from five rural areas with similar general demographic and geographic characteristics and were comparable in terms of SES and parental occupations. These factors are expected to mitigate some of the impact of lack of accounting for clustering, and the overall impact on the effect estimates is expected to be minimal.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and outcome. Study strengths include addressing potential key covariates, but it was limited by the cross-sectional study design and the group-level exposure data.

E.1.10. Soto-Barreras et al. (2019)

E.1.10.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 9–10 years
- ***Study area:*** Chihuahua, Mexico
- ***Sample size:*** 161 children
- ***Data relevant to the review:*** Water fluoride, urinary fluoride, exposure dose, and dental fluorosis index by IQ grade.
- ***Reported association with fluoride exposure:*** No: Results were not presented to evaluate an association between fluoride exposure and IQ but to compare fluoride levels within IQ grades. For this reason, the results of this study are not comparable to

other studies that evaluated IQ scores by fluoride exposure levels. No significant differences in measured fluoride levels across IQ grades were observed.

E.1.10.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information to inform the risk-of-bias evaluation because it was not necessary.
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were selected using a multistage cluster sampling. During the first stage, 13 public elementary schools were randomly selected from a pool of 73 using a cluster sample design. Secondly, only fourth-grade students were included. Authors stated that they wanted to keep the same grade level, but there were no specific details as to why fourth graders were selected as opposed to any other grade. Lastly, only children whose parents or guardians attended and responded to the survey were included. There is no information provided on how the 13 schools selected may have been similar to or different from the 60 schools not selected. There is no information provided on the number of children in the fourth grade to know participant rates. It was only noted that 245 children were examined, but 161 were included after the exclusion rules were applied. Inclusion and exclusion criteria are presented. Reasons for exclusion do not appear to be related to exposure or outcome. Characteristics of participants and non-participants are not compared; however, characteristics of the 161 included children were provided, and any differences were taken into account in the analysis.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposed groups were similar and were recruited using similar methods during the same time frame.
- Confounding:
  - *Rating:* Probably high risk of bias (−)
  - *Summary:* No covariates were considered when evaluating associations between fluoride exposure and intelligence; covariates were considered only when evaluating associations between fluoride levels and dental caries. According to Table 4 of the study, there was no significant association between IQ grade and age, sex, parental education, or SES status. No other information was reported or considered. There is no information on potential co-exposures. According to water quality maps, the arsenic prediction indicates a greater than 50% probability of exceeding the WHO guidelines for arsenic of 10 µg/L in areas of Chihuahua, Mexico.
  - *Potentially important study-specific covariates:* Arsenic.
    - *Direction/magnitude of effect size:* The impact on the direction and magnitude of effect size is unknown. There is potential for arsenic to occur in the study area, but it is not known how it relates to fluoride exposure. If they occur

Commented [A116]: Change made in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 5): The NTP response to NASEM's comments indicate they contacted investigators on specific issues, and that some study authors responded while others did not. Recommendation: NTP should provide an assessment of how non-response would affect the grading of these studies. This may be tied to the NASEM concern regarding transparency in the risk-of-bias assessment. [redacted] requests an appendix and language in the Discussion detailing which studies received a change to their risk-of-bias determination.

together in the water, it would likely bias the association away from the null; however, if they occur in different areas, there is potential to bias the association toward the null.

- *Basis for rating:* Probably high risk of bias based on indirect evidence that there is potential for exposure to arsenic that was not sufficiently addressed.

- Attrition:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* A total of 161 of 245 children were included in the study. Exclusion criteria are presented and are unrelated to outcome or exposure. For the 161 children, there are no missing outcome data.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
  - *Rating:* Probably low risk of bias (+); Probably high risk of bias (−)
  - *Summary:* **Urinary Fluoride (probably low risk of bias):** First morning void urine samples were collected based on NIOSH methods. Water samples were also stated to be collected, but it does not appear that methods followed any particular standard, and there is no indication that subjects were provided with collection containers. Analysis was based on a calibration curve using fluoride ion-selective electrode. QC methods were mentioned. Based on results, there were values below detection limits, but LODs or % below LOD were not reported.

    Daily fluoride exposure (probably high risk of bias): Daily fluoride exposure was based on the water fluoride level, drinking water consumption (based on parental report of how many glasses of water consumed), and body weight.

    - *Direction/magnitude of effect size:* Spot urine samples that did not account for dilution could have exposure misclassification. The misclassification is likely non-differential and is not likely to bias in any specific direction. Daily exposure was based partially on parental report of water consumption. The direction and magnitude of effect is unknown.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure. The daily fluoride exposure is probably high risk of bias because there is indirect evidence that the exposure was assessed using methods of unknown validity.
- Outcome:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Intellectual ability was evaluated using Raven's Colored Progressive Matrices by an independent examiner. Some details were provided, but it was not stated that the tests were assessed blind; however, there is no indication that subjects were from high fluoride areas, and the assessor would not have

knowledge of the urine or water fluoride levels. Results for children were converted into a percentile according to age (details not provided), and overall scores were assigned an intellectual grade of I to V as described in the report.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.

- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* The Kolmogorov-Smirnov test was used to determine variable distribution. The Kruskal-Wallis test was used to compare exposure levels between IQ grades with Dunn's post hoc test. Multivariate logistic regression was used to estimate the association between presence of dental caries and various risk factors. Fluoride levels in drinking water and urine and fluoride exposure dose were compared across intellectual grades. Children were from 13 schools selected via stratified cluster sample design. There was no adjustment for clustering at the school level or for the sampling design. Although the analysis used individual-level exposures rather than area-level exposures, if the exposure levels within a certain school were highly correlated (which might be expected), then the results might still be biased. The large number of clusters (13 schools) makes clustering less of a concern, and the impact on the effect estimates is expected to be minimal.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements and blinding of outcome assessor to participants' fluoride exposure, but it is limited by the cross-sectional study design, lack of accounting for urine dilution, and lack of consideration for potential exposures to arsenic in the study area. Although the study is considered to have low potential for bias overall, the focus of the study was to evaluate the relationship between fluoride exposure and lower rates of dental caries. In terms of evaluating an association between fluoride exposure and IQ scores, the study is limited by the way the data were reported.

E.1.11. Sudhir et al. (2009)

E.1.11.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 13–15 years
- ***Study area:*** Nalgonda district (Andhra Pradesh), India
- ***Sample size:*** 1,000 children
- ***Data relevant to the review:*** Mean IQ grade (not standard scores) or IQ distribution by water fluoride strata (<0.7, 0.7–1.2, 1.3–4.0, and >4.0 ppm).
- ***Reported association with fluoride exposure:*** Yes: Significant increase in mean and distributions of IQ grades (i.e., increase in proportion of children with intellectual impairment) with increasing drinking water fluoride levels.

E.1.11.2. Risk of Bias

- Author contacts:
  - Authors were contacted in September of 2017 for additional information related to risk-of-bias evaluation, but no response was received.
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children were selected from the same general population during the same time frame and were then broken down into nearly equal exposure groups. A cross-sectional study was conducted among 13–15-year-old school children of Nalgonda district, Andhra Pradesh, between August and October 2006. Data were collected from the school children who were lifelong residents of Nalgonda district, Andhra Pradesh, and who consumed drinking water from the same source during the first 10 years of life. A stratified random sampling technique was used. The entire geographical area of Nalgonda district was divided into four strata based on different levels of naturally occurring fluoride in the drinking water supply. Children were randomly selected from schools in the different strata. It was noted that the 1,000 selected children were equally divided among all four strata; however, each group did not have 250 children (rather, each had 243–267). Participation rates were not reported. Exclusion criteria included children who had a history of brain disease and head injuries, children whose intelligence had been affected by congenital or acquired disease, children who had migrated or were not permanent residents, children with orthodontic brackets, and children with severe extrinsic stains on their teeth. Age and sex data are presented in Table 1 of the study, but this information is not presented by the different fluoride groups.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and were recruited using the same methods during the same time frame.
- Confounding:
  - *Rating:* Probably low risk of bias (+)

- *Summary:* Data were collected using a self-administered questionnaire and clinical examination. The questionnaire requested information on demographic data (appears to cover age and sex), permanent residential address, staple food consumed, liquids routinely consumed, and aids used for oral hygiene maintenance (fluoridated or non-fluoridated). SES was measured using the Kakkar socioeconomic status scale (KSESS) with eight closed-ended questions related to parental education, family income, father's occupation, and other factors. All children were asked to fill out the form, and the answers obtained were scored using Kakkar socioeconomic status scoring keys. Based on this scoring, children were divided into three groups: lower class, middle class, or upper class. Age, sex, and SES were not found to be significantly associated with IQ. Other covariates, including smoking, were not addressed. Co-exposures such as arsenic and lead were not addressed; however, there is no indication that lead is a co-exposure in this population, and arsenic is not likely a major concern in this area based on water quality maps.
- *Potentially important study-specific covariates:* Key covariates age, sex, and measures of SES were similar between exposure groups; however, arsenic was not considered. Arsenic often occurs in the drinking water along with fluoride in some Indian populations; however, based on water quality maps, this does not appear to be an issue in the Nalgonda district of Andhra Pradesh. Iodine deficiencies are not mentioned.
  - *Direction/magnitude of effect size:* Conceptually, the presence of arsenic would potentially bias the association away from the null if present with fluoride. Deficiencies in iodine would likely bias the association away from the null if present in areas of high fluoride but toward the null if present in areas of non-high fluoride. Neither of these were considered issues in this study for reasons noted above.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the key covariates were considered, co-exposure to arsenic was likely not an issue in this area, and methods used for collecting the information were valid and reliable.

- Attrition:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results were available for the 1,000 children selected to participate.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence of no attrition.
- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children were placed into one of four strata based on the level of fluoride in drinking water. Collection of water samples was done in the districts. The placement into strata was based on fluoride levels obtained from documented records of the District Rural Water Works Department. Once the children were assigned to strata, it was confirmed that the fluoride level of their drinking water was within the strata assigned. This was done using the methodology followed in

the National Oral Health Survey and Fluoride Mapping 2002–2003. During the initial visits to the schools, the children were interviewed regarding their history of residence and source of drinking water from birth to 10 years. The first child meeting the criteria was given a bottle for water collection, and the next child was given a bottle for collection only if the water source was different from that of a previous child. Children were asked to collect a water sample from the source that was used in the initial 10 years of their life (and that sample was collected the next day). It was not reported whether all bottles were returned. The water samples collected were subjected to water fluoride analysis using an ion-specific electrode, Orion 720A fluoride meter at District Water Works, Nalgonda to confirm the fluoride levels in the water before commencement of clinical examination. LOD and QA/QC details were not reported.

- *Direction/magnitude of effect size:* There is some potential for exposure misclassification based on recall of the children on the source of water used in their first 10 years of life. The misclassification is likely non-differential and not likely to bias in any specific direction. Children who had changed water since birth were excluded, but it was not specifically noted that the fluoride in the water source was stable over the years.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- Outcome:
  - *Rating:* Probably high risk of bias (NR)
  - *Summary:* Raven's standard progressive matrices (1992 edition) was used to assess IQ. Raven's test is a standard test; although there is no information provided to indicate that the methods were reliable and valid in this study population, the test was created to be culturally fair (+ for methods). Blinding or other methods to reduce potential bias were not reported (NR for blinding). No response was received to an email request for clarification in September 2017. Overall rating for methods and blinding = NR.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that the outcome assessors were not blind to participants' fluoride exposure and could bias the results.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*

- *Statistical analyses:* Chi-square test and Spearman rank correlation were used to assess the association between four different fluoride levels and IQ grades. Area-level exposures were used. Clustering of children within the four areas was not accounted for in the analysis; however, because multiple villages were included in each fluoride exposure level, clustering was less of a concern and the impact on the effect estimates was expected to be minimal.
- *Other potential concerns:* None identified.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include verification of exposure measurements and consideration of key covariates, but it was limited by the cross-sectional study design and lack of information on blinding during outcome assessment.

E.1.12. Till et al. (2020)

E.1.12.1. Study Details

- ***Study design:*** Prospective cohort
- ***Population:*** MIREC participants (pregnant mothers and their children aged 3–4 years)
- ***Study area:*** 10 cities, Canada
- ***Sample size:*** 398 mother-child pairs (247 from non-fluoridated areas, 151 from fluoridated areas; 200 breastfed as infants, 198 formula-fed as infants)
- ***Data relevant to the review:*** Adjusted linear regression models evaluating associations between IQ and water fluoride concentration (with or without adjusting for maternal urine) in formula-fed or breastfed infants or fluoride intake from formula.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower performance IQ with water fluoride per 0.5-mg/L increase by breastfeeding status (adjusted βs = −9.26 formula-fed, −6.19 breastfed) and fluoride intake from formula (adjusted β = −8.76); significantly lower full-scale IQ with water fluoride per 0.5-mg/L increase in formula-fed children (adjusted β = −4.40); no significant changes in full-scale IQ for water fluoride in breastfed children or fluoride intake from formula-fed children; no significant changes in verbal IQ scores with fluoride exposure.

E.1.12.2. Risk of Bias

- Author contacts:
  - ~~Authors were not contacted for additional information because it was not necessary.~~Authors were not contacted for the 2020 publication. Authors were contacted in June 2019 for additional information on the Green et al. (2019) publication. Information obtained from that correspondence may have been used for additional information in the 2020 publication.

Commented [A117]: Change made in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 5): The NTP response to NASEM's comments indicate they contacted investigators on specific issues, and that some study authors responded while others did not. Recommendation: NTP should provide an assessment of how non-response would affect the grading of these studies. This may be tied to the NASEM concern regarding transparency in the risk-of-bias assessment. [redacted] requests an appendix and language in the Discussion detailing which studies received a change to their risk-of-bias determination.

- Population selection:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Pregnant women were recruited between 2008 and 2011 by the MIREC program from 10 cities across Canada. Inclusion and exclusion criteria were provided. Additional details were stated to be available in Arbuckle et al. (2013). A total of 610 children were recruited to participate in the developmental follow-up with 601 children completing all testing. The demographic characteristics of women included in the current analyses (n = 398) were not substantially different from the original MIREC cohort (n = 1,945) or the subset without complete water fluoride and covariate data (n = 203). A table of characteristics of the study population was provided. Approximately half of the children lived in non-fluoridated cities and half lived in fluoridated cities.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were similar and were recruited with the same methods during the same time frame.
- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariates were selected a priori that have been associated with fluoride, breast feeding, and children's intellectual ability. Final covariates included sex and age at testing, maternal education, maternal race, secondhand smoke in the home, and HOME score. City was considered but excluded from the models. Covariates that were not assessed include parental mental health, iodine deficiency/excess, parental IQ, and co-exposure to arsenic and lead. Co-exposure to arsenic is less likely an issue in this Canadian population because it receives water mainly from municipal water supplies that monitor for lead and arsenic, and the lack of information is not considered to appreciably bias the results. In addition, a previous study on this population (Green et al. 2019) conducted sensitivity analyses on co-exposures to lead and arsenic. Results from these sensitivity analyses support the conclusion that co-exposures to lead and arsenic are not likely a major concern in this study population.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect:* Not applicable.
  - *Basis for rating:* Probably low risk of bias based on direct evidence that key covariates were considered and indirect evidence that the methods used to collect the information were valid and reliable and co-exposures were not an issue.
- Attrition:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of 610 children, 601 (98.5%) in the MIREC developmental study who were ages 3–4 years completed the neurodevelopment testing. Of the 601 children who completed the neurodevelopmental testing, 591 (99%) completed the infant feeding questionnaire and 398 (67.3%) reported drinking tap water. It was noted that the demographic characteristics were not substantially different from the

original MIREC cohort or the 203 subjects without complete water fluoride or covariate data.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Information on breastfeeding was obtained via questionnaire at 30–48 months. Fluoride concentration in the drinking water was assessed by daily or monthly reports provided by water treatment plants. Water reports were first linked with mothers' postal codes, and the daily or weekly amounts were averaged over the first 6 months of each child's life. Additional details can be found in Till et al. (2018). Maternal urinary exposure was used to assess fetal fluoride exposure. Procedures can be found in Green et al. (2019).
    - *Direction/magnitude of effect size:* There is not any specific direction or magnitude of bias expected. Urinary fluoride levels are reflective of recent exposure. The possibility of exposure misclassification would be similar in all subjects and would be non-differential. For the fluoride intake from formula, exposure was based on the fluoride levels in the water at the residence and the proportion of time that the infant was not exclusively breastfed. This exposure misclassification would also be non-differential.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- Outcome:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Intelligence was tested using the Wechsler Preschool and Primary Scale of Intelligence III, which is considered a gold standard test. It is appropriate for both the study population and age group. It was not reported whether the evaluators were blind to the child's fluoride exposure status during the assessment. Although it is unlikely that the assessors had knowledge of the specific drinking water levels or maternal urine levels, there is potential that the outcome assessors had knowledge of the city the child lived in and whether the city was fluoridated or non-fluoridated. Correspondence with the study authors on the outcome assessment for Green et al. (2019) indicated that it was unlikely that the testers had knowledge of the city's fluoridation. The same is assumed here. Specific measurements included were identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- Selective Reporting:

  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Regression diagnostics were used to test assumptions for linearity, normality, and homogeneity. There were two potential influential observations (based on Cook's distance), and sensitivity analyses re-estimated the models without these two variables. Effect modification by breastfeeding status was evaluated. Interestingly, all regression coefficients were divided by 2 to represent change in IQ per 0.5-mg/L change in fluoride. One concern is posed by the lack of accounting for city in the regression models, ideally as a random effect. The authors explored including city as a covariate in the models; however, city was not included either because it was strongly multi-collinear with water fluoride concentration (VIF > 20) (model 1, with water fluoride concentration) or because fluoride intake from formula is a function of water fluoride concentration (assessed at the city level) and was therefore deemed redundant (model 2). However, the models use city-level water fluoride concentrations—and, in sensitivity analyses, adjust for maternal urinary fluoride—which warrants exploration of city as a random effect rather than a fixed effect (as would be the case by having it included as a covariate). Even including individual-level maternal urinary fluoride might not fully account for lack of a city effect, given that the subjects were from six different cities, with half of them fully on fluoridated water. Hence, even individual-level exposures are likely to be correlated at the city level. Based on a previous analysis (Green et al. 2019), it is unlikely that exclusion of city from models (as a fixed or random effect) would significantly impact the effect estimates.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, prospective cohort design, and consideration of key covariates.

E.1.13. Trivedi et al. (2012)

E.1.13.1. Study Details

- ***Study design:*** Cross-sectional

- ***Population:*** Children aged 12–13 years
- ***Study area:*** Kachchh, Gujarat, India
- ***Sample size:*** 84 children
- ***Data relevant to the review:*** Mean IQ scores and distribution by low and high fluoride villages.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower mean IQ score in the high fluoride villages (92.53 ± 3.13) compared with the low-fluoride villages (97.17 ± 2.54) in boys and girls combined (and by sex).

E.1.13.2. Risk of Bias

- Author contacts:
  - Authors were contacted in September of 2017 to obtain additional information for risk-of-bias evaluation. Additional information provided by the authors informed the rating decision for the following risk-of-bias domains: Selection, Attrition, Detection (exposure assessment), Detection (outcome assessment).
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There is insufficient information provided on the sampling methods to determine whether the populations were similar. Although it was noted that samples were obtained for groundwater quality from March to May of 2011, there is no indication that the children were selected at the same time or during a similar time frame. Correspondence with the author indicates that children were selected within a week of the water collection based on random selection of a school in the village. Study participants were selected from six different villages of the Mundra region of Gujarat, India. Subjects were grouped into high and low villages based on the level of fluoride in the drinking water of those villages. The number of subjects per village was not reported, but it was noted that there were 50 children in the low-fluoride group and 34 children in the high fluoride group. It is not clear whether the differences in numbers were based on different participation rates or whether there were fewer children in the high fluoride villages. Recruitment methods, including any exclusion criteria and participation rates, were not provided. SES was stated to be low and equal based on questionnaire information, but the results were not provided. It should also be noted that only regular students (having attendance more than 80%) of standard 6th and 7th grades were selected, but it was not noted whether attendance varied by village. Correspondence with the study author indicated that there was an average of 20 students per class with an average of 40 students per village. It appears that keeping the requirement of 80% attendance was a limiting factor that resulted in different numbers of children by area; however, this was applied similarly to both groups.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and recruited using the same methods during the same time frame.

Commented [A118]: Change made in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 5): The NTP response to NASEM's comments indicate they contacted investigators on specific issues, and that some study authors responded while others did not. Recommendation: NTP should provide an assessment of how non-response would affect the grading of these studies. This may be tied to the NASEM concern regarding transparency in the risk-of-bias assessment. [redacted] requests an appendix and language in the Discussion detailing which studies received a change to their risk-of-bias determination.

- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children were stated to be students of the 6th and 7th standard grades. Age was not addressed, but the children would all be of similar ages based on the grades included. Results were reported for males and females separately as well as combined. SES and iodine consumption were stated to be analyzed via a questionnaire and were standardized on the basis of the 2011 census of India. Although it was noted in the abstract that the SES was equal (no data provided), the study report did not mention the iodine results. Although arsenic and lead were not considered, the study authors provided physicochemical analyses for the water samples from the six different villages. While the authors did not specifically analyze lead or arsenic in the water samples, these physicochemical analyses suggest that differential lead or arsenic exposure was unlikely. Moreover, based on water quality maps, arsenic was not expected to be a major concern in this study area. According to the information from the water quality maps and the physiochemical analysis of the water provided, there is indirect evidence that neither arsenic nor lead were a concern in this study population.
  - *Potentially important study-specific covariates:* Key covariates age, sex, and measures of SES were similar between exposure groups; however, arsenic was not considered. Arsenic often occurs in the drinking water along with fluoride in some Indian populations; however, based on water quality maps, arsenic does not appear to be an issue in the study area.
    - *Direction/magnitude of effect size:* Conceptually, the presence of arsenic would potentially bias the association away from the null if present with fluoride, or toward the null if present in the reference group; however, for reasons noted above, arsenic is not considered a concern in this study population.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable, that potential co-exposures were not an issue, and that key covariates were addressed.
- Attrition:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results were provided for 84 children, but the methods do not indicate how many children were initially selected to participate, nor were any exclusion criteria provided. It was noted in the results that 84 children had their groundwater and urine tested, but it was not noted whether analyses were restricted to these children or whether exposures were assessed in all the children who had IQ measurements. Correspondence with the study author indicated that the main reason for exclusion was a <80% attendance rate, with fluoride and IQ measured on all 84 children who met the criteria.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence of no attrition.
- Exposure:

- *Rating:* Probably low risk of bias (+)
- *Summary:* Children in villages were grouped based on fluoride levels that were assessed in groundwater (low fluoride villages versus high fluoride villages). The average concentration of these levels was considered to be the levels in the drinking water with confirmation using urinary fluoride levels. The groundwater samples were selected to cover major parts of the taluka and represent overall groundwater quality. Ten samples were obtained from each village. Fluoride was measured in the groundwater using ion exchange chromatography. Although urine levels were also significantly higher in the high fluoride village, no information was provided on how or when the urinary samples were obtained or how they were measured. However, correspondence with the study author indicated that the groundwater and urine fluoride levels were available for all 84 children, indicating that the urine measures were available for the children that had IQ measures. The urine samples were stated to be collected at the same time the second water sample was collected.
  - *Direction/magnitude of effect size:* Fluoride levels were measured in both the drinking water and urine. Although there is some variability in the measurements, there is no overlap between the two groups, and the urine and drinking water levels in the children support each other. Any potential exposure misclassification would be non-differential, and the impact on the direction and magnitude of the effect size is unknown.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- Outcome:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Outcome methods were only noted to be reported in Trivedi et al. (2007), which was scored as follows: IQ was measured in the children of both areas using a questionnaire prepared by Professor JH Shah, copyrighted by Akash Manomapan Kendra, Ahmedabad, India, and standardized on the Gujarati population with a 97% reliability rate in relation to the Stanford-Binet Intelligence Scale (+ for methods). Blinding or other methods to reduce bias were not reported, but correspondence with the study author indicated that the teachers were blind to the status of fluoride. The teachers administered the tests in the presence of a research fellow. It is not completely clear who scored the tests, but it is assumed the teachers (+ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcomes were assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)

- *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
- *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.

- Other potential threats:
  - *Rating:* Probably high risk of bias (−)
  - *Summary:*
    - *Statistical analyses:* Mean IQ scores in low and high fluoride villages were compared using a t-test. Consideration of heterogeneity of variances was not reported. Results are reported as means and standard errors of the means, with p-values for significant differences. Area-level exposures were used. There was no accounting for clustering of children within the villages, and comparative analyses did not account for covariates. Urinary fluoride was not considered in the comparative analyses. The lack of individual exposure levels and the lack of accounting for clustering are likely to bias the standard error of the difference in mean IQ levels between the high- and low-fluoride villages and make the differences appear stronger than they actually are.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that the statistical analyses did not account for clustering, and this lack of accounting could bias the association. There were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and the addressing of potential key covariates, but the study was limited by the cross-sectional study design. Another limitation was the lack of accounting for clustering, which may bias the standard error of the differences, making the effect appear stronger than it actually is; however, this does not change the nearly 5-point difference in IQ scores between the two villages.

E.1.14. Wang et al. (2012)

E.1.14.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 8–13 years [possibly the same study population as Xiang et al. (2003a)]
- ***Study area:*** Wamiao and Xinhuai villages located in Sihong County, Jiangsu Province, China
- ***Sample size:*** 526 school children
- ***Data relevant to the review:*** Mean IQ and % low IQ (<80) by total fluoride intake.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower mean IQ in the endemic versus non-endemic regions, as reported in Xiang et al. (2003a); when the high-exposure group was broken into four exposure groups based on fluoride intake, a dose-dependent decrease in IQ and increase in % with low IQ was observed;

significant correlation between total fluoride intake and IQ ($r = -0.332$); for IQ <80, adjusted OR of total fluoride intake per 1 mg/(person/day) was 1.106 (95% CI: 1.052, 1.163).

E.1.14.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information because it was not necessary.
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study appears to have the same study population as Xiang et al. (2003a) and Xiang et al. (2011); however, it does not cite these studies as providing additional information, and the numbers of children differ; therefore, it may be a separate analysis on the same villages. The years of testing were not provided, so it cannot be determined whether study subjects were the same. Two villages, Wamiao and Xinhuai, located 64 km apart in Sihong County, Jiangsu Province, were selected for the study. Wamiao is a village in a region with severe endemic fluorosis, and Xinhuai is a village in a non-endemic fluorosis region. Neither village has fluoride pollution from coal or industrial sources. Villages were stated to be similar in terms of annual per capita income, transportation, education, medical conditions, natural environment, and lifestyle. All primary students ages 8–13 years currently in school in either village were surveyed with exclusions noted. Of 243 children from Wamiao, 236 (97.12%) were included, and of 305 children from Xinhuai, 290 (95.08%) were included. No table of subject characteristics was provided.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited using the same methods within the same time frame, with direct evidence that there was no difference in participation/response rates.
- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Logistic regression of low IQ rate and total fluoride intake adjusted for age and sex. Both villages had hand-pumped well water for drinking water, but the authors do not mention whether arsenic was also present in the drinking water. However, a publication by Xiang et al. (2013) in the same study areas indicates that Xinhuai (the low-fluoride area) had significantly higher arsenic levels compared with Wamiao (the endemic fluorosis area), which would bias the association toward the null. Areas were stated to be similar in annual per capita income, transportation, education, medical conditions, natural environment, and lifestyle; however, no details were provided. This study did not address other co-exposures, but other studies on populations in these villages (Xiang et al. 2003a; Xiang et al. 2011) indicate that iodine and lead are not concerns.
  - *Potentially important study-specific covariates:* Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based

on information provided in Xiang et al. (2013), arsenic concentrations were higher in the low-fluoride area compared with the high fluoride area. Because there were significant effects on IQ observed in the high fluoride areas, the impact of co-exposure to arsenic is less of a concern. The presence of arsenic in the control village may cause an underestimation of the effect of fluoride, but despite this potential impact, a significant association between fluoride exposure and IQ was reported.

- *Direction/magnitude of effect size:* Presence of arsenic in this study population would potentially bias the association toward the null.

- *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key covariates were considered, methods used for collecting the information were valid and reliable, and co-exposures to arsenic and lead and iodine deficiency were not attributing to the association observed in this study. The potential for bias toward the null combined with the reported significant association increases confidence in the observed effect.

- Attrition:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Data are reported for all 526 children noted to be included in the study. There is a slight discrepancy in the reported total number of children from the high-fluoride village and the number of participants from the high-fluoride village between this paper (236 participated of 243 total children) and the 2003 and 2011 publications on the same study population (222 of 238). This discrepancy is not explained but is not expected to appreciably bias the results.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
  - *Rating:* Probably low risk of bias (+); Probably high risk of bias (−)
  - *Summary:* **Water fluoride (+ probably low risk of bias):** Exposure was based on drinking water levels and fluoride intake. Residents in the Wamiao village were divided into five groups based on fluoride levels in the drinking water. Clean, dry polyethylene bottles were used to collect 50 mL of drinking water from each student's household, and fluoride content was measured.

    Total fluoride intake (− probably high risk of bias): Six families from each of the five Wamiao groups were randomly selected as dietary survey households. Intakes of various foods by each person at each meal and intakes of unboiled water, boiled water, and tea were surveyed for four consecutive days. Methods for food collection were described. Five representative households from each village were selected based on geographic location, population distribution, housing structure, and other conditions. Indoor air samples were collected once daily for five consecutive days; outdoor air was sampled at two points once daily for five days. Methods for determining fluoride content in samples were noted to follow

specific guidelines. Calculation of total fluoride intake was stated to follow Appendix A of the People's Republic of China Health Industry Standard with some details provided. Although it is assumed the method is valid, it was not detailed how each fluoride determination was made for each subject, and it appears that total fluoride intake was determined based on data from select subjects and not all subjects.

        - *Direction/magnitude of effect size:* There is potential for exposure misclassification based on calculating fluoride intake based on measurements from a few select subjects rather than all subjects. The potential impact on the direction and magnitude of effect size cannot be assessed based on the information provided.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure. The total fluoride intake is probably high risk of bias because there is indirect evidence that the exposure was assessed using methods of unknown validity.
- Outcome:
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* The IQ of each child was measured with the Combined Raven's Test for Rural China (CRT-RC) (++ for methods). The test was stated to be administered to the children independently in a school classroom under the supervision of three exam proctors. Testing methods, testing language, and testing conditions were all in strict accordance with the CRT-RC guidebook. Major testing personnel received necessary training by the Psychology Department of East China Normal University. The children undergoing IQ testing and the test scorers were kept double-blind throughout the testing process (++ for blinding). Overall rating = ++.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- Selective Reporting:
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
    - *Rating:* Probably low risk of bias (+)
    - *Summary:*
        - *Statistical analyses:* Logistic regression analysis was used to determine the odds of having low IQ with increasing fluoride intake. Analyses and methods

are not well described. There is no mention of what tests were used for the mean IQ comparison by village; however, statistical software (SPSS) was used, suggesting appropriate tests were applied. Simple linear regression analyses were conducted to evaluate associations between total fluoride intake and children's IQ or low IQ rate. There is no evidence that regression diagnostics were used to test model assumptions for linearity, normality, and homogeneity. Clustering at the village level was not accounted for in the analyses. The overall impact of these factors on effect estimates is expected to be minimal given the use of individual-level data and adjustment for important covariates.

    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements with blinding at outcome assessment, but is limited by the cross-sectional study design and lack of individual measurements to calculate fluoride intake. All key covariates were accounted for in the study design or analysis, but there is potential for the presence of arsenic to bias the association toward the null.

E.1.15. Wang et al. (2020b)

E.1.15.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** School children aged 7–13 years
- ***Study area:*** Tianjin City, China [possibly a subset of the children from Yu et al. (2018)]
- ***Sample size:*** 571 school children
- ***Data relevant to the review:*** IQ scores by urine and water fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant associations between IQ score and water fluoride (adjusted $\beta = -1.587$ per 1-mg/L increase) and urinary fluoride (adjusted $\beta = -1.214$ per 1-mg/L increase) in boys and girls combined based on both quartiles and continuous measures. No significant modification effect of sex.

E.1.15.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information because it was not necessary.
- Population selection:
  - *Rating:* Definitely low risk of bias (++)

- *Summary:* Subjects were from a cross-sectional study conducted in 2015, but no citation was provided on this cohort [presumably the Yu et al. (2018) cohort]. It was noted that the subjects in that cohort were from districts with historically high or normal fluoride levels. Subjects for this study were selected by using a stratified and multistage random sampling approach. Brief description was provided. The study area consisted of three historically high fluoride areas and four non-endemic areas. A flow diagram was provided for inclusion and exclusion, but this detail was given for all children and not by area. Therefore, it cannot be determined whether the participation differed by area. However, there was a 93% recruitment rate, and the 13 excluded due to missing data were not likely excluded due to exposure. Detailed characteristics of the study population are provided. Exclusion criteria included: "children who had congenital or acquired diseases affecting intelligence, or a history of cerebral trauma and neurological disorders, or those with a positive screening test history (like hepatitis B virus infection, Treponema palladium infection and Down's syndrome) and adverse exposures (smoking and drinking) during maternal pregnancy, prior diagnosis of thyroid disease, and children who had had missing values of significant factors (2.2%) were also excluded."
- *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were recruited using similar methods during the same time frame and that any differences between the exposed groups were accounted for in the statistical analyses.

- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Study authors noted that the study areas were not exposed to other neurotoxins such as lead, arsenic, or mercury nor were they iodine-deficient. Final models included age, sex, child's BMI, maternal and paternal education, household income, and low birth weight. The other covariates that were considered are unclear as the authors only noted that the covariates were selected based on current literature. Reasons for exclusion included history of disease affecting intelligence, history of trauma or neurological disorders, positive screening test history, or exposures such as smoking or drinking during pregnancy. Information was obtained by questionnaire or measurements. Covariates such as parental BMI, behavioral and mental health disorders, IQ, and quantity and quality of the caregiving environment were not considered.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Probably low risk of bias because there is direct evidence that the key covariates were considered and indirect evidence that the methods for collecting the information were valid and reliable and that co-exposure to arsenic was not an issue in this area.
- Attrition:
  - *Rating:* Definitely low risk of bias (++)

  - *Summary:* A detailed chart of the recruitment process is presented. The study had a 93% recruitment rate, and only 2.2% of subjects with missing data for certain covariates were excluded.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children provided spot urine samples, presumably at the time of examination. Water samples were randomly collected from public water supplies in each village. Fluoride concentrations were analyzed using fluoride ion-selective electrode according to the national standardized method in China. There is no indication of whether the urine samples accounted for dilution.
    - *Direction/magnitude of effect size:* Not accounting for dilution could cause some exposure misclassification. The impact on the direction and magnitude of effect size would depend on where the differences occurred.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.
- Outcome:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Assessments of IQ scores were conducted by graduate students at the School of Public Health, Tongji Medical College, Huazhong University of Science and Technology. Each team member was assigned a single task, meaning that only one person would have conducted the IQ tests. A Combined Raven's Test for Rural China was used. Therefore, the test was appropriate for the study population (++ for method). It was noted that the examiner was trained and blind to the exposure (++ for blinding). Overall = ++
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)

- *Summary:*
  - *Statistical analyses:* Logistic and multivariate regression models accounting for covariates were used. Results are presented as betas or ORs and 95% CIs. Regression diagnostics were conducted for all models, including examination of multicollinearity, heteroscedasticity, and influential observations. Mediation and interaction analyses were appropriate. There is no evidence that the stratified and multistage random sampling approach for subject selection was accounted for in the analyses by using sampling weights or accounting for clustering using random effect models; however, selected villages were similar in population and general demographic characteristics. Given the use of individual-level data and adjustment for important covariates, the impact on the regression coefficients is likely to be minimal.
  - *Other potential concerns:* None identified.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and no other potential threats of risk of bias were identified.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key covariates were considered in the study design or analysis.

E.1.16. Xiang et al. (2003a)

E.1.16.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 8–13 years
- ***Study area:*** Wamiao and Xinhuai villages located in Sihong County, Jiangsu Province, China
- ***Sample size:*** 512 school children
- ***Data relevant to the review:*** Comparison of IQ (mean and distribution) between Wamiao County (a severe endemic fluorosis area) and Xinhuai County (a non-endemic fluorosis area); additional breakdown of the Wamiao area into five water fluoride exposure groups.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower IQ scores observed with water fluoride levels of 1.53 mg/L or higher. Percentage of subjects with IQ scores below 80 was significantly increased at water fluoride levels of 2.46 mg/L or higher. Significant inverse correlation between IQ and urinary fluoride ($r = -0.164$). Mean IQ scores for children in the non-endemic region (100.41 ± 13.21) were significantly higher than the endemic region (92.02 ± 13.00).

E.1.16.2. Risk of Bias

- Author contacts:

- Authors were not contacted for additional information because it was not necessary.

- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Two villages, Wamiao and Xinhuai, located 64 km apart in Sihong County, Jiangsu Province, were selected for this study, which was conducted between September and December 2002. Wamiao is located in a severe fluorosis endemic area, and Xinhuai is located in a non-endemic fluorosis area. Neither village has fluoride pollution from burning coal or other industrial sources. All eligible children in each village were included; children who had been absent from either village for 2 years or longer or who had a history of brain disease or head injury were excluded. In Wamiao, 93% of the children (222 out of 238) were included in the study; in Xinhuai, 95% were included (290 out of 305). The children in Wamiao were divided into five subgroups according to the level of fluoride in their drinking water: <1.0 mg/L (group A), 1.0–1.9 mg/L (group B), 2.0–2.9 mg/L (group C), 3.0–3.9 mg/L (group D), and >3.9 mg/L (group E). Children in Xinhuai (0.18–0.76 mg/L in the drinking water) served as a control group (group F). Demographic characteristics are not presented, and statistical analyses are not adjusted, but mean IQ scores are stratified by age, sex, family income, and parental education.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited using the same methods within the same time frame, with direct evidence that there was no difference in participation/response rates.
- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although information was stated to be collected on personal characteristics, medical history, education levels of the children and parents, family SES, and lifestyle, only sex, age, family income, and parental education were considered. Potential co-exposures, such as arsenic, were not addressed. A separate publication in 2003 [(Xiang et al. 2003b), letter to the editor] indicated that blood lead levels were not significantly different between the two areas. Although arsenic was not addressed specifically in this publication, Xiang et al. (2013) measured both fluoride and arsenic in the Wamiao and Xinhuai areas. Xinhuai (the low-fluoride area) had significantly higher arsenic levels compared with Wamiao (the endemic fluorosis area). This is likely to bias the association toward the null; however, the study observed a significantly lower IQ score in the endemic fluorosis area. Iodine was tested in a subset of the children and found not to be significantly different between the two groups.
  - *Potentially important study-specific covariates:* Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on information provided in Xiang et al. (2013), arsenic concentrations were higher in the low-fluoride area compared with the high fluoride area. Because there were significant effects on IQ observed in the high fluoride areas, the impact

of co-exposure to arsenic is less of a concern. The presence of arsenic in the control village may cause an underestimation of the effect of fluoride, but despite this potential impact, there was still a significant association between fluoride exposure and IQ.

- *Direction/magnitude of effect size:* Presence of arsenic in this study population would potentially bias the association toward the null.

- *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key covariates were taken into account, methods used for collecting the information were valid and reliable, and co-exposures to arsenic and lead and iodine deficiency were not attributing to the effect observed in this area. The potential for bias toward the null, combined with the reported significant association increases confidence in the observed effect.

- Attrition:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are complete. IQ results were reported for all 512 children included in the study (222 in the endemic area and 290 in the nonendemic area).
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Exposure was based on drinking water and urinary levels of fluoride. The two study areas were selected to reflect a severe endemic area and a non-endemic area. Drinking water was collected from wells, and early-morning spot urine samples were collected from a randomly selected subsample of children. Both water and urine samples were measured using fluoride ion-selective electrode, but no quality control was discussed. Both absolute and creatinine-adjusted urine results were reported.
    - *Direction/magnitude of effect size:* There is potential for exposure misclassification because only current levels were assessed. Migration of subjects in or out of the area was not assessed, but the study authors noted that if the children had been absent from the village for 2 or more years, they were excluded. Misclassification would likely be non-differential, which could likely bias the association in either direction.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- Outcome:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* The IQ of each child was measured with the Combined Raven's Test for Rural China (CRT-RC) (++ for methods). The test was stated to be administered to the children independently in a school classroom, in a double-blind manner, under the supervision of an examiner and two assistants, and in

accordance with the directions of the CRT-RC manual regarding test administration conditions, instructions to be given, and test environment (++ for blinding). Overall rating = ++

- *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.

- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* There is no mention of the tests conducted, but data were stated to be analyzed using SAS, suggesting appropriate tests were applied. Results provided in the tables indicate that t-tests comparing IQ values between the villages (overall and by sex) were conducted, but it was not reported that heterogeneity of variance was assessed. In addition, correlations between IQ and age, family income, and parents' education level were tested with Pearson's correlation. There is no evidence that a test for trend was conducted to evaluate the stated "significant inverse concentration-response relationship between the fluoride level in drinking water and the IQ of children."
    - A potential concern raised by the NASEM (2020) committee's review was the lack of accounting for relationships in exposure between persons from the same village. Given only two villages were included and the analyses consisted of village-level comparisons (no use of individual-level covariate data), it is likely that the standard error of the difference in mean IQ between fluoride in water exposure groups will be biased, making differences appear stronger than they actually are. Without controlling for village effects and given the large differences in fluoride concentrations and IQ levels between villages, the apparent dose-response relationship could be due to a village effect in addition to a fluoride effect. However, a dose-response relationship is apparent within the "exposed" village, diminishing the concern for a village-only effect and likely minimizing the impact on the effect estimates.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that statistical analyses were appropriate and that there were no other threats of risk of bias.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and blinding of outcome assessor to exposure but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key covariates were considered in the study design or analysis, but there is potential for the presence of arsenic to bias the association toward the null.

E.1.17. Xiang et al. (2011)

E.1.17.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 8–13 years [same study population as Xiang et al. (2003a)]
- ***Study area:*** Wamiao and Xinhuai villages located in Sihong County, Jiangsu Province, China
- ***Sample size:*** 512 school children
- ***Data relevant to the review:*** Mean IQ scores and odds ratio for having an IQ <80 presented by serum fluoride quartiles.
- ***Reported association with fluoride exposure:*** Yes: Significant linear trend across quartiles of serum fluoride and children's IQ score <80 (adjusted ORs for Q1 and Q2; Q1 and Q3; and Q1 and Q4, respectively: 1; 2.22 [95% CI: 1.42, 3.47]; and 2.48 [95% CI: 1.85, 3.32]); significant effects observed at ≥0.05 mg/L serum fluoride.

E.1.17.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information because it was not necessary.
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study population was the same as that used in the Xiang et al. (2003a) study, but a few more measurements were available and different analyses were conducted. The comparison population was considered the same based on the study populations being recruited from similar populations, using similar methods, during the same time frame. Demographic characteristics were not provided.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited using the same methods within the same time frame, with direct evidence that there was no difference in participation/response rates.
- Confounding:
  - *Rating:* Probably low risk of bias (+)

- *Summary:* As was noted in the 2003 publication (Xiang et al. 2003a), information was collected on personal characteristics, medical history, education levels in the children and parents, family SES, and lifestyle. In the logistic regression model age and sex were adjusted for in the analysis. In the previous report, no significant associations were observed between groups for family income and parents' education (Xiang et al. 2003a). Urinary iodine and blood lead levels were also stated to be measured and were noted not to be significantly different between the groups. Although the iodine levels were reported in the previous publication, the lead levels were not and neither were the methods. Lead information is reported in a letter to the editor (Xiang et al. 2003b) and was not significantly different between the areas. Although arsenic was not addressed specifically in this publication, Xiang et al. (2013) measured both fluoride and arsenic in the Wamiao and Xinhuai areas. Xinhuai (the low-fluoride area) had significantly higher arsenic levels compared with Wamiao (the endemic fluorosis area). This is likely to bias the association toward the null; however, the study observed a significantly lower IQ score in the endemic fluorosis area and with increasing serum fluoride.
- *Potentially important study-specific covariates:* Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on information provided in Xiang et al. (2013), arsenic concentrations were higher in the low-fluoride area compared to the high fluoride area. Because there were significant effects on IQ observed in the high fluoride areas, the impact of co-exposure to arsenic is less of a concern. The presence of arsenic in the control village may cause an underestimation of the effect of fluoride, but despite this potential impact, there was still a significant association between fluoride exposure and IQ.
  - *Direction/magnitude of effect size:* Presence of arsenic in this study population would potentially bias the association toward the null.
- *Basis for rating:* Probably low of risk bias because there is indirect evidence that the key covariates were considered, methods used for collecting the information were valid and reliable, and co-exposures to arsenic and lead and iodine deficiency were not attributing to the effects observed in this area. The potential bias toward the null, combined with the reported significant association increases confidence in the observed effect.

- Attrition:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are reported for all 512 children noted to be included in the study.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Fluoride levels were measured in serum with a fluoride ion-selective electrode. A fasting venous blood sample was used. No details are provided on

validation (including correlation with drinking water levels) or QA. Children who did not reside in their village for at least 2 years were excluded. Results were provided in quartiles, but the authors combined the lower two quartiles. After combining the two lower quartiles, the exposure levels ranged from <0.05 mg/L (Q1 + Q2) to >0.08 mg/L (Q4).

- *Direction/magnitude of effect size:* Serum fluoride may not be the best estimate for exposure. There is potential for exposure misclassification because only current levels were assessed. Migration of subjects in or out of the area was not assessed, but the study authors noted that if the children had been absent from the village for 2 or more years, they were excluded. Misclassification would likely be non-differential, which could bias results in either direction.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- Outcome:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* IQ was assessed as part of the 2003 evaluation. IQ was measured with the Combined Raven's Test for Rural China, which is appropriate for this population (++ for methods). Although this study does not provide details, the original study article from 2003 provides specific details. The study authors indicate in the 2003 publication that the tests were conducted in a double-blind manner, and these are the same results and population (++ for methods). Overall rating = ++
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses conducted were appropriate for the study. Chi-square tests were used to compare categorical variables, and multiple logistic regression was used to evaluate the association between serum fluoride levels and risk of low IQ. A potential concern raised by the NASEM (2020) peer review was the lack of accounting for relationships in

exposure between persons from the same village. Although only two villages were included, in the analyses that consisted of village-level comparisons, it is likely that the standard error of the difference in mean IQ between villages is biased. This is less of a concern for the mean IQ comparisons across quartiles of serum fluoride levels and for the logistic regression analyses of risk of low IQ and individual-level serum fluoride levels. Without controlling for village effects and given the large differences in fluoride concentrations and IQ between villages, the apparent dose-response relationship could be due to a village effect in addition to a fluoride effect. However, the dose-response relationship is still present within the "exposed" village, diminishing the concern for a village-only effect and likely minimizing the impact on the effect estimates.

- *Other potential concerns:* None identified.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements with blinding at outcome assessment but is limited by the cross-sectional study design and use of serum concentrations. All key covariates were considered in the study design or analysis, but there is potential for the presence of arsenic to bias the association toward the null.

E.1.18. Yu et al. (2018)

E.1.18.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 7–13 years
- ***Study area:*** Tianjin City, China
- ***Sample size:*** 2,886 school children
- ***Data relevant to the review:*** IQ for normal (≤1 mg/L) versus high (>1 mg/L) water fluoride; betas for IQ score by water and urine fluoride groupings; ORs by IQ category using water and urine fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant difference in mean IQ scores in high water fluoride areas (>1.0 mg/L; 106.4 ± 12.3 IQ) compared to the normal water fluoride areas (≤1.0 mg/L; 107.4 ± 13.0). Distribution of IQ scores was also significantly different (p = 0.003). Every 0.5-mg/L increase in water fluoride (between 3.40 and 3.90 mg/L) was associated with a 4.29 decrease in IQ score (95% CI: −8.09, −0.48).

E.1.18.2. Risk of Bias

- Author contacts:
  - Authors were contacted in September 2018 to obtain additional information for the risk-of-bias evaluation. Additional information provided by the authors

Commented [A119]: Change made in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 5): The NTP response to NASEM's comments indicate they contacted investigators on specific issues, and that some study authors responded while others did not. Recommendation: NTP should provide an assessment of how non-response would affect the grading of these studies. This may be tied to the NASEM concern regarding transparency in the risk-of-bias assessment. [redacted] requests an appendix and language in the Discussion detailing which studies received a change to their risk-of-bias determination.

informed the rating decision for the following risk-of-bias domains: Detection (outcome assessment).

- Population selection:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* School children (2,886), aged 7–13 years, were recruited from the rural areas of Tianjin City, China. After exclusion, 1,636 children were assigned to the "normal-fluoride" exposure group, and 1,250 were assigned to the "high-fluoride" exposure group based on a cut-off water fluoride level of 1.0 mg/L. A multistage random sampling technique, stratified by area, was performed to select representative samples among local children who were permanent residents since birth. Detailed characteristics of the study population were provided. Exclusion criteria included: 1) children who had congenital or acquired diseases affecting intelligence, 2) children with a history of cerebral trauma and neurological disorders, 3) children with a positive screening test history (like hepatitis B virus infection, Treponema palladium infection and Down's syndrome), and 4) children with adverse exposures (smoking and drinking) during maternal pregnancy. A table of characteristics was provided by fluoride level with differences adjusted in the analysis.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were recruited using similar methods during the same time frame and that any differences between the exposed groups were considered in the statistical analyses.
- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Demographic data were collected by trained investigators during a face-to-face interview with the recruited children and their parents. Questionnaires were not stated to be validated. The developmental status of the children was further assessed by calculation of BMI, and all measurements were conducted by nurses based on recommended standard methods. Variables that presented differential distribution between the normal-fluoride and high-fluoride exposure groups were adjusted in the linear regression analysis of IQ data and included age, sex, paternal and maternal education levels, and low birth weight. Children exposed to smoking in utero were excluded from the study. Sensitivity analyses were conducted by modifying covariates adjusted in multivariable models among demographics (age and sex); development (BMI); socioeconomics (maternal education, paternal education, and household income); history of maternal disease during pregnancy (gestational diabetes, malnutrition, and anemia); and delivery conditions (hypoxia, dystocia, premature birth, post-term birth, and low birth weight). None of the study sites selected were in areas endemic for iodine deficiency disorders, nor were other potential neurotoxins like lead, arsenic, and mercury present. Variables such as parental BMI and behavioral and mental health disorders were not addressed.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.

        - *Direction/magnitude of effect size:* Not applicable.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that methods of obtaining the information were valid and reliable and direct evidence that all key covariates and co-exposures were considered.
- Attrition:
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* There were 1,636 children assigned to the "normal-fluoride" exposure group based on water fluoride and 1,250 children assigned to the "high-fluoride" exposure group. Exclusion from the original group of 2,886 children was adequately described. A total of 2,380 children provided urine samples. There is no indication that the data presented excludes any additional children or urine samples, but results do not indicate a sample size for all results.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* According to the annual surveillance data from the CDC, the drinking water sources and water fluoride concentrations in each village had remained at stable levels over the past decade. During the investigation, water samples were collected randomly from the public water supplies in each village. Spot (early-morning) urine samples from every child and water samples from each village were collected in pre-cleaned, labeled polythene tubes and transported to the lab within 24 hours while frozen. Samples were stored at −80°C until analysis. Concentrations of fluoride ions (mg/L) were analyzed using the national standardized ion-selective electrode method in China; the detection limit was 0.01 mg/L. Samples were diluted with an equal volume of total ionic strength adjusted buffer (TISAB) of pH 5–5.5 for optimal analysis. Double-distilled deionized water was used throughout the experiment. There is no reporting of any QC methods.
        - *Direction/magnitude of effect size:* Spot urine samples may lead to non-differential exposure misclassification. The large population size likely dilutes any potential effects of occasional misclassification. Because the drinking water sources of fluoride had been noted to be stable for the past decade and the children were 13 years or younger, there would only be exposure misclassification if there was a lot of migration between areas.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- Outcome:

- *Rating:* Definitely low risk of bias (++)
- *Summary:* IQ scores were measured using the second edition of the Combined Raven's Test–The Rural in China (CRT-RC2) for children aged 7–13 years (++ for methods). The test was completed by each participant within 40 minutes, according to the instruction manual. For each test, 40 children were randomly allocated to one classroom to take the test independently under the supervision of four trained professionals. There is no mention of whether the evaluators were blinded to the fluoride group of each child (normal vs. high fluoride) or whether there were steps taken to ensure consistency in scoring across the evaluators. It is also not clear whether the 40 children randomly assigned to the classroom were specific to the village or whether a local center was used. Correspondence with the study authors indicated that the four professionals worked together throughout the examination without knowledge of the child's fluoride exposure (++ for blinding).
- *Basis for rating:* Definitely low risk of bias based on the direct evidence that the outcome was assessed using instruments that were valid and reliable, and that the outcome assessors were blind to participants' fluoride exposure.

- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study. Univariate and multivariable piecewise linear regression models were used to estimate the associations between water fluoride or urinary fluoride levels and IQ scores. Multiple logistic regression analysis was used to evaluate the association between water or urinary fluoride levels and IQ degree using the normal intelligence group as the control. Sensitivity analyses were conducted. There is no evidence that residual diagnostics were used to examine model assumptions or that the complex sampling design (stratified multistage random sampling) was accounted for in the analysis using sampling weights and adjustment for clustering. The impact of these factors on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for numerous important covariates.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements with blinding at outcome assessment but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key covariates, including potential co-exposures, were considered in the study design or analysis.

E.1.19. Zhang et al. (2015b)

E.1.19.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 10–12 years
- ***Study area:*** Tianjin City, China
- ***Sample size:*** 180 children
- ***Data relevant to the review:*** IQ by control and high fluoride groups; IQ correlations with water, serum, or urinary fluoride levels; betas for IQ with urinary fluoride levels (by genotypes)
- ***Reported association with fluoride exposure:*** Yes: Significant correlation between IQ score and children's serum fluoride ($r = -0.47$) and urinary fluoride ($r = -0.45$); significant difference in mean IQ score for high-fluoride area (defined as >1 mg/L in drinking water; 102.33 ± 13.46) compared with control area (<1 mg/L; 109.42 ± 13.30).

E.1.19.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information because it was not necessary.
- Population selection:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were similar and recruited during the same time frame using the same methods. Authors recruited schoolchildren from a high fluoride area (1.40 mg/L) and a control area (0.63 mg/L) in Tianjin City, China. In accordance with the principles of matching social and natural factors such as educational standard, economic situation, and geological environments as much as possible, two areas with different fluoride concentrations in the groundwater were selected by a stratified cluster random sampling of this region. A total of 180 5th grade children aged 10 to 12 years from two primary schools located 18 km apart in the Jinnan District were recruited—Gegu Second Primary School (from an endemic fluorosis area) and Shuanggang Experimental Primary School (from a non-endemic fluorosis area). The areas are not affected by other drinking water contaminants, such as arsenic or iodine. All subjects were unrelated ethnic Han Chinese and residents in Tianjin with similar physical and mental health status. The authors excluded subjects with known neurological conditions, including pervasive developmental disorders and epilepsy. Descriptive statistics of the study

population are presented by exposure group in Table 1 of the study. A number of potential differences were considered in the statistical analyses.

- *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposure groups were similar and recruited using similar methods during the same time frame.

- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariates included in the statistical models were age, sex, educational levels of parents, drinking water fluoride (mg/L), and levels of thyroid hormones (T3, T4, and TSH). Authors report that the study areas were not affected by other contaminants such as arsenic or iodine, and residents were of similar physical and mental health status. Other important covariates (maternal demographics, smoking, reproductive health) were not considered. Covariate data were obtained from a study questionnaire.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and direct evidence that key covariates, including potential co-exposures, were considered.
- Attrition:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results are complete for the 180 children selected for the study.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- Exposure:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Drinking water samples (10 mL) were collected from the tube wells of each child's household. Three fasting venous blood samples were also collected. Urine samples were collected in the early morning before breakfast. Fluoride content in drinking water (W-F), serum (S-F), and urine (U-F) was measured using an ion analyzer EA940 with a fluoride ion-selective electrode (Shanghai Constant Magnetic Electronic Technology Co, Ltd, China), according to the China standard GB 7484-87. All reference solutions for the fluoride determinations were double-deionized water. Parallel samples were set for determination, and averages were taken. The quantitation limits of this method for W-F, S-F, and U-F were 0.2, 0.012, and 0.5 mg/L, respectively. Recovery rates for this method were in the range of 94.3%–106.4%. The intra- and inter-assay coefficients of variation for fluoride were 2.7% and 6.7%, respectively. Dilution of the urinary fluoride was not addressed.
    - *Direction/magnitude of effect size:* Not applicable.

  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposure was consistently assessed using well-established methods that directly measured exposure.
- Outcome:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* A Combined Raven's Test for Rural China (CRT-RC) was taken to evaluate the IQ of each child (++ for methods). The study report stated that all tests were administered at school by a trained examiner who was masked to participants' drinking water fluoride levels (++ for blinding). Overall rating for methods and blinding = ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All results outlined in the abstract, introduction, and methods sections were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Associations between serum and urinary fluoride levels and IQ score were estimated using general linear models and multivariate linear regression by COMT polymorphism. Normality (Kolmogorov-Smirnov test) was evaluated for all continuous variables. There is no evidence that residual diagnostics were used to examine model assumptions or that the complex sampling design (stratified multistage random sampling) was accounted for in the analysis using sampling weights and adjustment for clustering. The impact of these factors on the regression effect estimates is expected to be minimal given the use of individual-level data and adjustment for numerous covariates.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, blinding of outcome assessor to participants' fluoride exposure, and consideration of key covariates including potential co-exposures.

E.2. Other Neurodevelopmental Studies

E.2.1. Barberio et al. (2017b)

E.2.1.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Canadian Health Measures Survey (cycles 2 and 3) participants (children aged 3–12 years)
- ***Study area:*** general population of Canada
- ***Sample size:*** 2,221 children (1,120 from Cycle 2, 1,101 from Cycle 3)
- ***Data relevant to the review:*** Associations between learning disability or ADHD (Cycle 2 only) assessed by parent or child self-report and urinary fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant increase in adjusted OR for learning disability with unadjusted urinary fluoride per 1-µmol/L increase (1.02; 95% CI: 1.00, 1.03) when Cycles 2 and 3 were combined. No significant associations with creatinine-adjusted or specific gravity-adjusted urinary fluoride. No significant association between urinary fluoride and ADHD.

E.2.1.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information because it was not necessary.
- Population selection:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* The comparison groups were selected from Cycles 2 and 3 of the Canadian Health Measures Survey. This is a nationally representative sample of residents living in 10 provinces, with clear exclusion criteria provided. Exclusion represented only about 4% of the target population (all Canadian residents 3–79 years old living in 10 provinces). A table of characteristics of the study population is provided.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the subjects were recruited from the same population using the same methods during the same time frame, and exposure groups were similar.
- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study adjusted for sex, age (3–12 years old), household education, and household income adequacy. Variables to discern fluoride source, including drinking water and dental products, were also considered. Cycle 2 data also included adjustments for: 1) children for whom tap water (vs. bottled or other) was the primary source of drinking water at home or away from home and 2) children who had lived in their current home for 3 or more years. Covariates such as parental behavioral and mental health disorders, smoking, and nutrition

were not discussed. The study used data from the Canadian Health Measures Survey, which consists of a nationally representative sample of Canadians. Most Canadians (~89%) receive water from municipal water supplies, which monitor for levels of lead and arsenic. Therefore, co-exposure to lead and arsenic are less likely an issue in this population and the lack of information is not considered to appreciably bias the results.

- *Potentially important study-specific covariates:* All key covariates were considered in this study.
  - *Direction/magnitude of effect size:* Not applicable.
- *Basis for rating:* Probably low risk of bias based on direct evidence that key covariates were addressed and indirect evidence that the methods used to collect the information were valid and reliable and that co-exposures were not an issue.

- Attrition:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariate data were missing for less than 5% of all analyses, apart from household income; household income was reported for only 71%–77% of participants and was imputed for the remainder.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Estimates of urinary fluoride (µmol/L) from spot urine were available for a subsample of respondents. Analysis was performed under standardized operating procedures at the Human Toxicology Laboratory of the Institut National de Santé Publique du Québec (accredited under ISO 17025). Fluoride content of urine samples was analyzed using an Orion pH meter with a fluoride ion-selective electrode with limits of detection of 20 µg/L (Cycle 2) and 10 µg/L (Cycle 3). Urinary dilution was addressed by using creatinine-adjusted levels as well as specific gravity-adjusted levels. In Cycle 3 only, estimates of the fluoride concentration of tap water samples collected from randomly selected households were available. The subsample of households selected for tap water sample collection corresponded to the person-level urine fluoride subsample. Analysis of the fluoride concentration of tap water was performed using a basic anion exchange chromatography procedure, with a limit of detection of 0.006 mg/L. QC methods were not addressed.
    - *Direction/magnitude of effect size:* There is not any specific impact on the direction or magnitude of effect size expected. Urinary fluoride levels are reflective of a recent exposure. Having a single concurrent measurement may not be reflective of the exposure associated with the outcome, but if subjects lived in the same area throughout life, the exposure may be an adequate

representation. Although there is possible exposure misclassification, it would likely be non-differential.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- Outcome:
  - *Rating:* Probably high risk of bias (−)
  - *Summary:* The primary outcome variable, diagnosis of a learning disability by a health professional, was based on a single item from a household survey asked to all respondents: "Do you have a learning disability?" Answer options were: "yes," "no," "don't know," or the participant refused to answer. For Cycle 2, those who indicated having a learning disability were also asked what kind, with the answer options of: "ADD," "ADHD," "dyslexia," or "other." This question was omitted in Cycle 3, and the reason for omission was not described. Parents or guardians answered all questions for children aged 3–11 years, while children 12 years and older answered questions themselves. The self-reporting of a learning disability did not appear to have been confirmed by medical records or a health professional (− for methods based on self-report of diagnosis by a health care professional; also, in Cycle 3, no specific disabilities were described). Blinding was not a concern as spot urine samples were sent to a separate lab, and self-reports would not have knowledge of their urine or tap water exposure level (+ for blinding). Overall rating = −.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that the outcome was measured using an insensitive method in the study population.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods sections were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Logistic regression analyses, adjusted and unadjusted for covariates, examined the associations between fluoride exposure and diagnosis of learning disability. Analyses were performed for Cycle 2 only (urinary fluoride and type of learning disability diagnosis), Cycle 3 only (urinary fluoride, water fluoride, and learning disability diagnosis), and Cycles 2 and 3 combined. Analyses used survey weights and bootstrapped weights to ensure proper computation of variance estimates. Results are reported as unadjusted and adjusted ORs with 95% CIs.
    - *Other potential concerns:* None identified.

- *Basis for rating:* Definitely low risk of bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include individual exposure measurements and the consideration of key covariates, but was limited by the cross-sectional study design and insensitive outcome measures.

E.2.2. Bashash et al. (2017)

E.2.2.1. Study Details

- ***Study design:*** Prospective cohort
- ***Population:*** Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants (pregnant mothers and their children aged 4 or 6–12 years).
- ***Study area:*** Mexico City, Mexico
- ***Sample size:*** 299 mother-child pairs, of whom 287 had data for the general cognitive index (GCI).
- ***Data relevant to the review:*** Adjusted and unadjusted associations between GCI and maternal or child's urinary fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant association between maternal urinary fluoride and GCI score (adjusted β per 0.5 mg/L increase = −3.15; 95% CI: −5.42, −0.87). No significant associations with children's urinary fluoride.

E.2.2.2. Risk of Bias

- Author contacts:
  - Authors were contacted for additional information on whether clustering was addressed. The authors provided results from additional models with cohort as a random effect, which informed the rating decision for the following risk-of-bias domains: Other.~~Authors were not contacted for additional information because it was not necessary.~~
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Study participants were selected from two different cohorts from three hospitals in Mexico City that serve low-to-moderate income populations. One cohort was from an observational study of prenatal lead exposure and neurodevelopmental outcomes, and the other was from a randomized trial of the effect of calcium on maternal blood lead levels. The authors state that participants had no history of psychiatric disorders, high-risk pregnancies, gestational diabetes, illegal drug use, or continuous prescription drugs, but information on smoking habits was not included. Study populations appear to be similar, but there may be some differences because subjects were selected from two different cohorts that were recruited during slightly different time periods.

Commented [A120]: Change made in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 5): The NTP response to NASEM's comments indicate they contacted investigators on specific issues, and that some study authors responded while others did not. Recommendation: NTP should provide an assessment of how non-response would affect the grading of these studies. This may be tied to the NASEM concern regarding transparency in the risk-of-bias assessment. [redacted] requests an appendix and language in the Discussion detailing which studies received a change to their risk-of-bias determination.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar despite the subjects coming from different original study populations for whom different methods were used for recruitment.

- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Data were collected via questionnaire on maternal age, education, marital status at first prenatal visit, birth order, birth weight, gestational age at delivery, maternal smoking, maternal IQ, and HOME scores. All models were adjusted for gestational age at birth, sex, birth weight, birth order, age at testing, maternal marital status, smoking history, maternal age at delivery, maternal IQ, education, and cohort, with additional testing for children's urinary fluoride, mercury, lead, and calcium. Sensitivity analyses were additionally adjusted for HOME score. Covariates not considered included BMI, iodine deficiency, arsenic, and maternal mental health and nutrition. Arsenic is assumed not to be a potential co-exposure in this population as the study authors did not discuss it as an issue but did discuss other co-exposures. Arsenic is included in the water quality control program in Mexico City and is not considered a concern in this population.
  - *Potentially important study-specific covariates:* All key covariates were addressed.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Probably low risk of bias based on direct evidence that key covariates, including other potential co-exposures, were considered, and indirect evidence that the methods used to collect the information were valid and reliable and that arsenic was not likely to be an issue in this study population.
- Attrition:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although there was a large amount of attrition, the study authors clearly describe all reasons for attrition and also provide characteristics to compare those participants included to those excluded. There were some slight differences between those included and those excluded, but there is nothing to indicate that the attrition would potentially bias the results.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- Exposure:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Urinary fluoride concentrations were determined in spot urine samples (2nd morning void) collected from mothers (during at least one trimester) and children ages 6–12 years. Fluoride content was measured using ion-selective electrode-based assays. QC methods were described including between laboratory

correlations. All samples were measured in duplicate. Extreme outliers were excluded. Urinary dilution was addressed by using creatinine-adjusted levels.

- *Direction/magnitude of effect size:* Not applicable.

- *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- Outcome:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Outcome was assessed using the McCarthy Scales of Children's Abilities (MSCA) in 4-year-old children (translated into Spanish) and the Wechsler Abbreviated Scale of Intelligence (WASI) in 6–12-year-olds. The WASI is a well-established test, and the validity of both tests is well documented by the authors. Inter-examiner reliability was evaluated and reported with a correlation of 0.99 (++ for methods). The study report stated that psychologists were blind to the children's fluoride exposure (++ for blinding). Overall rating for methods and blinding = ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study. Statistical tests of bivariate associations (using Chi-square tests for categorical variables and analysis of variance [ANOVA]) were used to compare the means of the outcomes or exposures within groups based on the distribution of each covariate. Generalized additive models (GAMs) were used to estimate the adjusted association between fluoride exposure and measures of children's intelligence. Residual diagnostics were used to examine model assumptions and identify any potentially influential observations. Results are reported as adjusted effects and 95% CIs. In sensitivity analyses, regression models accounted for clustering at the cohort level by using cohort as a fixed effect in the models. Although using cohort as a random effect would be more appropriate, using individual-level exposure data and accounting for numerous important covariates in the models likely captured the cohort effect.

Commented [A121]: Change made in response to the [redacted] Reviewer comment below; see DocA1_Monograph for detailed response:

Reviewer comment (DocA1_Monograph, page 5): The NTP response to NASEM's comments indicate they contacted investigators on specific issues, and that some study authors responded while others did not. Recommendation: NTP should provide an assessment of how non-response would affect the grading of these studies. This may be tied to the NASEM concern regarding transparency in the risk-of-bias assessment. [redacted] requests an appendix and language in the Discussion detailing which studies received a change to their risk-of-bias determination.

Additional models with cohort as a random effect were also subsequently made available via personal communication with the study authors and showed similar results to the main model.

- *Other potential concerns:* None identified.

- *Basis for rating:* Definitely low risk of bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, blinding of outcome assessor to participants' fluoride exposure, and the prospective cohort study design.

E.2.3. Bashash et al. (2018)

E.2.3.1. Study Details

- ***Study design:*** Prospective cohort
- ***Population:*** ELEMENT participants (pregnant mothers and their children aged 6–12 years)
- ***Study area:*** Mexico City, Mexico
- ***Sample size:*** 210 mother-child pairs
- ***Data relevant to the review:*** Associations between ADHD and other attention/impulsivity scores and maternal urinary fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant associations between maternal urinary fluoride (per 0.5-mg/L increase) and Conners' Rating Scales-Revised (CRS-R) scores, including Cognitive Problems and Inattention Index (adjusted $\beta = 2.54$; 95% CI: 0.44, 4.63), DSM-IV Inattention Index (adjusted $\beta = 2.84$; 95% CI: 0.84, 4.84), DSM-IV ADHD Total Index (adjusted $\beta = 2.38$; 95% CI: 0.42, 4.34), and ADHD Index (adjusted $\beta = 2.47$; 95% CI: 0.43, 4.50).

E.2.3.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information because it was not necessary.
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Participants were a subset of mother-child dyads enrolled in various longitudinal birth cohort studies of the Early Life Exposure in Mexico to Environmental Toxicants (ELEMENT) project. Subjects were included from two of the four cohorts for which maternal urinary samples were available. Participants in cohort 2A were recruited between 1997 and 1999, and participants in cohort 3 were recruited from 2001 to 2003. Inclusion and exclusion criteria were applied consistently across the two cohorts. A table of subject characteristics was provided in the study, and any differences were considered in the analysis.

Study populations appear to be similar, but there may be some differences because subjects were selected from two different cohorts: one from an observational study on prenatal lead exposure and the other from a randomized trial on the effects of calcium on blood lead levels. In addition, they were recruited from slightly different time periods.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposed groups were similar, and any differences were considered in the analysis.

- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Questionnaires were used to collect information on maternal age, maternal education, history of smoking, and marital status during the first pregnancy visit. Child information at birth included birth weight, sex, birth order, and gestational age as calculated by the nurse. Mothers also responded to an SES questionnaire during the visit when the psychometric tests were administered. The Home Observation for Measurement of the Environment (HOME) score was evaluated in a subset of participants. Covariates were selected a priori. Models were adjusted for maternal age at delivery, years of education, marital status, smoking history, gestational age at birth, age at outcome assessment, sex, birth order, SES, cohort, and calcium intervention. Arsenic is included in the water quality control program in Mexico City and is not considered a concern in this population.
  - *Potentially important study-specific covariates:* None identified, although this study did not specifically address arsenic or other co-exposures. Bashash et al. (2017) addressed potential co-exposure to lead and mercury but did not address arsenic. Arsenic was potentially addressed as part of the water quality program in Mexico City.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Probably low risk of bias based on direct evidence that key covariates were addressed, and indirect evidence that the methods used to collect the information were valid and reliable and that arsenic and other potential co-exposures were not likely to be an issue in this study population.
- Attrition:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although there was a large amount of attrition from the original cohorts, it was unlikely related to outcome or exposure, and there were very little missing data from those included in the study. Of the 231 mothers with a minimum of one maternal urine fluoride measurement and matching outcome identified for the project, only 17 were excluded based on incomplete demographic and outcome information.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- Exposure:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Mothers provided at least one spot urine sample during pregnancy. As described in Bashash et al. (2017), urinary concentrations were determined on second morning void. Fluoride content was measured using ion-selective electrode-based assay. Bashash et al. (2017) describes QC methods. All samples were measured in duplicate, and extreme outliers were excluded. Urinary dilution was addressed by using creatinine-adjusted levels.
    - *Direction/magnitude of effect:* N/A
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- Outcome:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Behaviors associated with ADHD were assessed using the Spanish version of Conners' Rating Scales-Revised, which has been validated for the evaluation of ADHD. Mothers completed the CRS-R at the same follow-up visit in which the child completed the CPT-II tests. All tests were applied under the supervision of an experienced psychologist (++ for methods). Use of only parent reports and not teacher reports was noted by the authors as a study limitation because there is considerable variation between the two sources in terms of identifying ADHD-associated behaviors. Blinding was not reported, but it is unlikely that the mothers were aware of their urinary fluoride levels. Although mothers may have had knowledge that they were receiving fluoride through fluoridated salt or naturally occurring fluoride in their water, they would not have knowledge that this was relevant to the study purpose as the ADHD tests were conducted for the original cohort (as was acknowledged by the study authors in the discussion) (++ for blinding). Overall rating = ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*

- *Statistical analyses:* Bivariate analyses included Chi-square tests for categorical variables and ANOVA for continuous outcomes. Appropriate univariate statistics and transformations were performed before bivariate analyses. Residuals from fully adjusted linear regressions were checked and suggested skewness. Gamma regression with an identity link was used to examine the adjusted association between prenatal fluoride and each neurobehavioral outcome (instead of using log transformation). Generalized additive models were used to visually examine potential non-linearity. Sensitivity analyses examined impact of other covariates. Diagnostics tests were used to assess violations of the model assumptions and to identify remaining influential observations. The Benjamini-Hochberg false discovery rate (FDR) procedure was used to correct for multiple testing.
- *Other potential concerns:* None identified.

o *Basis for rating:* Definitely low risk of bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, blinding of outcome assessor to participants' fluoride exposure, and the prospective cohort study design.

E.2.4. Choi et al. (2015)

E.2.4.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** First-grade children (ages 6–8 years)
- ***Study area:*** Mianning County in southern Sichuan, China
- ***Sample size:*** 51 first-grade children
- ***Data relevant to the review:*** Associations between learning, memory, visual motor ability, motor ability, and manual dexterity with continuous urine or drinking water fluoride levels. Study also had information based on dental fluorosis score.
- ***Reported association with fluoride exposure:*** No: None of the outcomes were significantly associated with fluoride exposure.

E.2.4.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information because it was not necessary.
- Population selection:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were selected during the same time frame using the same methods. Fifty-one first-grade children residing in Mianning County in southern Sichuan, China were included in this pilot study. It is not specified whether the 51

children represented all first-grade children from this area or whether some refused to participate. Children who did not speak Chinese, were not students at the Primary School of Sunshui Village in Mianning County, or those with chronic or acute disease that might affect neurobehavioral function tests were excluded. Demographic characteristics are presented in Table 1 of the study, which indicates that subjects were similar. Covariates were adjusted for in the statistical analyses.

- *Basis for Rating:* Definitely low risk of bias based on direct evidence that the exposure groups were similar and were recruited within the same time frame using the same methods with no evidence of differences in participation/response rates.

- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The parents or guardians completed a questionnaire on demographic and personal characteristics of the children (sex, age at testing, parity, illnesses before age 3, and past medical history) and caretakers (age, parity, education and occupational histories, residential history, and household income). A 20-µL capillary blood sample was collected at the school by a Mianning County Center for Disease Control (CDC) health practitioner and tested for possible iron deficiency, which could be used as a covariate of neurodevelopmental performance. Covariates that were not assessed include maternal BMI, parental mental health, maternal smoking status, maternal reproductive factors, parental IQ, and HOME score. However, the study authors noted that confounding bias appeared to be limited due to the minimal diversity in the social characteristics of the subjects. The study authors indicated that CDC records documented that levels of other contaminants, including arsenic and lead, were very low in the area. Iodine differences were not specifically addressed, but there is no indication from the information provided that this might have been a concern.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Probably low risk of bias because there is direct evidence that the key covariates were considered and indirect evidence that co-exposure to arsenic was likely not an issue in this area and that methods used for collecting the information were valid and reliable.
- Attrition:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The majority of results were reported for the 51 children stated to be included in the pilot study. In Table 5 of the study, the N for each dental fluorosis category totals only 43, but the text indicates 8 children did not have a Dean Index because permanent teeth had not erupted.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were

documented when subjects were removed from the study or excluded from analyses.

- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study used three different measurements of fluoride exposure: well water fluoride concentrations from the residence during pregnancy and onwards, fluoride concentrations from children's first morning urine samples, and degree of children's dental fluorosis. Fluoride concentrations in community well water were measured and recorded by Mianning County CDC; specific methods were not reported, but standard methods were likely used because analyses were conducted by the CDC and were likely the same as those used to measure the fluoride in urine. Migration of subjects was noted to be limited. Well water fluoride concentrations of the mother's residence during pregnancy and onward were used to characterize a child's lifetime exposure. To provide a measure of the accumulated body burden, each child was given a 330-mL (11.2-oz) bottle of Robust© distilled water (free from fluoride and other contaminants) to drink the night before the clinical examinations, after emptying the bladder and before bedtime. The first urine sample was collected at home the following morning, and the fluoride concentration was determined on a 5-mL sample using an ion-specific electrode at the Mianning CDC. There is no indication that urinary fluoride levels accounted for dilution, nor was it clear that the method of administering water to the children and collection methods sufficiently controlled for differences in dilution. One of the investigators, a dentist, performed a blinded dental examination on each child's permanent teeth to rate the degree of dental fluorosis using the Dean Index. The Dean Index is commonly used in epidemiological studies and remains the gold standard in the dentistry armamentarium. The Index has the following classifications: normal, questionable, very mild, mild, moderate, and severe. Quality control (QC) procedures are not reported but were likely appropriate.
    - *Direction/magnitude of effect size:* Current levels were used to assess lifetime exposure. This is likely to be a non-differential exposure misclassification, and direction of bias is unknown. Because subject migration appears to be limited, it is likely that the current fluoride levels are adequate reflections of past exposure. Dental fluorosis would be an indicator that exposure occurred in the past, and there was a fair correlation between degree of dental fluorosis and current urine and water fluoride levels, with both increasing with increasing levels of dental fluorosis.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.
- Outcome:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study authors adopted culture-independent tests considered feasible for children aged 6 to 8 years. The Wide Range Assessment of Memory

and Learning (WRAML) was used for the assessment of memory and learning. Three subtests were also used. The Finger Windows subtest assesses sequential visual memory. The Design Memory subtest assesses the ability to reproduce designs from memory following a brief exposure. The Visual Learning subtest assesses the ability to learn the locations of pictured objects over repeated exposures. The Wechsler Intelligence Scale for Children-Revised (WISC-IV) included digit span for auditory span and working memory and block design for visual organization and reasoning. The grooved pegboard test assesses manual dexterity. The tests used have been validated on a Western population. Although there is no information provided to indicate that the tests were validated on the study population, the study authors indicated that the tests were culture-independent (+ for methods). Blinding of the outcome assessors or steps to minimize potential bias was not reported. However, it is unlikely that the assessors had knowledge of the individual exposure as children all came from the same area, and water and urine levels were tested at the CDC (+ for blinding). Overall = +.

  - *Basis for rating:* Probably low risk of bias based on indirect evidence that all outcomes were assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.

- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses were appropriate. Multiple regression models evaluated the associations between exposure indicators and test scores after adjusting for covariates. Specific regression models are not described or referenced, just stated to be "standard regression analysis with confounder adjustment." The distributions of fluoride concentrations in urine and water were skewed and were log10-transformed to approximate a Gaussian distribution (test not specified). Results were reported as adjusted effects and 95% Cis. There was no evidence that residual diagnostics were used to examine model assumptions; however, the impact on the effect estimates is expected to be minimal.
    - *Other potential concerns:* It should be noted that this study was a pilot study and, therefore, had a relatively small sample size (i.e., 51 children).

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in the confounding, exposure, and outcome domains. Study strengths include individual fluoride measurements with blinding at outcome assessment likely. All key covariates and many other covariates were considered in the study design or analysis.

E.2.5. Li et al. (2004) [translated in Li et al. 2008a]

E.2.5.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Full-term, normal neonates 24–72 hours old from healthy mothers
- ***Study area:*** Zhaozhou County, Heilongiang Province, China
- ***Sample size:*** 91 neonates (46 males and 45 females)
- ***Data relevant to the review:*** Comparison of neurobehavioral capacity between children in the high-fluoride area compared to the control area.
- ***Reported association with fluoride exposure:*** Yes: Significant differences in neurobehavioral assessment total scores between high-fluoride (36.48 ± 1.09) and control (38.28 ± 1.10) groups; significant differences in total neurobehavioral capacity scores as measured by non-biological visual orientation reaction and biological visual and auditory orientation reaction between the two groups (11.34 ± 0.56 in controls compared to 10.05 ± 0.94 in high-fluoride group).

E.2.5.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information because it was not necessary.
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There is indirect evidence that the exposure groups were similar. Participants were recruited during the same time frame using the same methods. From 2002 to 2003, 273 neonates were born in a hospital in Zhaozhou County, China. Ninety-one of 273 full-term neonates (46 males, 45 females) were randomly selected. Mothers ranged in age from 20 to 31 years, met multiple health criteria, and had not changed residence during pregnancy. Authors report that the two study groups were located in the same area with similar climate, living habits, economic and nutritional conditions, and cultural backgrounds, but do not provide these data in the manuscript. There is no statistically significant difference in the mode of delivery, birth weight, infant length, or sex. Subjects were separated into exposure groups after random selection.
  - *Basis for Rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited within the same time frame

using the same methods with no evidence of differences in participation/response rates.

- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* No covariates were specifically considered in the analysis. The study authors note similarities in characteristics in the two populations (i.e., living habits, economic and nutritional conditions, and cultural backgrounds) but do not provide these data nor do they indicate which specific characteristics were considered. There were no significant differences in infant sex, birth method, gestational age, or infant weight and length. All tests were conducted when children were 1–3 days old. No potential co-exposures were discussed. Although arsenic is considered a potential issue in China, water quality maps indicate that there is a 25%–50% probability that the drinking water in that area exceeds the WHO guideline for arsenic of 10 µg/L.
  - *Potentially important study-specific covariates:* Key covariates, including age, sex, and measures of socioeconomic status (SES), were similar between exposure groups; however, arsenic was not considered. Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on water quality maps, arsenic does not appear to be an issue in Zhaozhou County of the Heilongjiang Province. Iodine deficiencies are not mentioned.
    - *Direction/magnitude of effect size:* Conceptually, the presence of arsenic would potentially bias the association away from the null if it were present with fluoride. Deficiencies in iodine would potentially bias the association away from the null if it were present in areas of higher fluoride but toward the null if it were present in areas of lower fluoride. Neither of these are considered a concern in this study for reasons detailed above.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the key covariates were considered, co-exposure to arsenic was likely not an issue in this area, and methods used for collecting the information were valid and reliable.
- Attrition:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Although authors did not discuss why only 91 of the 273 neonates available were randomly selected, results were available for all 91 subjects.
  - *Basis for rating:* Definitely low risk of bias based on results being available for all subjects.
- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were split into control and high-fluoride groups based on fluoride levels in their places of residence. Although the levels were provided (1.7–6.0 mg/L for the high-fluoride group compared to 0.5–1.0 mg/L for the control group), it was not reported how or when these levels were measured. Urine was collected when women were hospitalized but before labor began. Urine samples were sent to a specific lab for measurement using fluoride ion-selective

electrode. It was noted that this procedure strictly followed the internal controls of the laboratory, indicating quality control. Level of detection (LOD) was not provided. Urinary fluoride levels were significantly higher in the high-fluoride mothers (3.58 ± 1.47 mg/L) compared to the control-group mothers (1.74 ± 0.96 mg/L). There was indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure. Although results were mainly based on exposure area, they were supported by urine data, making exposure misclassification less of a concern.

  - *Direction/magnitude of effect size:* There is high variability in both water fluoride and urine fluoride in the subjects from the high-exposure area. Although there is no overlap in the water fluoride levels in the exposure areas, there is some overlap in the urine concentrations in the mothers from the two areas. This may reflect the single measurement and pose no specific bias, or it could indicate that some mothers in the high-fluoride area have lower fluoride exposure, which could bias the association toward the null.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.

- Outcome:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* A standard neonatal behavioral neurological assessment method was carried out by professionals in the pediatric department working in a neonatal section trained specifically for these programs and passing the training exams (+ for methods). The examinations were carried out 1 to 3 days after delivery. Because urine samples were collected on the day of delivery and sent to a separate laboratory, it is likely that the outcome assessors were blind. Although the subjects were separated by fluoride exposure area, it is not likely that the professionals were aware of the exposure as the tests were conducted in the hospital (+ for blinding).
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study authors reported numerous outcomes in sufficient detail; however, because a list of outcomes tested was not provided, there is no direct evidence that all were reported.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that all the study's measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*

- *Statistical analyses:* Statistical analyses are described only as a t-test. Consideration of heterogeneity of variance was not reported. Results are reported as mean and standard deviations of neurological scores. Maternal urinary fluoride levels were used only to compare exposures between exposed and control groups. Infants in the control group were from four villages, and those in the exposed group were from five villages within the same district. Infants were randomly selected before they were assigned to exposed or control groups. In the comparisons, there was no accounting for clustering at the village level. It is likely that the standard error of the difference in mean neurobehavioral assessment scores between the high fluoride group and control group will be biased, making differences appear stronger than they actually are. However, the use of multiple villages per exposure group is likely to mitigate some of the impact of this lack of accounting for clustering, and the overall impact on effect estimates is expected to be minimal.
- *Other potential concerns:* It should be noted that although the study states that subjects were randomly selected, it is unclear why only 91 subjects were included and whether they were randomly selected to obtain equal numbers in the high-fluoride and control groups.

o *Basis for rating:* Probably low risk of bias based on indirect evidence that statistical analyses were appropriate and that there were no other potential threats of risk of bias.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in the confounding, exposure, and outcome domains. Study strengths include individual fluoride measurements to support the differences in the two areas. Tests were noted to be conducted at the hospital, providing indirect evidence that blinding was not a concern during the outcome evaluation. Although there was some potential for bias due to the lack of accounting for arsenic or iodine deficiencies, co-exposure to arsenic was likely not a major concern according to groundwater quality maps.

E.2.6. Riddell et al. (2019)

E.2.6.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Canadian Health Measures Survey (Cycles 2 and 3) participants (children aged 6–17 years)
- ***Study area:*** General population, Canada
- ***Sample size:*** 3,745 children
- ***Data relevant to the review:*** Adjusted odds ratios for ADHD and attention symptoms per 1 unit increase in urinary fluoride by water fluoride in the tap water or community fluoridation status.
- ***Reported association with fluoride exposure:*** Yes: Significantly increased odds of ADHD diagnosis (adjusted OR = 6.10; 95% CI: 1.60, 22.8) or hyperactivity/inattentive symptoms (adjusted β = 0.31; 95% CI: 0.04, 0.58) per 1-mg/L increase in tap water fluoride. In addition, a significant association between

ADHD diagnosis (adjusted OR = 1.21; 95% CI: 1.03, 1.42) or hyperactivity/inattentive symptoms (adjusted β = 0.11; 95% CI: 0.02, 0.58) and community water fluoridation status. No significant associations with urinary fluoride levels.

E.2.6.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information because it was not necessary.
- Population selection:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were part of Cycles 2 and 3 of the Canadian Health Measures Survey. This is a nationally representative sample of residents living in 10 provinces. Specific inclusion criteria were provided. This study was restricted to children 6–17 years of age with different fluoride measurements that consisted of three participant samples. One of the samples was available only in Cycle 3.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were similar and were recruited with the same methods during the same time frame.
- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariates included in all models included age at testing, sex, ethnicity, BMI, parents' education, total household income, exposure to cigarette smoke inside the home, and log-transformed concurrent blood lead levels. Covariates such as parental behavioral and mental health disorders, quantity and quality of caregiving environment, and co-exposure to arsenic were not discussed. The study used data from the Canadian Health Measures Survey, which consists of a nationally representative sample of Canadians. Most Canadians (~89%) receive water from municipal water supplies, which monitor for levels of arsenic. Therefore, co-exposure to arsenic is not likely an issue in this population. Rationale for selection of covariates was based on relationship to ADHD diagnosis and to fluoride metabolism based on literature review and consultation with an ADHD expert. There is no information of the source of data for covariates, but it is likely the questionnaires from the Canadian Health Measures Survey, which are considered standardized and validated.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key covariates were considered, co-exposure to arsenic was likely not an issue, and methods used for collecting the information were valid and reliable.
- Attrition:
  - *Rating:* Probably low risk of bias (+)

- *Summary:* There is no information indicating that there were any data excluded due to missing covariates. All exclusions of children were described and reasonable (i.e., drinking bottled water when considering city fluoridation as a measure of fluoride exposure). Outliers were stated to be excluded, but methods for determining this were provided, and it was noted that the outliers were 0.27% of the values.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* **Urinary Fluoride:** Spot urine samples were collected under normal non-fasting conditions and analyzed using an Orion pH meter with a fluoride ion-selective electrode after being diluted with an ionic adjustment buffer. Analysis was performed at the Human Toxicology Laboratory of the Institut National de Santé Publique du Québec. The precision and accuracy of the fluoride analyses, including quality control and quality assurance, were described by Health Canada (2015). The limits of detection were 20 µg/L for Cycle 2 and 10 µg/L for Cycle 3 with no values below detection. Fluoride levels were adjusted for specific gravity.

    Water Fluoride in Tap Water: Tap water was collected at the subjects' homes in Cycle 3 only. Samples were analyzed for fluoride concentrations using anion exchange chromatography procedure with an LOD of 0.006 mg/L. Values below the LOD were imputed with LOD/square root(2). Of the 980 samples, 150 (15%) were below detection.

    Chlorinated Water Fluoride Status: This was determined by viewing reports on each city's website or contacting the water treatment plant (provided in supplemental material). Children were excluded if they drank bottled water, had a well, had a home filtration system, lived in the current residence for 2 years or less, or lived in an area with mixed city fluoridation.

    - *Direction/magnitude of effect size:* There is not any specific impact on the direction or magnitude of effect size expected. Urinary fluoride levels are reflective of a recent exposure, but the study authors adjusted to account for dilution. The possibility of exposure misclassification would be similar in all subjects and would be non-differential. There is less potential for exposure misclassification due to tap water or chlorinated water fluoride status, since children who drank bottled water were excluded and children who had a home filtration system were excluded from the chlorinated water status.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- Outcome:
  - *Rating:* Probably high risk of bias (−)

- *Summary:*

  Strengths and Difficulties Questionnaire (SDQ): The questionnaire was administered to youths under 18 years. Children aged 6–11 years had SDQ ratings provided by parents and guardians, but youths aged 12–17 years completed the questionnaire themselves. Tests consist of 25 items with a 3-point scale. Items were divided into five subscales: emotional problems, conduct problems, hyperactivity-inattention, peer problems, and prosocial behavior. The current study used only the hyperactivity-inattention subscale. Validation of this method was not reported (− for methods).

  ADHD: Ninety percent of youths with ADHD are diagnosed after age 6. For children aged 6–11 years, ADHD diagnosis was provided by parents, but youths aged 12–17 years completed the questionnaire themselves. Cycle 2 asked "Do you have a learning disability?"; if the subject answered "yes," he/she was asked to specify the type (four options were available and described). In Cycle 3, parents were asked directly whether they had ADHD, and children 12 years and older were asked whether they had a physician diagnosis of ADHD and, if so, what subtype (− for methods because different methods were used, and only the children 12 years and older in Cycle 3 were asked specifically about a doctor's diagnosis). Both were measured in both cycles. Blinding is likely not an issue as subjects would not have knowledge of the urine or tap water fluoride levels. However, they would likely have knowledge of the city.

- *Basis for rating:* Probably high risk of bias based on indirect evidence that the outcome was assessed using insensitive methods that varied based subject age.

- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods sections were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Robust logistic regression was used to examine the association between fluoride exposure and ADHD diagnosis, adjusting for covariates. Box-Tidewell tests were used to check the linearity of the relationship with the continuous predictors. Linear regression was used for the SDQ scores using Huber-White standard errors. Multicollinearity was evaluated using variance inflation factor (VIF) statistics. Outliers with high studentized residuals, high leverage, or large Cook's distance values were removed from all analyses with urinary fluoride. All regressions were tested for interactions between fluoride exposure and age and between fluoride exposure and sex. Sensitivity analyses were conducted to test the different

survey cycles. There is no mention of adjustment for the complex survey design using survey weights or bootstrapped weights to ensure appropriate calculation of the estimated variances; however, the overall impact on effect estimates is expected to be minimal.

- *Other potential concerns:* None identified.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include individual exposure measurements and the addressing of key covariates, but was limited by the cross-sectional study design and insensitive outcome measures.

E.2.7. Rocha-Amador et al. (2009)

E.2.7.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 6–11 years
- ***Study area:*** Durango, Mexico
- ***Sample size:*** 80 children
- ***Data relevant to the review:*** Associations between visuospatial organization and visual memory (using the Rey-Osterrieth Complex Figure Test, children's version) and urinary fluoride levels in the children.
- ***Reported association with fluoride exposure:*** Yes: Significant correlation between urinary fluoride and visuospatial organization ($r = -0.29$) and visual memory ($r = -0.27$) scores. No significant correlations with arsenic.

E.2.7.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information because it was not necessary.
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were from the same population and were recruited during the same time frame using the same methods. Although this study compared three sites with antecedents of environmental pollution to mixtures of either F-As, Pb-As, or DDT-PCBs, authors evaluated each contaminant separately. The only area of interest with F and As contamination is in Durango state (5 de Febrero) where drinking water is polluted naturally with F and As at levels exceeding 6 and 19 times, respectively, the World Health Organization (WHO) limits (WHO 2008). Children attending public schools were screened through personal interviews for study eligibility. Inclusion criteria were children between 6 and 11 years old, living in the study area since birth, whose parents signed the agreement to

participate. Children with a neurological disease diagnosed by a physician and reported by the mother were excluded from the study. The final sample for the F-As group was 80. Participation rates were not reported. Selected demographic characteristics are presented in Table 1 of the study.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the populations were similar and recruited during the same time frame using the same methods.

- Confounding:
  - *Rating:* Probably high risk of bias (−)
  - *Summary:* Covariates included blood lead (PbB), age, sex, and height-for-age z-scores; only age had significant associations and was included in the final analysis. Arsenic was also assessed and analyzed separately from fluoride. Arsenic in urine was analyzed by atomic absorption spectrophotometer coupled to a hydride system (Perkin-Elmer model AAnalyst 100). Although the model did not adjust for arsenic, arsenic in the F-As group was not associated with either outcome; therefore, arsenic co-exposure is not considered a major concern in this study. PbB was analyzed with a Perkin-Elmer 3110 atomic absorption spectrophotometer using a graphite furnace. Authors note that the mean blood lead level in the F-As study area was 5.2 µg/dL, and 8% of the children had values above the reference value of 10 µg/dL. PbB was stated not to affect results and was not included in the final analysis. Other covariate data were obtained during the study interview. Father's education was provided and, in the F-As group, was stated to range from 0–16 years, but this was not considered. Maternal education, smoking, and SES were also not considered. The authors provide an SES score of 5.9 ± 1.4 for the 5 de Febrero region (the fluoride region). It is not clear whether this would vary by fluoride or arsenic levels.
  - *Potentially important study-specific covariates:* SES.
    - *Direction/magnitude of effect size:* There are insufficient data to determine the impact on the magnitude or direction of effect size. The impact on the direction of the association would likely depend on the association between fluoride exposure and SES.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that the SES was not considered in the study design or analysis and may have varied by fluoride levels.
- Attrition:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are complete. All 80 participants stated to be the final sample for the site of interest (F-As) were included in all analyses.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- Exposure:
  - *Rating:* Probably low risk of bias (+)

- *Summary:* Fluoride in urine (FU) was analyzed according to method 8308 (''fluoride in urine'') from the National Institute for Occupational Safety and Health (NIOSH 1984) with a sensitive specific ion electrode. As a quality control check, reference standard ''fluoride in freeze dried urine'' (NIST SRM 2671a) was analyzed. The accuracy was 97.0% ± 6.0%. Levels of FU and AsU were adjusted for urinary creatinine, which was analyzed by a colorimetric method (Bayer Diagnostic Kit, Sera-Pak1 Plus). However, details on the collection methods were not reported.
  - *Direction/magnitude of effect size:* Spot urine samples in a small sample size (i.e., 80 children) may have some exposure misclassification. Adjusting for dilution reduces the potential for misclassification based on differences in dilution. Exposure misclassification would likely be non-differential.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- Outcome:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* IQ was assessed through the Rey-Osterrieth Complex Figure Test (ROCF). This is a less well-established method, although the authors provide citations suggesting it has been validated and standardized for the Mexican population (+ for methods). According to the study report, the neuropsychologist who administered the test was blinded to all exposure types and levels (++ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used log-transformed exposure variables (although rationale was not provided). Crude and partial correlations were calculated to evaluate associations between serum fluoride levels and TOCF scores. There is no other description of the regression model, and regression diagnostics to evaluate model assumptions are not presented; however, the overall impact on effect estimates is expected to be minimal.

- *Other potential concerns:* None identified.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements and blinding of outcome assessor to participants' fluoride exposure, but it is limited by the cross-sectional study design, lack of consideration of SES in the study population, co-exposure with arsenic, and use of spot samples in a small population.

E.2.8. Valdez Jimenez et al. (2017)

E.2.8.1. Study Details

- ***Study design:*** Prospective cohort
- ***Population:*** Infants aged 3–15 months
- ***Study area:*** Durango City and Lagos de Moreno, Jalisco, Mexico
- Sample size: 65 infants
- ***Data relevant to the review:*** The Bayley Scales of Infant Development II was used to assess Mental Development Index scale and the Psychomotor Development Index scale in children aged 3 to 15 months and evaluated for associations with first and second trimester maternal urine fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant association between log10-mg/L maternal urinary fluoride and MDI score during first trimester (adjusted $\beta = -19.05$; SE = 8.9) and second trimester (adjusted $\beta = -19.34$; SE = 7.46). No association between maternal fluoride during any trimester and Psychomotor Developmental Index (PDI).

E.2.8.2. Risk of Bias

- Author contacts:
  - Authors were not contacted for additional information because it was not necessary.
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were recruited from two endemic areas in Mexico. The study authors do not provide information on the similarities or differences between the two areas, nor do they indicate whether there were different participation rates. However, recruitment methods were the same. Women receiving prenatal care in health centers located in Durango City and Lagos de Moreno, Jalisco, Mexico were recruited in 2013–2014. Participation rates are not likely to be an issue as characteristics were similar between those who participated and those who did not. Although the authors did not provide characteristics by area, the characteristics provided do not indicate any differences that may be biased by the

selection. Considering the age range for the non-participants, the mean age for non-participants appears to be incorrect (or the age range is incorrect); however, there does not appear to be a difference that would potentially indicate selection bias.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited with the same methods in the same time frame, with no evidence of differences or issues with participation/response rates.

- Confounding:
  - *Rating:* Probably high risk of bias (−)
  - *Summary:* Questionnaires were used to obtain information about sociodemographic factors, prenatal history, mother's health status before pregnancy (e.g., use of drugs, vaccines, diseases), and the type of water for drinking and cooking. The marginalization index (MI) was obtained from the National Population Council (CONAPO). Two additional surveys were conducted during the second and third trimester of pregnancy to get information about the mother's health, pregnancy evolution, and sources of water consumption. A survey was also conducted to get information about childbirth (type of birth, week of birth, weight and length of the baby at birth, Apgar score and health conditions of the baby during the first month of life). This information was corroborated with the birth certificate. Linear regression models included gestational age, children's age, marginality index, and type of drinking water. Bivariate analyses were conducted on the other factors, including sex, prior to conducting multivariable regression models. Some important covariates were not considered, including parental mental health, IQ, smoking, and potential co-exposures. Water quality maps indicate a potential for arsenic to be present in the study area.
  - *Potentially important study-specific covariates:* Arsenic is a potential co-exposure in this area of Mexico.
    - *Direction/magnitude of effect size:* If arsenic were present as a co-exposure, it would likely bias the association away from the null.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that there is a potential for co-exposure with arsenic that was not addressed.
- Attrition:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Out of the 90 women selected for inclusion in the study, 65 approved the participation of their infants. The authors provide a table of characteristics between women who consented to their children's cognitive evaluation and those who participated only in biological monitoring. There were no significant differences between the groups. There were fewer women who provided urine during the second and third trimesters. All specified children are included in the relevant analyses.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were

documented when subjects were removed from the study or excluded from analyses.

- Exposure:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Fluoride exposure was assessed through morning urine samples and water fluoride levels collected from the children's homes. Sampling methodology was appropriately documented, and water levels were quantified through specific ion-sensitive electrode assays. QC was described, and accuracy was >90%. Urinary fluoride was corrected by specific gravity.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- Outcome:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Neurodevelopment was assessed with the Bayley Scales of Infant Development II (BSDI-II) that was noted to be reliable and valid for evaluating children from 3 months to 5 years of age. The average age of children assessed was 8 months, with a range of 3–15 months) (++ for methods). The study report stated that a trained psychologist who was blinded about the mother's fluoride exposure evaluated the infants at home (++ for blinding). Overall rating for methods and blinding = ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.
- Selective Reporting:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported. Table 4 of the study displays only data for trimesters 1 and 2. Although third trimester data were collected, they were not reported, likely because they were available for only 29 subjects. No discussion of this was provided.
  - *Basis for rating:* Probably low risk of bias because, although it appears some data were not reported, it is likely because there were insufficient data and not because the authors were selectively reporting the results.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used log10-transformed exposure variables. Normality, homoscedasticity, and linearity assumptions were tested and satisfied for MDI and PDI scores. Bivariate analyses included

correlations, t-tests, and ANOVA. Multiple linear regression models by the first and second trimester of pregnancy were used to evaluate the association between maternal fluoride exposure and MDI and PDI scores. The best-fit model was selected using a "stepwise method," and the best-fit line was evaluated using "the curve fitting method." It is not further specified or cited what these methods entailed. Best-fit or goodness-of-fit statistics are not reported. It is unclear how a best-fit model could be selected when the authors state that all models adjusted for the same set of covariates regardless of significance, and these covariates also appear in the final model—presumably the best-fit model. It is unlikely that a stepwise method would retain all those covariates unless they were forced in the model. Residual analysis was conducted to assess model validity; however, there is no description of the results of the residual analysis. Nonetheless, the impact on effect estimates is expected to be minimal.

- *Other potential concerns:* No other potential concerns were identified. In the peer-review report, NASEM (2020) cited the following as potential concerns: "the large difference in numbers of males and females in the offspring (20 males, 45 females), and apparently incorrect probabilities were reported for age differences between participants and nonparticipants, high rates of cesarean deliveries and premature births among participants (degree of overlap not reported), and incorrect comparisons of observed prematurity rates with national expected rates." However, these concerns were taken into consideration in other domains (***Selection, Confounding***).

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements and blinding of outcome assessor to participants' fluoride exposure, but it is limited by the cross-sectional study design and lack of accounting for potential co-exposures to arsenic.

E.2.9. Wang et al. (2020a)

E.2.9.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** School children aged 7–13 years
- ***Study area:*** Tongxu County, China
- ***Sample size:*** 325 school children
- ***Data relevant to the review:*** Associations between ADHD and other measures of learning disability with urine fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant association between psychosomatic problems and urinary fluoride (per 1-mg/L increase; adjusted $\beta = 4.01$ [95% CI: 2.74, 5.28]) and increased risk of a T-score >70 with urinary fluoride (per 1-

mg/L increase; adjusted OR = 1.97 [95% CI: 1.19, 3.27]). No significant associations with ADHD or other measures of learning disability.

E.2.9.2. Risk of Bias

- Author contacts:
  - Authors were contacted in July of 2020 to obtain additional information for risk-of-bias evaluation. No response was received.
- Population selection:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were recruited in 2017 from Tongxu County, China. Children were selected from four randomly selected primary schools in the area. Selection was based on specified inclusion rules. It was noted that the living habits and diets of the participants from the four schools were well matched, but details were not provided. The area did not have industrial pollution within 1 km of the living environment of the children, and it was noted that the children were not exposed to other neurodevelopmental toxicants (lead, cadmium, arsenic, or mercury). A table of subject characteristics was provided in the study but not by school or exposure. This was a pilot study, and it was not explicitly stated whether all eligible subjects participated in the study. There is no information on participation rates or whether they varied by school.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposed groups were recruited using similar methods during the same time frame and that any differences between the exposed groups were considered in the statistical analyses.
- Confounding:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* It was noted that subjects were well matched in terms of living habits and diets, but there were no specifics provided. It was noted that there was no industrial exposure or exposure to other neurotoxins such as lead, cadmium, arsenic, or mercury. Covariates were collected using a standardized and structured questionnaire completed by the children and their guardians under the direction of investigators, but reliability or validity of the questionnaire was not reported. Information collected included age, sex, weight, height, parental education level, and parental migration (or work as migrant workers). IQ scores evaluated by the Combined Raven's Test–the Rural in China were used to represent basic cognitive function. Models were adjusted for age, BMI, sex, mother and father migration, and urinary creatinine. Adjustments were not made for parental education, race/ethnicity, maternal demographics (e.g., maternal age, BMI), parental behavioral and mental health disorders (e.g., ADHD, depression), smoking (e.g., maternal smoking status, secondhand tobacco smoke exposure), reproductive factors (e.g., parity), iodine deficiency/excess, maternal (and paternal) IQ, quantity and quality of caregiving environment (e.g., HOME score), or SES other than parental migration. There is no evidence to suggest that SES

would differ substantially among the four rural schools in the same area of China that were randomly selected.

- *Potentially important study-specific covariates:* SES.
  - *Direction/magnitude of effect size:* The impact on the direction and magnitude of effect size are unknown. It was noted that the subjects were matched in terms of living habits and diet, and this could be an indication that SES was not different among the groups, but details were not provided.
- *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key covariates were considered, that the methods for collecting the information were valid and reliable, and that co-exposure to arsenic was not an issue in this area.

- Attrition:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are complete. It was noted that there were 325 subjects included, and results were available on all subjects.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- Exposure:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Spot urine samples were collected from each child in the early morning into cleaned polyethylene tubes. Fluoride concentrations were measured using fluoride ion-selective electrode [with reference to Ma et al. (2017); however, that reference cites Zhou et al. (2012)]. Therefore, no QC methods or LODs were available. Fluoride concentrations were creatinine-adjusted.
    - *Direction/magnitude of effect size:* Spot urine samples account for only recent exposure. Although this could cause some exposure misclassification, the number of subjects should help dilute any issues with the non-differential misclassification.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.
- Outcome:
  - *Rating:* Probably high risk of bias (NR)
  - *Summary:* Children's behavior was assessed by the Chinese version of Conners' Parent Rating Scale-Revised (CPRS-48). The homogeneity reliability of Cronbach α in the Chinese version of CPRS-48 was 0.932, the correlation of Spearman-brown split-half was 0.900, and the retest reliability of total score was 0.594. Raw scores for each subscale were converted into sex- and age-adjusted T-scores within a mean ± standard deviation (SD) of 50 ± 10. The guardians independently completed the CPRS-48 according to the instruction manual under the direction of trained investigators (++ for methods). Blinding is not reported. Although it is unlikely that the outcome assessors were aware of the fluoride

levels in the urine, it is unclear whether subjects were selected based on areas with endemic fluoride or whether parents were aware of fluoride concentrations in the areas (NR for blinding). Overall rating for methods and blinding = NR.

- *Basis for rating:* Probably high risk of bias based on no information provided to indicate that the outcome assessors were blind to the participants' fluoride exposure.

- Selective Reporting:
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- Other potential threats:
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Multiple linear regression models were used to assess the association between urinary fluoride exposure and each behavioral outcome. Logistic regression was used to assess the risk of behavioral problems (T-scores >70) due to fluoride exposure. Sensitivity analyses were performed, with models adjusting for combinations of age, BMI, sex, mother migrated, father migrated, and urinary creatinine levels. Regression diagnostics to evaluate model assumptions are not described; however, the overall impact on effect estimates is expected to be minimal.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and no other potential threats of risk of bias were identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include individual exposure measurements, but it is limited by the cross-sectional study design and lack of details on blinding of the outcome assessment. All key covariates were considered in the study design or analysis.

Appendix F. Mechanistic Data from Animal Studies

Table of Contents

Figures

A number of animal studies were available that presented mechanistic data in several effect categories (see Figure F-1). Limiting the data to studies with at least one exposure at or below 20 ppm fluoride drinking water equivalents (gavage and dietary exposures were backcalculated into equivalent drinking water concentrations for comparison) still provided a sufficient number of studies for evaluation of several mechanistic endpoints while allowing for a more focused look at exposure levels most relevant to human exposures. The following sections summarize the mechanistic data by effect category. Although there is some evidence of consistency in mechanistic effects, overall these data are insufficient to increase confidence in the assessment of findings from human epidemiological studies.

	Dose Level	
Mechanism	**All**	**≤20 ppm**
Biochemical (brain/neurons)	66	25
Neurotransmitters	53	23
Oxidative stress	78	25
Histopathology	67	30
Apoptosis/cell death	19	7
Inflammation	7	5
Thyroid	50	17
Other	3	2

Figure F-1. Number of Animal Mechanistic Studies for Fluoride by Mechanistic Category and Exposure Level

An interactive version of Figure F-1 and additional study details are available at https://public.tableau.com/app/profile/ntp.visuals/viz/FluorideTableauDashboards/ReadMe. The number of studies that evaluated mechanistic effects associated with at least one exposure at or below 20 ppm fluoride is tabulated in the "≤20 ppm" column. The total number of studies per mechanistic category is summarized in the "All" column.

F.1. Neurotransmitters

Neurotransmitter and biochemical changes in the brain and neurons were considered the mechanistic areas with the greatest potential to demonstrate effects of fluoride on the brain of animals in the lower dose range and provide evidence of changes in the brain that may relate to lower IQ in children (see Figure F-2). Twenty of 23 neurotransmitter studies assessed changes in brain cholinesterase activity associated with fluoride exposure at or below 20 ppm fluoride. Acetylcholine is a major neurotransmitter involved in learning, memory, and intelligence (Chen 2012; Gais and Schonauer 2017). AChE is responsible for the breakdown of acetylcholine in the synapses of nerve cells. Changes in cholinesterase, acetylcholine, or AChE could be related to effects on memory. Evidence of an effect varied among the low risk-of-bias studies that assessed changes in cholinesterase or acetylcholine (n = 11 drinking water studies) (Adedara et al. 2017a; Akinrinade et al. 2015a; Baba et al. 2014; Chouhan et al. 2010; Gao et al. 2008a; Gao et al. 2009; Khan et al. 2017; Liu et al. 2010; Mesram et al. 2016; Nkpaa and Onyeso 2018; Sun et al. 2000 [translated in Sun et al. 2008]), with the majority reporting evidence of an effect that is considered inconsistent with the phenotypic outcome (see Quality Assessment of Individual

Studies section for methods on determining which studies pose low risk of bias). Decreases in cholinesterase will cause increases in acetylcholine, which can have a positive effect on learning and memory; however, long-term decreases in cholinesterase can lead to secondary neuronal damage occurring in the cholinergic region of the brain (Chen 2012).

Five of the 11 studies with low risk of bias (Adedara et al. 2017a; Baba et al. 2014; Gao et al. 2009; Khan et al. 2017; Nkpaa and Onyeso 2018) found statistically significant decreases in cholinesterase or AChE in brain homogenates (with some brains dissected into specific regions prior to homogenizing) with fluoride concentrations in drinking water at or below 20 ppm, and four of the five studies found statistically significant decreases in cholinesterase or AChE below 10 ppm. The five studies were conducted in rats (Wistar or Sprague-Dawley) with exposure ranging from 28 days to 6 months. An additional 2 out of 11 studies (Akinrinade et al. 2015a; Gao et al. 2008a) reported decreases in brain homogenate AChE at concentrations at or below 20 ppm fluoride in drinking water, but statistical significance was not reached. These studies were also conducted in rats with exposure for 30 days or 3 months. Gao et al. (2008a) reported a dose-dependent decrease in brain homogenate AChE in the low (5 ppm fluoride) and high (50 ppm fluoride) treatment groups compared with the control group, but the decrease was statistically significant only in the high-dose group. Similarly, Akinrinade et al. (2015a) observed a dose-dependent decrease in percent intensity of AChE immunohistochemistry in the prefrontal cortex associated with 2.1 and 10 ppm sodium fluoride in drinking water, but neither result was statistically significant. Gao et al. (2009) found lower brain homogenate AChE levels in the 5-ppm animals compared with the 50-ppm animals; therefore, the results were not always dose-dependent.

Relative to the above-mentioned studies, 2 of the 11 low risk-of-bias studies observed opposite effects on brain cholinesterase levels. Sun et al. (2000) [translated in Sun et al. (2008)] observed a significant increase in brain cholinesterase in Kunming mice associated with fluoride drinking water concentrations from 10 to 100 mg/L but did not observe a dose response. Chouhan et al. (2010) did observe a dose-related increase in AChE levels in brain homogenate of Wistar rats with sodium fluoride concentrations of 1 to 100 ppm for 12 weeks and noted statistically significant results at 1, 50, and 100 ppm but not at 10 ppm.

Mesram et al. (2016) did not assess changes in AChE but observed a significant decrease in acetylcholine levels in cerebral cortex homogenate through 30 days of age in rats treated in utero with 20 ppm sodium fluoride, which may suggest an increase in AChE levels. Likewise, Liu et al. (2010) did not assess changes in AChE but measured nicotinic acetylcholine receptors (nAChRs) in brain homogenate of rats following drinking water fluoride exposure, which the authors stated could modulate physiological and pharmacological functions that are involved in learning- and memory-related behaviors. Significant decreases in the protein expressions of nAChR subunits at 2.26 ppm fluoride were observed; however, the corresponding receptor subunit mRNAs did not exhibit any changes (Liu et al. 2010).

The studies that assessed other neurotransmitters of the brain and neurons were too heterogeneous or limited in number to make any determination on mechanism, even before limiting the review of the data to low risk-of-bias studies. There were only five studies that evaluated dopamine and/or metabolites (Banala et al. 2018; Chouhan et al. 2010; Reddy et al. 2014; Sudhakar and Reddy 2018; Tsunoda et al. 2005). Four of the studies observed decreases in dopamine levels in the brain with exposures of less than 20 ppm fluoride (Banala et al. 2018;

Chouhan et al. 2010; Reddy et al. 2014; Sudhakar and Reddy 2018); however, the fifth study (Tsunoda et al. 2005) observed increased dopamine and metabolites at fluoride exposures below 20 ppm (with statistical significance achieved only for the metabolite homovanillic acid in one brain region). No differences from the control group were observed at levels above 20 ppm fluoride. Other neurotransmitters were evaluated at or below 20 ppm fluoride exposure, but generally only in a couple of studies.

F.2. Biochemistry (Brain/Neurons)

Similar to the above, the endpoints measured in brain biochemistry studies were too heterogeneous or limited in number to make any determination on potential relevance of mechanism, even before limiting the review of the data to low risk-of-bias studies (see Figure F-2). Endpoints related to biochemical changes in the brain or neurons included carbohydrate or lipid changes, RNA or DNA changes, changes in gene expression, or changes in protein expression. For the most part, only a single study was available for any given endpoint. The largest body of evidence on biochemistry was on protein level in various brain regions. Eleven low risk-of-bias studies were identified that evaluated protein levels; however, few studies evaluated the same proteins or areas of the brain. In the few cases in which the same protein was evaluated, results were not always consistent. These data are insufficient to increase confidence or support a change to hazard conclusions.

F.3. Histopathology

Histological data can be useful in determining whether effects are occurring in the brain at lower fluoride concentrations; however, author descriptions of these effects may be limited, thereby making it difficult to directly link histological changes in the brain to learning and memory effects. Histopathology of the brain was evaluated in 31 studies with concentrations at or below 20 ppm fluoride, of which 15 were considered low risk-of-bias studies (Adedara et al. 2017b; Akinrinade et al. 2015a; Bhatnagar et al. 2002; Bhatnagar et al. 2011; Chouhan et al. 2010; Guner et al. 2016; Jia et al. 2019; Jiang et al. 2014; Lou et al. 2013; McPherson et al. 2018; Mesram et al. 2016; Nageshwar et al. 2018; Niu et al. 2018; Pulungan et al. 2016; Zhao et al. 2019). In all but one low risk-of-bias study [Pulungan et al. (2016); gavage], animals were exposed to fluoride via drinking water. All low risk-of-bias studies were conducted in rodents, and all but three were conducted in rats (Wistar [seven studies], Sprague-Dawley [four studies], Long-Evans hooded [one study]). Overall, the low risk-of-bias studies that evaluated histopathology in the brain had low potential for bias for key questions regarding randomization and exposure characterization; however, eight studies were rated as probably high risk of bias for the key risk-of-bias question regarding outcome assessment based on lack of reporting of blinding of outcome assessors and/or inadequate description of outcome measures or lesions. Moreover, low image quality in some of the studies hampered the ability to verify the quality of the data. Further technical review of the 15 low risk-of-bias studies was conducted by a board-certified pathologist. Based on confidence in the results for each study, the technical reviewer further categorized the low risk-of-bias studies as studies with higher or lower confidence in the outcome assessment, which is reflected in the following summary of the brain histopathology results. Main limitations of the histopathology data identified by the pathologist included lack of information on methods of euthanasia and fixation. Perfusion fixation is generally considered the

best practice for lesions of the central nervous system in addition to complete fixation of the brain prior to its removal from the skull (Garman et al. 2016). Four of the low risk-of-bias studies reported that they used this method (Bhatnagar et al. 2002; Bhatnagar et al. 2011; McPherson et al. 2018; Pulungan et al. 2016). Two of the low risk-of-bias studies handled the brains before fixation was complete, which can produce artifacts that can resemble dead neurons (Nageshwar et al. 2018; Zhao et al. 2019). Fixation and brain removal details were inadequately described in the remaining low risk-of-bias studies.

Although there was heterogeneity in the endpoints reported (e.g., cell size, shape, and counts; nuclei fragmentation; increased vacuolar spaces) and some variation in the consistency of the evidence based on the area of the brain evaluated, the majority of the low risk-of-bias studies (11 of 14 drinking water studies) found some histological change in the brain of rats or mice treated with fluoride at concentrations at or below 20 ppm, of which 8 studies reported histological changes in the brain at or below 10 ppm. Histological changes in the hippocampus (one of the areas of the brain most evaluated for histological changes) associated with fluoride exposure at or below 20 ppm were reported in three of four low risk-of-bias studies with higher confidence in the outcome assessment (Bhatnagar et al. 2002; Bhatnagar et al. 2011; Guner et al. 2016) and in three of four low risk-of-bias studies with lower confidence in the outcome assessment (Jiang et al. 2014; Nageshwar et al. 2018; Niu et al. 2018). McPherson et al. (2018) was the only drinking water study (with higher confidence in the histopathology outcome assessment) that did not observe any histological changes in hippocampus at 10 or 20 ppm fluoride in male Long-Evans hooded rats exposed in utero through adulthood (>PND 80). Although there are too few studies to definitively explain the inconsistency in results, McPherson et al. (2018) also did not observe any associations between fluoride exposure and impairments to learning and memory, which is inconsistent with the majority of developmental exposure studies that observed learning and impairments associated with fluoride exposure for other strains of rats. Similarly, histological changes in the cortex were reported in three of the four low risk-of-bias drinking water studies with higher confidence in the outcome assessment (Akinrinade et al. 2015a; Bhatnagar et al. 2011; Chouhan et al. 2010) and in three of four low risk-of-bias studies with lower confidence in the outcome assessment (Lou et al. 2013; Mesram et al. 2016; Nageshwar et al. 2018).

Histological changes were also consistently reported in other areas of the brain in studies with higher confidence in the outcome assessment, including the amygdala, caudate putamen, cerebellum, and hypothalamus, although each of these areas of the brain was evaluated in only one low risk-of-bias study (Bhatnagar et al. 2011; Guner et al. 2016). Pulungan et al. (2016), one of two low risk-of-bias studies with higher confidence in the outcome assessment that did not report histological changes in the brain, observed a decreasing trend in the number of pyramidal cells in the prefrontal cortex with increasing dose, but this was not changed at concentrations below 20 ppm (the study administered sodium fluoride via gavage; the 5-mg/kg/day dose was considered equivalent to 15.3 ppm fluoride in drinking water), nor were any of the results statistically significant.

F.4. Oxidative Stress

Oxidative stress is considered a general mechanistic endpoint that cannot be specifically linked to neurodevelopmental or cognitive effects in humans; however, like histopathology, it may help in identifying changes in the brain occurring at lower concentrations of fluoride. Oxidative stress

in the brain was evaluated in 25 studies that examined concentrations at or below 20 ppm fluoride, of which 15 studies had low potential for bias (Adedara et al. 2017a; Adedara et al. 2017b; Akinrinade et al. 2015b; Bartos et al. 2018; Chouhan and Flora 2008; Chouhan et al. 2010; Gao et al. 2008b; Gao et al. 2009; Guner et al. 2016; Khan et al. 2017; Mesram et al. 2016; Nageshwar et al. 2018; Nkpaa and Onyeso 2018; Shan et al. 2004; Zhang et al. 2015a). All of the low risk-of-bias studies were conducted in rats (mainly Wistar or Sprague-Dawley) and administered fluoride via drinking water with exposure durations ranging from 28 days to 7 months. Although there was heterogeneity in the endpoints reported (i.e., varying measures of protein oxidation, antioxidant activity, lipid peroxidation, and reactive oxygen species [ROS]) and some variation in the consistency of the evidence based on the endpoint, the majority of the studies (13 of 15) (Adedara et al. 2017a; Adedara et al. 2017b; Akinrinade et al. 2015b; Bartos et al. 2018; Gao et al. 2008b; Gao et al. 2009; Guner et al. 2016; Khan et al. 2017; Mesram et al. 2016; Nageshwar et al. 2018; Nkpaa and Onyeso 2018; Shan et al. 2004; Zhang et al. 2015a) found evidence of oxidative stress in the brains of rats treated with fluoride at concentrations at or below 20 ppm, of which 10 studies reported oxidative stress in the brain below 10 ppm fluoride. The most consistent evidence of oxidative stress in the brain was reported through changes in antioxidant activity. Eleven of the 12 low risk-of-bias studies that evaluated antioxidant activity reported an effect at concentrations at or below 20 ppm (Adedara et al. 2017a; Adedara et al. 2017b; Akinrinade et al. 2015b; Bartos et al. 2018; Gao et al. 2008b; Gao et al. 2009; Guner et al. 2016; Khan et al. 2017; Mesram et al. 2016; Nageshwar et al. 2018; Nkpaa and Onyeso 2018). Decreases in antioxidant activity using measures of superoxide dismutase (SOD) activity were reported in seven of eight low risk-of-bias studies (Adedara et al. 2017a; Adedara et al. 2017b; Akinrinade et al. 2015b; Khan et al. 2017; Mesram et al. 2016; Nageshwar et al. 2018; Nkpaa and Onyeso 2018), and, among these seven studies, all that also measured changes in catalase (CAT) activity (n = 6 studies) also reported decreased activity (Adedara et al. 2017a; Adedara et al. 2017b; Khan et al. 2017; Mesram et al. 2016; Nageshwar et al. 2018; Nkpaa and Onyeso 2018). A decrease in total antioxidant capacity (T-AOC) as a measure of antioxidant activity was also consistently reported in two low risk-of-bias studies (Gao et al. 2008b; Gao et al. 2009), and a decrease in glutathione peroxidase (GPx) activity was reported in two of three low risk-of-bias studies (Adedara et al. 2017b; Nkpaa and Onyeso 2018).

Relative to the above-mentioned studies, 2 of the 15 low risk-of-bias studies (Chouhan and Flora 2008; Chouhan et al. 2010) did not observe statistically significant effects on oxidative stress in the brain with concentrations at or below 20 ppm fluoride; however, the measure of oxidative stress evaluated in Chouhan and Flora (2008) and Chouhan et al. (2010) (glutathione [GSH] to oxidized glutathione [GSSG] ratio as an indication of antioxidant activity and ROS levels) were not evaluated in any other low risk-of-bias study. Chouhan and Flora (2008) observed a dose-dependent increase in ROS levels associated with 10, 50, and 100 mg/L sodium fluoride in drinking water; however, results were not statistically significant at any dose. In Chouhan et al. (2010), the levels of ROS were significantly higher at 50 ppm sodium fluoride in drinking water, but statistical significance was not met at doses below 20 ppm fluoride (1 and 10 ppm sodium fluoride) or at 100 ppm sodium fluoride; yet, hydrogen peroxide levels as a measure of ROS were found to be significantly increased at 15 ppm sodium fluoride in drinking water in studies conducted by another group of authors (Adedara et al. 2017a; Adedara et al. 2017b).

F.5. Apoptosis/Cell Death

Seven low risk-of-bias studies were identified that evaluated apoptosis with concentrations at or below 20 ppm fluoride. Results from these studies were inconsistent and were insufficient for evaluating fluoride-induced apoptosis. These data are insufficient to increase confidence or support a change to hazard conclusions.

F.6. Inflammation

Five low risk-of-bias studies were identified that evaluated potential effects of fluoride on inflammation with concentrations at or below 20 ppm. The inflammation markers were too heterogeneous or limited in number to make any determination on potential relevance of mechanism, even before limiting the review of the data to low risk-of-bias studies. These data are insufficient to increase confidence or support a change to hazard conclusions.

F.7. Thyroid

Seventeen studies were identified that evaluated potential effects of fluoride on the thyroid with concentrations at or below 20 ppm (see Figure F-1). These animal thyroid data are not further described because this endpoint has been directly evaluated in a number of human studies that have failed to identify consistent evidence to suggest that thyroid effects are a requisite mechanism by which fluoride causes neurodevelopmental or cognitive effects in humans.

Mechanism	Mechanism Subcategory	Direction of Effect ↑	↓	NS	Grand Total
Biochemistry (brain/neurons)	Carbohydrate/lipid-related		1	1	2
	Gene expression		1		1
	Protein levels associated with brain function	8	7	7	11
	Other biochemical	2			2
Neurotransmitters	Cholinesterase	2	7	3	11
	Dopamine and metabolites		1		1
	Other neurotransmitters	2	2	1	2
Oxidative stress	Antioxidant activity	1	10	4	12
	Lipid peroxidation byproduct	7		4	11
	Protein oxidation	2		1	2
	ROS	2		2	4

Figure F-2. Number of Low Risk-of-bias Animal Studies That Evaluated Biochemical, Neurotransmission, and Oxidative Stress Effects at or below 20 ppm by Mechanism Subcategory and Direction of Effect

An interactive version of Figure F-2 and additional study details are available at https://public.tableau.com/app/profile/ntp.visuals/viz/FluorideTableauDashboards/ReadMe. This figure displays study counts for low risk-of-bias studies, as these counts are most relevant to the text in this section. Counts for high risk-of bias studies or all studies combined can be accessed in the interactive figure. Study counts are tabulated by significance—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with at least one endpoint in the mechanistic subcategory with significantly increasing results at fluoride exposure levels of ≤20 ppm. These columns are not mutually exclusive (i.e., a study may report on multiple endpoints with varying results within a single mechanistic subcategory and therefore may be reflected in the counts for the "↑", "↓", and NS columns but would be counted only once in the Grand Total column). Endpoints, species, strain, sex, and exposure duration are available for each study in the interactive figure.

Appendix G. Protocol History and Revisions

Date	Activity or Revision
December 14, 2016	**Draft evaluation protocol reviewed:** sent to technical advisors for peer review
April 10, 2017	**Draft human risk-of-bias protocol reviewed:** sent to technical advisors for peer review
May 2, 2017	**Draft animal risk-of-bias protocol reviewed:** sent to technical advisors for peer review
June 2017	**Evaluation protocol finalized:** Review protocol finalized for use and posting
May 29, 2019	**Revised protocol:** Revised review protocol posted
September 16, 2020	**Revised protocol:** Revised review protocol posted

Appendix H. Supplemental Files

The following supplemental files are available at https://ntp.niehs.nih.gov/go/785076 or as *Sup01_Monograph_NASEM_Feb_2021.pdf.*

H.1. Protocol

NTP Protocol for Systematic Review of Human, Animal, and Mechanistic Evidence - Second Revision (September 16, 2020)
ntpprotocol_revised20200916_508.pdf

NTP Protocol for Systematic Review of Human, Animal, and Mechanistic Evidence - First Revision (May 29, 2019)
protocol_fluoridemay2019_508.pdf

NTP Protocol for Systematic Review of Human, Animal, and Mechanistic Evidence (June 2017)
protocol_fluoridejune2017_508.pdf

H.2. Response to NASEM Committee Letter Report

NASEM Committee Letter Report and Response for Monograph Only
Sup01_Monograph_NASEM_Feb_2021.pdf

In February 2022, the [redacted] provided comments to NIEHS/DNTP on the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* and a draft manuscript on a meta-analysis of fluoride exposure and IQ in children. This document contains a subset of the overall [redacted] comments related to the prepublication 2022 NTP Monograph along with the NIEHS/DNTP responses. The monograph-related comments from the [redacted] are reproduced here in black text, and the NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] and [redacted]:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to this [redacted] would be marked "*see DocA1_Monograph for detailed response*").
- [redacted] For comments related to DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

February 1, 2022

[redacted]

Feedback to NTP/NIEHS regarding:
Fluoride state of the science document
Fluoride and IQ Meta-analysis manuscript

Fluoride state of the science document:

A1.1: Issue: Keeping findings in context
As NASEM noted in their review of the 2019 Draft Monograph, "the context into which the monograph falls calls for much more carefully developed and articulated communication on this issue." [redacted] fully concurs with this recommendation and with NASEM's 2019 assessment that "NTP needs to state clearly that the monograph is not designed to be informative with respect to decisions about the concentrations of fluoridethat are used for water fluoridation."

NTP stated in the revised draft of the monograph that the evidence of "effects on cognitive neurodevelopment are inconsistent, and therefore unclear" at the levels typically found in drinking water in the US. NASEM agreed with this assessment, statingthat "[m]uch of the evidence presented in the report comes from studies that involve relatively high fluoride concentrations. Little or no conclusive information can be garnered from the revised monograph about the effects of fluoride at low exposure concentrations (less than 1.5 mg/L)."

[redacted] is extremely concerned that the revised 2021 NTP report and the meta-analysis omit this important context that was previously included. Without clarification, readers may interpret that exposure to fluoride at any concentration is associated with lower IQ,a conclusion that is not borne out by the available science or the findings of the systematic review.

Recommendation: [redacted] requests NTP include a statement in the systematic review abstract and fulltext, as well as the meta-analysis, like that found in the 2020 draft monograph: "When focusing on findings from studies with exposures in ranges typically foundin drinking water in the United States (0.7 mg/L for optimally fluoridated community water systems) that can be evaluated for dose response, effects on cognitive neurodevelopment are inconsistent, and therefore unclear."

Response: Disagree (edited for clarity)

- As pointed out, the language the [redacted] recommends was in the 2020 draft NTP Monograph, and that earlier version of the document included a draft

dose-response meta-analysis. Similar language was also included in the 2019 draft NTP Monograph that was based on a qualitative look at the shapes of dose-response curves in papers that reported dose-response analyses of fluoride exposures and children's IQ. The prepublication 2022 NTP Monograph focuses on the question of whether fluoride from all sources can affect neurodevelopmental outcomes and is written to avoid giving the mistaken impression that this systematic review is focused only on drinking water. It is true that our stated confidence assessment is based primarily on studies with total exposures higher than those generally associated with consumption of optimally fluoridated water in the United States. However, the confidence assessment also includes findings from studies with fluoride exposures that are similar to those associated with optimally fluoridated water supplies in the United States. In addition, no studies examining individual-level exposures have been conducted in the United States. As demonstrated in Green et al. (2019), who used repeated individual urinary measurements, drinking water measures likely capture only a portion of a person's total exposure to fluoride, as personal preferences and habits may increase total exposures to unknown levels. Therefore, we do not consider it appropriate to put the conclusions in the context of what is recommended for optimally fluoridated community water systems in the United States. However, to provide the context that the confidence conclusions are primarily based on studies that included total exposures that approximate or are higher than 1.5 mg/L, the following statement was added to the abstract and summary of the prepublication 2022 NTP Monograph.

"This review finds, with moderate confidence, that higher fluoride exposure (e.g., represented by populations whose total fluoride exposure approximates or exceeds the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride) is consistently associated with lower IQ in children. More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ."

A1.2: Issue: Limitations section

In its response letter, NASEM requested adding clarifying information in the manuscript. NTP itemized items in the state-of-the-science manuscript on limitations of the evidence base and the systematic review. However, these limitations do not address the following issues comprehensively:

Transparency regarding adherence to OHAT protocol.

Recommendation: NTP should specify the areas where they departed from the OHAT protocol.

Response: Agree (change made)

- Edits were made to provide more specificity in the monograph and protocol with respect to detailing all aspects of the systematic review. However, we submit that the systematic review described in the 2021 draft NTP Monograph and the prepublication 2022 NTP Monograph fully adhered to the pre-published, project-specific protocol for this systematic review (https://ntp.niehs.nih.gov/go/785076) with additional specificity or methodological details described in the *Methods* section of the monograph. The [redacted] comment appears to be referring to the use of the methods as described in the OHAT handbook relative to the methods as described in the specific protocol for this systematic review. To clarify that the protocol describes all the methods for the monograph, additional detail was added to the *Foreword* and *Methods* sections of the prepublication 2022 NTP Monograph. This includes the following text in the *Methods* section:

  "NTP conducts systematic reviews following prespecified protocols that describe the review procedures selected and applied from the general methods outlined in the OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673). The protocol describes project-specific procedures tailored to each systematic review that supersede the methods in the OHAT Handbook."

- The methodological details described in our specific protocol with additional specificity described in the *Methods* section of the monograph (e.g., the decision not to consider data from studies that reported only thyroid gland size or goiters) provide a level of documentation that meets or exceeds standard practice for documentation in the field.

A1.3: Potential bias as there was no systematic selection of Chinese databases to be searched. Two databases were selected because they contained studies of which the authors were aware.

Recommendation: As NASEM noted, this introduces potential bias. [redacted] suggests this be added as a limitation.

Response: Agree (change made)

- We accepted this recommendation in the prepublication 2022 NTP Monograph and the following statement has been added to the *Limitations of the Systematic Review* section of the systematic review.

  "The supplemental literature search for non-English-language studies not indexed in traditional databases supports the comprehensive nature of the literature search strategy for this systematic review. In the absence of guidance on the most complete non-English-language databases that may contain health studies of fluoride, a standard systematic review approach for database selection was followed whereby a set of exemplar documents, called 'seed studies' were used. Databases were selected

that identified non-English-language studies of fluoride that we were aware of and were not captured in searches of databases from the main literature search. This informed approach influenced the selection process; however, this is not considered a limitation because it provided an objective measure by which to compare databases."

A1.4: Some included studies with complex sample designs did not report if they used population weights to generate estimates.

Recommendation: In addition to listing this as a limitation, NTP should identify these studies in the body of the report.

Response: Disagree (no change)

- *Appendix E* provides extensive details on the risk-of-bias assessment for the low risk-of-bias studies. The risk-of-bias assessment included evaluation of whether complex sampling designs were accounted for with the use of sampling weights or adjustments for clustering.
- We have addressed these issues in the meta-analysis.

A1.5: Clustering: NASEM identified that in some population studies, participants livingin the same communities were assigned the same measure of fluoride exposure without considering the effect in the data analysis. These correlations may artificially increase the statistical power.

Recommendation: Limitations should note the studies where clustering was a potential threat and specifically whether the investigators addressed this.

Response: Agree (change made)

- In response to comments from the NASEM Committee, we revised text in *Appendix E* of the prepublication 2022 NTP Monograph (previously *Appendix 4* in the 2020 draft NTP Monograph) to note specifically whether each low risk-of-bias study applied an analytic approach that addressed clustering when that was a feature of the study design. Our risk-of-bias assessment carefully considered study-specific failures to account for sampling strategy or clustering in determining potential for bias.

A1.6: The NTP response to NASEM's comments indicate they contacted investigators on specific issues, and that some study authors responded while others did not.

Recommendation: NTP should provide an assessment of how non-response would affect the grading of these studies. This may be tied to the NASEM

concern regarding transparency in the risk-of-bias assessment. [redacted] requests an appendix and language in the Discussion detailing which studies received a change to their risk-of-bias determination.

Response: Disagree (edited for clarity)

- The risk-of-bias rating explanations provided in the HAWC web-based evaluation platform via URL links and *Appendix E* of the prepublication 2022 NTP Monograph (previously *Appendix 4* of the 2020 draft NTP Monograph) previously noted whether an author responded to our inquiry for more information and whether the response impacted the risk-of-bias rating. To provide information more clearly on author inquiries and how information received by the authors was used in the risk-of-bias analysis, we have made updates to the HAWC study profiles for each human study and to *Appendix E*. When author inquiries were conducted, they are noted in the study profiles. If an author did not respond, it is noted in the study profile (e.g., "No response was received to email request for clarification"). If an author responded and provided additional information that informed a rating decision in the risk-of-bias analysis, it is now noted in the HAWC study profiles and *Appendix E* which risk-of-bias questions were impacted.

A1.7: Meta-Analysis: The meta-analysis, originally requested by NASEM to obtain measures of association and sensitivity analysis across selected studies wasremoved to be published separately.

Recommendation: The meta-analysis should be reinserted into the NTP draft document. Moving the meta-analysis to a separate document makes it difficult forthe reader to understand and interpret the conclusions of the systematic review. Further, the meta-analysis, when reinserted into the NTP systematic review, should address NASEM's critiques of the September 2020 draft (abstracted below):

Response: Disagree (no change)

- With removal of the hazard assessment from the 2020 draft NTP Monograph, our focus shifted to providing a qualitative confidence assessment of the relevant literature of fluoride exposure and neurodevelopmental and cognitive health effects in children and adults, which is presented in the prepublication 2022 NTP Monograph. In contrast, the updated meta-analysis manuscript provides a quantitative assessment of the studies examining fluoride exposure and IQ in children. After considering the scope and nature of the NASEM Committee's comments, we determined that the confidence assessment of the complete evidence base on neurodevelopmental and cognitive health effects in children and adults is a broad and distinct issue from the specific focus of the meta-analysis on IQ in children without the hazard assessment section to integrate these different analyses. In addition, we

concluded that the topic is of such high public health importance that the integration of the confidence assessment of the complete evidence base on neurodevelopmental and cognitive health effects would be better done as a collective effort by the public health community in a larger conversation about the appropriate method and timing of population exposures to fluoride to benefit oral health.

A1.8: Issue: New evidence

Two studies (Ibarluzea et al., 2021 and Aggeborn & Ohman, 2021) published in 2021 were not included in the systematic review or meta-analysis. These studies have comparable methods to other included studies.

Recommendation: The Ibarluzea and Aggeborn & Oehman studies should be evaluatedand included when assessing the evidence, similar to the 15 additional studies from theChinese databases. [redacted] also recommends NTP include a comparison between Ibarluzea et al., 2021, and Green et al., 2019, because both studies investigate fluoride exposures at levels used for water fluoridation.

Response: Disagree (edited for clarity)

- Adding only these two studies beyond the literature cutoff date for the prepublication 2022 NTP Monograph without adding other studies that have been published since the cutoff date would introduce bias and subject us to potential criticism of cherry-picking studies. With respect to Aggeborn and Öhman (2021), we examined this study when it was published as a non-peer-reviewed white paper in 2017 and it was excluded then as non-peer-reviewed reference. It was determined that including this study would not impact our confidence conclusions for multiple reasons: 1) it is a high risk-of-bias ecological study; 2) it would also fall under other neurodevelopmental (non-IQ) studies; and 3) it uses cognitive tests that are not specified.
- With respect to Ibarluzea et al. (2021), we acknowledge the potential interest in this specific study, and we added the following text and footnote to the prepublication 2022 NTP Monograph:

  "Note that NTP is aware of a conference abstract by Santa-Marina et al. on a Spanish cohort study that looked at fluoride exposure and neuropsychological development in children (Santa-Marina et al. 2019). The evaluation team conducted a targeted literature search in April 2021 to see whether the data from this study had been published. When no publication was found, the evaluation team contacted the study authors to inquire about the publication of their data. The response from the study authors indicated that the study report was being finalized but had not yet been sent to a journal for review; therefore, it was not considered here.[9]"
- Relevant text from the footnote below:

"[9]NTP is aware that this study was published after April 2021 (Ibarluzea et al. 2021) and, therefore, is not included in this monograph because it is beyond the dates of the literature search. Even if it had been published earlier, the study would not have contributed to the body of evidence on children's IQ because the authors assessed other neurodevelopmental or cognitive effects, specifically the association between fluoride exposure and neuropsychological development in children aged 1 year using the Mental Development Index (MDI) of the Bayley Scales of Infant Development and in children aged 4 years using the General Cognitive Index (GCI) of the McCarthy Scales of Children's Abilities (MSCA)."

In May 2022, the [redacted] provided comments to NIEHS/DNTP on the prepublication 2022 *NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* by email. The [redacted] comments are reproduced here in black text, and the NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to Agencies for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] and [redacted]:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to this [redacted] would be marked *"see DocA2_Monograph for detailed response"*).
- [redacted] For comments related to DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

[redacted],

A2.1: Thank you for forwarding the prepublication draft of the NTP SoS report on fluoride and the comments and responses to [redacted] feedback shared with NTP in February. [redacted] am writing to share specific feedback on the "[redacted] comments and response-final.pdf" that you shared last week.

Under this issue of "keeping findings in context," your response including the following new statement to be added in the abstract and summary of the SoS document:

"This review finds, with moderate confidence, that higher fluoride exposure (e.g., represented by populations whose total fluoride exposure approximates or exceeds the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride) is consistently associated with lower IQ in children. More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ."

For your consideration, [redacted] proposes the following revised text with justification for the revision provided by the numbered notes below.

***Proposed revision**: This review finds, with moderate confidence, that high fluoride exposure is associated with lower IQ in children (1). Studies of fluoride exposure at levels typically found in drinking water in the United States are inconclusive (2). More studies are needed before determining the effect of lower fluoride exposure on children's IQ. (3)*

Justification:

(1) Except for Green 2019 and Bashash 2017, all studies in the systematic review were cross-sectional and compared populations at higher fluoride levels than those used in community water fluoridation (CWF). In some studies, comparable fluoride levels to CWF were included as controls, which as a subset of studies show a positive correlation between higher fluoride exposures and IQ scores.

(2) Two prospective studies, Green 2019 and Ibarluzea 2022, measured fluoride exposures at comparable levels to CWF in the U.S. These two studies reported opposite effects of fluoride on IQ levels. A third study (Bashash 2017) is not equivalent to Green 2019 and Ibarluzea 2022 because the primary source of F exposure in Mexico is fluoridated salt, which is not available in the U.S. Even if the Bashash 2017 study is included, the conclusion would still be that these effects remain "inconclusive."

(3) Based on (2) above, the NTP conclusion would support additional prospective studies to clarify the directionality of any potential effect of lower fluoride exposure on IQ.

This is [redacted] only feedback on your comments/response document. [redacted] appreciate your time and effort to respond to [redacted] original feedback.

Please do not hesitate to contact [redacted] if you have questions or need clarification.

Best regards,

[redacted]

References

Bashash M, Thomas D, Hu HH, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Ettinger AS, Wright R, Zhang Z, Liu Y, Schnaas L, Mercado-Garcia A, Tellez-Rojo MM, Hernandez-Alva M. Prenatal fluoride exposure and cognitive outcomes in children at 5 and 6–12 years of age in Mexico. 2017. Environmental Health Perspectives; 125:097017.

Green R, Lanphear B, Homung R, Flora D, Martinez-Mier A, Neufeld R, Ayotte P, Muckle G, Till C. 2019. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. JAMA Pediatrics 173:940–948.

Ibarluzea J, Gallastegi M, Santa-Marina L, Jiménez Zabala A, Arranz E, Molinuevo A, Lopez-Espinosa M-J, Ballester F, Villanueva CM, Riano I, Sunyer J, Tardon A, Lertxundi A. 2022. Prenatal exposure to fluoride and neuropsychological development in early childhood:1-to 4 years old children. Environmental Research; 207:112181.

Response: Disagree (no change)

- We appreciate the continued dialog over the concluding statement that appears in several places in the monograph, including the *Summary* and *Abstract*. We consider the suggested three-part revision to the concluding statement to convey essentially the same information as our existing text but overemphasizes water levels of fluoride rather than total fluoride exposure. Considering the justifications offered for the change, we agree with justification (1) concerning the characterization of the relevant database. However, we note in the prepublication 2022 NTP Monograph on page 12 that, while we were aware of the Ibarluzea et al. (2021) study (identified as Ibarluzea 2022 in the [redacted] comment above) from following preliminary reports that appeared as meeting abstracts, the final publication date was beyond the literature cutoff date for the monograph. Even if the Ibarluzea et al. (2021) study had been published earlier, it would not have contributed to the body of evidence on children's IQ because the authors assessed other neurodevelopmental or cognitive effects, specifically the association between fluoride exposure and neuropsychological development in children aged 1 year using the Mental Development Index (MDI) of the Bayley Scales of Infant Development and in children aged 4 years using the General Cognitive Index (GCI) of the McCarthy Scales of Children's Abilities (MSCA). That factor aside, we have stressed in the monograph that our conclusions apply to total fluoride exposures rather than to exposures exclusively through drinking water. Although we tend to agree that "studies of fluoride exposure at levels typically found in drinking water in the United States are inconclusive," Green et al. (2019) was the only high quality prospective study included in the prepublication 2022 NTP Monograph that evaluated a population exposed to fluoride in drinking water at levels typically found in drinking water in the United States. Therefore, it is more accurate to state, as we currently have, that:

  "More studies are needed to fully understand the potential for lower fluoride exposure [referring to the parenthetical from the previous sentence: (e.g., represented by populations whose total fluoride exposure approximates or exceeds the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride)] to affect children's IQ."

In November 2021, the [redacted] provided comments to NIEHS/DNTP on the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* (the NTP Monograph) and a draft manuscript on a meta-analysis of fluoride exposure and IQ in children (the meta-analysis manuscript). NIEHS/DNTP prepared responses and shared those responses back to [redacted] in April 2022.

In July 2022, the [redacted] provided two sets of comments to NIEHS/DNTP, again on the NTP Monograph (prepublication 2022 version) and the meta-analysis manuscript.

- The first set of [redacted] comments was provided as a new layer of input on top of the original [redacted] comments (from November 2021) and NIEHS/DNTP responses. This document contains a subset of the overall [redacted] comments (from November 2021 and July 2022) related to the NTP Monograph along with the NIEHS/DNTP responses. The monograph-related comments from the [redacted] are reproduced below in black text and the NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.
- The second set of [redacted] comments was provided in track changes embedded in the prepublication 2022 NTP Monograph in Microsoft Word. See file "DocB2_Monograph" for the second set of [redacted] comments and NIEHS/DNTP responses.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] and [redacted]:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to this [redacted] would be marked "*see DocB1_Monograph for detailed response*").
- [redacted] For comments related to DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted].

- The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

[redacted] comments from November 2021 and July 2022

Summary of [redacted] comments on the *"Draft NTP monograph on the state of the science concerning fluoride exposure and neurodevelopmental and cognitive health effects: a systematic review"* ("SoS document") and draft Taylor *et al. Association between fluoride exposure and children's intelligence: A systematic review and meta- analysis* manuscript ("meta-analysis document")

B1.1:

- **[redacted] comment on SoS document (November 2021):** It would be helpful if the Abstract was clear in the Discussion that the conclusion about effects on IQ in children was derived from high human exposures (higher than US exposures) without getting into more hazard conclusions or assessments.

  Response: Agree (edited for clarity)

  - While it is correct that much of the literature evaluating exposures to fluoride with respect to reduced cognition in children likely involves exposures to amounts assumed to be in excess of what are consumed in the United States, there is actually very little direct U.S. exposure information on which to base this. As we discuss in the monograph, fluoride is found in water, certain foods, dental products, some pharmaceuticals, etc., and individual behaviors are likely to be an important determinant of actual total fluoride exposures. Green et al. (2019) is the study most likely to approximate U.S. exposures because Canada has the same optimal fluoridation level (0.7 mg/L in drinking water) as the United States. The individual exposure levels reported in Green et al. (2019), as documented by repeated urinary measurements, suggest widely varied exposures from optimally fluoridated drinking water combined with fluoride from other sources. Additionally, fluoride in drinking water from wells in certain parts of the United States are known to exceed artificially-fluoridated water levels in the Canadian cities studied by Green et al. (2019). To clarify that our moderate confidence conclusion is primarily based on studies with total fluoride exposure that approximates or exceeds what is generally associated with consumption of optimally fluoridated water in the United States, the *Abstract* section was revised as follows:

    "This review finds, with moderate confidence, that higher fluoride exposure (e.g., represented by populations whose total fluoride exposure approximates or exceeds the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride) is consistently associated with lower IQ in children."

B1.2:

- **[redacted] comment on SoS document (November 2021):** [redacted] suggest that any public communication make this point about exposure clear.

  Response: Agree (no change requested)

  - We agree that public communication concerning total exposures to fluoride is essential. In our assessment of studies relying only on drinking water levels as an exposure metric,

we find that the data concerning levels below the WHO safe water guidelines are inconsistent and unclear with respect to effects on children's cognition.

B1.3:

- **Follow-up [redacted] comment on SoS document (July 2022):** The [redacted] comment requesting public communication was focused on the issue that studies were conducted on populations with higher exposures from water than are routinely found in the United States; the above response indicated that they agree that public communication concerning total exposures is essential. This is a different point. Has the request that there be public communication about higher exposure from fluoride in water been adequately addressed?

  Response: Disagree (no change)

  - The comment implies that our conclusions are based solely on "studies [that] were conducted on populations with higher exposures from water than are routinely found in the United States." This implication is not accurate. It is true that our stated confidence assessment is based primarily on studies with total exposures higher than those generally associated with consumption of optimally fluoridated water in the United States. However, the confidence assessment also includes findings from studies with fluoride exposures that are similar to, or lower than, those associated with optimally fluoridated water supplies in the United States. In addition, as mentioned in a previous response, there is very little direct information on actual total fluoride exposure in the United States. As demonstrated in Green et al. (2019), who used repeated individual urinary measurements, drinking water measures likely capture only a portion of a person's total exposure to fluoride as personal preferences and habits may increase total exposures to unknown levels. Therefore, this document, as well as any associated communication, focuses on total fluoride exposures from all sources, not just drinking water. We acknowledge that these complexities in the data, along with a lack of direct U.S. exposure data, are important to convey in public communications. We point this out in the prepublication 2022 NTP Monograph and will communicate this as a limitation of the database.

B1.4:

- **[redacted] comment on SoS document (November 2021):** [redacted] has followed the ongoing discussion between NTP and NASEM related to the alleged impacts on development and cognition, and specifically has reviewed the literature reports from Basham, et al., (2017) and Green et al., (2019). [redacted] concerns related to the study designs and the utility and accuracy of the urinary fluoride measurements have previously been communicated. [redacted] reiterate that actual exposure to fluoride and serum fluoride levels were not measured during these investigations.

  Response: Agree (no change)

  - We agree that actual exposures (as opposed to exposure estimates) have not been measured in any of the studies we have reviewed. Although human serum levels tend to reflect fluoride levels in water (WHO 2002), they vary widely during the day, and only rarely were they measured or reported in the literature we evaluated. In short, these concerns along with others contributed to our conclusion of only moderate confidence in an association between fluoride exposures and children's cognitive neurodevelopment in the prepublication 2022 NTP Monograph.

B1.5:

- [redacted] **comment on SoS document (November 2021):** The revised NTP monograph seems to address concerns from prior comments as NTP removed the hazard assessment and is now calling this a "state of the science" document. However, the meta-analysis that NTP removed from the original monograph is now being published independently. Although it will be in a scientific review publication [Note: journal name deleted by NIEHS/DNTP], [redacted] think that this may raise questions regarding exposure levels and neurodevelopmental effects, as the publication does not seem to put the exposure levels into context.

  Response: Agree (no change)

  - The revised meta-analysis manuscript will address exposure issues in a like manner as outlined above for the prepublication 2022 NTP Monograph.

B1.6:

- [redacted] **comment on SoS document (November 2021):**

  In September 2020, [redacted] conveyed the following concerns:

  > *The dosimetry of exposure plays a central role in this assessment. In this regard, there is some concern regarding the classifications of risk of bias that were ascribed to some on the studies on question #8 "Can [redacted] be confident in the exposure characterization?". Specifically, concerns were raised on how studies such as those below, where inadequate analytical detail is provided by the authors, can be classified as "Probably low risk of bias". Analytical (and consequently, exposure assessment) bias can only be assumed to be low following an appropriate evaluation of the analytical procedures, including data on the validation of the methodologies in the laboratory where the analyses were conducted (e.g. accuracy, precision). In light of this, it seems to us that the following studies (not necessarily a comprehensive list) fall short of warranting such a classification.*
  >
  > *Saxena et al. (2012)*
  >
  > *"The fluoride levels were analyzed by a fluoride ion selective electrode, Orion 9609BN (Thermo Fisher Scientific Inc., West Palm Beach, United States). (…) The fluoride content in the urine was determined using a fluoride ion selective electrode, Orion 9609BN (Thermo Fisher Scientific Inc., West Palm Beach, United States)."*
  >
  > *Seraj et al. (2012)*
  >
  > *"The fluoride and iodine in the drinking water were analyzed by SPADNS (Sulfophenylazo dihydroxynaphthalenedisulfonate) method, Utilizing 4000 UV-Vis spectrophotometer (Hach Company, Germany) in the environmental health engineering laboratory of Public Health School of Tehran University of Medical Sciences."*
  >
  > *Xiang et al. (2011)*
  >
  > *"F levels in serum were measured with a F ion selective electrode*

These issues seem to remain partially unaddressed in the current version of the document. For example, in the Saxena *et al.* (2102) and Xiang *et al.* (2011) studies, the basis for rating maintains its exposure classification of *"Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure."* That conclusion remains unclear when all that the authors indicate is that they used a fluoride-selective electrode without additional details, including for example, calibration, assessment of linearity, intra- and inter-day precision and accuracy, inclusion of quality controls, etc. As raised in [redacted] prior comments, analytical (and consequently, exposure assessment) bias can only be assumed to be low following an appropriate evaluation of the analytical procedures and the authors do not provide any such detail.

Response: Disagree (no change)

- The comment suggests that studies that described analytical methods without explicit methodological details on calibration and intra- and inter-day precision should not be rated as *probably low risk of bias* for exposure assessment. However, this seems to reflect a misunderstanding between criteria for *definitely low risk of bias* (which requires **direct** evidence of both a well-established method to measure fluoride and a QC procedure including such things as recovery rates, blanks, or reference standards) and *probably low risk of bias* (which requires that a well-established method to measure fluoride was used and **indirect** evidence that typical procedures were followed). The risk-of-bias ratings in question were judged to be *probably low risk of bias* as is consistent with the protocol.

B1.7:

Follow-up [redacted] comment on SoS document (July 2022): Which criteria were considered to ascertain if specific analytical methods to directly measure fluoride were "well established", and what protocol was followed to determine if the cited methods fulfilled such criteria?

Response: Agree (no change requested)

- "Well established" methods denote accepted methods for measuring fluoride levels. As noted in the protocol (https://ntp.niehs.nih.gov/go/785076), the preferred analytical method is the ion selective electrode method. However, use of other standard methods such as NIOSH Method 8308 or other governmental standard methods were considered well established. Any study noting that they used these methods was rated *probably low risk of bias* for exposure. In order to be rated *definitely low risk of bias* for exposure, a study also had to provide a detailed description of QC procedures (i.e., direct evidence) that were followed, including such things as use of recovery rates, blanks, or reference standards.

B1.8:

- [redacted] **comment on SoS document (July 2022):** The authors are using an old version of the PRISMA flow diagram - The 2020 PRISMA flow diagram can be found here: https://prisma-statement.org/prismastatement/flowdiagram.aspx

Response: Disagree (edited for clarity)

- The [redacted] comment refers to "*Figure 2. Study Selection Diagram*" in the prepublication 2022 NTP Monograph, which follows or exceeds the Preferred Reporting Items for

Systematic reviews and Meta-Analyses (PRISMA) standards. Figure 2 is a static reference flow diagram that includes all of the information suggested in the PRISMA standards. In addition to Figure 2, we have included an interactive reference flow diagram (https://hawcproject.org/summary/visual/assessment/405/Figure-2/) that far exceeds the PRISMA reporting standards, allowing readers to identify and link to all of the included studies as well as all of the studies excluded at each stage of reference screening for eligibility. The prepublication 2022 NTP Monograph follows PRISMA reporting standards as outlined in the OHAT handbook and our evaluation-specific protocol (https://ntp.niehs.nih.gov/go/785076). Although there are slight design differences, the study selection diagram contains all of the recommended reporting elements. Therefore, while we agree that Figure 2 resembles an older version of the PRISMA diagram, the interactive reference flow diagram exceeds the 2020 PRISMA recommendations (Page et al. 2021) and follows the evaluation-specific protocol. We have updated the PRISMA reference in the monograph from Moher et al. (2009) to Page et al. (2021) to clarify that we are following the current recommendations.

Note: The [redacted] comment on the overall confidence in the meta-analysis and subgroup analyses for the mean-effects meta-analyses are not reproduced here as they are not directly relevant to the prepublication 2022 NTP Monograph. See Doc08_Meta-analysis for the meta-analysis-relevant comments and responses.

B1.9:

- [redacted] **comment on SoS and meta-analysis documents (July 2022):** The [redacted] raised concerns regarding exposure measurement in previous comments. The current Discussion sections in each document cover some exposure measurement limitations but may not sufficiently address [redacted] previous comments or other important issues potentially impacting individual and group urinary fluoride measurement, such as variation in period of urine collection, variations/transient increases in excretion, variations in clearance times, as well as total fluoride exposure by age, sex, developmental stage, and over time.

  Response: Disagree (no change)

  - In responses to earlier comments from [redacted], we have pointed out reasons as to why we consider these concerns to be overstated and speculative. For example, for a study to be considered lower risk of bias for exposure, we required creatinine or specific gravity adjustments for measurements of urinary fluoride. We also cited studies reporting reasonable agreements between 24-hr urine and repeated-volume-corrected spot urine fluoride levels in the monograph.
  - We also note that there is no evidence to suggest that that the factors cited above (e.g., variation in period of urine collection) could account for the consistent direction of the association between fluoride in urine and children's IQ observed across the body of evidence. To do so, these factors would need to affect both fluoride exposure and children's IQ. If such evidence was provided, we would assess it.

In July 2022, the [redacted] provided comments to NIEHS/DNTP on the prepublication 2022 *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review*. These [redacted] comments were provided in tracked changes embedded in a Microsoft Word version of the monograph. The full text of [redacted] comments has been reproduced below verbatim in black text along with the specific sentence referred to by [redacted] as quotes under a heading for the specific section of the monograph (e.g., "Abstract"). Responses have been added in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] and [redacted]:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to [redacted] would be marked "*see DocB2_Monograph for detailed response*").
- [redacted] For comments related to DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

B2.1: ***Abstract page xii:*** *"Fluoride is a common exposure in our environment that comes from a variety of sources and is ~~widely~~ promoted for its dental and overall oral health benefits."* [Text in red-strikethrough font deleted by [redacted] without further comment.]

Response: Disagree (no change)

- We consider this sentence to be correct as written in the *Abstract* section.

B2.2: ***Abstract page xii:*** *"The evidence reviewed at that time was from dental and skeletal fluorosis-endemic regions of China with fluoride levels in water typically > XX mg/L."* [Text in red font inserted by [redacted] without further comment.]

Response: Disagree (no change)

- Additional detail to characterize the evidence cited in the NAS 2006 document is not necessary to support the statement and is beyond the focus of the paragraph. We note that dental and/or skeletal fluorosis has been reported in areas where high levels of fluoride are found in coal, as well as in areas with high levels in drinking water.

B2.3: ***Abstract page xiii****: "This review finds, with moderate confidence, that higher fluoride exposure (e.g., ~~represented by populations whose total fluoride exposure approximates or~~ exceed~~ings~~ the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride) is consistently associated with lower IQ in children."* [Text in red font inserted and red-strikethrough font deleted by [redacted].]

[redacted] **comment:** [redacted] suggest that this text be struck here, and elsewhere in the State of the Science (SOS) document as it implies that all fluoride exposures are known for all studies, and this is not the case.
The SOS reports states that drinking-water fluoride levels may underestimate exposure to fluoride.

Response: Disagree (no change)

- The parenthetical expression provides an example of what is meant by "higher fluoride exposure." This example was added to the prepublication 2022 NTP Monograph in response to many earlier comments that requested this clarification. The statement correctly emphasizes that total fluoride exposure from all sources is the important consideration, and it does not imply that all fluoride exposures are known for all studies.

B2.4: ***Abstract page xiii****: "More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ"*

[redacted] **comment***:* [redacted] suggest providing more detail in the abstract for added context. For example:
Associations between lower total fluoride exposure and children's IQ remain unclear. No population-level analysis was presented of effects from exposures to 0.7 mg/L, the U.S. Public Health Service recommended fluoridation level for community water systems for prevention of

dental decay. More studies at lower exposure levels are needed to fully understand potential associations in ranges typically found in the United States. Approximately 0.5 percent of community water systems in the U.S. have naturally occurring levels of fluoride > 1.5 mg/L water, the WHO guideline.

Response: Disagree (no change)

- The suggested text contains errors. First, the publication by Green et al. (2019) did examine the IQ of children in Canada exposed to fluoride in both non-fluoridated communities and fluoridated communities with 0.7 mg/L in drinking water (the same recommended fluoridation level for community water systems in the United States). Individual exposure levels of women living in these optimally fluoridated cities in Canada, as documented by repeated urinary measurements, suggest widely varied total exposures from water combined with fluoride from other sources. Many of these urinary fluoride measurements exceed those expected from consuming fluoride in water alone that contains 1.5 mg/L fluoride or less. The Bashash et al. (2017) study also provided information from a population in Mexico whose urinary fluoride exposures were comparable to those identified in the Green et al. (2019) study. Both studies are reviewed in our monograph and contribute to our confidence conclusions. Second, the suggested text implies that there are populations that could be studied where exposure to fluoride is only through drinking water. Our document stresses that fluoride exposures are from multiple sources, and that our confidence statements apply to total fluoride exposure, rather than exposures from drinking water alone. While we cite the number of people in the United States provided fluoridated drinking water from community water systems at >1.5 mg/L in the *Introduction* section, we consider the last line of the suggested addition again misdirects the reader from the emphasis on total fluoride exposure as the critical exposure metric.

B2.5: ***Introduction page 1****: "Fluoride is a common exposure in our environment from a variety of sources and is ~~widely~~ promoted for its dental and overall oral health benefits."* [Text in red-strikethrough font deleted by [redacted] without further comment.]

Response: Disagree (no change)

- We consider this sentence from the *Introduction* section to be correct as written.

B2.6: ***Introduction page 1****: "This level is the maximum amount of fluoride ~~contamination~~ (naturally occurring, not from water fluoridation) that is allowed in water from public water systems and is set to protect against increased risk of skeletal fluorosis, a condition characterized by pain and tenderness of the major joints."* [Text in red-strikethrough font deleted by [redacted] without further comment.]

Response: Disagree (no change)

- The official designation of the regulatory limit is the Maximum Contaminant Level, or MCL. Therefore, the sentence is correct as written.

B2.7: ***Introduction page 1****: "~~Commonly cited health~~ Health concerns ~~related~~ cited in relation to fluoride are bone fractures and skeletal fluorosis, lower intelligence quotient (IQ) and other neurological effects, cancer, and endocrine disruption."* [Text in red text inserted and red-strikethrough font deleted by [redacted] without further comment.]

Response: Agree (change made)

- The proposed edit was accepted as suggested.

B2.8: ***Methods page 10****: "The reference lists of all included studies; relevant reviews, editorials, and commentaries; and the Fluoride Action Network website ~~(http://fluoridealert.org)~~ were manually searched for additional relevant publications."* [Text in red-strikethrough font deleted by [redacted].]

[redacted] **comment**: The URL was already noted above.

Response: Agree (change made)

- The URL was deleted as suggested, pending implementation of any final formatting standards for NTP monographs.

B2.9: ***Methods page 17:*** *"Analytical methods to measure fluoride in biological or water samples also varied, some of which included atomic absorption, ion-selective electrode methods, colorimetric methods, or the hexamethyldisiloxane microdiffusion method."*

[redacted] **Comment:** Concerns were raised previously about the quality control of the analytical measurements.

Response: Disagree (no change)

- [redacted] has suggested that studies describing analytical methods without explicit methodological details on calibration and intra- and inter-day precision should not be rated as *probably low risk of bias* for exposure assessment. However, as we noted elsewhere, this seems to reflect a misunderstanding between criteria for *definitely low risk of bias* (which requires **direct** evidence of both a well-established method to measure fluoride and a QC procedure including such things as recovery rates, blanks, or reference standards) and *probably low risk of bias* (which requires that a well-established method to measure fluoride was used and **indirect** evidence that typical procedures were followed). The text in question is in the *Methods* section on risk-of-bias considerations for human studies, where the basic approach was outlined following prespecified criteria provided in the protocol (https://ntp.niehs.nih.gov/go/785076). Note that *Appendix E* provides details on the risk-of-bias ratings (including justifications) for each of the low risk-of-bias studies.

B2.10: ***Methods p 19:*** *Duan et al. (2018) reported a significant non-linear dose- response relationship between fluoride dose and intelligence with the relationship stated as most evident with exposures from drinking water above 4 mg/L (or 4 ppm) fluoride.*

[redacted] **comment:** Was a threshold considered?

Response: No change requested

- Duan et al. (2018) did not discuss specific considerations of threshold.

B2.11: ***Results p 41:*** *"Similarly, water fluoride concentrations for pregnant women from fluoridated areas (mean water fluoride levels of 0.59 ± 0.08 mg/L) versus pregnant women from non-fluoridated areas (mean water fluoride levels of 0.13 ± 0.06 mg/L) were associated with a significant 5.29-point decrease in IQ score per 1-mg/L increase in fluoride in both boys and girls combined (95% CI: −10.39, −0.19; p-value <0.05) (Green et al. 2019)."*

[redacted] **comment:** Is this actually observed (a 1 mg/L difference in fluoride concentrations leading to a 5.29 point decrease in IQ), or is this a predicted hypothetical effect from a model? If this is a modeled result rather than an observed result, should this be stated more clearly? Same comment may apply broadly.

Response: Disagree (edited for clarity)

- Green et al. (2019) presented the observed results from an adjusted linear regression analysis of water fluoride concentrations (mg/L) and children's FSIQ scores, providing the coefficient corresponding to a 1-mg/L difference in fluoride exposure. These are measured data, rather than hypothetical numbers.
- The sentence in question has been edited for clarity as follows:

  "Similarly, based on drinking water concentrations, a 1-mg/L increase of fluoride in drinking water was associated with a significant 5.29-point decrease in IQ score in both boys and girls combined (95% CI: −10.39, −0.19; p-value <0.05) (Green et al. 2019)."

B2.12: ***Results page 47:*** ***"****The results from 53 studies with high potential for bias that evaluated IQ in children also consistently provide supporting evidence of decrements in IQ associated with exposures to fluoride."*

[redacted] **comment:** At any exposure level? Please qualify.

Response: Disagree (no change)

- As discussed in the monograph, there is moderate confidence from low risk-of-bias studies of an association between higher fluoride exposure and lower IQ in children when total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L. Although the high risk-of-bias studies cover a range of fluoride exposure levels from below 1 mg/L to over 4 mg/L, we have not determined a level of confidence for specific exposure levels in the high risk-of-bias studies; therefore, we cannot further qualify this statement.

B2.13: ***Discussion page 78:*** *"Associations between lower total fluoride exposure [e.g., represented by populations whose total fluoride exposure was lower than the WHO Guidelines for Drinking- water Quality of 1.5 mg/L of fluoride (WHO 2017)] and children's IQ remain unclear."*

[redacted] **comment:** This language was commented on earlier. How is exposure in water correlated with overall exposure in this sentence?

Response: No change requested

- The statement relies on empirical observations of a close correspondence between drinking water concentrations and urinary fluoride concentrations first described prior to significant additional fluoride exposures from other sources such as dental products (see Kumar et al. (2017) [DOI 10.1007/s13201-016-0492-2] as an example). Our assessment of confidence in the association between higher fluoride exposure and lower children's IQ is supported by studies that report total fluoride exposures as represented by urinary measurements.

B2.14: ***Summary page 81:*** ***"*** *This review finds, with moderate confidence, that higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] is consistently associated with lower IQ in children."*

[redacted] **comment:** Relation between total fluoride exposure and 1.5 mg/l level is not clear.

Response: Disagree (no change)

- See response to the preceding comment.

B2.15: ***Summary page 81:*** ***"****This review finds, with moderate confidence, that higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] is consistently associated with lower IQ in children. Populations with exposure to >1.5 mg/L naturally occurring fluoride in water represent 0.59% of the U.S. population."* [Text in red font inserted by [redacted].]

[redacted] **comment:** Might want to add, if correct, primarily from very small water systems – please check EPA website.

Response: Disagree (no change)

- The proposed reference to naturally occurring fluoride in water comes from a periodic survey of community water systems in the United States carried out by the Centers for Disease Control and Prevention. However, the moderate confidence is based entirely on studies outside the United States. The proposed sentence also focuses on fluoride in water systems rather than "total fluoride exposure" as in the current text. In addition, it is not clear what [redacted] is referring to as "very small water systems" or for what purpose this additional information would serve.

References

Kumar S, Lata S, Yadav J, Yadav JP. 2017. Relationship between water, urine and serum fluoride and fluorosis in school children of Jhajjar District, Haryana, India. Applied Water Science. 7:3377–3384. doi: 10.1007/s13201-016-0492-2.

In June 2022, the [redacted] provided comments to NIEHS/DNTP on the prepublication 2022 *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review*. These [redacted] comments were embedded in a PDF version of the monograph. The full [redacted] have been reproduced below verbatim along with the specific monograph text referred to by [redacted] in quotes and the section and page number of the monograph (e.g., "Summary page 81"). Responses have been added in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] comments and [redacted] comments:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to this [redacted] would be marked "*see DocC_Monograph for detailed response*").
- [redacted] For comments related to DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

[redacted] Comments on the NTP State of the Science document

C.1: ***Abstract page xii:*** *"A systematic review protocol was developed and utilized following the standardized OHAT systematic review approach for conducting literature-based health assessments."*

[redacted] **Comment:** See NASEM comment about the protocol on p. 4 of Response to Fluoride NASEM Letter 10.5.2021:"Although the statement clarifies the general role of the handbook, the committee finds that it does not address the committee's previous recommendation to set the expectation for how closely the process described in the handbook will be followed in the protocol and in the eventual systematic review. For example, the handbook section "Key Questions and Analytical Framework" that guides development of the population, exposure, comparator, and outcomes (PECO) statement is not included in the fluoride protocol or the revised monograph. As the committee recommended in its previous review, NTP should treat each systematic review protocol as a stand-alone document that contains all the information necessary for understanding of the planning and conduct of the review, and these expectations should be explicitly stated in the protocol. The committee did not find that revisions of the protocol adequately addressed this recommendation.

Response: Disagree (edited for clarity)

- The response here mirrors the response to the original NASEM Committee comment. We appreciated the desire of the NASEM Committee for more specificity in the protocol with respect to laying out all aspects of the systematic review; however, we respectfully submit that the specificity and level of detail provided in the protocol meet, and in many aspects exceed, standard practice in the field. The following text was added to the *Methods* section of the prepublication 2022 NTP Monograph to further clarify the role of the OHAT handbook and the stand-alone nature of the protocol.

  "The protocol served as the complete set of methods followed for the conduct of this systematic review. The OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673) is a source of general systematic review methods that were selected and tailored in developing this protocol. Options in the OHAT handbook that were not specifically referred to in the protocol were not part of the methods for the systematic review."

  Optional approaches in the OHAT handbook such as the formulation of "key questions" or an "analytical framework" were not necessary and not conducted for the fluoride systematic review.

C.2: ***Abstract page xii:*** *"The current bodies of experimental animal studies and human mechanistic evidence do not provide clarity on the association between fluoride exposure and neurocognitive or neurodevelopmental human health effects."*

[redacted] **comment:** IMPT Conclusion- the lack of biological plausibility or mechanistic evidence is critical and should be weighted more heavily in assessments of human impact.

Response: Disagree (no change)

- As pointed out in the *Discussion* section of the monograph, "*Mechanistic studies in humans provide some evidence of adverse neurological effects of fluoride. However,*

these studies were too heterogenous and limited in number to make any determination on biological plausibility." As indicated in the *Limitations of the Evidence Base* subsection, "*The understanding of the specific molecular events responsible for fluoride's adverse effects on neurobehavioral function is poor.*" The prepublication 2022 NTP Monograph clearly states that the moderate confidence expressed in the association between higher fluoride exposures and children's IQ is based on the consistent pattern of findings in human epidemiological studies. It is not unusual to observe and appreciate the potential for human health effects of a given exposure in the absence of understanding the mechanistic events responsible.

C.3:

[redacted] **comment:** More recent studies have shown "The discrepancy between experimental and epidemiological evidence may be reconciled with deficiencies inherent in most of these epidemiological studies on a putative association between fluoride and intelligence, especially with respect to adequate consideration of potential confounding factors, e.g., socioeconomic status, residence, breast feeding, low birth weight, maternal intelligence, and exposure to other neurotoxic chemicals. In conclusion, based on the totality of currently available scientific evidence, the present review does not support the presumption that fluoride should be assessed as a human developmental neurotoxicant at the current exposure levels in Europe."

Guth S et al. Toxicity of fluoride: critical evaluation of evidence for human developmental neurotoxicity in epidemiological studies, animal experiments and in vitro analyses. Archives of Toxicology (2020) 94:1375–1415.

Response: Disagree (no change)

- We appreciate the thoughts and opinions of Guth et al. (2020) on the collective potential deficiencies in the body of human evidence on whether fluoride acts as a developmental neurotoxicant. However, we point out that their publication fails to provide a critical evaluation of the majority of the studies they cite, instead focusing on Green et al. (2019) and Broadbent et al. (2015), two studies that differ appreciably in design and quality, to support their case. Nonetheless, the overall conclusion by Guth et al. (2020), "The available epidemiological evidence does not provide sufficient arguments to raise concerns with regard to CWF in the range of 0.7–1.0 mg/L, and to justify the conclusion that fluoride is a human developmental neurotoxicant that should be categorized as similarly problematic as lead or methylmercury at current exposure levels," is not at odds with our conclusion.

C.4: ***Abstract page xiii:*** *"This review finds, with moderate confidence, that higher fluoride exposure (e.g., represented by populations whose total fluoride exposure approximates or exceeds the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride) is consistently associated with lower IQ in children."*

[redacted] **comment**: Based upon the standards for "Quality" criteria, the confidence estimate appears overstated. www.gradeworkinggroup.org

Response: Disagree (no change)

- Our approach for assessing confidence in the body of evidence is indeed a GRADE-based methodology, as described in the *Methods* subsection of the monograph, *Confidence Rating: Assessment of the Body of Evidence*, with additional details provided in the protocol for this systematic review (https://ntp.niehs.nih.gov/go/785076) and the OHAT handbook (https://ntp.niehs.nih.gov/go/ohathandbook). Also note that the GRADE working group has updated its terminology over the years to refer to "certainty" or confidence in the body of evidence to avoid confusion with the term "quality." The principal benefits of GRADE-based approaches are the consistency of the steps and transparency in the process of developing confidence ratings and documenting the scientific bases for these judgements. In the GRADE-based method, confidence is typically assessed separately for each outcome (e.g., IQ) because confidence in the body of evidence often varies between outcomes and age groups (e.g., children versus adults). There are three separate subsections in the *Results*—one for *IQ in Children*, one for *Other Neurodevelopmental or Cognitive Effects in Children*, and another for *Cognitive Effects in Adults*—where the body of evidence is described and critically assessed specifically to develop confidence ratings. The *Confidence Assessment of Findings on IQ in Children* subsection describes in detail how the data support moderate confidence in the body of evidence that higher exposure to fluoride is associated with lower IQ in children. Each of the GRADE-based factors considered for potentially increasing (e.g., dose response) or decreasing (risk of bias) confidence in the body of evidence is described along with evidence-based support of judgements for each factor.

C.5: [redacted] **comment**: The data do not support the assertion of an effect below 1.5 mg/L. Therefore, all conclusory statements in this document should be explicit that any findings from the included studies only apply to water fluoride concentrations above 1.5 mg/L.

Response: Disagree (no change)

- We do not agree with this comment. Our assessment considers fluoride exposures from all sources, not just water. As discussed in the prepublication 2022 NTP Monograph, because fluoride is also found in certain foods, dental products, some pharmaceuticals, and other sources, individual behaviors are likely an important determinant of actual exposures. Even in the optimally fluoridated cities in Canada studied by Green et al. (2019), individual exposure levels, as documented by repeated urinary measurements, suggest widely varying total exposures from water combined with fluoride from other sources. For example, some urinary fluoride measurements exceed those that would be expected from consuming water that contains fluoride at 1.5 mg/L. While much of the literature evaluating exposures to fluoride and reduced cognition in children involves total exposures to amounts assumed to approximate or exceed the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride, we make it clear that our assessment considers total fluoride exposure from all sources, not just drinking water alone.

C.6: ***Abstract page xiii:*** *"More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ."*

[redacted] **comment:** Suggest alternative language: There is a need to develop basic guidelines for designing and conducting prospective population-based (epidemiological) fluoride studies relevant to diverse communities and at fluoride exposure levels (0.7 mg/L) recommended in the United States.

Response: Disagree (no change)

- This statement may well be correct, but the recommendation is beyond the stated objective and *Specific Aims* of our systematic review as described in the protocol and document. The prepublication 2022 NTP Monograph does not attempt to establish support for this point.

C.7: ***Preface page xiv:*** *"Overall the revised monograph seems to include a wealth of evidence and a number of evaluations that support its main conclusion, but the monograph falls short of providing a clear and convincing argument that supports its assessments…. Thus, NTP has removed the hazard assessment step and retitled this systematic review of fluoride exposure and neurodevelopmental and cognitive health effects as a "state-of-the-science" document to indicate the change."*

[redacted] **comment:** If the monograph does not provide a clear and convincing argument in support of its assessments--- removing one step doesn't strengthen the validity of the assertions. Note: many conclusions in this SOS monograph seem to reflect the hazard conclusions from the previous version of the monograph

Response: Disagree (no change)

- The comment refers to a statement from the NASEM Committee review of the 2020 draft NTP Monograph that reached a hazard conclusion that fluoride was *presumed to be a cognitive neurodevelopmental hazard to humans*. This text from the *Preface* explains that the hazard assessment step was removed from the current prepublication 2022 NTP Monograph to address the NASEM Committee's comment on "clear and convincing argument" for the NTP hazard conclusions in the 2020 draft NTP Monograph. The goal of the current, extensively revised monograph is to provide a comprehensive assessment of the scientific literature on fluoride as an important resource to inform its safe and appropriate use. The prepublication 2022 NTP Monograph includes a number of additional studies and provides the most complete and transparent critical assessment of the human epidemiological literature to date.

C.8: ***Preface page xiv:*** *"This state-of-the-science document does not include the meta-analysis of epidemiological studies or hazard conclusions found in previous draft monographs; however, it provides a comprehensive and current assessment of the scientific literature on fluoride as an important resource to inform safe and appropriate use."*

[redacted] **comment**: This SOS report includes Appendix A which presents all the data and analysis from the meta-analysis found in previous draft monograph... therefore, the weakness identified by NASEM: "but the monograph falls short of providing a clear and convincing argument that supports its assessments persists. " [Note that the word "persists" should be outside the quote.]

Response: Disagree (no change)

- As stated in this quote from the *Preface*, the hazard conclusions were removed to address the NASEM Committee's comment on "clear and convincing argument" for the NTP hazard conclusions in the 2020 draft NTP Monograph. *Appendix A* (Data Figures: Neurodevelopmental or Cognitive Effects and Outcomes) presents results of low risk-of-bias studies evaluated in the prepublication 2022 NTP Monograph that formed the basis for the confidence statements reached for children's IQ studies, children's other cognition and neurobehavior studies, and adult cognition studies. *Appendix A* does not include results of the meta-analysis and is not meant to provide a clear and convincing argument. The main text in conjunction with *Appendix E* provide support for our assessment and overall confidence rating.

C.9: [redacted] **comment**: Removing the meta-analysis as a response to NASEM comments would not remedy the shortcoming they cited: "but the monograph falls short of providing a clear and convincing argument that supports its assessments?

Response: Disagree (no change)

- This comment is addressed in the previous two responses. Again, it is inappropriate to attribute the NASEM Committee's comments on the prior 2020 draft NTP Monograph to the prepublication 2022 NTP Monograph.

C.10: ***Introduction page 1:*** *"Commonly cited health concerns related to fluoride are bone fractures and skeletal fluorosis, lower intelligence quotient (IQ) and other neurological effects, cancer, and endocrine disruption."*

[redacted] **comment**: This statement is inflammatory. It is not a reflection of the current state of the science on this issue. However, these assertions that have been made by the Fluoride Action Network and are not evidenced-based. Ref: Osteoporos Int. 2008 Mar;19(3):257-68. Epub 2007 Aug 15.

Effects of treatment with fluoride on bone mineral density and fracture risk--a meta-analysis

P Vestergaard 1, N R Jorgensen, P Schwarz, L Mosekilde

Affiliations expand PMID: 17701094

Response: Agree (change made)

- We disagree that the statement is inflammatory and also note that the sentence is historically accurate in the scientific literature. Nonetheless, we have removed the reference to bone fractures. The sentence has been revised to read as follows:

  "Health concerns cited in relation to fluoride are skeletal fluorosis, lower intelligence quotient (IQ) and other neurological effects, cancer, and endocrine disruption."

C.11: ***Introduction page 3:*** *"However, the NASEM Committee's reviews (NASEM 2020; 2021) of the 2019 and 2020 drafts of the monograph indicated that, "Overall the revised monograph seems to include a wealth of evidence and a number of evaluations that support its main conclusion, but the monograph falls short of providing a clear and convincing argument that supports its*

assessments….” For this reason, our methods were revised to remove the hazard assessment step (i.e., the section “Integrate Evidence to Develop Hazard Identification Conclusions” and the associated section “Translate Confidence Ratings into Level of Evidence for Health Effect”).”

[redacted] **comment**: If NASEM stated that the monograph did not provide clear and convincing argument that supports its assessment, then removing the hazard assessment would not change the strength of the evidence. Also, why remove the meta-analysis from the context of the report?

Response: Disagree (no change)

- The hazard assessment step, leading to the hazard conclusion referred to in the NASEM comment, integrates information across evidence streams after each evidence stream is assigned a confidence rating. As previously pointed out, the current document stops at the confidence rating step and provides a transparent assessment of studies to support confidence ratings for children’s IQ, children’s other cognitive and neurobehavioral outcomes, and adult cognition. The finding of moderate confidence in the body of evidence concerning the association between higher fluoride exposures and lower IQs in children is unchanged from earlier drafts of the monograph.
- The meta-analysis was removed for separate publication because we did not consider it necessary to reach a confidence rating for children’s IQ in the prior drafts of the monograph. Indeed, the current draft of the meta-analysis is careful to point out that the collective quantitative assessment of the children’s IQ studies is based on a systematic review that supported moderate confidence in the association between higher fluoride exposures and deficits in children’s IQ.

C.12: ***Introduction page 3:*** *“Identify literature that assessed neurodevelopmental and cognitive health effects, especially outcomes related to learning, memory, and intelligence, following exposure to fluoride in human, animal, and relevant in vitro/mechanistic studies.”*

[redacted] **comment**: If an element or substance is known to be in the environment, then EVERYTHING would qualify as "following" an exposure... therefore, this would be measuring prevalence of learning, memory, and intelligence in the study population.

Response: Disagree (no change)

- As defined in the PECO (Population, Exposure, Comparator and Outcome) statement in Table 1 of the monograph, exposure to fluoride in human studies is based on administered dose or concentration, biomonitoring data (e.g., levels in urine, blood, other specimens), environmental measures (e.g., levels in air, water), or job title or residence. Furthermore, the temporality of the exposure preceding outcome can be established by study design (e.g., prospective cohort) or analysis (e.g., prevalence of dental fluorosis in children, limiting study populations to children who lived in the same area for long periods of time).

C.13: ***Methods page 6:*** *“Comparable populations not exposed to fluoride or exposed to lower levels of fluoride (e.g., exposure below detection levels).”*

[redacted] **comment**: how would you determine "detection" levels if you are not measuring the dose of exposure? Not clear how comparable popltns *(sic)* of not exposed are equivalent or appropriate to use in lieu of "exposed to lower levels of fluoride"

Response: Disagree (edited for clarity)

- The PECO (Population, Exposure, Comparator and Outcome) statement in Table 1 of the monograph defines the parameters for human studies to be included in the systematic review. The exposure requirements are stated as studies with exposure to fluoride based on "*administered dose or concentration, biomonitoring data (e.g., urine, blood, other specimens), environmental measures (e.g., air, water levels), or job title or residence*" with additional specifications outlined in the text. Similarly, the comparator requirements indicate that a study must include a comparator population in addition to individuals or populations exposed to fluoride. Following the standard approach for epidemiological studies, a comparable population must be either not exposed to fluoride or exposed to lower levels of fluoride. The parenthetical "*(e.g., exposure below detection limits)*" was not intended to define lower levels of fluoride and was not used as such for the literature search or elsewhere in the evaluation. The "*exposure below detection limits*" phrase was meant as an example where measures below detection limits would be considered a population not exposed to fluoride. We moved the phrase as follows and inserted the following footnote on how the criteria were used in the evaluation.

  "Comparable populations not exposed to fluoride (e.g., exposure below detection levels) or exposed to lower levels of fluoride"

  Footnote: *Note: The "(e.g., exposure below detection limits)" was moved after "populations not exposed to fluoride" to reflect how it was used in the literature search and elsewhere in this systematic review.*

C.14: ***Methods page 8:*** *"Studies identified from other sources or manual review that might impact conclusions are considered under "references identified through other sources" in Figure 2."*

[redacted] **comment**: what is the relationship of the 11 studies identified through these means proportional to those included from the database searches? it is concerning that there were 11 studies that were NOT identified in the database search that may have been important... does this represent literature/study selection bias?

Response: Disagree (edited for clarity)

- The monograph provides information as to why the references identified by other sources were not captured in the database searches. Note that many of the studies initially identified by other sources were non-English-language studies, and we recognized that additional targeted search strategies were required to identify non-English-language studies for this review. The supplemental search of Chinese databases was designed to address these challenges. This is described in the monograph as follows:

  "These 11 studies (9 human and 2 animal studies) were not identified through the electronic database searches, as they were not indexed in any of the electronic databases searched. Note that the supplemental search of non-English-language

databases was designed in part to identify non-English-language studies that are not indexed in traditional bibliographic databases such as PubMed. It was successful in this goal, as multiple studies that were initially only identified through "other sources" were subsequently captured in the supplemental Chinese database search, leaving only 11 as identified through other sources."

- Note that all 11 studies were published in non-Western journals. Regarding the source for identifying these 11 studies, the monograph describes the sources as "*…identified by technical advisors or obtained by manually searching the Fluoride Action Network website or reviewing reference lists of published reviews and other included studies*." Of the nine human studies identified through other sources, five were identified via their inclusion in the Choi et al. (2012) meta-analysis, and four were only identified in the FAN database (three of which were Indian studies).
- Regarding the impact of these 11 studies on the systematic review, only 1 of the 11 studies was a low risk-of-bias IQ study in children, and this study was included in the 19 low risk-of-bias studies upon which the moderate confidence rating for the IQ-in-children body of evidence is based. The omission of this single study would not impact the moderate confidence rating. Of the remaining 10 studies, 7 were high risk-of-bias studies of IQ in children and 1 was a high risk-of-bias study of adults. The inclusion or omission of the 7 (out of 53) high risk-of-bias IQ-in-children studies or the 1 (out of 8) high risk-of-bias adult studies would not impact any confidence conclusions in the monograph. Similarly, the two experimental animal studies would not impact the evaluation as the animal evidence was considered inadequate.
- The text identified by [redacted] was edited for clarity replacing "might impact conclusions" with "satisfy the PECO criteria for inclusion" as follows:

  "Studies identified from other sources or manual review that satisfy the PECO criteria for inclusion are considered under "references identified through other sources" in Figure 2."

C.15: ***Methods page 9:*** *"Although additional studies were identified, data that would materially advance the animal and mechanistic findings were not identified; therefore, these studies were not extracted nor were they added to the draft."*

[redacted] **comment:** However, why select older studies when more current ones are available?

Response: Disagree (no change)

- The comment refers to the decision to not update the experimental animal and animal mechanistic study sections in the 2020 draft NTP Monograph or the prepublication 2022 NTP Monograph. The review did not select older studies when newer studies were available. As stated, when the literature review was updated through May 2020, newer literature was scanned for information that could materially extend the findings of experimental animal or mechanistic studies, which previously had been determined to be inadequate to affect the confidence level based on the human studies. Newer studies did not materially extend the earlier findings. Furthermore, consideration of the newer studies did not change the determination that these data were inadequate; therefore, they were not extracted or added to the document.

C.16: ***Methods page 9:*** *"A secondary goal was to examine whether the non-English-language studies on the Fluoride Action Network website (http://fluoridealert.org/)—a site used as another resource to identify potentially relevant studies because it is known to index fluoride publications—had been selectively presented to list only studies reporting effects of fluoride."*

[redacted] **comment:** what is the conclusion on whether they are selectively presented?

Response: No change requested

- We saw no indication that the studies were selectively presented on the FAN website.

C.17: ***Methods page 9:*** *"Studies identified that evaluated primary neurodevelopmental or cognitive outcomes were included and either translated or reviewed by an epidemiologist fluent in Chinese."*

[redacted] **comment:** There are many Chinese dialects and thus, interpretations and translations may vary. Are these made available?

Response: Disagree (no change)

- Almost all the Chinese-language literature used in this evaluation was read and underwent data extraction by an epidemiologist fluent in Chinese. There were no instances where this was hindered by dialect. As indicated in the *Risk-of-bias Considerations for Humans Studies* section on pages 14-15 of the prepublication 2022 NTP Monograph, some of the Chinese-language literature was also available as English translations, and these studies are listed in *Appendix C*, Section 2.1, *List of Included Studies*. In addition, an interactive version of the study selection diagram is publicly available in the Health Assessment Workspace Collaborative that can be used to search for individual studies and their bibliographic information (https://hawcproject.org/summary/visual/assessment/405/Figure-2/). Most of the Chinese-language studies were determined to be high risk of bias. For papers that were considered potentially low risk of bias based on the English translation, the accuracy of the translation was verified by the epidemiologist fluent in Chinese.

C.18: ***Methods page 9:*** *"Supplemental Chinese Database Literature Search."*

[redacted] **comment:** Why not just utilize what was identified through Main Literature Database Search?

Response: Disagree (no change)

- A principal tenant of a systematic review is to find all literature related to the question being addressed. Many of the studies on populations exposed to fluoride are from China, which has a large scientific literature and studies may be published in non-English-language journals that are not indexed in U.S. databases. Therefore, we developed the supplemental literature search of Chinese databases to address this potential issue. A number of non-English studies were also identified by searching other

sources. For example, the reference lists of all included studies and relevant reviews/meta-analyses were manually searched for additional relevant references.

C.19: ***Methods page 10:*** *"The reference lists of all included studies; relevant reviews, editorials, and commentaries; and the Fluoride Action Network website (http://fluoridealert.org/) were manually searched for additional relevant publications."*

[redacted] **comment:** Inclusion of these "other resources" likely bias the environmental scan given that FAN posts studies skewed toward detrimental effects?

Response: Disagree (no change)

- As previously stated, we have no tangible evidence to support this assertion. However, to address "other resources" as described in the quoted sentence, our search included the manual scanning of reference lists of all included studies, relevant reviews, etc., not just the FAN website.

C.20: ***Methods page 14:*** *"Quality Assessment of Individual Studies"*

[redacted] **comment:** Consider this heading might be renamed Risk of Bias Assessment since this section is more about Risk-of-bias than Quality and since risk of bias is not the same as "quality".

""Quality" as used in GRADE means more than risk of bias and so may also be compromised by imprecision, inconsistency, indirectness of study results, and publication bias." https://pubmed.ncbi.nlm.nih.gov/21208779/ <http://cccrg.cochrane.org/author-resources. Version 3.0 December 2016

Response: Disagree (no change)

- [redacted] is correct that, historically, GRADE used the term "quality" to refer to more than risk of bias and that use of the term included imprecision, risk of bias, and other factors for the evaluation of the body of evidence. However, it is precisely this confusion that caused GRADE to move away from the term "quality" for the GRADE framework for assessing certainty in the evidence. GRADE has updated its use of terminology over the years to refer to "certainty" or confidence in the body of evidence where it once used the term "quality." Unfortunately, GRADE maintains both terms, which may have led to [redacted] confusion. We have been very careful with our terminology and use the term "quality" when describing our overall process for assessing individual studies, not just risk of bias in an effort to reach those who are unfamiliar with risk of bias and to reflect that our method does consider factors that are not strictly risk of bias (e.g., methodological considerations under outcome assessment). In addition, our approach specifically uses "assessment of individual studies" for quality and risk of bias to avoid confusion with the evaluation of the body of evidence when rating confidence.

C.21: ***Methods page 15:*** *"The remaining studies (i.e., other than the high risk-of-bias studies) were considered to have lower potential for bias (i.e., low risk of bias) and to be of high quality."*

[redacted] **comment:** An assessment and determination of low bias should not automatically translate to a study being classified as high quality. Suggest that these should be handled/assessed separately and a separate set of criteria to determine the quality. It does not appear as though *any* low risk of bias studies were rejected for "quality" reasons

Response: Disagree (no change)

- Please see the previous response, as [redacted] appears to be referring to the historical use of the term "quality" to refer to more than just risk of bias. However, it is precisely this confusion that caused GRADE to move away from the term "quality" over the years and to instead use "certainty" or confidence in the evidence where it once used the term "quality."
- As explained in this text from the section on *Quality Assessment of Individual Studies* in the *Methods* section of the monograph, the terms "high quality" and "low risk of bias" are being used synonymously in the prepublication 2022 NTP Monograph.

C.22: ***Methods page 16:*** *Exposure was assessed using a variety of methods in the human body of evidence thereby introducing heterogeneity across the selected studies and complicating the comparison across study findings.* [Text in red font added by reviewer.]

[redacted] **comment:** A consistent critique of the evidence-base is the heterogeneous measures of fluoride exposure, the absence of precise dose measurement, and measurement methods that do not allow an evaluation of cumulative fluoride exposure.
These weaknesses in exposure estimates have the potential to produce misclassification bias.

Response: Disagree (no change)

- We discuss the limitations of the different types of fluoride exposure assessment methods, including those mentioned in this comment, in our *Risk of Bias Considerations for Human Studies* section of the monograph. Any potential misclassification bias in exposure measurement methods, and likely direction of bias, would have been described in the risk of bias assessment (*Appendix E*). We noted no evidence that cumulative exposure was necessary as cognitive deficits were identified in children of all ages tested.

C.23: ***Methods page 16:*** *"Despite potential issues with spot urine samples, if authors made appropriate efforts to reduce the concern for bias (e.g., accounting for dilution), studies that used this metric were generally considered to have probably low risk of bias for exposure."*

[redacted] **comment:** The approach to classifying studies for risk-of-bias using spot urine samples solely based upon the study authors' "appropriate efforts" is concerning. The assertion of correlation between spot urine and 24-hour samples is not scientifically sound--- The Zohouri study referenced here as substantiating evidence is based upon a sample size of n=7 children aged 1-3 years.

Response: Disagree (no change)

- The conditions under which spot urine samples can be considered to support a determination of *probably low risk of bias* are discussed in the *Exposure* subsection of *Risk-of-bias Considerations for Human Studies*. The comment does not provide reasons to explain why the approach described in that subsection is not scientifically sound.

C.24: ***Methods page 19:*** *"Probably Low risk of bias: There is indirect evidence of low risk-of-bias practices, OR it is deemed that deviations from low risk-of-bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias."*

[redacted] **comment:** Cases of unknown or undocumented bias were considered as indirect evidence and may therefore have been misclassified as Low risk rather than 'Unknown/undocumented" which would in other analyses be considered high risk.

Response: Disagree (no change)

- It is a misstatement that cases of unknown or undocumented bias were considered as indirect evidence. If information to make a risk-of-bias judgement was not available, it was categorized as "not reported," which is equivalent to *probably high risk of bias*.

C.25: ***Methods page 20:*** *"Furthermore, the review did not exclude subjects exposed in occupational settings. All exposure levels and scenarios encountered in human studies are considered direct (i.e., applicable, generalizable, and relevant to address the objective of the assessment); therefore, a downgrade for indirectness would not be applied to bodies of evidence from human studies."*

[redacted] **comment:** The explanation of "Indirectness" assigning study subjects as "all humans"- provides NTP the ability to disregard "indirectness" in its totality as a Quality criterion. This creates a scenario where for example, a study of 90 year old retired Chinese coal miners could be considered "direct" evidence of exposure applying to children in the United States or elsewhere.

Response: Disagree (no change)

- As is made clear in the monograph, independent confidence statements are provided for our assessment of fluoride exposures in relation to children's IQ, children's cognitive and neurobehavior outcomes other than IQ, and adult cognition. Therefore, the hypothetical study of 90-year-old Chinese coal miners would be considered in relation to other studies on adults, and not children.

C.26: ***Results page 53:*** *"We conclude that there is moderate confidence in the body of evidence that higher fluoride exposure is associated with lower IQ in children."*

[redacted] **comment:** Based upon the standards for Quality criteria , the confidence estimate is overstated. www.gradeworkinggroup.org

Response: Disagree (no change)

- This comment is essentially the same as one [redacted] made previously on the *Abstract* section; therefore, we present a brief version of that response here. Our approach for assessing confidence in the body of evidence is indeed a GRADE-based methodology as described in the *Methods* subsection *Confidence Rating: Assessment of the Body of Evidence*, the protocol, and the OHAT handbook. The principal benefits of GRADE-based approaches are the consistency of the steps and transparency in the process of developing confidence ratings and documenting the scientific bases for these judgements. The *Confidence Assessment of Findings on IQ in Children* section of the monograph describes in detail how the data support moderate confidence in the body of evidence that higher exposure to fluoride is associated with lower IQ in children.

C.27: ***Results page 54:*** *" The initial moderate confidence rating is based on 15 of the 19 low risk-of-bias studies that have 3 of the 4 key study design features shown in Figure 1 (i.e., exposure occurred prior to outcome, individual-based outcomes were evaluated, and a comparison group was used)."*

[redacted] **comment:** Meeting these three design features does not reflect the "quality" of the studies and therefore, calls into question the classification as Moderate confidence.

Response: Disagree (no change)

- [redacted] is referring to one step in the process of assessing confidence in the body of evidence. We agree that this single sentence alone does not reflect the entire approach. The sentence preceding the one identified by [redacted] outlines the steps, explaining that, "*This confidence rating was reached by starting with an initial confidence rating based on key study design features of the body of evidence and then considering factors that may increase or decrease the confidence in that body of evidence*." The following sentences discuss how each of the factors is considered for the body of evidence on IQ studies in children and refers to Figure 1 in the *Methods* section. As is clearly stated in the *Methods* subsection on *Organizing and Rating Confidence in Bodies of Evidence* on pages 19-22 and illustrated in Figure 1, the key study design features are used to set an initial confidence rating, which is then subjected to potential upgrades or downgrades for all the factors discussed (e.g., risk of bias, consistency, indirectness, imprecision, publication bias, magnitude of effect). These factors are considered collectively when determining the final level of confidence in the evidence base.

C.28: ***Results page 54:*** *"**Unexplained inconsistencies:** The data are consistent, and there was no downgrade for this factor. Eighteen of the 19 low risk-of-bias studies reported associations between higher fluoride levels and lower IQ scores in children. These studies were conducted in 5 different countries on more than 7,000 children from 15 different study populations. There is consistency in results across prospective and cross-sectional study designs. There is also consistency in results across studies using different fluoride exposure measures, including urinary and drinking water fluoride. The one study that did not observe an association did not provide results in a comparable manner and therefore this body of evidence is not considered to have unexplained inconsistencies."*

[redacted] **comment:** However, exposure measures, tests and scales used across the studies were not consistent or standardized. A visual assessment of Figures A1 through A5 appears that a claim of consistency is not supported by the evidence presented. The majority of the findings presented fall in the null effect range and positive findings often overlap with reference findings.

Response: Disagree (no change)

- We disagree with this comment. Support for these statements on the consistency of the data are based on our detailed assessment of the IQ studies as described in the *Results* section on pages 40-47, which summarizes the results of the low risk-of-bias IQ studies. Consistency refers to the direction of the association between fluoride exposure (at any level) and children's IQ. Figures A1–A5 are referenced in the text; however, our statement of consistency was not developed by visual inspection of these figures.

C.29: ***Results page 54:*** *"**Indirectness:** IQ in humans is a direct measure of the association of interest; therefore, no adjustment in confidence is warranted."*

[redacted] **comment:** The definition exempts NTP from all criticism related to indirectness among all included human studies. This in turn allows NTP to include additional studies outside of the initial search criteria established in their protocol, e.g. including two non-English, databases, as well as FAN identified literature and considering those studies as "direct evidence" and relevant to U.S. populations.

Response: Disagree (no change)

- The statement concerning indirectness simply states that studies that directly measure IQ in children are not downgraded for indirectness because they directly measure the outcome of interest in children. We disagree that this this provides us license to expand our search strategy (which was conducted in response to a NASEM Committee recommendation). We also point out in the *Limitations of the Evidence Base* section that the absence of any comparable studies on U.S. populations is a data gap.

C.30: ***Results page 54:*** *"There is no evidence of imprecision that would warrant a downgrade. Eighteen studies reported lower IQ with higher fluoride, and no issues with imprecision were identified to challenge the significance of the effect estimate."*

[redacted] **comment:** Imprecision clearly evident upon visual inspection of Figures A1 through A5, which frequently shows wide and overlapping confidence intervals. Therefore, this discussion item should be revised.

Response: Disagree (edited for clarity)

- As stated above, we disagree with this comment based on our detailed assessment of the IQ studies as described in the *Results* section on pages 40-47, which summarizes the results of the low risk-of-bias IQ studies. However, we agree that the referenced statement could be more precise with respect to the criteria for considering a downgrade in confidence based on imprecision as outlined in the *Methods* section on page 21 of the monograph. Therefore, this text was edited to read as follows:

"There is no evidence of serious imprecision that would warrant a downgrade. Eighteen low risk-of-bias studies reported lower IQ with higher fluoride, and no issues with imprecision were identified to challenge the significance of the response estimates."

C.31: ***Results page 54:*** *"**Publication bias:** There is no strong evidence of publication bias; therefore, no downgrade was applied for publication bias."*

[redacted] **comment:** NTP does not seem to have adjusted its methodology in response to NASEM critique p.22 NASEM review response (Sept 2021): "In addition to what it has presented, it should mention the weaknesses of the tests used to evaluate that bias. One weakness is that the evaluation of the funnel plot involves mostly a subjective interpretation, which can be especially troublesome when the number of studies is small. Another weakness is the possibility that positive results from the funnel plot and the Egger and Begg tests might be caused by something other than publication bias. In addition, NTP uses the phrase "eliminating publication bias" when it refers to the results of the trim and fill analyses (see, for example, NTP 2020a, p. 49). However, because the tests for publication bias are not 100% specific, it is not known exactly what is being eliminated by the trim and fill process. The committee suggests that a better phrase might be "adjusting for possible publication bias." In summary, acknowledging the weaknesses of the tests that were used to evaluate publication bias would make the report more transparent."

Response: Agree (no change)

- This [redacted] comment is no longer relevant to the prepublication 2022 NTP Monograph because the phrase quoted by NASEM is not in the monograph. However, this NASEM Committee's comment has been fully addressed in the meta-analysis manuscript (e.g., phrasing has been revised and a limitation has been added). Please see our full response to this NASEM comment in document "Sup01_Meta-analysis_NASEM_Feb_2021."

C.32: ***Results page 55:*** *"The magnitude of effect size and the overall strength and quality of the human literature base provide moderate confidence in the body of evidence that higher exposure to fluoride is associated with lower IQ in children (see the Discussion section for strengths and limitations of the evidence base)."*

[redacted] **comment:** The data do not support the assertion of an effect below 1.5 mg/L. Therefore, all conclusory statements in this document should be explicit that any findings from the included studies only apply to water fluoride concentrations above 1.5 mg/L.

Response: Disagree (no change)

- We do not agree with this comment. As explained in a previous response, our assessment considers fluoride exposures from all sources, not just water. As discussed in the prepublication 2022 NTP Monograph, because fluoride is also found in certain foods, dental products, some pharmaceuticals, and other sources, individual behaviors are likely an important determinant of actual exposures. Even in the optimally fluoridated cities in Canada studied by Green et al. (2019), individual exposure levels, as documented by repeated urinary measurements, suggest widely varying total exposures from water combined with fluoride from other sources. For example, some urinary

fluoride measurements exceed those that would be expected from consuming water that contains fluoride at 1.5 mg/L. While much of the literature evaluating exposures to fluoride and reduced cognition in children involves total exposures to amounts assumed to approximate or exceed the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L for fluoride, we make it clear that our assessment considers total fluoride exposures from all sources, not just drinking water alone.

C.33: ***Results page 61:*** *"Altogether, the results from eight of nine low risk-of-bias studies (three prospective cohort studies and five cross-sectional studies from seven different study populations) provide evidence of significant associations between fluoride exposure and cognitive neurodevelopmental outcomes in children other than decrements in IQ (see Figure A-9 through Figure A-11) (Barberio et al. 2017b; Bashash et al. 2017; Bashash et al. 2018; Li et al. 2004 [translated in Li et al. 2008a]; Riddell et al. 2019; Rocha-Amador et al. 2009; Valdez Jimenez et al. 2017; Wang et al. 2020a)."*

[redacted] **comment:** In this section, this summary statement without further explanation is misleading. Elsewhere in this document the authors indicate that the data regarding ADHD effects contains significant heterogeneity regarding methods and outcomes and thereby precludes conclusions about ADHD and other attention-related disorders.

Response: Disagree (edited for clarity)

- We disagree that further explanation is required in this section. The *Overall Findings* subsection in the *Summary of Results* is consistent with how we present the other summaries for each body of evidence. We describe the results of individual studies included in the body of evidence, but not the reasoning for the confidence rating, which is discussed later in the section (on page 66). We are removing the word "altogether" from the sentence below to lessen expectations for a confidence-level statement in this section.

  "The results from eight of nine low risk-of-bias studies (three prospective cohort studies and five cross-sectional studies from seven different study populations) provide evidence of significant associations between fluoride exposure and cognitive neurodevelopmental outcomes in children other than decrements in IQ (see Figure A-9 through Figure A-11) (Barberio et al. 2017b; Bashash et al. 2017; Bashash et al. 2018; Li et al. 2004 [translated in Li et al. 2008a]; Riddell et al. 2019; Rocha-Amador et al. 2009; Valdez Jimenez et al. 2017; Wang et al. 2020a)."

C.34: ***Results page 64:*** *"As discussed above, there are nine studies considered to have low risk of bias when assessed across all risk-of-bias domains."*

[redacted] **comment:** There remain other confounders not considered by NTP that are known to contribute to neurodevelopmental effects; This calls into question the appropriateness of NTP's determination of low risk-of-bias looking at only three covariates (age, sex, and SES). This determination is especially risky given the geographic heterogenicity of the studies referenced. NTP considered additional covariates later in this report although limited to only three in children studies. Reference: page 70

"potential concern for bias regarding covariates not being addressed, including possible co-exposures in occupational studies (e.g., aluminum) and smoking." These covariates also apply to potential for children study outcomes from parental and environmental exposure in the home, school, or community.

Response: Disagree (no change)

- It is inaccurate to say that the determination of low risk of bias was looking at only three covariates (age, sex, and SES). These three covariates were identified as "key" across all studies of fluoride exposure and any neurodevelopmental and cognitive outcome. To be assigned a rating of *probably low risk of bias* for the confounding domain, studies were required to address the three key covariates **in addition** to any other covariates considered important for the specific study population and outcome. Additional covariates considered important for this evaluation, depending on the study population and outcome, included race/ethnicity; maternal demographics (e.g., maternal age, body mass index [BMI]); parental behavioral and mental health disorders (e.g., attention deficit hyperactivity disorder [ADHD], depression); smoking (e.g., maternal smoking status, secondhand tobacco smoke exposure); reproductive factors (e.g., parity); nutrition (e.g., BMI, growth, anemia); iodine deficiency/excess; minerals and other chemicals in water associated with neurotoxicity (e.g., arsenic, lead); maternal and paternal IQ; and quantity and quality of caregiving environment (e.g., Home Observation Measurement of the Environment [HOME] score).

C.35: ***Results page 65:*** *"Seven of the nine studies [i.e., all low risk-of-bias studies except Barberio et al. (2017b) and Riddell et al. (2019)] used appropriate methods for measuring other neurodevelopmental effects in the study population, and blinding of outcome assessors was either reported or not a concern in eight of the nine studies [i.e., all with the exception of Wang et al. (2020a)]."*

[redacted] **comment:** Lack of evidence regarding "blinding" is not sufficient to warrant assumption of study's low-risk. Further, this does not appear to conform with accepted scientific rigor of study design and implementation.

Response: Disagree (no change)

- Blinding of outcome assessors was considered for each study and is described in the HAWC database. A low risk-of-bias rating for the outcome assessment domain is based on (1) whether appropriate methods were used to measure the outcome and, when methods had any subjectivity, whether (2) the outcome was assessed blind. The quote provided above indicates that, of the nine studies, two did not use appropriate methods and a third study did not report information on blinding. Although the quote doesn't specifically state that these three studies received a rating of "probably high" for outcome assessment, the next paragraph in the monograph clearly provides this information. The initial paragraph that is quoted summarizes concerns with the outcome assessment in the group of studies, and the next paragraph provides more detail.

C.36: ***Results page 66:*** *"The high-quality studies (i.e., studies with low potential for bias) provide ~~evidence~~ assumption of an association between fluoride exposure and other cognitive neurodevelopmental effects, including lower neurobehavioral scores in infants, cognitive effects other than IQ in children, and increased attention-related disorders including ADHD in children. However, due to limitations in the data set, including the heterogeneity in the outcomes assessed, a limited number of directly comparable studies, and differences in outcome assessment methods even when studies evaluated similar outcomes, there is low confidence based on this body of evidence that fluoride exposure is associated with other cognitive neurodevelopmental effects in children."* [Text in red font inserted and red-strikethrough font deleted by [redacted].]

[redacted] **comment:** This [second] sentence directly contradicts the leading, first sentence in this paragraph. The first sentence by itself is misleading. Note: red text indicates editorial change.

Response: Disagree (edited for clarity)

- The characterization of the available data as an "assumption" is not accurate so we maintained the original word (i.e., "evidence"). We also disagree that it contradicts the previous sentence. However, to clarify the first sentence, we inserted the word "some" before "evidence" as follows:

  "The high-quality studies (i.e., studies with low potential for bias) provide some evidence of an association between fluoride exposure and other cognitive neurodevelopmental effects, including lower neurobehavioral scores in infants, cognitive effects other than IQ in children, and increased attention-related disorders including ADHD in children."

C.37: ***Results page 73:*** *Serum epinephrine and norepinephrine were significantly increased in a fluoride-endemic region (it was not reported whether subjects were children or adults) compared with a non-endemic region (Chinoy and Narayana 1992). Serum adrenaline and noradrenaline were significantly increased in adults in a fluoride-endemic area (fluoride in the drinking water ranged from 1.0–6.53 ppm) compared with a control area (fluoride in the drinking water ranged from 0.56–0.72 ppm) (Michael et al. 1996)."*

[redacted] **comment:** For the non-health professional reader, the use of different nomenclature for the same neurotransmitter is confusing.

Response: Agree (edited for clarity)

- We agree that epinephrine and norepinephrine are the same as adrenaline and noradrenaline, and we have edited the second sentence as follows:

  "A separate study reported that serum epinephrine and norepinephrine (referred to as adrenaline and noradrenaline in the study) were significantly increased in adults in a fluoride-endemic area (fluoride in the drinking water ranged from 1.0–6.53 ppm) compared with a control area (fluoride in the drinking water ranged from 0.56–0.72 ppm) (Michael et al. 1996)."

C.38: ***Results page 74:*** *"Serum AChE was significantly reduced in children from a high fluoride region compared with a lower fluoride region (Singh et al. 2013)."*

[redacted] **comment:** A lay reader would not know what AChE is- question relevance.

Response: Disagree (no change)

- AChE is defined as Acetylcholinesterase in the *Mechanistic Data in Humans* section on page 72, which is the first time it is used in this section. On page F-2 we outline the relevance of AChE for neurological effects.

C.39: ***Results page 74:*** *"Aborted fetuses from high fluoride areas in China were found to have histological changes in the brain and significant changes in neurotransmitter levels compared with a control area (Du et al. 1992 [translated in Du et al. 2008]; Yu et al. 1996 [translated in Yu et al. 2008])."*

[redacted] **comment:** Consider that the studies cited are high risk of bias (per NTP), this sentence is out of place and may be seen as inflammatory without adding value for the SOS and therefore, recommend deletion.

Response: Disagree (no change)

- This sentence is in a paragraph that describes mechanistic data among high risk-of-bias studies. This study has been mentioned in all prior drafts of the monograph and has not garnered criticism from any other reviewer.

C.40: ***Results page 75:*** *"Although any effects in the brain or neurological tissue at lower concentrations of fluoride may support reduced IQ in humans, it may be difficult to distinguish the potential effects of fluoride on learning and memory functions from other neurological or general health outcomes."*

Reviewer comment: This statement is "fishing" and unsupported. Recommend deleting.

Response: Disagree (no change)

- This sentence states a fact, and we object to the characterization of the sentence as "fishing." Similar to the previous comment, this statement has been in all prior drafts of the monograph and has not garnered criticism from any other reviewer.

C.41: ***Discussion page 76:*** *"Altogether, the results from eight of nine high-quality studies (three prospective cohort and five cross-sectional studies from seven different study populations) provide some evidence that fluoride is associated with other cognitive or neurodevelopmental outcomes in children. The data also suggest that neurodevelopmental effects occur in very young children. However, the number of studies is limited, and there is too much heterogeneity in the outcomes measured and methods used to directly compare studies of any one outcome. Additional studies on outcomes such as attention-deficit hyperactivity disorder (ADHD) and other attention-related disorders, where there is some evidence of an effect of fluoride exposure, would be necessary to critically assess the data."*

[redacted] **comment:** This appears contrary to the preceeding sentence and is not valuable to present description of the details of low-confidence studies.

Response: Disagree (edited for clarity)

- We have edited the description of the body of evidence to include low- and high-quality studies, as follows:

  "The literature in children was separated into studies assessing IQ and studies assessing other cognitive or neurodevelopmental outcomes. There is low confidence in the body of evidence from studies that evaluate fluoride exposure and other cognitive or neurodevelopmental outcomes in children. This body of evidence is made up of nine high-quality studies (three prospective cohort and six cross-sectional studies from seven different study populations) and six low-quality studies. Eight of the nine high-quality studies provide some evidence that fluoride is associated with other cognitive or neurodevelopmental outcomes in children. The data also suggest that neurodevelopmental effects occur in very young children. However, the confidence in this body of evidence is low because the number of studies is limited, and there is too much heterogeneity in the outcomes measured, ages assessed, and methods used, to directly compare studies of any one outcome. Additional studies on outcomes such as attention-deficit hyperactivity disorder (ADHD) and other attention-related disorders, where there is some evidence of an effect of fluoride exposure, would be necessary to critically assess the data."

C.42: ***Discussion page 76:*** *"This review finds, with moderate confidence, that high fluoride exposure ~~is~~ may be associated with lower IQ in children."* [Text in red font inserted and red-strikethrough font deleted by [redacted].]

[redacted] **comment:** The data do not support the assertion of an effect below 1.5 mg/L. Therefore, all conclusory statements in this document should be explicit that any findings from the included studies only apply to water fluoride concentrations above 1.5 mg/L.

Response: Disagree (no change)

- As stated earlier, we disagree with this comment because it refers only to water fluoride concentrations. As explained in previous responses, our assessment considers fluoride exposures from all sources, not just water. As we discussed in the prepublication 2022 NTP Monograph, because fluoride is also found in certain foods, dental products, some pharmaceuticals, and other sources, individual behaviors are likely an important determinant of actual exposures. Even in the optimally fluoridated cities in Canada studied by Green et al. (2019), individual exposure levels, as documented by repeated urinary measurements, suggest widely varying total exposures from water combined with fluoride from other sources. Many, but not all, of these measurements exceed those that would be expected from consuming water that contains fluoride at 1.5 mg/L. While much of the literature evaluating exposures to fluoride and reduced cognition in children involves total exposures to amounts assumed to approximate or exceed the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L for fluoride, we make it clear that our assessment considers total fluoride exposures from all sources, not just drinking water alone.

C.43: ***Discussion page 76:*** *"The association between higher fluoride exposure and lower IQ in children was consistent across different study populations, study locations, study quality/risk-of-bias determinations, study designs, exposure measures, and types of exposure data (group-level and individual-level)."*

[redacted] **comment:** Exposure measures, tests and scales used across the studies were *not* consistent or standardized.
A visual assessment of Figures A1 through A5 appears that a claim of consistency is not supported by the evidence presented. The majority of the findings presented fall in the null effect range and positive findings often overlap with reference findings.

Response: Disagree (no change)

- We disagree. The statement refers to the consistency in the direction of the association between fluoride exposure and IQ in children across studies. Our support for this statement is found in subsections of the *Results* covered on pages 40-47, which summarizes the results of the low risk-of-bias children's IQ studies.

C.44: ***Discussion page 76:*** *"There were 19 low risk-of-bias studies that were conducted in 15 study populations, across 5 countries, and evaluating more than 7,000 children."*

[redacted] **comment:** Note: The studies included in this analysis spanned five different countries, different exposure measures and collection methods, types of exposure data, etc. and cannot be considered "consistent".

Response: Disagree (no change)

- As stated above, consistency refers to the direction of the association between fluoride exposure and children's IQ across the studies.

C.45: ***Discussion page 77:*** *"This review found that the quality of exposure assessment has improved over the years. More recent studies by Valdez Jimenez et al. (2017), Bashash et al. (2017), and Green et al. (2019) used individual measures of urinary fluoride, either maternal urine collected prenatally or children's urine, which confirmed the association between higher total fluoride exposure and lower children's IQ and other cognitive neurodevelopmental effects."*

Review comment: The association is not causal and therefore, suggest changing the word [confirmed] to "support"... the association..

Response: Agree (change made)

- We agree and have changed "*confirmed*" to "*support*".

C.46: ***Discussion page 77:*** *"Studies using different types of exposure measures reported similar findings of an association, which strengthens confidence in earlier studies that reported IQ deficits with increasing group-level fluoride exposure."*

[redacted] **comment:** This statement appears to imply that there are no studies with negative findings and the sentence asserts a dose-response relationship for which data are incomplete.

Response: Agree (change made)

- We agree and have changed "*increasing group-level fluoride exposure*" to "*high group-level fluoride exposure*".

C.47: ***Discussion page 77:*** *"A few studies also support the possibility of heightened sensitivities to the detrimental cognitive effects of fluoride exposure in individuals with certain genetic polymorphisms in dopamine receptor D2 or catechol-O-methyltransferase (Cui et al. 2018; Zhang et al. 2015b), potentially impacting dopamine catabolism and receptor sensitivity."*

[redacted] **comment:** This statement presumes a conclusion that detrimental cognitive effects of fluoride exposure are proven.

Response: Disagree (no change)

- We disagree. The statement refers to the possibility of a greater sensitivity to fluoride exposures in individuals with certain genetic polymorphisms and identifies this as warranting further study.

C.48: ***Discussion page 77:*** *"Differential exposures to fluoride and genetic susceptibilities of children to fluoride may represent special situations that would appear to warrant further research."*

[redacted] **comment:** This sentence as written promotes the premise that fluoride is proven that children have genetic susceptibilities by the suggestion that further research is needed (without reference/citation)

Response: Disagree (no change)

- We disagree. See the response to the prior comment, and note our references in the text to studies by Cui et al. (2018) and Zhang et al. (2015).

C.49: ***Discussion page 78:*** *"Reported responses to fluoride exposure are consistent in studies of both low and high quality."*

[redacted] **comment:** Of the 72 studies, there are some with equivocal or contrary results.

Response: Agree (edited for clarity)

- Text has been changed to read: "*Reported associations between higher fluoride exposure and lower children's IQ are consistent in the vast majority of studies of both low and high quality.*"

C.50: ***Discussion page 78:*** *"Reported responses to fluoride exposure are consistent across different study populations, study designs, and exposure measures."*

[redacted] **comment:** This statement is contrary to the earlier declarations of significant heterogeneity across studies and thus, high risk of bias.

Response: Disagree (edited for clarity)

- The point of the comment is unclear. There is heterogeneity in outcomes among studies that assessed neurodevelopmental outcomes other than IQ (see Figure 3 in the prepublication 2022 NTP Monograph), limiting the evidence base for any one outcome such as ADHD. The quoted statement is referring to the evidence base of 72 IQ studies. The consistency in direction of the association in the studies with heterogeneity in methods of exposure and outcome assessment, in 5 different countries, and accounting for a wide variety of covariates all serve to rule out the possibility that there is a common factor other than fluoride exposure that can account for this outcome.
- The statement in question was revised as follows:

  "Reported associations between higher fluoride exposure and lower children's IQ are consistent across different study populations, study designs, and exposure measures."

C.51: ***Discussion page 78:*** *"A wide variety of important covariates are either addressed by study design or captured across the evidence base, with no consistent patterns that would suggest an alternative explanation."*

[redacted] **comment:** see previous section comments:
Inadequate consideration of all relevant covariates; versus only three covariates used as criteria for inclusion.

Response: Disagree (no change)

- It is inaccurate to say that only three covariates were used as criteria for inclusion. See the previous response to the comment on "***Results page 64***" that describes the importance of the three "key" covariates and lists the other relevant covariates considered important depending on the specific study population and outcome.
- In addition, the text on pages 47-49 in the *Confounding for IQ Studies in Children* section and Figure 6 in the prepublication 2022 NTP Monograph provide a more detailed characterization of the consideration of covariates.

C.52: ***Discission page 78:*** *"Studies rarely separated the results by sex or provided information to indicate that sex was not a modifying factor."*

[redacted] **comment:** More recent publications indicate differences in response by sex.
Community Water Fluoridation: A Review of Neurological and Cognitive Effects. Ottawa: CADTH; 2019 Oct. (CADTH rapid response report: summary with critical appraisal).
ISSN: 1922-8147 (online)
Ibarluzea, Jesú., Gallastegi, M., Santa-Marina, L., Jiménez Zabala, A., Arranz, E., Molinuevo, A., Lopez-Espinosa, M.-J., Ballester, F., Villanueva, C.M., Riano, I., Sunyer, J., Tardon, A., Lertxundi, A., Prenatal exposure to fluoride and neuropsychological development in early childhood: 1-to 4

years old children., *Environmental Research* (2021), doi: https://doi.org/10.1016/j.envres.2021.112181.

Response: Agree (no change)

- We agree that the studies identified by [redacted] evaluated responses by sex; however, the statement that "*Studies rarely separated the results by sex*" still applies.

C.53: ***Discussion page 78:*** *"Associations between lower total fluoride exposure [e.g., represented by populations whose total fluoride exposure was lower than the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] and children's IQ remain unclear. More studies at lower exposure levels are needed to fully understand ~~potential~~ whether there are associations in ranges typically found in the United States (i.e., <1.5 mg/L in water). However, it should be noted that, as of April 2020, CWS supplying water with ≥1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population (~1.9 million people) (CDC Division of Oral Health 2020)."* [Text in red font inserted and red-strikethrough font deleted by [redacted].]

[redacted] **comment:** The last sentence of this bullet should be removed as it is not relevant to Limitations of the Evidence Base.

Response: Disagree (no change)

- The fact that there is a significant number of people in the United States served by drinking water sources containing >1.5 mg/L naturally occurring fluoride supports the need to more fully understand the potential impact of total exposure to fluoride in these areas along with areas with lower, more typical levels of fluoride in drinking water.

C.54: ***Discussion page 79:*** *"Failure to address important covariates was an issue for many studies."*

Reviewer comment: This was also true for the low risk of bias studies

Response: Disagree (no change)

- We disagree. As detailed in other responses above, the *Confounding for IQ Studies in Children* section on pages 40-47 and Figure 6 in the prepublication 2022 NTP Monograph provide a more detailed characterization of the consideration of covariates.

C.55: ***Discussion page 79:*** *"Studies conducted in areas with high, naturally occurring fluoride levels in drinking water often did not account for potential exposures to arsenic or iodine deficiencies in study subjects in areas where these substances were likely to occur."*

[redacted] **comment:** This is also relevant for several of the low-risk studies that may have accounted for arsenic in water but not levels in foods like rice, or accounted for other environmental toxins and parental exposures.

Response: Disagree (no change)

- We disagree. Arsenic in drinking or ground water was considered an important potential co-exposure for all low risk-of-bias studies and is extensively discussed in the monograph text and in *Appendix E*. While arsenic in rice was not considered, it is highly unlikely that differential exposures to arsenic in rice would correlate with higher exposures to arsenic across the five countries and cultures represented in the database.

C.56: ***Discussion page 80:*** *"This systematic review has few limitations."*

[redacted] **comment:** This is a gratuitous interpretation of the limitations of this body of evidence. Please see previous comments.

Response: Disagree (no change)

- We disagree. The section on limitations of the systematic review deals with deviations from best practices in performance of these types of literature reviews. The statement is followed by details and explanation. There are, in fact, few limitations.

C.57: ***Discussion page 80:*** *"In addition, the systematic review covered a wide range of study designs, populations, and measures of fluoride exposure."*

[redacted] **comment:** A wide range of study designs, populations, and measures of exposure in a systematic review are limitations, not a strengths.

Response: Disagree (no change)

- We disagree. While this statement may be true for systematic reviews of clinical studies, for environmental epidemiological studies, a "wide range of study designs, populations, and measures of exposure" demonstrate the robustness of the findings to alternative explanations.

C.58: ***Discussion page 80:*** *"The supplemental literature search for non-English-language studies not indexed in traditional databases supports the comprehensive nature of the literature search strategy for this systematic review."*

[redacted] **comment:** Dependent upon the quality of the supplemental databases and additional studies considered, this could add to bias rather than mitigating it. Example: Chinese translation may differ by dialects.

Response: Disagree (no change)

- We disagree. The potential for differences in Chinese dialects is not a substantive concern for why a systematic review would not conduct a comprehensive literature search. The Chinese literature used in this evaluation was reviewed and extracted by an epidemiologist fluent in Chinese, and there were no instances where this was hindered by dialect. In addition, the data in most scientific literature are presented in tabular formats.

C.59: ***Discussion page 80:*** *"This informed approach influenced the selection process; however, this is not considered a limitation because it provided an objective measure by which to compare databases."*

[redacted] **comment:** This is a significant limitation in that the pre-planned study protocol using a Main Literature Search was modified with the equivalence of a "convenience sample" of added literature (non-English language databases and "Other" literature from Fluoride Action Network-FAN) . This method introduces enormous potential for selection bias that may interfere with an objective analysis and impact conclusions.

Response: Disagree (no change)

- The reason for expanding the literature search to include Chinese databases was to address possible issues of a "convenience sample" after having identified a large number of publications through other sources. While FAN was one of the sources, it was not the only source searched. As the majority of the studies identified from other sources were from Chinese publications, the additional search was in Chinese databases. This is not considered a limitation because, while there was potential for bias for the selection of studies by FAN, searching the databases independently and applying the criteria used for this assessment removed the potential bias.

C.60: ***Summary page 81:*** *"There is, however, a large body of evidence on IQ effects in children."*

[redacted] **comment:** This appears to equate that the "large body of **evidence**" is synonymous with "proof" and none of the studies with designed to demonstrate causality

Response: Disagree (no change)

- This statement does not refer to proof, a term which does not appear in the prepublication 2022 NTP Monograph. Subsequent sentences that follow this quote in the monograph of low confidence for other neurodevelopmental and cognitive effects and moderate confidence that higher fluoride exposure is consistently associated with lower IQ in children do not equate to proof.

C.61: ***Summary page 81:*** *"There ~~is also some evidence~~ is low confidence in the literature for effects that fluoride exposure is associated with other neurodevelopmental and cognitive effects~~; although~~ because of the heterogeneity of the outcomes~~, there is low confidence in the literature for these other effects~~."* [Text in red font was inserted by [redacted] and text in red-strikethrough font was deleted by [redacted] without comment.]

Response: Disagree (no change)

- Although we are open to editing that improves clarity, the proposed text does not, in our opinion, provide an improvement and the repetition of the word "effects" was confusing. We maintained the original text for clarity of the paragraph with the addition of the word "higher" before fluoride in response to a comment from a separate reviewer:

"There is also some evidence that higher fluoride exposure is associated with other neurodevelopmental and cognitive effects; although, because of the heterogeneity of the outcomes, there is low confidence in the literature for these other effects."

C.62: ***Summary page 81:*** *"This review finds, with moderate confidence, that higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] is consistently associated with lower IQ in children."*

[redacted] **comment:** This sentence is an over-interpretation of the underlying science. Ref. (Letter to Dr. Woychik, 12/17/20 December 17, 2020

Response: Disagree (no change)

- We disagree. The statement refers to confidence due to consistency of an association between higher fluoride exposure and lower IQ in children across studies. Our support for this statement is found in subsections of the *Results* covered on pages 40-47, which summarizes the results of the low risk-of-bias children's IQ studies.

C.63:

[redacted] **comment:** This statement does not reflect the literature that demonstrates a positive or neutral association of fluoride exposure and IQ.(ref. **:** Community Water Fluoridation: A Review of Neurological and Cognitive Effects. Ottawa: CADTH; 2019 Oct. (CADTH rapid response report: summary with critical appraisal).
ISSN: 1922-8147 (online): and Ibarluzea, Jesú., Gallastegi, M., Santa-Marina, L., Jiménez Zabala, A., Arranz, E., Molinuevo, A., Lopez-Espinosa, M.-J., Ballester, F., Villanueva, C.M., Riano, I., Sunyer, J., Tardon, A., Lertxundi, A., Prenatal exposure to fluoride and neuropsychological development in early childhood: 1-to 4 years old children., *Environmental Research* (2021), doi: https://doi.org/10.1016/j.envres.2021.112181.)

Response: Disagree (no change)

- The CADTH report was not included in our evaluation because it is a review. Furthermore, it appears that the CADTH report was designed to evaluate only one study and this study is included in our evaluation. The inclusion criteria were limited to studies published after 2017 with fluoride levels below 1.5 mg/L in drinking water. The CADTH report did not evaluate study quality and reached different conclusions than our much more comprehensive systematic review.
- In addition, we comment on the Ibarluzea et al. (2021) study in a footnote to the *Screening of the May 2020 Literature Search Update* section of the prepublication 2022 NTP Monograph. Specifically, we state that it was published after April 2021 and, therefore, is not included in the monograph because it is beyond the dates of the literature search. Even if it had been published earlier, the study would not have contributed to the body of evidence on children's IQ because the authors assessed other neurodevelopmental or cognitive effects.

C.64: ***Summary page 81:*** *"More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ."*

[redacted] **comment:** Suggest replacing last sentence with: There is a need to develop basic guidelines for designing and conducting prospective population-based (epidemiological) fluoride studies in the United States relevant to diverse communities and at fluoride exposure levels (0.7 mg/L) recommended in the United States.

Response: Disagree (no change)

- As stated earlier, this statement may well be correct, but our document does not attempt to establish support for this point.

C.65: ***Appendix E page E-2****: "E.1. IQ Studies"*

[redacted] **comment:** Several studies to not appear to address the consumption of rice which is known in many regions to contain extremely high levels of arsenic.
Feng Liang, Yulan Li, Guilin Zhang, Mingguang Tan, Jun Lin, Wei Liu, Yan Li & Wenwei Lu (2010) Total and speciated arsenic levels in rice from China, Food Additives & Contaminants: Part A, 27:6, 810-816, DOI: 10.1080/19440041003636661

Response: Disagree (no change)

- Although it is true that we did not consider arsenic levels in the rice directly, areas that are known to have high arsenic levels in the water based on water quality maps were identified to address arsenic. As these are also the areas that are likely to have high arsenic in the rice, this comment is considered addressed by the water quality maps. As noted above, it is highly unlikely that differential exposures to arsenic in rice would correlate with higher exposures to arsenic across the five countries and cultures represented in the database.

C.66: ***Appendix E page E-18:*** *"E.1.6. Green et al. (2019)"*

[redacted] **comment:** A review of this study published in 2019 states that the conclusion of maternal exposure to higher levels of fluoride during pregnancy was associated with lower IQ levels is not supported by the data. The difference in maternal "exposure" between non-fluoridated and fluoridated groups was minimal and "adjusted estimates with a limited set of covariates showed no statistically significant association between an increase of 1 mg/L in MUFSG and FSIQ, PIQ or VIQ in all children"
Community Water Fluoridation: A Review of Neurological and Cognitive Effects. Ottawa: CADTH; 2019 Oct. (CADTH rapid response report: summary with critical appraisal).
ISSN: 1922-8147 (online)

Response: Agree (no change)

- While it is correct that the CADTH reported that, "adjusted estimates with a limited set of covariates showed no statistically significant association between an increase of 1 mg/L in MUFSG and FSIQ, PIQ, or VIQ," relying solely on an arbitrary classification of results into "significant" and "non-significant" (typically based on a p-value) is

unnecessary and can be damaging to a valid interpretation of data. The estimates of the magnitude of effect and uncertainty surrounding the estimates are more important for scientific inference and sound judgement (Greenland et al. 2016).

- If CADTH says that the difference in maternal exposure between non-fluoridated and fluoridated groups was minimal, it is inaccurate. Green et al. (2019) reports that the mean maternal urinary fluoride concentration was significantly higher among women who lived in communities with fluoridated drinking water (0.69 ± 0.42 mg/L) compared with women who lived in communities without fluoridated drinking water (0.40 ± 0.27 mg/L; $p < 0.001$).
- We also point out other statements of results from the Green et al. (2019) paper as cited in the CADTH report: "In boys, every 1 mg/L increase in mother's urine fluoride levels was associated with a 4.49 point lower intelligence quotient score [95% CI, -8.38, -0.60; p=0.02]. Every 1 mg increase in daily fluoride intake of mothers corresponded with 3.66 points lower in total children's intelligence quotient score [95% CI, -7.16, -0.15; p=0.04]." Interestingly, the statistically significant p-values that corresponded with these results were not included in the reporting of these findings [brackets with 95% CIs and p-values added by DNTP]. We fully report Green et al. (2019) study findings in Table 6 of the prepublication 2022 NTP Monograph, including those in [redacted] comment, and discuss this study on page 41 and in *Appendix E*. We also report findings by Till et al. (2020), another study based on the Maternal-Infant Research on Environmental Chemicals cohort in Canada.

C.67: ***Appendix E page E-92:*** *"The area did not have industrial pollution within 1 km of the living environment of the children, and it was noted that the children were not exposed to other neurodevelopmental toxicants (lead, cadmium, arsenic, or mercury)."*

[redacted] **comment:** Study did not take into account food exposures

Response: Agree (no change)

- We agree that the study does not take into account food exposures. However, as noted above in regard to arsenic exposure via food, there is no known food exposure that would differentially occur along with fluoride that would be a potential concern. The fact that the study addressed the environment where many of the subjects were growing their own food indicates that there is little concern for other neurodevelopmental toxicants in a manner that would bias the direction of effect across the body of literature.

References

Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG. 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 31:337–350. doi: 10.1007/s10654-016-0149-3

In June 2022, the [redacted] provided comments to NIEHS/DNTP on the prepublication 2022 *NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* and a draft manuscript on a meta-analysis of fluoride exposure and IQ in children. This document contains a subset of the overall [redacted] comments related to the monograph along with the NIEHS/DNTP responses. The monograph-related comments from the [redacted] are reproduced here in black text, and the NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph" The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] comments and [redacted]:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to this [redacted] would be marked "*see DocD_Monograph for detailed response*").
- [redacted] For comments related to DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

[redacted] Review

June 2022

Note: The [redacted] provided six comments on the meta-analysis manuscript that are not reproduced here as they are not directly relevant to the prepublication 2022 NTP Monograph. See "Doc07a_Meta-analysis" for the meta-analysis-relevant comments and responses.

NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review

D.1:

1) This should undergo journal peer-review in order to published.

Response: Disagree (no change)

- The prepublication 2022 NTP Monograph is a revised document that was developed considering the comments of multiple rounds of external peer review, including two rounds of review by a NASEM Committee. In addition, this document has undergone an additional review by five independent experts, all of whom agreed with the conclusions of the systematic review. We consider the monograph to have undergone sufficient expert review to warrant publication.

D.2:

2) The abstract is anchored on the WHO fluoride levels and does not seem to indicate the U.S. approach to fluoride levels, which could create communication confusion in the United States.

Response: Disagree (no change)

- We have chosen to focus on the WHO guideline value for fluoride in drinking water because our moderate confidence assessment was primarily based on studies with total fluoride exposures approximating or higher than that provided by the 1.5 mg F/L drinking water level considered safe by WHO. We emphasize total fluoride exposure because some studies we reviewed that showed deficits in IQ, including Green et al. (2019), were performed in optimally fluoridated areas (0.7 mg F/L); however, based on scatterplots of urinary levels, it was apparent that some children—or mothers during their pregnancy—were exposed to fluoride from sources in addition to drinking water. We have no basis on which to state that our findings are not relevant to some children or pregnant people in the United States because exposure measurements in the United States are not well studied or reported.

In July 2022, the [redacted] provided comments to NIEHS/DNTP on the prepublication 2022 *NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* and a draft manuscript on a meta-analysis of fluoride exposure and IQ in children. This document contains all of the [redacted] comments along with the NIEHS/DNTP responses because they are relevant to the monograph. The specific comments from [redacted] are reproduced here in black text and the NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] comments and [redacted] comments:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to [redacted] would be marked "*see DocE_Monograph for detailed response*").
- [redacted] For comments related to DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

[redacted]

Responses to the National Toxicology Program/NIEHS Fluoride Exposure Manuscripts

June [July] 20, 2022

E.1:

[redacted] have provided comments on the NTP State of the Science Monograph and the meta-analysis manuscript, as requested by leadership of the National Toxicology Program (NTP) and the National Institute of Environmental Health Sciences (NIEHS). The consensus from [redacted] is that the documents are well-constructed, comprehensive, and very interesting.

Response: No change requested

- We appreciate the comments and the careful review.

Additional remarks and suggestions are the following:

E.2:

- Consider mentioning existing recommendations from the US Preventive Services Task Force (USPSTF), the American Academy of Pediatric Dentistry (AAPD), and the American Academy of Pediatrics (AAP) to help anchor the background. Resources to review and possibly incorporate include the Prevention of Dental Caries in Children Younger Than 5 Years: Screening and Interventions, Oral Health Practice Tools, and Fluoride Therapy. Parents and providers would want to be well informed before implementing recommendations, especially when considering the benefit/harm balance. It is believed that potential harms of fluoride supplementation re: child IQ have not been considered as potential harms in the recommendations from USPSTF, AAPD, and AAP which is worth noting.

  Response: Disagree (no change)

  - We appreciate the suggestion to provide a more complete outline of current recommendations for uses of fluoride. The suggestion implies that our documents may be sufficient to inform personal and public health decisions concerning appropriate fluoride use. While we consider our documents to be valuable contributions to these decisions, we suggest a more appropriate use would be to stimulate and inform a public health service-wide reconsideration of the potential hazards of total fluoride exposure in relation to its benefits to oral health.

E.3:

- Highly consider reaching out to AHRQ to review these documents. AHRQ routinely commissions systematic reviews, including meta-analysis, from its Evidence-based Practice Centers. It could warrant a revision or update to USPSTF recommendations on fluoride supplementation. If there's actual equipoise, it could also be an interesting question for ODP Pathways to Prevention.

  Response: No change requested

  - We appreciate the suggestion to provide our systematic review documents to AHRQ for further consideration, and we recognize that their review could warrant a revision or

update to the USPSTF recommendations on fluoride. Similar to the previous [redacted] comment, this suggestion implies that our documents may be sufficient to inform personal and public health decisions concerning appropriate fluoride use. However, we suggest a more appropriate use of our documents would be to stimulate and inform a Public Health Service-wide reconsideration of the potential hazards of total fluoride exposure in relation to its benefits to oral health.

E.4:

- Clarify the implications of the timing and duration of exposure on neurocognitive development. This is purportedly a report about neurocognitive development, but there is very light treatment of development in the text. The only references recognized were in the limitations listed in the discussion:
    - "No studies are available to evaluate fluoride exposure over a child's lifetime and neurodevelopmental or cognitive changes over time.
    - The database does not allow for comparison of ages and possible changes at different developmental stages in children to assess if there is a delay in development or if associations persist.
    - The database does not allow for establishing clear correlations between prenatal and postnatal exposures."

  Other than these bullets, there was no discussion of developmental impact. The only distinction made is dividing the studies of children or adults, and "children" include all studies up to age 17. Particularly for bullet 2, given that the studies examine different age ranges, this seems like an omission, even if the discussion is only to outline how strong conclusions can't be drawn, as they do for other topics.

  Response: Agree (no change)
    - We agree that additional information concerning the timing of fluoride exposure and associated potential cognitive effects is a critically important data gap in understanding relative hazards from fluoride to brain development. Some investigators have examined effects of prenatal versus postnatal exposures to fluoride on children's IQ (e.g., Till et al. 2020), and further studies of neurobehavioral hazards in relation to oral health benefits from in utero exposures appear warranted. Therefore, we agree there is limited discussion of these topics, but other than noting this deficiency, we have little further to add from our assessment of the current literature.

In May 2022, the [redacted] provided comments to NIEHS/DNTP on the prepublication 2022 *NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review*. These [redacted] comments were embedded in a PDF version of the monograph. The full [redacted] comments have been grouped by topic and are reproduced below verbatim along with the specific monograph text referred to by [redacted] in quotes and the section and page number of the monograph text (e.g., "Summary page 81"). Formatting has been applied to aid in reading. Responses have been added in blue text following each of the comments beginning with the word "**Response**" in bold font. Comments related to a particular topic were grouped together under headings for Comments on health benefits, Comments related to U.S. water fluoridation levels, Comments on the NASEM review of the 2020 draft NTP Monograph, and Other comments. In several cases, the same or similar comments are identified by [redacted] and, therefore, a "Collective Response" is provided.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] comments and [redacted] comments:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to [redacted] would be marked "*see DocF_Monograph for detailed response*").
- [redacted] For comments related to DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.

- The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

F.1: Comments on health benefits

Abstract page xii: "Fluoride is a common exposure in our environment that comes from a variety of sources and is widely promoted for its dental and overall oral health benefits."

[redacted] **Comment:** Add at least one more sentence on this. Maybe give an example

Abstract page xii: *"Existing animal studies provide little insight into the question of whether fluoride exposure affects IQ."*

[redacted] **Comment:** Need to add something about the known public health benefits for a more well-rounded picture/for context.

Discussion page 76: *"The potential health benefits of fluoride with respect to oral health are acknowledged but are not the focus of this review."*

[redacted] **Comment:** Please confirm with [redacted] for references, but [redacted] believe there is a significant amount of data out there to suggest that it's more than just "potential health benefits." Recommend expanding upon this a bit more to describe some of the health benefits that have been shown.

Summary page 81: *"More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ."*

[redacted] **Comment:** Again, worth stating/reinforcing the overall benefits and public health value of fluoride.

Collective Response: Disagree (edited for clarity)

- We cite the dental and overall oral health benefits of fluoride in the *Introduction* section; however, we have been careful to not give the incorrect impression in the *Abstract* or *Discussion* sections that we are providing any assessment of oral health benefits or weighing hazards versus benefits of fluoride exposures in the monograph. In addition, we agree that the benefits are substantive; therefore, we removed "potential" so that the reference to "health benefits" is more positively worded.

Comments related to US water fluoridation levels

F.2: ***Abstract page xiii:*** *"This review finds, with moderate confidence, that higher fluoride exposure (e.g., represented by populations whose total fluoride exposure approximates or exceeds the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of*

fluoride) are consistently associated with lower IQ in children. More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ."

[redacted] **Comment:** Need to put this in context with the US levels (0.7 mg/L).

Response: Disagree (no change)

- Throughout the monograph we have stressed that drinking water is only one source of fluoride exposure, typically comprising 30–70% of a person's total exposure. In the *Introduction* section, we provide information about other sources of fluoride exposure and also identify the EPA drinking water standard, the WHO drinking water quality guideline, and the U.S. Public Health Service (PHS) recommendation for artificial water fluoridation. We have chosen to refer to the WHO drinking water quality guideline for fluoride in the *Abstract* section and in other places in the monograph because, in our overall assessment of the epidemiology literature, it represents a useful total fluoride exposure equivalent metric, above which we have moderate confidence in an association with lower IQs in children. Several of the highest quality studies showing lower IQs in children were done in optimally fluoridated (0.7 mg/L) areas in Canada, but the individual exposure information in those studies, as documented by repeated urinary measurements, suggests widely varying total fluoride exposure from drinking water combined with exposures from other sources. For example, many urinary fluoride measurements exceed those that would be expected from consuming water that contains fluoride at 1.5 mg/L. Additionally, according to the CDC, over a million people in the United States are exposed to naturally occurring fluoride at >2 mg/L in their drinking water. For these reasons, we have chosen not to make specific reference to the PHS fluoridation recommendation in the *Abstract* section.
- The existing text that addresses this comment is in the *Introduction* section on page 1 of the monograph.

F.3: ***Introduction page 2:*** *"In 2016, the National Toxicology Program (NTP) 2016, NTP published a systematic review of the evidence from experimental animal studies on the potential effects of fluoride exposure on learning and memory (NTP 2016). That systematic review found a low-to-moderate level of evidence that deficits in learning and memory occur in experimental animals exposed to fluoride."*

[redacted] **Comment:** What levels? High? Above 0.7 mg/L?

Response: Disagree (no change)

- The large experimental animal database includes studies at many exposure levels as outlined in the 2016 monograph. As will be reiterated later, it is difficult to globally characterize exposures that vary in type and duration as "high" without providing any context. Certainly, most experimental animal studies employed exposures that were higher than 0.7 mg/L in water, or its equivalent,

but whether this is "high" with respect to lifetime human exposures requires consideration of other factors.

F.4: ***Introduction page 2:*** *"As the NTP (2016) report on the experimental animal evidence focused on learning and memory and developed confidence ratings for bodies of evidence by life stage of exposure (i.e., exposure during development or adulthood), this monograph also evaluates two different age groups in humans (i.e., children and adults) with a focus on cognitive neurodevelopmental effects in children and cognitive effects in adults in order to address potential differences in health impacts based on time frame of exposure (i.e., during development or during adulthood)."*

[redacted] **Comment:** And levels?

Response: Disagree (no change)

- Although exposure levels were not the specific driver for evaluating two different age groups in humans, the monograph provides extensive information about the fluoride exposure levels measured in the epidemiological studies.

  For example, fluoride levels can be found in the existing "*Exposure Measures and Summary Statistics*" column of Tables 6, 7, and 8.

F.5: ***Results page 40:*** *"All three prospective cohort studies found an association between increasing maternal or child fluoride exposure and lower IQ in children (Bashash et al. 2017; Green et al. 2019; Till et al. 2020). Two of the studies (Green et al. 2019; Till et al. 2020) were based on the same Canadian study population, but one evaluated prenatal fluoride exposure and the other evaluated postnatal fluoride exposure."*

[redacted] **Comment:** What level? Above WHO guidelines?

Response: Disagree (no change)

- As indicated earlier, the Canadian studies were performed in both non-fluoridated and optimally fluoridated (0.7 mg F/L) areas, which is less than half the WHO drinking water guideline.

F.6: ***Results page 41:*** *"An increase of 0.5 mg/L of maternal urinary fluoride was associated with a 2.5-point decrease in IQ score [95% CI: −4.12, −0.59] in boys and girls combined (see Figure A-8)."* [Bashash et al. 2017]

[redacted] **Comment:** From what starting point?

Response: Disagree (no change)

- Both the Green et al. (2019) and Bashash et al. (2017) studies report fluoride exposures at the individual study participant level; therefore, the data are continuous, with a "starting point" below the level of analytical detection.

F.7: ***Results page 41:*** *"In the Maternal-Infant Research on Environmental Chemicals cohort, consisting of 10 cities in Canada, Green et al. (2019) also reported inverse associations between IQ scores in children and multiple measures of prenatal fluoride exposure, including maternal urinary fluoride, maternal fluoride intake, and water fluoride concentrations. Green et al. (2019) observed a statistically significantly lower IQ for boys associated with maternal urinary fluoride averaged across trimesters (4.49-point decrease in IQ score [95% CI: −8.38, −0.60; p-value = 0.02] per 1-mg/L increase in maternal urinary fluoride); however, results were not significant in boys and girls combined (1.95-point decrease in IQ [95% CI: −5.19, 1.28]) and were positive but not significant in girls (2.40-point increase in IQ [95% CI: −2.53, 7.33])."*

[redacted] **Comment:** Again, helpful to say at what levels to know starting point.

Response: Disagree (no change)

- As stated above, the data were reported at the individual level and therefore were continuous data with no "starting point."

F.8: ***Results page 42:*** *"Overall, the cross-sectional studies consistently provide evidence that fluoride exposure is associated with lower IQ scores in children."*

[redacted] **Comment:** High?

Response: Disagree (edited for clarity)

- Table 1 provides the PECO (Population, Exposure, Comparator, and Outcome) statement that indicates that all studies were included that had exposed populations compared with populations not exposed or exposed to lower levels of fluoride. Fluoride levels in the studies are reported in Table 6 and include a range of exposures.
- We have chosen not to place the modifier of "high" in this instance because there is no consistent definition of "high" that applies across all studies. Instead, we have added the word "higher" to specify "higher fluoride exposure" within a given study.

F.9: ***Results page 44:*** *"Two studies (Cui et al. 2018; Zhang et al. 2015b) observed associations between lower IQ in children and exposure to fluoride, with variations in results in subpopulations of children with different polymorphisms (see Figure A-7)."*

[redacted] **Comment:** Level?

Response: Disagree (no change)

- As stated above, the exposure levels for these studies are found in Table 6, and they seem generally in line with other studies.

F.10: ***Results page 45:*** *"As described in this section, the body of evidence for studies assessing the association between fluoride exposure and IQ in children consistently provides evidence of an association between higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] and lower IQ in children; however, there is less certainty in the evidence of an association in populations with lower fluoride exposures."*

[redacted] **Comment:** Again, this needs to be made clear above because it is not.

Response: Disagree (no change)

- This comment is addressed in an earlier response to a similar comment on the *Abstract* section, but additionally, we consider this point to be made clearly and is an accurate reflection of our confidence in the association between total fluoride exposures approximating or exceeding the WHO water quality guideline of 1.5 mg/L and lower IQ in children.

F.11: ***Results page 45:*** *"Based on the qualitative review of these studies, the evidence of an association between fluoride exposure below 1.5 mg/L and lower IQ in children appeared less consistent than results of studies at higher exposure levels."*

[redacted] **Comment:** This is a very important point.

Response: Agree (no change requested)

F.12: ***Results page 45:*** *"In the five studies that reported results by sex separately, consistent findings of lower IQ associated with fluoride exposure were generally reported for both sexes."*

[redacted] **Comment:** High?

Response: Disagree (edited for clarity)

- As stated earlier, the comparisons in the epidemiological studies are between populations that had a range of fluoride exposures that could be compared with similar populations with lower or no fluoride exposures. We have chosen not to place the modifier of "high" in this instance or in many of the other 18 instances where this is suggested because there is no consistent definition of "high" that applies across all cases.

  Instead, we have added the word "higher" after the phrase "associated with" to specify "higher fluoride exposure" in this instance, and in several other instances where the addition of "higher" is appropriate, as indicated in subsequent responses to [redacted] comments.

F.13: ***Results page 46:*** *"In summary, it is unclear whether one sex is more susceptible to the effects of fluoride exposure than the other due to the limited number of studies that analyzed*

exposure and outcome by sex and the lack of a consistent pattern of findings that one sex is more susceptible."

[redacted] **Comment:** High?

Response: Disagree (no change)

- This statement refers to whether there is evidence to support that one sex is more susceptible to neurodevelopmental and cognitive effects of fluoride at any level, not just "high" levels of fluoride exposure. Therefore, as stated earlier, there is no consistent definition of "high" that can be applied across these studies.

F.14: ***Results page 48:*** *"All 15 studies observed an association between lower IQ and fluoride exposure."*

[redacted] **Comment:** High?

Response: Disagree (edited for clarity)

- We have added the word "higher" after the phrase "lower IQ and" to specify "higher fluoride exposure."

F.15: ***Results page 48:*** *"The other study did not address arsenic co-exposure and, as noted above, was conducted in an area that had potential for arsenic exposure to occur (Soto-Barreras et al. 2019); it is also the only low risk-of-bias study that did not observe an association between lower IQ and fluoride exposure (see Appendix E for further discussion of the risk-of-bias concern regarding arsenic for this study)."*

[redacted] **Comment:** High?

Response: Disagree (edited for clarity)

- We have added the word "higher" near the end of the sentence to specify "higher fluoride exposure."

F.16: ***Results page 61:*** *"Altogether, the results from eight of nine low risk-of-bias studies (three prospective cohort studies and five cross-sectional studies from seven different study populations) provide evidence of significant associations between fluoride exposure and cognitive neurodevelopmental outcomes in children other than decrements in IQ (see Figure A-9 through Figure A-11) (Barberio et al. 2017b; Bashash et al. 2017; Bashash et al. 2018; Li et al. 2004 [translated in Li et al. 2008a]; Riddell et al. 2019; Rocha-Amador et al. 2009; Valdez Jimenez et al. 2017; Wang et al. 2020a)."*

[redacted] **Comment:** High?

Response: Disagree (edited for clarity)

- We have added the modifier "higher" after the word "between" to specify "higher fluoride exposure and cognitive…"

F.17: ***Results page 63:*** *"In summary, the high-quality studies (i.e., studies with low potential for bias) provide evidence of an association between fluoride exposure and neurodevelopmental and cognitive effects in children other than IQ; however, the body of evidence is limited by heterogeneity in the outcomes evaluated and few directly comparable studies."*

[redacted] **Comment:** High

Response: Disagree (edited for clarity)

- We have added the word "higher" after the word "between" to specify "higher fluoride exposure..."

F.18: ***Results page 63:*** *"High risk-of-bias studies (n = 6) also provide some evidence of associations between fluoride exposure and neurodevelopmental or cognitive effects in children other than effects on IQ, but the results are inconsistent and address different outcomes (Jin et al. 2016; Li et al. 1994 [translated in Li et al. 2008b]; Malin and Till 2015; Morgan et al. 1998; Mustafa et al. 2018; Shannon et al. 1986)."*

[redacted] **Comment:** High?

Response: Disagree (no change)

- This statement refers to high risk-of-bias studies examining evidence of an association between any level of fluoride exposure—not just "high" levels of fluoride exposure—and neurodevelopmental or cognitive effects in children. Therefore, as stated earlier, there is no consistent definition of "high" exposure that can be applied across these studies.

F.19: ***Results page 66:*** *"The high-quality studies (i.e., studies with low potential for bias) provide evidence of an association between fluoride exposure and other cognitive neurodevelopmental effects, including lower neurobehavioral scores in infants, cognitive effects other than IQ in children, and increased attention-related disorders including ADHD in children."*

[redacted] **Comment:** High?

Response: Disagree (no change)

- This statement refers to high-quality studies evaluating the association between any level of fluoride exposure and other cognitive neurodevelopmental effects. As stated earlier, there is no consistent definition of "high" that can be applied across these studies.

F.20: ***Results page 69:*** *"The results from five out of eight high risk-of-bias studies provide evidence of cognitive impairment in adults associated with exposure to fluoride; however, there was heterogeneity in the outcomes assessed, a limited number of directly comparable studies, and some variability in results (e.g., variation in IQ results across studies)."*

[redacted] **Comment:** High levels of?

Response: Disagree (edited for clarity)

- We have added "higher levels of" after the phrase "associated with" to specify "higher levels of exposure to fluoride..."

F.21: ***Results page 71:*** *"Eight low risk-of-bias studies that evaluated fluoride exposure and mechanistic data in humans were considered potentially relevant to neurological effects."*

[redacted] **Comment:** High?

Response: Disagree (no change)

- These eight low risk-of-bias studies evaluated associations between any level of fluoride exposure—not just "high" fluoride exposure—and mechanistic data in humans.

F.22: ***Results page 71:*** *"Barberio et al. (2017a) evaluated associations between fluoride and TSH levels in children and adults combined and found no relationship between fluoride exposure (measures in urine and tap water) and TSH levels."*

[redacted] **Comment:** High?

Response: Disagree (no change)

- Barberio et al. (2017a) evaluated associations between any level of fluoride exposure—not just "high" fluoride exposure—and TSH levels in children and adults.

F.23: ***Results page 72:*** *"Among high risk-of-bias studies (see Figure D-19 and Figure D-20), varying results were reported in 11 studies that evaluated associations between fluoride exposure and thyroid hormones, and a few of these studies (Lin et al. 1991; Wang et al. 2001; Yang et al. 1994 [translated in Yang et al. 2008]) were complicated by high or low iodine in the high fluoride area."*

[redacted] **Comment:** High?

Response: Disagree (no change)

- These studies evaluated associations at any level of fluoride exposure, not just "high" exposures.

F.24: ***Results page 72:*** *"When considering associations between fluoride and TSH, T3, and T4 levels together, studies that evaluated changes in all three thyroid hormones reported varying combinations of increases, decreases, or no changes in levels across the three hormones, although among the eight low and high risk-of-bias studies that evaluated associations between fluoride exposure and TSH, T3, and T4 levels and reported increases in TSH levels in children, seven of the eight studies found no alterations in T3 levels (one study found an increase in T3), and six of the eight studies found no alterations in T4 levels (two studies found an increase in T4)."*

[redacted] **Comment:** High?

Response: Disagree (no change)

- These studies evaluated associations at any level of fluoride exposure, not just "high" exposures.

F.25: ***Discussion page 76:*** *"This systematic review evaluated the available animal and human literature concerning the association between fluoride exposure and cognitive neurodevelopment."*

[redacted] **Comment:** High?

Response: Disagree (no change)

- The original intent of the systematic review was to evaluate the available animal and human literature concerning the association between any level of fluoride exposure and cognitive neurodevelopment.

F.26: ***Discussion page 76:*** *"There is <u>low confidence</u> in the body of evidence from studies that evaluate fluoride exposure and other cognitive or neurodevelopmental outcomes in children."*

[redacted] **Comment:** High?

Response: Disagree (no change)

- The body of evidence includes studies that evaluated the association between any level of fluoride exposure—not just "high" fluoride exposures—and other cognitive and neurodevelopmental outcomes in children.

F.27: ***Discussion page 76:*** *"Most of the epidemiological studies (n = 72) assessed the association between fluoride exposure and IQ in children."*

[redacted] **Comment:** High?

Response: Disagree (no change)

- The body of evidence includes studies that evaluated the association between any level of fluoride exposure—not just "high" fluoride exposures—and IQ in children.

F.28: ***Discussion page 76:*** *"This review finds, with moderate confidence, that fluoride exposure is associated with lower IQ in children."*

[redacted] **Comment:** High?

Response: Disagree (edited for clarity)

- We have added the word "higher" to specify "higher fluoride exposure."

F.29: ***Discussion page 78:*** *"Reported responses to fluoride exposure are consistent in studies of both low and high quality."*

[redacted] **Comment:** High?

Response: Disagree (edited for clarity)

- We have added the word "higher" to specify "higher fluoride exposure" across all the low- and high-quality studies.

F.30: ***Summary page 81:*** *"There is also some evidence that fluoride exposure is associated with other neurodevelopmental and cognitive effects; although, because of the heterogeneity of the outcomes, there is low confidence in the literature for these other effects."*

[redacted] **Comment:** High?

Response: Disagree (edited for clarity)

- We have added the word "higher" to specify "higher fluoride exposure."

F.31: ***Results page 50:*** *Figure 6 – heading row.*

[redacted] **Comment:** Might be helpful to add fluoride levels here.

Response: Disagree (no change)

- Fluoride levels are provided in Table 6.

F.32: ***Results page 53:*** *"We conclude that there is moderate confidence in the body of evidence that higher fluoride exposure is associated with lower IQ in children."*

[redacted] **Comment:** Defined as...?

Response: Disagree (no change)

- We have adequately qualified this statement throughout the monograph.

Comments on the NASEM review

F.33: ***Introduction page 2:*** *"A committee convened by the National Academies of Sciences, Engineering, and Medicine (NASEM) reviewed earlier drafts of this monograph (September 6, 2019, and September 16, 2020) (NASEM 2020; 2021)."*

[redacted] **Comment:** and?

Response: No change requested

- The sentence in the monograph that follows the one quoted above states, "*The current document incorporates changes stemming from those reviews*..." and will provide a link to the NTP website for the NASEM Committee comments and responses. Additional information on the NASEM review is also provided in the *Preface*.

Other comments

F.34: ***Introduction page 3:*** *"The overall objective of this evaluation was to undertake a systematic review to develop NTP human health hazard identification conclusions on the association between exposure to fluoride and neurodevelopmental and cognitive effects based on assessing levels of evidence from human and non-human animal studies with consideration of the degree of support from mechanistic data."*

[redacted] **Comment:** "Conclusions" sounds so fixed. Recommend revising.

Response: Disagree (no change)

- The original intent of the monograph was in fact to reach conclusions about the purported association between fluoride exposures and cognitive and neurodevelopmental effects in humans. This was established and posted in the systematic review protocol for this evaluation. Following the NASEM review, the draft hazard conclusions were removed, and the focus of the monograph was revised to present and transparently evaluate the evidence base for cognitive and neurodevelopmental effects associated with fluoride exposures and provide our level of confidence in the evidence base. Thus, use of the word "conclusions" in the context of the original objective is appropriate.

F.35: ***Introduction page 3:*** *"In addition, a meta-analysis of the epidemiological studies examining children's IQ in relation to fluoride exposure added to the 2020 draft in response to NASEM comments (NASEM 2020) will be published separately and is not part of this document."*

[redacted] **Comment:** [redacted] would not presume that this will happen. Revise.

Response: Agree (change made)

- The text in question was revised to indicate the intent to publish separately.

F.36: ***Methods page 19:*** *"The meta-analysis conducted in association with this systematic review further informs this issue and will be published separately."*

[redacted] **Comment:** Would revise because it is being submitted for publication and also [redacted] would not link it to further informing this since there are still questions about the meta-analysis.

Response: Agree (change made)

- Although we do not have remaining questions about the meta-analysis, the text in question was revised to indicate the intent to publish separately.

F.37: ***Introduction page 3:*** *"Assess the internal validity (risk of bias) of individual studies using pre-defined criteria."*

[redacted] **Comment:** Who developed them? How/with what in mind?

Response: No change requested

- To avoid bias, systematic reviews are conducted following a detailed protocol developed prior to conducting the evaluation. The protocol is customized to the specifics of the research question—in this case, to evaluate evidence of potential neurodevelopmental and cognitive health effects associated with exposure to fluoride. The risk-of-bias criteria were developed by subject matter experts with backgrounds in epidemiology, neurotoxicity, fluoride, and public health to address these specifics as part of the protocol. To support the rigor and transparency of the systematic review process, the protocol was reviewed and revised based on input from appropriate experts listed in the *About this Review* section of the monograph (Thomas Webster, PhD, Boston University; Joseph Braun, PhD, Brown University; Gail Wasserman, PhD, Columbia University; Marie Sutton, PhD, Dublin Health Research Board, and Thomas Zoeller, PhD, University of Massachusetts).

F.38: ***Methods page 16:*** *"To be assigned a rating of probably low risk of bias for the key risk-of-bias question regarding the confounding domain, studies were not required to address every important covariate listed; however, studies were required to address the three key covariates for all studies, the potential for co-exposures, if applicable (e.g., arsenic and lead, both of which could affect cognitive function), and any other potential covariates considered important for the specific study population and outcome."*

[redacted] **Comment:** Why three?

Response: No change requested

- When the protocol was developed, age, sex, and socioeconomic status were selected as the key covariates for assessing potential bias in the confounding domain based on an assessment of the fluoride epidemiological literature. These three covariates were considered important potential confounders (i.e., associated with both fluoride exposure and neurological outcomes) for all studies (e.g., they apply across different geographic locations and study populations and outcomes). Other covariates were considered important in the context of individual studies because of their study population or geographic location (e.g., co-exposure to arsenic in studies conducted in China). As stated in response to the previous question, the protocol was reviewed and revised based on input from appropriate experts listed in the *About this Review* section of the monograph including environmental epidemiologists, neurotoxicologists, and researchers with fluoride expertise.

F.39: ***Results page 28:*** *"Because IQ tests should be culturally relevant, the tests used often differed between studies, reflecting adjustments for the range in populations studied (e.g., western vs. Asian populations). In some cases, different IQ tests were used to study similar populations. Overall, these studies used IQ tests that were population- and age-appropriate."*

[redacted] **Comment:** Are there well-accepted methods for standardizing across these different tests? How was that considered?

Response: No change requested

- There are well-accepted methods for standardizing results across these different tests for quantitative analysis. For example, the standardized mean difference is used as a summary statistic in meta-analysis when the studies all assess the same outcome (e.g., IQ) but measure it in a variety of ways (e.g., Chinese Standardized Raven Test, Wechsler Intelligence Scale for Children). While this does not apply to the prepublication 2022 NTP Monograph, which is a qualitative assessment of the literature rather than a quantitative assessment, we are using this method in our meta-analysis of fluoride exposure and children's IQ.

F.40: ***Results page 48:*** *"Co-exposures to arsenic and lead were not considered a concern in 18 of 19 low risk-of-bias studies [i.e., all except for Soto-Barreras et al. (2019)] because the studies addressed the potential co-exposures, the co-exposures were not considered an issue in the study population, or the impact of the potential bias on the results was not a concern."*

[redacted] **Comment:** How?

Response: Disagree (no change requested)

- *Appendix E* provides detailed information on all of the low risk-of-bias studies, including details on the risk-of-bias analysis, risk-of-bias ratings, and the basis for each rating.

F.41: ***Discussion page 77:*** *"Thus, drinking water fluoride levels may, but usually do not, reflect total fluoride exposure. This could be a potential limitation in studies that rely on water fluoride data to assess fluoride exposure (in particular, earlier studies). However, because water is only part of a person's total exposure to fluoride, this limitation would likely result in an underestimate of exposure to fluoride."*

[redacted] **Comment:** May be worth expanding on this more because the values reported in the studies may be underestimates, so the association with IQ may only be at values higher than what is reported.

Response: Agree (no change)

- We agree with the comment without changes to the text. Indeed, this is the basis for the emphasis on total fluoride exposures, rather than emphasis on drinking water exposures only as reflected in the U.S. PHS fluoridation recommendation. We also recognize that when fluoride levels in drinking water are high, the contribution of fluoride exposures from other sources to total exposures tends to be less, and vice versa.

F.42: ***Discussion page 77:*** *"To decrease an exclusively formula-fed infant's exposure to fluoride, for the purpose of reducing risk of dental fluorosis, the Centers for Disease Control and Prevention recommends using low-fluoride bottled water to mix with infant formula (CDC 2015)."*

[redacted] **Comment:** This seems out of place.

Response: Disagree (no change)

- The sentence in question follows two sentences on water consumption as part of total fluoride exposure and fluoride exposure in bottle-fed infants as follows:

  "It is worth noting that there are circumstances wherein typical children's water consumption considered with water fluoride levels may substantially underestimate total fluoride exposure. One example is bottle-fed infants wherein nutrition is provided by powdered formula that is rehydrated with fluoridated water (Till et al. 2020)."

- These statements provide additional context, and this information will be available in our communication Q&As, so it would be provided to the interested public if they enquire.

F.43: ***Discussion page 78:*** *"Studies rarely separated the results by sex or provided information to indicate that sex was not a modifying factor."*

[redacted] **Comment:** Also seemed like there was a lot of variability in how much SDoH (e.g., SES status) were considered. These are important factors that could influence IQ fairly heavily.

Response: Disagree (no change)

- Among the 19 low risk-of-bias children's IQ studies, only one did not properly account for SES status. To properly account for SES, a study was required to consider (e.g., adjust for in the statistical model) for one or more measures of SES (see below for what qualified as a measure of SES). Measures of SES were assessed following a consistent predefined approach. As described in the protocol (https://ntp.niehs.nih.gov/go/785076), to be assigned a rating of *probably low risk of bias* for the key risk-of-bias question regarding the confounding domain, studies were required to address the three key covariates (i.e., age, sex, and SES). The acceptable measures of SES, as outlined on page 58 of the protocol included, but were not limited to, maternal education, household income, marital status, and crowding. The protocol was reviewed and revised based on input from appropriate experts listed in the *About this Review* section of the prepublication 2022 NTP Monograph including environmental epidemiologists, neurotoxicologists, and researchers with fluoride expertise.
- Figure 6 in the prepublication 2022 NTP Monograph presents the different measures of SES included in the low risk-of-bias IQ studies conducted in children. These measures included SES scaled scores, household/family income, child education, caretaker/parental education, and occupation/employment. Despite these different measures of SES, the findings of these studies consistently provided evidence of an inverse association between higher fluoride exposure and lower IQ in children. This consistency in the direction of the association strengthens our confidence in the body of evidence.

F.44: ***Discussion page 78:*** *"However, it should be noted that, as of April 2020, CWS supplying water with ≥1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population (~1.9 million people) (CDC Division of Oral Health 2020)."*

[redacted] **Comment:** Elaborate a little further.

Response: Agree (change made)

- The preceding sentences to this statement in the monograph outline a limitation in the evidence base at lower total fluoride exposure levels (citing total fluoride exposure lower than the WHO guideline of 1.5 mg/L), specifically that more studies at lower exposure levels are needed to fully understand potential associations at fluoride levels in drinking water typically found in the United States (<1.5 mg/L). The CWS data indicate that the moderate confidence in the association between higher fluoride exposure and lower IQ is relevant, at least to children living in areas of the United States where fluoride in drinking water is

known to be at or above 1.5 mg/L. This is only compounded by additional exposures to fluoride from other sources.

- We have elaborated on why this information is important in the revised text provided below:

  "Associations between lower total fluoride exposure [e.g., represented by populations whose total fluoride exposure was lower than the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] and children's IQ remain unclear. More studies at lower exposure levels are needed to fully understand potential associations in ranges typically found in the United States (i.e., <1.5 mg/L in water). However, it should be noted that, as of April 2020, CWS supplying water with ≥1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population (~1.9 million people) (CDC Division of Oral Health 2020). This indicates that the moderate confidence in the association between higher fluoride exposure and lower IQ is relevant, at a minimum to children living in these areas of the United States where fluoride in drinking water is known to be at or above 1.5 mg/L. This is only compounded by additional exposures to fluoride from other sources."

F.45: ***Discussion page 80:*** *"This systematic review has few limitations. The human body of evidence included a large database of observational studies. Most of the observational studies were cross-sectional; however, 12 of these were considered to provide sufficient evidence that exposure occurred prior to the outcome. In addition, the systematic review covered a wide range of study designs, populations, and measures of fluoride exposure."*

[redacted] **Comment:** This is important and [redacted] think should be explained further.

Response: Disagree (no change)

- This point is expanded upon in the *Strengths of the Evidence Base* subsection of the *Discussion*. The consistency of findings across different study populations, study designs, exposure measures, and outcomes in five different countries, provides considerable support to our overall assessment.
- More detailed information is available in the sections of the monograph that have been organized by study design, including the *Results by Study Design – Prospective Studies* and *Results by Study Design – Cross-sectional Studies* sections.

F.46: ***Discussion page 80:*** *"This is not considered a limitation because these studies did not include specific information on thyroid hormones that could indicate a mechanism for thyroid involvement in neurodevelopment."*

[redacted] **Comment:** So why list it in the limitation? But also, it is a limitation in that the original studies didn't gather this, so this review couldn't assess.

Response: Disagree (no change)

- To support the transparency of systematic review methods, we documented that the approach to evaluating studies of thyroid effects changed based on expert input during the review. We understand this is a minor point and does not perfectly fit in this section. However, we considered the content is best presented here and better to report than to omit.

F.47: ***Discussion page 76:*** *"First, there are fewer limitations and greater confidence in the results of the high -quality studies. Second, there is a relatively large number of high -quality studies (n = 19), such that the body of evidence from these studies could be used to evaluate confidence in the association between fluoride exposure and changes in children's IQ."*

[redacted] **Comment:** It's still only 7k or so kids in total though, which may not be viewed as a large enough sampling.

Response: Disagree (no change)

- We disagree that 7,000 children is not a large enough sampling. For perspective, one might consider the NHANES assessments. The sample size for NHANES is 5,000 people (all ages) and is considered a representative sample of the U.S. population. The number of participants who provide biomonitoring samples is about 1/3 of that total, so it is recommended at least 4 years of data (two NHANES cycles) be combined to obtain a sample size with an acceptable level of reliability for most of the sampling domains. For example, analyses of serum fluoride levels in children (ages 6–19 years) using NHANES data typically combine two cycles with data for ~3,000–4,000 children. NHANES also assesses urinary fluoride levels, but the CDC does not make these data available to the public.

F.48: ***Summary page 81:*** *"There is, however, a large body of evidence on IQ effects in children."*

[redacted] **Comment:** Large? Maybe growing instead.

Response: Agree (no change)

- We agree that the body of evidence is a large, consistent, and growing database.

In November 2021, [redacted] received: 1) the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Sy*stematic Review, 2) a copy of the NASEM Committee's comments on the Sup04_2020_draft_NTP_Monograph with NIEHS/DNTP responses (draft version of Sup01_Monograph), and 3) the [redacted] instructions. The instructions consisted of a preface, charge, instructions for the review, and a series of specific peer-review questions grouped by the following three topics: General Comments, Human Studies, and Studies in Non-Human Animals.

[redacted] were asked to provide their substantive scientific and technical comments and suggestions within the [redacted] form. In addition, they were asked whether they "Agree", "Agree in principle", or "Do not agree" with each NTP conclusion on confidence in a body of evidence.

The [redacted] instructions and specific peer-review questions are reproduced in the pages that follow in black text. [redacted] comments and responses to each question are also provided in black text starting with the words "[redacted] **comments**" in bold font. The NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] comments and [redacted] comments:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to this [redacted] would be marked "*see DocG_Monograph or detailed response*").
- [redacted] For comments DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

Preliminary comments on the draft NTP monograph prepared by the peer review [redacted] are noted below.
These preliminary comments are not binding and should not be construed to represent NTP determination or policy.

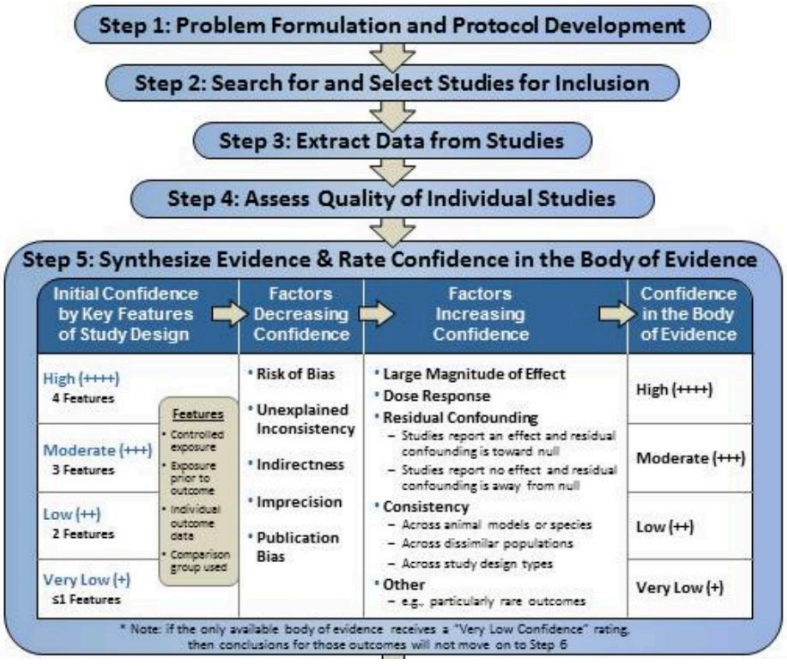
National Toxicology Program
NTP Monograph Letter Peer-Review Panel
Draft NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review

National Institute of Environmental Health Sciences
Research Triangle Park, NC

December 31, 2021

Fluoride State of the Science Document Review Form
[redacted]

Preface*:*
The objective of this evaluation was to conduct a systematic review of the published literature regarding the potential for exposure to fluoride to affect neurodevelopment and cognition in humans. The evaluation presented in the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* represents a comprehensive and current assessment. The methods used are from the Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration, which presents a seven-step framework for systematic review and evidence integration. Please note: this evaluation stops at step 5 of the systematic review process and does not proceed to step 6 to translate the confidence rating for the body of evidence into a level of evidence for health effects (see Figure 2 from the handbook).



Charge:

(1) Comment on the technical accuracy and whether the draft *NTP Monograph on State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* is clearly stated, and objectively presented.

(2) Determine whether the scientific evidence supports the NTP's confidence ratings for the bodies of evidence regarding neurodevelopmental and cognitive health effects associated with exposure to fluoride.

Instructions for Review:
All materials for this review are available in the Electronic Council Book (ECB). You will receive the specific URL and a password for accessing the ECB.

This evaluation identified 159 human studies relevant for assessing neurological health effects of exposure to fluoride; however, many studies included only secondary outcomes (e.g., 55 studies of thyroid hormones that were investigated as a potential mechanism). The scientific evidence in children and adults was evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood. Several studies evaluated learning and memory (n = 8 studies) or other cognitive developmental effects (e.g., total neurobehavioral scores and total mental capacity index in children, cognitive impairment in adults; n = 14 studies). Sixty-six human studies investigated IQ in children. Nineteen of the 66 IQ studies were determined to have low potential for bias and therefore, were categorized as "low risk of bias". Please give special attention to our assessment of these 19 studies.

- The 19 studies are available as PDFs and organized alphabetically in a folder on the ECB.
- All other studies are provided in the Health Assessment Workspace Collaborative, or HAWC database under the "studies list" tab, also organized alphabetically. You will also be provided a username and password for HAWC that will give your [redacted] permissions to access the PDFs in HAWC along with visualizations and other study information for this project at the following link (https://hawcproject.org/study/assessment/405/).

Please provide your substantive scientific and technical comments and suggestions within this [redacted] form. Identify and provide the rationale or scientific support for proposed changes or suggestions where possible.

If necessary, you can also provide additional editorial comments and recommendations for improving the report outside your specific charge questions (this form) within the draft report itself. Please note that only those comments included on the [redacted] form will be considered part of NTP's peer review report.

A. General Comments

1. Please comment on whether the scientific information presented in the draft monograph, including presentation of data in tables and figures, is technically correct, and clearly and objectively presented. Please suggest any improvements.

2. Please identify any information that should be added or deleted.

B. Human studies

I. Fluoride exposure and children's IQ

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on measures of IQ in children.

G.1: [redacted] Comments: Regarding the overall search (for all topics) in the Chinese databases (page 10), the strategy as described is unacceptable and flies in the face of the goal of systematic review. The statements "A primary goal of the screening of the newly-retrieved human references in the supplemental search of Chinese databases was to identify null, or no-effect, studies" and "Null studies that were identified were translated and included." A plain reading of these statements suggests a high degree of bias by the researchers such that evidence of an association (not null studies) were omitted. All studies should be found and included, regarding of findings.
That said, it may be that your aim was to identify studies missed due to reporting or publication bias. If that is the case, this should be stated as the primary goal. However, it might just be better to drop this concept altogether, since the purpose of searching the non-English databases was to capture studies that have not made it into (primarily Western) databases. This is sufficient explanation**.**

Response: Agree (change made)

- We agree that the search should be independent of the study findings (i.e., not dependent on whether a study found an effect of fluoride), and the primary literature search was conducted without bias for study findings in each case. However, the large number of studies from our primary literature search reporting a negative association between fluoride exposure and children's IQ raised questions about possible publication bias. In the search of Chinese databases, we conducted the literature search independent of study findings, but we initially gave translation priority to studies that appeared to show no association. Although this was done to address potential publication bias, we agree that this was not appropriate and therefore have taken additional steps to translate and extract data from all non-English studies identified from the Chinese database searches that were not previously included. Therefore, the statements about null or no-effect studies no longer apply and have been deleted from the Sup02_2022_Prepublication_NTP_Monograph. In addition, we updated the text in the *Literature Search* section to reflect that the search of Chinese databases was conducted to identify studies that may have been missed in previous searches because non-English language studies are not always indexed in the main databases used for this systematic review.

G.2: [redacted] **Comments:** As part of eligibility criteria, what is a "case study" and how does it differ from a case report? Did you have a minimum sample size? The best [redacted] could glean is >1. Table 4 implies that case series were included. Does case series mean a single group study (all had the same exposure) or a series of cases? In either case, how are these relevant? Please make explicit what the difference is between a cohort study and a cross-sectional study. I'm assuming you require cohort studies to be longitudinal, but this should be stated explicitly. Do you have a minimum duration of follow-up to count as a longitudinal study?
All of this can/should be addressed by adding Study Design rows to Tables 1-3.

Response: Agree (change made)

- Common definitions for study designs as used in environmental health research were followed. Rather than adding study design rows to Tables 1–3 of the Sup02_2022_Prepublication_NTP_Monograph, which would suggest exclusion of studies based on study design, we have added descriptions of the cohort, case-control, cross-sectional, and case report/case series study designs based on the NRC Report on Environmental Epidemiology (NRC 1997) as footnotes to Table 4 in the Sup02_2022_Prepublication_NTP_Monograph, as follows:

  "[c]Cohort studies are observational studies in humans that examine a cohort prospectively or retrospectively over time. Although cohort studies may include longitudinal analyses, it is not a prerequisite of the cohort study design.

  [d]Case-control studies are observational studies in humans that compare exposures of individuals who have a specific health effect or disease with exposures of controls who do not have the effect or disease. Controls generally come from the same population from which the cases were derived.

  [e]Cross-sectional studies are observational studies in humans that examine the relationship between exposures and outcomes or health effects assessed contemporaneously. Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).

  [f]A case report (or case study) is a descriptive study of a single individual or small group in which the study of an association between an observed effect and a specific environmental exposure is based on clinical evaluations and histories of the individual(s). A case series study in environmental epidemiology is designed to share health-related events on a collection of case reports on subjects with the same or similar health outcome(s) and environmental exposure(s)."

- The following text was also added to the Study Selection section of the Sup02_2022_Prepublication_NTP_Monograph to clarify the terms "case study" and "case report."

  "Although there are various definitions of 'case study' and 'case report,' the terms are used here to refer to publications designed to share health-related events on a single subject or patient with a disease, diagnosis, or specific outcome in the presence of a specific exposure (see Table 4 for study design definitions)."

- To be relevant for this review, a case series study would be a similar publication designed to share health-related events on a collection of case reports on subjects with neurological measures and exposure to fluoride. We did not require a

minimal sample size. It is useful to note that no case series studies were identified in the searches conducted for this systematic review.

- Regarding the cohort study design, throughout the monograph, we refer to cohort studies as prospective or retrospective, as appropriate. We did not impose a requirement that cohort studies be longitudinal, since longitudinal analyses (which have repeated measurements of the outcome over time) can be part of a cohort study but are not a prerequisite of the cohort study design. We also did not impose any restrictions on the duration of follow up. The new footnotes added to Table 4 clarify that longitudinal analyses are not required in the cohort study design.

G.3: [redacted] **Comments:** Why would you exclude conference abstracts, theses, dissertations, and other non-peer reviewed reports, but include unpublished data?

Response: Edited for clarity

- NTP only includes publicly accessible information in its evaluations. This information is typically based on studies published in peer-reviewed journals. NTP, however, can consider unpublished data or data presented in the grey literature (e.g., conference reports where the complete study details and data are available) that have not undergone peer review, provided the owners of the data are willing to have the study details and results made publicly accessible. NTP would organize a peer review of any submitted unpublished data. The Sup02_2022_Prepublication_NTP_Monograph has been revised to state that no unpublished data were included in the monograph:

  "Although no unpublished data were included in the review, unpublished data were eligible for inclusion, provided the owner of the data was willing to have the data made public and peer reviewed [see protocol (https://ntp.niehs.nih.gov/go/785076) for more details]."

G.4: [redacted] **Comments:** Regarding stopping screening as early as you did, it is not clear why it is acceptable to miss 1 or 2 relevant human studies.

Response: No change requested

By using SWIFT Active Screener software to screen the initial literature search results, we avoided the need to manually screen over 13,000 abstracts. As outlined in the Sup02_2022_Prepublication_NTP_Monograph and systematic review protocol (https://ntp.niehs.nih.gov/go/785076), in addition to the screening of bibliographical databases, several additional methods to identify relevant literature were also employed. These included publicly posting the literature search results and asking [redacted] at each stage whether they were aware of any additional relevant articles, screening the reference lists of reviews and included papers for possible articles, and conducting updated literature searches as outlined in response to a previous comment by [redacted]. The use of SWIFT Active Screener was estimated to result in the potential to miss one or two relevant human studies with primary neurodevelopmental or cognitive outcomes. The savings in time and impact were weighed against the potential impact of missing 1 or 2 studies relative to the nearly 100 human epidemiological studies identified with primary neurodevelopmental or cognitive outcomes, and this tradeoff was deemed to be acceptable.

G.5: [redacted] **Comments:** Assuming it is true, please clarify that the non-English databases were double-screened in full.

Response: Agree (no change)

- Yes, screening at both the title and abstract and full-text levels was conducted by two reviewers independently and in duplicate for the main databases and the non-English-language databases. The *Screening Process* section of the Sup03_2021_draft_NTP_Monograph and the Sup02_2022_Prepublication_NTP_Monograph state the following, which applies to all screening that was conducted during the evaluation:

  "References retrieved from the literature search were independently screened by two trained screeners at the title and abstract level to determine whether a reference met the evidence selection criteria."

  "Studies that were not excluded during the title and abstract screening were further screened for inclusion with a full-text review by two independent reviewers…"

G.6: [redacted] **Comments:** Based on the Methods section, it appears that "high quality" pediatric studies from prior to 2015 would have been excluded in the current analyses. As written (e.g., on page 14 at the end of the Data Extraction section), it seems that older data were simply ignored (without justification). However, the Results (e.g., Figure 4) includes older studies.

Response: Disagree (edited for clarity)

- This was not the intent of the identified text. As stated in the text at the end of the *Data Extraction* section of the Sup02_2022_Prepublication_NTP_Monograph, excluded studies referred to experimental animal studies that were previously reviewed and included in the NTP 2016 assessment of the experimental animal literature available at that time. The prepublication Sup02_2022_Prepublication_NTP_Monograph has been edited as follows to clarify that the statement applies to animal studies:

  "Although literature search activities for the current assessment identified experimental animal studies prior to 2015, the current assessment did not re-evaluate animal studies published prior to 2015 because these were reviewed in the NTP (2016) assessment."

G.7: [redacted] **Comments:** The description of the translated Chinese articles (page 15) needs to be written in the active voice to better describe who was confirming the accuracy of the translation and how. What about other languages (and what were they)?

Response: Agree (edited for clarity)

- We have revised the text to clarify that all translated studies were originally published in Chinese and edited the following text to clarify who confirmed the accuracy of the translations for the five low risk-of-bias studies.

  "Therefore, in order to assess whether the lack of information relevant to key risk-of-bias concerns was the result of a loss in translation, the original Chinese publications and the translated versions of the five studies that had the most potential for being included in the low risk-of-bias group of studies were reviewed by a team member with Chinese as first language to determine whether the

translations were accurate and whether any of the risk-of-bias concerns could be addressed (An et al. 1992; Chen et al. 1991 [translated in Chen et al. 2008]; Du et al. 1992 [translated in Du et al. 2008]; Guo et al. 1991 [translated in Guo et al. 2008a]; Li et al. 2009). For all five studies, the translations were determined to be accurate, and there was no impact of the translations on the key risk-of-bias concerns."

G.8: [redacted] **Comments:** The section "Considerations for Pursuing a Narrative or Quantitative Evidence Synthesis" (page 19) does not describe the actual considerations of the logic for the approach taken, but just summarizes previous meta-analyses (which don't clearly belong in the Methods section). Why is the NTP's meta-analysis not included here? Why are animal and in vivo studies not summarized? Why are adult studies not fully evaluated? Why are not low risk of bias studies not fully evaluated?

Response: Disagree (edited for clarity)

- The decision to pursue a narrative evidence synthesis rather than a meta-analysis was made while preparing the 2019 draft NTP Monograph because our goal of generating a document to support a hazard assessment did not require a quantitative estimate of hazard (e.g., numeric estimate of IQ points lost per mg F/L of drinking water or urine). To clarify the timeline of draft monographs and important decision points on content of the systematic review to address the objectives, we have added a new table (Table B-1 in *Appendix B* of the Sup02_2022_Prepublication_NTP_Monograph).

  Briefly, comments received from the NASEM Committee that reviewed the 2019 draft NTP Monograph (NTP, 2019) recommended that we perform a meta-analysis and indicated that the outcome would be critical to reaching a hazard conclusion. We therefore performed a meta-analysis, which included a *dose-response meta-analysis*, in the revised Sup04_2020_draft_NTP_Monograph (NTP, 2020). In its review of that Sup04_2020_draft_NTP_Monograph, the NASEM Committee again stated that the document fell short of supporting our hazard call, and the Committee also had additional recommendations to improve the meta-analysis.

  After reflecting on the NASEM Committee comments on the Sup04_2020_draft_NTP_Monograph, we decided to remove the evidence integration step from the systematic review of the literature and instead issue the report (after further independent peer review) as a document outlining the state of the science on the literature examining the association between fluoride exposure and potential deficits in neurodevelopment and cognition. This change is outlined in the *Preface* to the Sup02_2022_Prepublication_NTP_Monograph. Removing the evidence integration step from the systematic review precluded a determination of an overall hazard call. We then decided to revise and submit the meta-analysis as a separate peer-reviewed publication because it was no longer needed in an evaluation of confidence in the database of human evidence. An additional consideration was that the meta-analysis, and in particular the *dose-response meta-analysis*, were performed only on the studies addressing fluoride exposure in relation to deficits in children's IQ, rather than on other neurological outcomes in children or on cognition in adults. The separate meta-analysis considers comments from the NASEM Committee in its revisions.

- In the 2019 draft NTP Monograph, the in vivo experimental animal data were summarized in the main body of the text in greater detail and were considered inadequate to inform whether fluoride exposure is associated with cognitive effects in humans. Following the recommendation from the NASEM Committee review, the experimental animal section was removed from the monograph. Although we attempted to address some of the critical deficiencies identified in the animal body of evidence through further use of in-house studies, we still consider the experimental animal data to be inadequate to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans. Therefore, we have not focused on updating this large literature base in the Sup02_2022_Prepublication_NTP_Monograph.
- Experimental animal mechanistic studies are summarized in *Appendix 5* of the Sup03_2021_draft_NTP_Monograph (which is now *Appendix F* in the Sup02_2022_Prepublication_NTP_Monograph). In vitro studies were not summarized because it was considered unlikely that this literature would provide sufficient information to inform an action of fluoride on neurodevelopment considering the large number of epidemiological studies that directly addressed the question. The following text in the *Methods* section of the Sup02_2022_Prepublication_NTP_Monograph has been revised to clarify this point:

  "In vitro studies were evaluated, although data were not extracted from these studies as none of the findings were considered informative with respect to biological plausibility."
- With respect to adult studies and full assessment of all low risk-of-bias studies, adult and child low risk-of-bias studies were evaluated, and results are reported in the Sup02_2022_Prepublication_NTP_Monograph and the supporting HAWC database. The higher quality (low risk-of-bias) studies were considered and described in the greatest detail because these studies formed the basis for the confidence ratings. The lower quality (high risk-of-bias) studies in children were considered to provide support for the higher quality (low risk-of-bias) studies.

G.9: [redacted] **Comments:** Page 22, can you provide a brief explanation for why the 15 additional identified references were missed by your literature searches?

Response: Agree (change made)

- We have added a footnote to the *Literature Search Results* section to clarify why the references identified by other sources were not captured in the database searches. In brief, 11 of the 15 references identified through other sources were not indexed in the bibliographic databases searched and therefore were not captured by the database searches. Many of the studies initially identified by other sources were non-English-language studies, and we recognized that additional targeted search strategies were required to identify non-English-language studies for this review. The supplemental search of Chinese databases was designed to address these challenges. Upon further review, we have clarified that four of the references in question were captured in the Chinese database searches, and we have made this correction to the text and study flow diagram.

We were unable to identify the remaining 11 studies in any database searches. The following footnote was added to the monograph:

"These 11 studies (9 human and 2 animal studies) were not identified through the electronic database searches, as they were not indexed in any of the electronic databases searched. Note that the supplemental search of non-English-language databases was designed in part to identify non-English-language studies that are not indexed in traditional bibliographic databases such as PubMed. It was successful in this goal, as multiple studies that were initially only identified through "other sources" were subsequently captured in the supplemental Chinese database search, leaving only 11 as identified through other sources."

- Regarding the impact of these 11 studies on the systematic review, only 1 of the 11 studies was a low risk-of-bias IQ study in children, and this study was included in the 19 low risk-of-bias studies on which the moderate confidence in the IQ-in-children body of evidence is based. The omission of this single study would not impact the moderate confidence rating. Of the remaining 10 studies, 7 were high risk-of-bias IQ-in-children studies and 1 was a high risk-of-bias adult study. The omission of the 7 (out of 53) high risk-of-bias IQ-in-children studies or the 1 (out of 8) high risk-of-bias adult studies would not impact any confidence conclusions in the monograph. Similarly, the two experimental animal studies would not impact the evaluation as the animal evidence was considered inadequate.

G.10: [redacted] **Comments:** The numbers of abstracts in Figure 2 do not align with the text.

Response: Agree (change made)

- We have made this correction.

G.11: [redacted] **Comments:** Figure 2, it would be better to separate out conference abstracts for "Other".

Response: Disagree (no change)

- The "other" excluded studies category includes conference abstracts among several other publication types that are not useful for a systematic review because they do not have complete presentations of data, methods, and results. In addition, the interactive version of the study flow diagram (Figure 2; https://hawcproject.org/summary/visual/assessment/405/Figure-2/) allows the reader to select the "other" category of excluded references, display bibliographic information for each study, and determine which references were conference abstracts.

G.12: [redacted] **Comments:** The top line of the excluded box should state that the 333 were from the original (pre-2020) search.

Response: Agree (edited for clarity)

- Footnotes to Figure 2 were revised (https://hawcproject.org/summary/visual/assessment/405/Figure-2/) to clarify that the first group of 333 studies excluded at the full-text level (footnoted with a single asterisk) were from all literature searches conducted during the review, and the second group of 156 studies excluded at the full-text level (footnoted with two asterisks) were from the 2020 literature search update only.

*"*Studies from all literature searches conducted during the review excluded at the full-text level for pre-established criteria…*

***Studies excluded from the 2020 literature searches for reasons other than pre-established criteria…"*

G.13: [redacted] **Comments:** On page 24, [redacted] have trouble with the Results statement "Congenital neurological malformations and neurological complications of fluorosis are not considered further due to the limited number of studies and the heterogeneity of outcomes evaluated in these studies." This belongs in the Methods, complete with a full explanation for criteria used to or not to report/consider.

Response: Disagree (edited for clarity)

- The *Results* section is the most appropriate location for this statement. As part of the systematic review, our study selection process identified studies with several different neurological endpoints. As stated in the *Methods* section of the monograph, the grouping of health effect results was performed based on the type and extent of data identified through the literature search and was not planned a priori. Although the process for deciding which groupings of health effects to synthesize and whether to synthesize all groupings of health effects was described in the protocol, the specific decisions were made based on the results of the literature search and selection. Following our review of the full body of epidemiological literature, we decided that the most appropriate focus of the monograph would be on neurodevelopment and cognition. Congenital neurological malformations and neurological complications of fluorosis were not considered further due to the limited number of studies and the heterogeneity of outcomes evaluated in these studies. The most appropriate location for this statement is in the *Results* section, as this decision was made as part of the evidence synthesis and was not an a priori methods decision; however, the referenced sentence has been edited to state that a few studies on these other health outcomes were identified, which is more in line with other statements in the *Results* section.

2. Comment on whether the approach used to assess risk of bias was clearly described and appropriately applied to the set of studies designated as "low risk of bias."

G.14: [redacted] **Comments: The approach used to assess risk of bias was clearly described and appropriate applied.**

Response: No change requested

- No response necessary.

G.15: [redacted] **Comments:** Detailed descriptions of assessment of risk of bias in animal studies (page 14) should be omitted since this was not done. It is confusing for the reader to repeatedly read methods that are not relevant to this review.

Response: Disagree (no change)

- We appreciate [redacted] comment but consider it appropriate to provide details on the methods for evaluating risk of bias in animal studies. Risk-of-bias assessment was conducted on the animal studies and was considered in determining the confidence in the animal data and in deciding whether to reach a

hazard conclusion. The risk-of-bias assessment of animal studies was a factor in determining that the animal studies were inadequate to inform the question of an association between fluoride exposures and neurodevelopmental and cognitive effects in humans and contributed to the decision to focus the systematic review on the large evidence base of human studies.

G.16: [redacted] **Comments:** Regarding the GRADE assessment of imprecision, please provide more detail about your thresholds between precise and imprecise. For example, what 95% CI would indicate imprecise?

Response: Agree (change made)

- The consideration of precision is outlined in the protocol (https://ntp.niehs.nih.gov/go/785076) with additional perspective provided in the OHAT handbook (https://ntp.niehs.nih.gov/go/ohathandbook). The GRADE-based assessment was used to determine whether there was no serious imprecision (essentially that data were precise), serious imprecision, or very serious imprecision. There are two principal considerations in reaching a judgement of no serious imprecision (see Table 12 of the OHAT handbook; https://ntp.niehs.nih.gov/go/ohathandbook). First, there are no or minimal indications of large standard deviations (i.e., SD > mean). In addition, for ratio measures, the ratio of the upper to lower 95% CIs for most studies (or meta-estimate) is <10 and, for absolute measures (e.g., percent control response), the absolute difference between the upper and lower 95% CIs for most studies (or meta-estimate) is <100.
- We have revised the sentence in the *Methods* section of the Sup02_2022_Prepublication_NTP_Monograph about the consideration of imprecision as follows:

  "There is no evidence of serious imprecision that would warrant a downgrade. Eighteen low risk-of-bias studies reported lower IQ with higher fluoride, and no issues with imprecision were identified to challenge the significance of the response estimates."

3. Comment on assessment of the human studies with regard to:

a) How findings from individual studies designated as "low risk of bias" were interpreted.

G.17: [redacted] **Comments:** A small point, but [redacted] think the description of 19 studies somewhat exaggerates the size of the body of evidence, since these studies were conducted in 15 study populations. For example, on page 36, it is unclear why the two articles by Green and Till should get double the weight (2 vs. 1 study) simply because the authors chose to publish 2 (vs. 1) articles.

Response: Disagree (edited for clarity)

- We disagree that the description of 19 studies exaggerates the size of the body of evidence. It is made clear in the Sup03_2021_draft_NTP_Monograph and the Sup02_2022_Prepublication_NTP_Monograph that these 19 studies come from 15 study populations. We also are transparent about the fact that the three prospective cohort studies are based on two study populations. We have added clarifying text in the first paragraph of the *Results by Study Design – Prospective*

Cohort Studies section (see quote below) that, while Green et al. (2019) and Till et al. (2020) use the same study population, the exposure measures used are different between the two publications, thus warranting consideration as separate studies.

"Two of the studies (Green et al. 2019; Till et al. 2020) were based on the same Canadian study population, but one evaluated prenatal fluoride exposure and the other evaluated postnatal fluoride exposure. Green et al. (2019) included maternal urinary fluoride, maternal fluoride intake, and water fluoride concentrations, while Till et al. (2020) used fluoride intake from formula or water concentrations in formula-fed versus breastfed infants."

G.18: [redacted] **Comments:** The overall conclusion is appropriate, but the organization of results is problematic and suggests possible bias and cherry-picking by the reviewers. [redacted] don't think there was bias or cherry picking, but the text raises these concerns. For example, on page 36, the statement that 18 of 19 studies provide consistent evidence is a misleading truism, since the 19th study was omitted for being inconsistent. Better to talk about all 19 studies (or 15 study populations) and go from there. There was consistency across 19 studies, with only a single study finding no association.
The strength of findings *must* be across *all* 19 studies, not just the 18 studies that directly support the conclusion.

Response: Disagree (no change)

- Although we agree that the higher-quality studies provide consistent evidence of an inverse association (and statements like this appear in the monograph), we decided that, in sentences that note the number of studies, our phrasing that "18 of 19 studies provide consistent evidence" is more appropriate and transparent since 18 of 19 studies provide evidence of an inverse association and the remaining study does not.

G.19: [redacted] **Comments:** It is unclear whether you are using the term prospective (cohort study) to mean prospective (as opposed to retrospective, it's correct meaning) or longitudinal (as opposed to cross-sectional). Please use the correct term. In any case, the reader needs to know both whether studies were prospective or retrospective and whether studies were longitudinal or cross-sectional.

Response: Agree (change made)

- Throughout the monograph, we refer to cohort studies (either prospective or retrospective) with consistent terminology. We have added definitions from NRC (1997) as footnotes to Table 4 to clarify study design types in the document. The following text is an example with the description of cohort studies:

  "[c]Cohort studies are observational studies in humans that examine a cohort prospectively or retrospectively over time. Although cohort studies may include longitudinal analyses, it is not a prerequisite of the cohort study design."

G.20: [redacted] **Comments:** Throughout, it is important to talk about *higher* exposure to fluoride, not just exposure to fluoride. Everyone is exposed to fluoride so describing the at-risk group as being exposed to fluoride is meaningless (and confusing).

Response: Agree (change made)

- We recognize [redacted] suggestion concerning use of the term "higher" when describing fluoride exposure. Many of the epidemiological studies evaluated involved simple comparisons between groups of children exposed to higher versus lower levels of fluoride, but there were wide variations in the actual fluoride exposure levels that comprised higher and lower, and some of the levels overlapped from study to study. In response to this request and requests of other [redacted], we have carefully reviewed the terminology used in the Sup02_2022_Prepublication_NTP_Monograph and have added the "higher" qualifier when appropriate to describe fluoride exposures, and have provided a benchmark (1.5 mg F/L WHO Guidelines for Drinking-water Quality) to aid in describing total exposure above which moderate confidence was determined for children's IQ studies.

G.21: [redacted] **Comments:** The Results about the cross-sectional studies (page 38) state that "the consistent results across multiple metrics increase our confidence in the data." Based on the appropriate description in the Methods on Page 21, upgrading based on Consistency "does not apply in this evaluation".

Response: Agree (edited for clarity)

- Confidence in bodies of evidence is evaluated under the GRADE-based system in the NTP OHAT approach by considering a specific set of factors that may either decrease or increase the confidence rating. Consistency of results is considered in two of these confidence factors which separately address different aspects: 1) unexplained inconsistency of results that may reduce confidence, and 2) consistency of results across multiple animal species that may increase confidence. [redacted] points out one part of the guidance on consistency under the *Methods* section *Factors to Consider for Potential Upgrading* that generally only applies to animal studies; however, there is also a section under *Factors to Consider for Potential Downgrading* that applies to both human and animal studies. The text in question in the *Results* section was revised to clarify that the consistency of the results across multiple metrics contributes to (rather than increases) the confidence in the body of evidence. As explained in that downgrading section, the consistency of results across human studies is used as the reason not to downgrade for unexplained inconsistency, and therefore was not considered further as a potential upgrade for essentially the same characteristics of the body of evidence. As explained in the upgrading section, an upgrade is typically applied when there are data reporting consistent results across multiple animal species. Upgrading for the human studies does not apply in this evaluation.

G.22: [redacted] **Comments:** Missing is a clear description or analysis across studies of what constitutes higher exposure levels that are associated with lower IQ. Page 40 starts a description of Exposure Levels, but lacks any quantitative description of high (or low) exposure. While [redacted] understand that a better analysis may arise from the future meta-analysis, there should be enough data in Table 6 to allow a more coherent summary of exposure level thresholds analyzed.
The *Summary of Key Findings for Low Risk-of-bias Children's IQ Studies* on Page 42 (and again on page 48) describes higher exposure as ≥1.5 mg/L, but other than a mention of the 2016

report, this threshold is not described or presented in the Results. To the reader, this threshold is unsupported by the included studies.

Response: Agree (change made)

- We agree that the Sup03_2021_draft_NTP_Monograph could have provided additional context for the use of the WHO Guidelines for Drinking-water Quality value of 1.5 mg F/L, so we have added language to address this in the Sup02_2022_Prepublication_NTP_Monograph. For example, the first sentence in the *Summary of Key Findings for Low Risk-of-bias Children's IQ Studies* section has been revised to state:

  "In summary, the high-quality studies (i.e., studies with low potential for bias) consistently demonstrate lower IQ scores with higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)]."

- In addition, it is useful to point out that, in the Sup04_2020_draft_NTP_Monograph, which included the draft meta-analysis, we presented information addressing the statistical significance of the findings from drinking water studies where the median exposed group-level exposures included all studies, those less than 4 ppm, those less than 2 ppm, and those less than 1.5 ppm. These cut-off points were described as being selected because they represent various regulatory rather than biological thresholds. In this respect, please also see response to [redacted] comment above regarding "higher" exposure to fluoride.

G.23: [redacted] **Comments:** Figure 6 has a header "Reported Effect of Fluoride" with just Soto Barreras with "No". However, Table 6 reports the findings for Soto Barreras as no significant difference. Maybe what is meant is Reported Significant Association Between Fluoride Exposure and IQ.

Response: Agree (no change)

- We agree with [redacted] description of Figure 6 and Table 6, but the notations in Figure 6 and Table 6 are correct. As summarized in Table 6, no significant differences in measured fluoride levels across IQ grades were observed. This Table 6 summary coincides with Figure 6, which dichotomously summarizes (i.e., with a 'yes' or 'no') that the study did not report an association with fluoride. Note that *Appendix E* (previously *Appendix 4* in the Sup03_2021_draft_NTP_Monograph) provides further detail on the study results, as follows:

  *"**Reported association with fluoride exposure:** No: Results were not presented to evaluate an association between fluoride exposure and IQ but to compare fluoride levels within IQ grades. For this reason, the results of this study are not comparable to other studies that evaluated IQ scores by fluoride exposure levels. No significant differences in measured fluoride levels across IQ grades were observed."*

G.24: [redacted] **Comments:** Also note that observational studies (with rare exceptions) do not provide evidence of an "effect", only of an association. Please use the term judiciously or not at all.

Response: Agree (change made)

- Edits have been made throughout the Sup02_2022_Prepublication_NTP_Monograph to use the terms 'effect,' 'association,' and 'correlation' consistently and most appropriately. For example, the following description of Table 6 was revised to replace 'effect' with 'association':

  "The purpose of the table is to summarize key findings (independent of whether an association is indicated) from each study…"

G.25: [redacted] **Comments:** Did the reviewers consider that the relatively small Soto Barrera study was "negative" because it was underpowered, not because it was inconsistent with the other studies?

Response: No change requested

- The study is noted as not observing evidence of an association and is not referred to as "negative" in the text. Because of the way the data are provided (i.e., the study evaluated fluoride levels within IQ grade and not whether IQ changed with increasing fluoride exposure), a comparison with the rest of the body of evidence cannot be made. Soto-Barreras et al. (2019) does not provide the data in a way that would allow evaluation of the association between fluoride exposure and IQ. Because the scope of this section is to present the observed IQ effects in children, we refrain from suggesting reasons for non-significance such as sample size.

b) How the overall set of confounders across the body of evidence from children's IQ studies was considered and presented.

G.26: [redacted] **Comments**: Overall, the confounders are considered and presented reasonably well.

Response: No change requested

- No response necessary.

G.27: [redacted] **Comments:** On page 42 and following, it would be clearer to separately and clearly discuss each of the 3 studies at increased risk confounding bias (Cui, Ding, Soto Barreras). The reasons these three studies were downgraded are buried in the text and unclear. [redacted] gleaned why Soto Barreras was at high risk, but I'm unclear why the other two studies are high risk.

Response: Disagree (no change)

- Overall, as described in the *Confounding for IQ Studies in Children* section, we determined that bias due to confounding was not considered to be a concern in the body of evidence, and the potential for the consistency in results to be attributable to bias due to confounding in the 19 low risk-of-bias studies was considered low. Therefore, we still consider that the most appropriate organization for the section on confounding for low risk-of-bias IQ-in-children

studies is to first summarize the strengths in the body of evidence regarding the potential for bias due to confounding. We also consider *Appendix E* in the Sup02_2022_Prepublication_NTP_Monograph (previously *Appendix 4* in the Sup03_2021_draft_NTP_Monograph), which the section in question directly refers to for further details for these three studies, to be the most appropriate place in the monograph to fully discuss the potential for bias due to confounding for these three studies (as well as all low risk-of-bias studies).

G.28: [redacted] **Comments:** It is unclear why Soto Barreras is considered to be low risk of bias (overall) if they did not account for arsenic in a high-exposure area. This seems like a major flaw. [redacted] did not find any description in the main part of the results (pages 28-41) that discuss this study and why it's included. Although, there's the unclear, unreferenced statement (page 36) that "Only one study did not observe evidence of an association between fluoride exposure and IQ; however, results were not provided in a manner that allowed for a direct comparison with other low risk-of-bias studies."

Response: Agree (change made)

- We have added text to further explain why the concern over co-exposure to arsenic in Soto-Barreras et al. (2019) would not result in the study being considered high risk of bias overall. In the Sup02_2022_Prepublication_NTP_Monograph, the following text has been added to the *Confounding for IQ Studies in Children* section for low risk-of-bias studies: "*Although Soto-Barreras et al. (2019) did not discuss arsenic, there is no direct evidence that arsenic was present in the study area.*" We have also added a direct link to *Appendix E*, which discusses the concern for this study in greater detail. In order for Soto-Barreras et al. (2019) to be considered high risk of bias overall due to the arsenic concern alone, there would need to be direct evidence that arsenic was driving the results (in this case, biasing the association toward the null). We do not have direct evidence of that for this study.

G.29: [redacted] **Comments:** The Outcome Assessment for IQ Studies section (page 48) is unclear. This problem occurs in much of the write up, where it is unclear what studies are being referred to. It states that 18 of 19 studies were low risk ("used appropriate methods for measuring IQ"), but does not indicate which study did not use appropriate methods or what the problem is. At the end of the paragraph there's a sentence about Sudhir not reporting blinding, but the paragraph starts by saying that "blinding of outcome assessors was not a concern).

Response: Agree (change made)

- We have made edits throughout the Sup02_2022_Prepublication_NTP_Monograph to further clarify the studies to which we are referring. In the referenced section, we have revised the text to clarify that, "*All 19 low risk-of-bias studies used appropriate methods for measuring IQ in the study population being assessed, and blinding of outcome assessors was not a concern in 18 of the 19 studies [i.e., all low risk-of-bias studies except Sudhir et al. (2009)].*"

c) How the confidence rating in the body of evidence was developed and supported.

G.30: [redacted] **Comments:** It's unclear what is meant (in Table 6 and scattered throughout the Results) that there was "No statistical adjustment for confounders" but then in Figure 6 (and also in the text) all studies "consider" the potential confounders age, sex, and SES.

Response: Agree (change made)

- Table 6 reports the statistical adjustment for covariates for each publication. Figure 6 takes it a step further and reports whether a potential covariate was a concern as a potential confounder. We updated the footnote to Figure 6 to clarify what we mean by "consider."

G.31: [redacted] **Comments:** An extension of a prior comment, on page 48 (Confidence Assessment of Findings on IQ in Children), the review does not provide evidence to support the statement that the high fluoride exposure should be interpreted as "mainly greater than the WHO Drinking Water Quality Guideline [≥1.5 mg/L]".

Response: Agree (change made)

- Although the specific statement referred to no longer appears in the *Confidence Assessment of Findings on IQ in Children* section, reference to the WHO Guidelines for Drinking-water Quality provides important clarification and context that the findings of this systematic review are on total fluoride exposure and there is uncertainty at lower exposure levels. As discussed in the Sup02_2022_Prepublication_NTP_Monograph, there is moderate confidence from low risk-of-bias studies of an association between higher fluoride exposure and lower IQ in children when total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg F/L. The statement reflects the collective assessment of fluoride exposure in the low risk-of-bias studies that form the basis of our confidence assessment for an association between higher fluoride exposure and lower IQ in children. It also relies on empirical observations of a close correspondence between drinking water concentrations and urinary fluoride concentrations (see Kumar et al. [2017] as an example). Our assessment of confidence in the association between higher fluoride exposure and lower children's IQ is supported by studies that report total fluoride exposure as represented by urinary measurements.
- For example, the publication by Green et al. (2019) examined the IQ of children in Canada exposed to fluoride in both non-fluoridated communities and fluoridated communities with 0.7 mg/L in drinking water (the same recommended fluoridation level for community water systems in the United States). As reported in that publication, individual exposure levels of women living in optimally fluoridated cities in Canada, as documented by repeated urinary measurements, suggest widely varied total exposure from water combined with fluoride from other sources. Many of these urinary fluoride measurements exceed those expected from consuming fluoride in water alone that contains 1.5 mg/L fluoride or less. The Bashash et al. (2017) study also provided information from a population in Mexico whose urinary fluoride exposures were comparable to those identified in the Green et al. (2019) study. Both studies are reviewed in the monograph and contribute to our confidence conclusions as stated in the revised *Abstract* and *Summary* sections of the Sup02_2022_Prepublication_NTP_Monograph.

"This review finds, with moderate confidence, that higher fluoride exposure (e.g., represented by populations whose total fluoride exposure approximates or exceeds the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride) is consistently associated with lower IQ in children. More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ."

G.32: [redacted] **Comments:** Similarly, the description around "Dose-response" on page 49 is not clearly supported by the text of the Results section. There is no clear dose-response section of the Results where related findings are described. The Results text mostly summarizes as "high" or "exposure" or in some instances association with a 1-mg/L increase or the equivalent.

Response: Agree (edited for clarity)

- As a result of removing the meta-analysis, which includes a *dose-response meta-analysis*, from the Sup03_2021_draft_NTP_Monograph, the referenced description around "Dose-response" is no longer necessary and has been removed from the Sup02_2022_Prepublication_NTP_Monograph.

G.33: [redacted] **Comments:** The "Consistency" section on page 49 should not discuss the consistency across studies. This was addressed in Unexplained Inconsistencies on the prior page. Do not confuse the two issues for the reader.

Response: Agree (change made)

- We agree that the text on consistency of findings across studies in humans was addressed under unexplained inconsistency. Therefore, we accepted the suggestion and deleted the sentence marked by [redacted] "*The high quality studies demonstrate a consistent pattern of findings that fluoride exposure is associated with lower IQ scores in children"* from the *Consistency* bullet in the *Confidence Assessment of Findings on IQ in Children* section. The *Consistency* bullet in the Sup02_2022_Prepublication_NTP_Monograph has been revised to state:

  *"**Consistency**: The consideration of a potential upgrade for consistency in the methods is primarily for non-human animal evidence, where it would be applied to address increased confidence for consistent effects across multiple non-human animal species. For human evidence, it is generally not applied, and the data would only be considered in deciding whether to downgrade for unexplained inconsistency. Therefore, no upgrade is applied for consistency."*

G.34: [redacted] **Comments:** A clearer statement, up front, is needed that the starting point for confidence is "moderate" and why this is the case. [redacted] think you're trying to say this with "The initial moderate confidence rating in the body of evidence" on page 48, but this sentence is unclear. I'm still unsure if "initial" here means where the GRADE confidence rating starts before assessing the evidence. Why start at moderate? The Methods section does not describe this concept.

Response: Agree (change made)

- The *Confidence Assessment of Findings on IQ in Children* section has been rearranged in the Sup02_2022_Prepublication_NTP_Monograph as suggested to start with the initial confidence rating and how initial confidence rating is determined. Then, the OHAT approach is outlined for considering factors that may

upgrade or downgrade confidence to reach a final rating on confidence in the body of evidence.

G.35:

4. ***NTP concludes a rating of moderate confidence in the body of evidence for lower IQ in children associated with fluoride exposure***.

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the moderate confidence rating.

II. Fluoride exposure and non-IQ neurodevelopmental or cognitive effects in children

G.36:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on non-IQ neurodevelopmental or cognitive effects in children.

[redacted] **Comments: Issues discussed above also pertain to this section (e.g., number of articles vs. number of study populations). No additional issues.**

Response: Disagree (edited for clarity)

- As described in response to [redacted] comment on B.I.3.a, we clearly specify the number of studies and study populations. We have also added clarifying text to the *Results by Study Design – Prospective Cohort Studies* section that, while Green et al. (2019) and Till et al. (2020) use the same study population, the exposure measures used are different between the two publications, thus warranting consideration as separate studies (see below).

  "Two of the studies (Green et al. 2019; Till et al. 2020) were based on the same Canadian study population, but one evaluated prenatal fluoride exposure and the other evaluated postnatal fluoride exposure. Green et al. (2019) included maternal urinary fluoride, maternal fluoride intake, and water fluoride concentrations, while Till et al. (2020) used fluoride intake from formula or water concentrations in formula-fed versus breastfed infants."

G.37:

2. Comment on whether the approach used to assess risk of bias for studies in children on non-IQ neurodevelopmental or cognitive effects was clearly described and appropriately applied.

[redacted] **Comments: Sufficiently well described and appropriately applied, with caveats related to issues raised in IQ section.**

Response: No change requested

- No response necessary.

3. Comment on assessment of the human studies with regard to:

G.38:

a) How findings from individual "low risk of bias" studies were interpreted.

[redacted] **Comments: Sufficiently well described and appropriately applied, with caveats related to issues raised in IQ section.**

Response: No change requested

- No response necessary.

G.39:

b) How the confidence rating in the body of evidence was developed and supported.

[redacted] **Comments:** Sufficiently supported, but it's unclear why the same format used for the IQ studies (pages 48-49) is not used here (page 59).

Response: Agree (edited for clarity)

- Due to the limitations of the data set, including the heterogeneity of outcomes assessed, a limited number of directly comparable studies, and differences in outcome assessment methods even when studies evaluated similar outcomes, we chose not to describe the confidence assessment in the same format and level of detail as in the IQ section. We have added text to the *Confidence Assessment of Findings on Other Neurodevelopmental Effects in Children* section to further clarify this point.

G.40:

4. ***The NTP concludes a rating of low confidence in the body of evidence for decreases in measures of other neurodevelopmental or cognitive effects in children associated with fluoride exposure***.

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

III. Fluoride exposure and cognitive effects in adults

G.41:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on cognitive effects in adults.

[redacted] **Comments:** No concerns with methods or search specific to this topic.

Response: No change requested

- No response necessary.

G.42:

2. Comment on whether the approach used to assess risk of bias for studies in adults on cognitive effects was clearly described and appropriately applied.

[redacted] **Comments:** Risk of bias of the two studies clear and mostly appropriately applied.

Response: No change requested

- No response necessary.

G.43:

3. Comment on assessment of the human studies with regard to:

a) How findings from individual studies were interpreted.

[redacted] **Comments:** The analysis between the two studies may be too simplistic. The French study was done in adults with not very high exposures to fluoride. In contrast the Chinese study compared adults with skeletal fluorosis (suggesting very high exposure) with others. It may be inaccurate to suggest that these two studies were not consistent. They may (consistently) show that relatively low exposures (even if above recommended) are not associated with cognitive outcomes, but very high exposures are.
This gets a [redacted] comments before about a lack of analysis regarding doses, dose effects, or thresholds.

Response: Agree (change made)

- We recognize that it may be inaccurate to suggest that these two studies were not consistent. We have revised the first sentence in the *Summary of Results* to say:

  "Results from two low risk-of-bias studies in adults did not provide enough evidence to evaluate consistency when assessing evidence for a potential association between fluoride exposure and cognitive impairment (based on the MMS Examination)".

G.44:

b) How the confidence rating in the body of evidence was developed and supported.

[redacted] **Comments:** Adequate

Response: No change requested

- No response necessary.

G.45:

4. ***The NTP concludes a rating of low confidence in the body of evidence for changes in cognitive effects in adults with fluoride exposure***.

X Agree
Although, better would be to say insufficient evidence and leave it at that. As per the NTP system, there is low confidence in a lack of a conclusion.

☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

C. Studies in non-human animals

G.46:

The NTP agrees with the comments of the NASEM committee (NASEM 2020, 2021) concerning the overall poor quality of the experimental animal database on fluoride exposure and neurodevelopmental effects, with many studies suffering from major reporting deficiencies. As indicated above, the monograph focuses on the large human epidemiology database because it directly addresses the question of whether fluoride affects human neurodevelopment.

Therefore, based on the recommendations of the NASEM committee, the experimental animal section and risk of bias details have been removed from this monograph and ***the NTP concludes that the scientific evidence from experimental animal data are inadequate to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans***.

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the inadequate designation.

D. Additional Comments:

G.47:

1. [redacted] **comments:** A clearer statement is needed up front, ideally in the Introduction about what topics were covered by full systematic review (which are a small subset of topics of interest) and why. It's very confusing to read through repeated descriptions of topics which are not being reviewed. As an example, it's disconcerting to repeatedly see that thyroid function is an outcome of interest (without an explanation as to why this is of interest to a review of neurodevelopmental and cognitive health effects) and then to come across the statement (page 13) that "Thyroid data were … not extracted." It's difficult to pick out and follow the reasoning for excluding most topics from full evaluation. The timing of and reasoning for the decisions to focus the systematic review on just "high quality" pediatric studies is unclear.

Response: Agree (change made)

- We have taken steps to increase clarity in the Sup02_2022_Prepublication_NTP_Monograph regarding the purpose of exploring thyroid function and the exclusion of some topics for full evaluation. Firstly, we have added a footnote to the *Introduction* section to clearly explain the interest in thyroid function as a potential mechanism for neurodevelopmental effects. The footnote reads:

  "The current review has evaluated the fluoride literature with an eye toward potential thyroid effects because a large literature base has accumulated examining the interaction of fluoride with iodine uptake by the thyroid gland and consequential effects on synthesis of thyroid hormones, which are recognized to play significant roles in neurodevelopment in utero and during early childhood. This literature, along with a detailed proposed mechanism of action, was recently reviewed by Waugh (2019)."

- Secondly, details have been added to the *Data Extraction* methods discussion to further clarify why data on specific endpoints were not considered informative to the systematic review and did not undergo full data extraction or study quality evaluation, as follows:

  "Data for primary and secondary outcomes, as well as thyroid hormone level data, were extracted from human studies. Studies evaluating only goiters or thyroid size were not extracted because they do not provide specific information on thyroid hormone levels that would inform whether a thyroid-mediated mechanism was involved in fluoride-associated changes in neurodevelopment."

- Thirdly, the sentence that [redacted] identifies ("Thyroid data were … not extracted.") has been revised to clarify that animal thyroid data were not extracted, whereas human thyroid data were extracted.
- We reviewed the document *Introduction*, *Objective*, *PECO Statements*, *Literature Search*, and *Data Extraction* sections to assure there is clarity in how the objectives were addressed in the systematic review. This includes decisions that were made based on the types of data that were identified in the literature searches. However, we disagree with [redacted] suggestion that these details should be moved to the *Introduction*, which should describe the objectives not the actual results of the search or specifics of data extraction.
- We consider that the monograph appropriately and clearly addresses the reasoning to focus the systematic review on the IQ studies in children, to consider all of the studies (both high quality and low quality), and to primarily base the confidence rating of moderate on the high-quality studies. The monograph states in the *Health Outcome Categories for Neurodevelopmental and Cognitive Effects* section and the *Human Neurodevelopmental and Cognitive Data* section that the vast majority of the human studies evaluated IQ in children; therefore, the discussion of cognitive neurodevelopmental effects in children focuses on IQ studies. In addition, the *IQ in Children* section explains the reasoning for focusing on high-quality studies by stating:

  "All available studies were considered in this evaluation; however, review of the body of evidence focused on the high-quality, low risk-of-bias studies for two main reasons. First, there are fewer limitations and greater confidence in the results of the high-quality studies. Second, there are a relatively large number of high-quality studies (n = 19), such that the body of evidence from these studies could be used to evaluate confidence in the association between fluoride exposure and changes in children's IQ. Therefore, the remainder of the discussion on IQ in children focuses on the 19 studies with low risk of bias. The high risk-of-bias studies are discussed briefly relative to their overall support of findings from the low risk-of-bias studies."

G.48:

2. [redacted] **comments:** Of note, the Introduction and Methods do not explain why thyroid function was evaluated. This was only (partially done) on page 63.

Response: Agree (change made)

- We have added a footnote to the *Introduction* section of the Sup02_2022_Prepublication_NTP_Monograph to explain the focus on potential thyroid effects. The footnote reads:

  "The current review has evaluated the fluoride literature with an eye toward potential thyroid effects because a large literature base has accumulated examining the interaction of fluoride with iodine uptake by the thyroid gland and consequential effects on synthesis of thyroid hormones, which are recognized to play significant roles in neurodevelopment in utero and during early childhood. This literature, along with a detailed proposed mechanism of action, was recently reviewed by Waugh (2019)."

G.49:

3. [redacted] **comments:** It is odd that the Discussion presents the topics in almost reverse order from the Results. [redacted] would expect to start with human evidence, children first, and IQ, then neurocog, then adults, then animal, etc.

Response: Agree (no change)

- We agree that it may be unusual to present the topics in the *Discussion* in a different order from the *Results*, and we had initially arranged the *Discussion* section in the order that [redacted] suggests; however, the current order was ultimately chosen for several reasons. While the epidemiological evidence for adverse effects of fluoride exposures on children's cognition appears strong, our overall confidence in the database is judged to be moderate. The deficiencies in the experimental animal data and the lack of a clear mechanistic understanding of how fluoride may be producing these effects are important to keep in mind, and they tended to be lost when placed later in the *Discussion*. Also, separating the main discussion of the epidemiological findings from the *Strengths of the Evidence Base* and *Limitations of the Evidence Base* sections was not considered optimal. Thus, we prefer to maintain the current *Discussion* structure.

G.50:

4. [redacted] **comments:** Although, [redacted] mentioned the issue before, it is notable that the Discussion does not address the evidence regarding dose effect or threshold.

Response: No change requested

- Earlier drafts of the monograph that were reviewed by the NASEM Committee included a more prominent discussion of dose effects, and the meta-analysis requested by the Committee dealt with this issue directly. However, the systematic review was designed to simply address the question of whether there is evidence for an association between fluoride exposure and cognitive neurodevelopment irrespective of dose. We have always considered these as two separate questions and found that combining them into one document ultimately detracted from an unbiased independent assessment of either. Thus, we deemphasized references to current water fluoridation practices in the Sup03_2021_draft_NTP_Monograph and the Sup02_2022_Prepublication_NTP_Monograph and have addressed the concept of thresholds by applying several data modeling approaches to the children's IQ data in a meta-analysis manuscript to be published separately.

References

Kumar S, Lata S, Yadav J, Yadav JP. 2017. Relationship between water, urine and serum fluoride and fluorosis in school children of Jhajjar District, Haryana, India. Applied Water Science. 7:3377–3384. doi: 10.1007/s13201-016-0492-2.

National Research Council (NRC). 1997. Environmental Epidemiology: Volume 2: Use of the Gray Literature and Other Data in Environmental Epidemiology. National Research Council (US) Committee on Environmental Epidemiology; National Research Council (US) Commission on Life Sciences. Washington (DC): National Academies Press (US).

In November 2021, [redacted] received: 1) the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review,* 2) a copy of the NASEM Committee's comments on the 2020 draft NTP Monograph with NIEHS/DNTP responses (draft version of Sup01_Monograph), and 3) the [redacted] instructions. The instructions consisted of a preface, charge, instructions for the review, and a series of specific peer-review questions grouped by the following three topics: General Comments, Human Studies, and Studies in Non-Human Animals.

[redacted] were asked to provide their substantive scientific and technical comments and suggestions within the [redacted] form. In addition, they were asked whether they "Agree", "Agree in principle", or "Do not agree" with each NTP conclusion on confidence in a body of evidence.

[redacted] instructions and specific peer-review questions are reproduced in the pages that follow in black text. [redacted] comments and responses to each question are also provided in black text starting with the words "[redacted] **comments**" in bold font. The NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] comments and [redacted] comments:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph~~Track Changes 2022 NTP Monograph~~ that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to this [redacted] would be marked "*see DocH_Monograph for detailed response*").
- [redacted] For comments DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

Preliminary comments on the draft NTP monograph prepared by the peer review [redacted] are noted below.
These preliminary comments are not binding and should not be construed to represent NTP determination or policy.

National Toxicology Program
NTP Monograph Letter Peer-Review Panel
Draft NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review

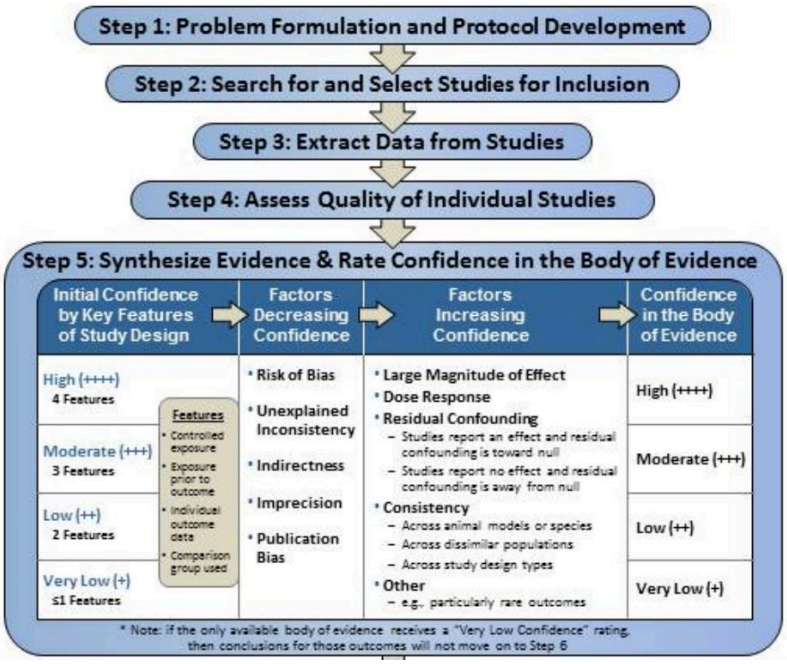
National Institute of Environmental Health Sciences
Research Triangle Park, NC

January 2, 2022

Fluoride State of the Science Document Review Form

[redacted]

Preface*:*
The objective of this evaluation was to conduct a systematic review of the published literature regarding the potential for exposure to fluoride to affect neurodevelopment and cognition in humans. The evaluation presented in the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* represents a comprehensive and current assessment. The methods used are from the Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration, which presents a seven-step framework for systematic review and evidence integration. Please note: this evaluation stops at step 5 of the systematic review process and does not proceed to step 6 to translate the confidence rating for the body of evidence into a level of evidence for health effects (see Figure 2 from the handbook).



Charge:

(1) Comment on the technical accuracy and whether the draft *NTP Monograph on State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* is clearly stated, and objectively presented.

(2) Determine whether the scientific evidence supports the NTP's confidence ratings for the bodies of evidence regarding neurodevelopmental and cognitive health effects associated with exposure to fluoride.

Instructions for Review:
All materials for this review are available in the Electronic Council Book (ECB). You will receive the specific URL and a password for accessing the ECB.

This evaluation identified 159 human studies relevant for assessing neurological health effects of exposure to fluoride; however, many studies included only secondary outcomes (e.g., 55 studies of thyroid hormones that were investigated as a potential mechanism). The scientific evidence in children and adults was evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood. Several studies evaluated learning and memory (n = 8 studies) or other cognitive developmental effects (e.g., total neurobehavioral scores and total mental capacity index in children, cognitive impairment in adults; n = 14 studies). Sixty-six human studies investigated IQ in children. Nineteen of the 66 IQ studies were determined to have low potential for bias and therefore, were categorized as "low risk of bias". Please give special attention to our assessment of these 19 studies.

- The 19 studies are available as PDFs and organized alphabetically in a folder on the ECB.
- All other studies are provided in the Health Assessment Workspace Collaborative, or HAWC database under the "studies list" tab, also organized alphabetically. You will also be provided a username and password for HAWC that will give you [redacted] permissions to access the PDFs in HAWC along with visualizations and other study information for this project at the following link (https://hawcproject.org/study/assessment/405/).

Please provide your substantive scientific and technical comments and suggestions within this [redacted] form. Identify and provide the rationale or scientific support for proposed changes or suggestions where possible.

If necessary, you can also provide additional editorial comments and recommendations for improving the report outside your specific charge questions (this form) within the draft report itself. Please note that only those comments included on [redacted] form will be considered part of NTP's peer review report.

A. General Comments

1. Please comment on whether the scientific information presented in the draft monograph, including presentation of data in tables and figures, is technically correct, and clearly and objectively presented. Please suggest any improvements.

H.1:

[redacted] **Comments:** In general the scientific information presented, including the data in tables and figures, is technically correct and clearly and objectively presented. Specific comments regarding the general evaluation of studies:

The use of the term "correlation" is confusing ([redacted] have marked this several times in the document and tables, but there are also other occurrences). Correlation is generally used to denote a correlation coefficient (either Pearson or Spearman); however, [redacted] believe it has been used to denote the estimated regression coefficients (more on this below). [redacted] would recommend changing the terminology for clarity.

Response: Agree (change made)

- The term has been changed to "association" when a regression coefficient was used, and we verified the accurate use of the term "correlation" in the text. We use "correlation" when Pearson correlation coefficients were reported or when discussing relationships between fluoride levels in various matrices (e.g., "correlations between urinary fluoride and fluoride in the drinking water").

H.2:

[redacted] **Comments:** The term neurologic to refer to outcomes such as anxiety and aggression (and other neurobehavioral outcomes) is not quite correct. Neurologic would refer to outcomes such as tremor or other objective neurological signs. The more correct term would be neurobehavioral. [redacted] have marked some of this in the text.

Response: Agree (change made)

- The term "neurologic" has been changed to "neurobehavioral" (or other appropriate text) in several places of the Sup02_2022_Prepublication_NTP_Monograph to address [redacted] feedback. In addition, footnotes 2, 5, and 6 were added to clarify changes to specific aims and the PECO statement.

H.3:

[redacted] **Comments:** Use of the term "gender" to denote sex differences is not in line with current usage. Gender refers to the socially constructed variable, while sex refers to the biological variable. Please change.

Response: Agree (change made)

- The Sup02_2022_Prepublication_NTP_Monograph has been revised to change "gender" to "sex" in this context.

H.4:

[redacted] **Comments:** The estimated regression coefficients from the studies need to be presented more clearly. For example, many times there is no reference, e.g. increase (or decrease) in score per 1 mg/L F in urine. Further, for the presentation of odds ratios, it is not clear what the dichotomous (or categorical) outcome variable is (e.g. IQ below 50). These suggestions are for clarity as well as for correctness.

Response: Agree (change made)

- We have reviewed the monograph and made changes, as needed, to ensure the following: 1) all coefficients have the exposure unit associated with the change (e.g., added "1 mg/L increase" of fluoride if it was not previously specified); 2) the direction of change is clear (e.g., added a minus sign or language to indicate an increase or decrease, as appropriate); and 3) for odds ratios, the occurrence of the outcome relative to exposure is clear. In addition, we have updated units of change in effect estimate per change in fluoride exposure or added cutoffs for categorical outcomes in Tables 6, 7, and 8. We added similar clarifications to figures in *Appendix A* by modifying figure titles to clearly reflect the type of effect estimate presented (e.g., correlation coefficient) or adding figure notes to highlight categorical cutoffs.

H.5:

[redacted] **Comments:** RE: confounding and covariates. Recent thinking regarding confounding requires the use of directed acyclic graphs to define variables which are theoretically confounders (based on previous literature). Thus, some clarification is needed on how the set of three important confounders were selected, i.e. sex, child age and a measure of socioeconomic status. Indeed, based on literature from other potential neurotoxins (e.g. lead, polychlorinated biphenyls, phthalates) it seems as though child sex would be an effect modifying variable, not a confounder (child sex would not be related, for example to exposure status under any definition of confounding). Variables such as arsenic or lead exposure would be co-exposures, and might be considered, for example, in an exposure mixtures analysis in future studies. Regarding the other variables that were listed in the confounding section, a case would need to be made that they are true potential confounders and not just covariates related to the outcome (e.g. maternal BMI). It is possible that some of these variables are related to socioeconomic status, such as HOME score, nutrition variables, but that would need to be documented.

Response: Agree (change made)

- We agree that there had been areas of the text where we had conflated the meaning of the word "confounder" with "covariate." [redacted] gives the example of sex, which could be a confounder or an effect measure modifier. To address the comment, we have changed the word "confounder" to "covariate" throughout the Sup02_2022_Prepublication_NTP_Monograph where necessary. An important covariate could be a potential confounder, a potential effect modifier, or co-exposure. We continue to specify that arsenic and lead are potential co-exposures and agree that future studies should consider conducting exposure mixture analyses where appropriate.
- As described in the protocol, age, sex, and socioeconomic status (SES) were identified as key covariates in the confounding risk-of-bias domain as they are potential confounders or effect modifiers in any human study of fluoride and cognitive neurodevelopmental health effects, whereas the other important covariates may be specific to the study population and/or outcome.
- We disagree that biological sex is not a potential confounder for several reasons: (1) sex has historically been considered an important potential confounder in the literature (see Table 6 in the Sup02_2022_Prepublication_NTP_Monograph) (Lash

et al. 2021; Gochfeld 2017); (2) sex is an important risk factor for neurodevelopmental and cognitive outcomes (Cowell and Wright 2017); and (3) potential sex-related ingestion and dietary differences are realistic in observational studies (D'Amico et al. 2020; Keller et al. 2019).

- In addition, the text in the *Confounding* methods discussion identifies the types of covariates that are related to SES (e.g., maternal education, household income, marital status, crowding). Figure 6 indicates which low risk-of-bias IQ-in-children studies considered the caregiving environment (e.g., HOME score) as a measure of SES. Additionally, a footnote has been added to Figure 6 in the Sup02_2022_Prepublication_NTP_Monograph that lists all other measures related to SES that were considered in the low risk-of-bias IQ-in-children studies (i.e., SES scaled scores, household/family income, child education, caretaker/parental education, and occupation/employment). The notes column of Figure 6 also documents which studies considered nutritional variables.

H.6:

[redacted] **Comments:** [redacted] commend the team immensely for all the work they did to account for arsenic in drinking water, to determine whether it is an important co-exposure.

Response: No change requested

- No response necessary.

H.7:

[redacted] **Comments:** Dose response: The authors correctly point out that many of the studies dealt with exposure levels considered relatively high, at least relative to the EPA drinking water standard and secondary standard. Further, in the group of studies considered to be low risk of bias, exposure was generally considered either on an arithmetic – but sometimes on a logarithmic scale (if quantitative), or based on a dichotomous variable of fluorosis, a manifestation of continuous high exposure, or whether study participants lived in an area known to have high levels of exposure. Thus, the conclusion of moderate confidence that fluoride is associated with deficits in IQ scores in children needs to be couched for these higher exposure levels, as there are few studies that provide evidence of this for exposures in the low range. This is not to say that there is no association at these lower levels, there may very well be an association; just that these results cannot be generalized to lower levels of exposure. This is true with other neurotoxins as well, for example, we know that the associations between lead and IQ scores is even steeper at the lower levels of exposure, but early studies where exposure was high were not able to discern those associations.

Response: Agree (change made)

- With respect to contextualizing the exposure levels for which we are providing a confidence rating, we had attempted to do this in a manner that satisfied requests from some reviewers during earlier iterations of the monograph. We then received comments from other reviewers asking what is meant by an exposure that is characterized as "high." A further complicating factor occurs when dealing with group-level exposures based on, for example, drinking water concentrations

where both exposed and reference groups are comprised of individuals who have fluoride exposures that are above and below the median level. At times, these exposures from other sources can be substantial and result in overlap between groups. As an illustration, Figure 2 in Green et al. (2019) compares maternal urinary fluoride measurements during pregnancy from communities with or without artificial water fluoridation. The urinary fluoride levels are generally higher in artificially fluoridated communities compared with non-fluoridated communities, but there is overlap. In this instance, we do not have sufficient information to identify a level below which there is no potential for fluoride to affect neurodevelopment or cognition; therefore, we have chosen to further characterize the findings by incorporating the term "higher" only when its meaning in the specific context is clear. For example, to clarify that our moderate confidence conclusion is primarily based on studies with total fluoride exposure that approximates or exceeds what is generally associated with consumption of optimally fluoridated water in the United States, the *Abstract* was revised as follows:

"This review finds, with moderate confidence, that higher fluoride exposure (e.g., represented by populations whose total fluoride exposure approximates or exceeds the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride) is consistently associated with lower IQ in children."

H.8:

[redacted] **Comments:** Cumulative exposure: The authors should make clear that exposure during gestation likely implies that there is continuing exposure in the post natal period. Further, these two exposure periods are likely highly correlated, making conclusions regarding a critical period of exposure difficult. The converse is also true – i.e. if exposure is measured in the post natal or childhood period, and especially if it is from drinking water, then there was likely exposure in the prenatal period as well.

Response: Agree (change made)

- We agree that prenatal and postnatal exposures are likely correlated; however, there are not enough data available for us to evaluate cumulative exposure in revisions to the monograph. We added the following *Limitations of the Evidence Base* in the *Discussion* section:

  "No studies are available to evaluate lifelong exposure in adults, or fluoride exposure over a child's lifetime and neurodevelopmental or cognitive changes over time.

  The database does not allow for comparison of ages and possible changes at different developmental stages in children to assess if there is a delay in development or if associations persist.

  The database does not allow for establishing clear correlations between prenatal and postnatal exposures."

2. Please identify any information that should be added or deleted.

H.9:

[redacted] **Comments:** In general the report is comprehensive and includes all necessary material. Hence, [redacted] have no major additions or deletions. [redacted] have one small addition, which would be a discussion of the toxicokinetics of fluoride – this is necessary because

the half life is relatively short, and a spot measure (or even several spot measures) in urine (or serum) would not entirely represent exposure history. Indeed, it may be the case here that chronic exposure to drinking water with high levels of fluoride may be a better marker of long term exposure.

Response: Agree (change made)

- We have added a brief discussion on fluoride toxicokinetics at the beginning of the *Exposure* section of the *Risk-of-bias Considerations for Human Studies* section of the Sup02_2022_Prepublication_NTP_Monograph, as follows:

  "Fluoride ion is rapidly absorbed from the gastrointestinal tract and is rapidly cleared from serum by distribution into calcified tissues and urinary excretion (IPCS 2002). There is general consensus that the best measures of long-term fluoride exposure are bone and/or tooth measurements, and other than measures of dental fluorosis, these were not performed in any of the studies reviewed in this document. Prolonged residence in an area with a given fluoride content in drinking water has been considered in many studies as a proxy for long-term exposure."

- In very general terms this may be a useful addition, although none of the studies included measurements of fluoride concentrations in bone or teeth, which are considered to be integrative measures of long-term exposures.

- [redacted] is correct that serum levels are not considered reliable reflections of chronic fluoride exposures as serum fluoride is rapidly cleared to calcified tissues at a rate that changes depending on the prior fluoride loading of the particular tissue. As we note in the monograph, urinary volume-corrected spot or 24-hour collections are considered reasonably good measures of exposure, although they represent a balance of recent intake, movement into and out of calcified tissues, and excretion. Repeated urinary measures during pregnancy are reported to have reasonable reproducibility over time, although in one study by Thomas et al. (2018), urinary fluoride concentrations tended to increase until about week 23 and then decline thereafter.

- Although all of the exposure measures used in this series of studies have some advantages and disadvantages, confidence in the association between measures of higher fluoride exposure and lower IQ was increased by the consistency in findings irrespective of the measure of fluoride exposure. See the *Exposure Measure and Study Population Factors* section of the Sup02_2022_Prepublication_NTP_Monograph where we describe the consistency of the direction of effect in the children's IQ studies across exposure metrics.

B. Human studies

I. Fluoride exposure and children's IQ

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on measures of IQ in children.

H.10:

[redacted] **Comments:** The approach used to select human studies on the associations between exposure to fluoride and neurodevelopment is appropriate. [redacted] would, however,

like further details on how differences between reviewers were reconciled. For example, what were the kappa statistics or intraclass correlation coefficients on the title/abstract review prior to reconciliation? Did the degree of agreement warrant further training of the reviewers? The same concerns regarding reliability can be made for the data extraction.

Response: Disagree (no change)

- We have not found Kappa Statistics or other measures of inter-rater agreement to be useful in obtaining agreement in the screening process. Instead, we have implemented a process that emphasizes training, pilot testing, and group discussion to assure consistency of approach. As described in the *Methods* section, screening is conducted by two independent screeners at both the title/abstract and full-text steps. When conflicts arise, they are resolved through discussion between the two screeners and consultation with a senior team member, if necessary, to reach consensus. Our protocol also describes that training and pilot testing phases are conducted on small sets of studies to assure consistency of approach in applying the PECO criteria and inclusion/exclusion guidance. When questions arise, we address them as a group so that all screeners develop a consistent approach. The process emphasizes inclusion of studies if there is any uncertainty at the title/abstract stage. At the full-text stage, we confirm that studies indeed have original data and meet the PECO criteria, so there is little uncertainty at that step. Studies either have the relevant data or they do not. In addition, while cross-screener agreement within a project team is essential when each reference is screened by a single reviewer, the issue has a much smaller potential impact when two screeners review each study in duplicate, as in this systematic review.
- The review process for data extraction involves a quality control (QC) review rather than extraction in duplicate. Data extraction is conducted by a single extractor followed by QC review because we have not found added value or reliability with independent data extractions. The QC review is conducted for all data extracted into HAWC (https://hawcproject.org/assessment/405/), the web-based content management system for our systematic reviews.

H.11:

[redacted] **Comments:** The use of the SWIFT-Active Screener is well described and addresses the concerns in the prior review.

Response: No change requested

- No response necessary.

H.12:

[redacted] **Comments:** The supplemental search of the non-English language databases is appropriate. However, what is the rationale for saying that they were used primarily to identify null or no-effect studies? Does that mean that if a study was identified that showed an association it was not abstracted? Please be a bit more clear on this.

Response: Agree (change made)

- Although extraction of studies identified from the Chinese database searches was previously focused on no-effect studies, we have taken steps to translate and extract data from all non-English studies identified from the Chinese database

searches. Therefore, the statement about null or no-effect studies no longer applies and has been deleted.

2. Comment on whether the approach used to assess risk of bias was clearly described and appropriately applied to the set of studies designated as "low risk of bias."

H.13:

[redacted] **Comments:** The focus on confounding, exposure characterization and outcome assessment are, as indicated, the key components of evaluating observational research. The other parameter is whether the participants represent the population from which they are recruited, i.e. selection bias. In prospective cohort studies this is not an issue, as the population is really the combination of those exposed and non-exposed. For cross sectional studies, this is a bit trickier, as the participants may reflect a select group within the overall population. For studies based on national or regional registries, such as the Canadian studies, this is less of a problem, but for others there is the possibility of bias, and the direction of such bias is difficult to predict. As [redacted] looked at the studies, the vast majority do not address this issue, but [redacted] believe that it is worth a discussion or at least a mention that the possibility of selection bias is real.

Response: Agree (change made)

- We agree with [redacted] that selection bias is an important consideration in risk-of-bias evaluations. We have edited the following text in the *Methods* section to clarify that, in addition to the three key risk-of-bias questions, the answers to the other risk-of-bias questions were considered in assessing potential bias, including selection bias.

  "The other risk-of-bias questions, including selection of study participants, were also considered and were used to identify any other risk-of-bias concerns that may indicate serious issues with a study that could cause it to be considered high risk of bias."

- *Appendix E* includes a detailed summary of the population selection and the basis for the ratings for selection bias and exposure characterization. All 19 low risk-of-bias IQ-in-children studies and 9 other neurobehavioral studies in children were rated either *probably low risk of bias* or *definitely low risk of bias* due to selection bias. Generally speaking, these studies provide direct or indirect evidence that exposure groups were similar and were recruited within the same timeframe using the same methods with no differences in participation/response rates (i.e., either direct evidence of similar participation/response rates or no evidence of differences in participation/response rates). Differences in the subjects across exposure groups were noted and addressed in the analysis.

H.14:

[redacted] **Comments:** For confounding, please see [redacted] remarks above. [redacted] do think that biological sex needs to be considered an effect modifier as in other studies of neurotoxins and neurodevelopmental outcomes. Further, as indicated later in the monograph, at times the choice of confounders needs to be study and area specific, so this should also be mentioned in this section. Finally, for the arsenic variable, as [redacted] indicated above [redacted] really appreciate the efforts made in defining this. However, please justify the choice of

10µg/L as the cutpoint – while it is the WHO guideline it is quite possible that there are neurodevelopmental effects with concentrations under this level.

Response: Agree (change made)

- We agree that biological sex should be considered a potential effect modifier in addition to (not instead of) a potential confounder. Please see previous response for details on our rationale and how text was revised to address [redacted] comments on confounding.
- Regarding choice of important covariates being study- and area-specific, we consider what we currently state in the *Methods* section to address [redacted] suggestion*:*

  "Additional covariates considered important for this evaluation, depending on the study population and outcome, included…" and, "To be assigned a rating of probably low risk of bias for the key risk-of-bias question regarding the confounding domain, studies were not required to address every important covariate listed; however, studies were required to address the three key covariates for all studies, the potential for co-exposures, if applicable (e.g., arsenic and lead, both of which could affect cognitive function), and any other potential covariates considered important for the specific study population and outcome."
- As for the choice of 10 µg/L as the cutoff point, [redacted] is correct that we chose this based on the WHO guideline (WHO 2017). We agree that it is possible there may be neurodevelopmental effects at concentrations below 10 µg/L; however, we have no basis on which to select a lower cutoff point. Note that we had initially added a statement to the Sup02_2022_Prepublication_NTP_Monograph stating that "*arsenic may be associated with neurodevelopment effects at concentrations below 10 µg/L*" in response to this reviewer's comment; however, as we were unable to support this statement with a reference, it has been removed.

H.15:

[redacted] **Comments:** Exposure characterization: This is well described. As [redacted] mention above, missing is a discussion regarding the toxicokinetics of fluoride, to allow the reader to make decisions on how good the spot urine samples are in reflecting cumulative exposure. [redacted] understand that there is a high correlation between the spot urine samples and 24 hour collections (with and without correction for dilution) but this still does not give day to day, week to week, or season to season variation.

Response: Agree (change made)

- As described in a previous comment, we have added a brief discussion on fluoride toxicokinetics at the beginning of the *Exposure* section of the *Risk-of-bias Considerations for Human Studies* section.
- With respect to variations in fluoride exposures over time, we agree that additional study of these variations would be interesting; however, our assumption is that individual exposure to fluoride is relatively consistent because it reflects personal preferences and habits (e.g., daily water consumption, tea consumption, dental product use).

H.16:

[redacted] **Comments:** A further concern with exposure assessment brought up in the previous review concerns the issue of clustering with regard to exposure. The authors of the monograph do a very nice job of addressing this issue as it was raised in the prior review, but pointing to the sensitivity analyses. [redacted] only concern remaining is that this is mentioned up front when the exposure characterization is discussed in the methods.

Response: Agree (change made)

- To address this suggestion, we have provided an additional sentence in the *Methods* section where risk-of-bias considerations for exposure are discussed.

  "Ideally, these studies would still need to consider and adjust for area-level clustering; however, these concerns are captured in evaluations of other potential threats to internal validity."

H.17:

[redacted] **Comments:** Finally, some measure of agreement between the raters on their bias assessment would be a good addition.

Response: Disagree (no change)

- While we appreciate this comment, we have not found measures of inter-rater agreement (e.g., kappa statistics) to be useful in this process and instead have implemented a process that emphasizes pilot testing to develop a consistent approach and group discussion when there are questions in the rating. In addition, to further support consistency, a senior member of the team served as one of the risk-of-bias assessors for all of the studies. In addition, while cross-reviewer agreement within a project team is essential when each reference is assessed by a single reviewer, the issue has a much smaller impact when two screeners review each study in duplicate, as in the current systematic review. We consider that the most important issue for consistency is to reach collective agreement, and the final risk-of-bias ratings reflect that agreement.

3. Comment on assessment of the human studies with regard to:

a) How findings from individual studies designated as "low risk of bias" were interpreted.

H.18:

[redacted] **Comments:** In general, studies designated as "low risk of bias" were interpreted correctly. [redacted] have a few suggestions as to how to clarify many of the points made.

While the results are generally consistent (table 6) it would be useful to present the results based on the exposure metric used. For example, studies using fluoride concentrations in "high" and "low" areas could be grouped together to illustrate the change in IQ points. Additionally, the actual IQ test used could also be used to group studies within exposure metric. There are clear differences in the scoring for the Raven and the WASI/WPPSI, for example and these are hard to tell from the presentation.

Response: Disagree (edited for clarity)

- We considered several ways to organize the table and each way has its benefits and drawbacks. There are limitations to a static table, which is why we are increasing our use of interactive tools and platforms to visualize data. For the purpose of this document, we consider the current organization to be most clear and appropriate for providing a quick summary of study characteristics and key findings per study. We have edited the paragraph that precedes Table 6 to clarify that the Table 6 organization is by country and then by year.
- Note that we considered [redacted] suggestion to group studies using fluoride concentrations in "low" and "high" areas together to illustrate the change in IQ scores. While an association is consistently observed when comparing low to high fluoride areas, comparing changes in IQ scores across these studies is challenging due to the variability in the exposure levels that are considered "low" and "high." There are no consistent definitions of "low" and "high" that apply across all cases. For this reason, we do not find this suggested organizational structure for Table 6 to be a more effective presentation of the data. We also considered [redacted] suggestion to group studies by IQ test; however, as the Raven's tests were almost exclusively conducted in China, India, and Iran, the current organization by country, to a large extent, also organizes the studies by IQ test. Therefore, we find the current structure accommodating for focusing on results by IQ test.

H.19:

[redacted] **Comments:** At times, associations are presented as different when other covariates are controlled. [redacted] presume that these assessments were made by inspection of the results in the studies, but should either be backed up with statistical testing or admitted that they were made by inspection. For example, in table 6 the study by Rocha-Amador, et al states that the estimated associations between fluoride and the full scale IQ (WISC) were smaller when arsenic was controlled, the estimated betas are given, but there is no indication whether the differences are statistically different.

Response: Disagree (edited for clarity)

- When study authors present associations between fluoride exposure and IQ that differ when other covariates are included, we reported the results as described by the study authors. We did not perform additional testing to support the author's reporting of results as this is beyond the scope of the assessment.
- The statement that [redacted] notes for Rocha-Amador et al. (2007) and the association with arsenic was misinterpreted. The purpose of the statement was to note that the association between arsenic exposure and children's IQ was smaller in magnitude than the association between fluoride exposure and children's IQ, not that the association with fluoride was smaller after controlling for arsenic. The revised text in Table 6 of the Sup02_2022_Prepublication_NTP_Monograph reads as follows:

  "Significant associations between log-transformed fluoride and IQ scores (full-scale IQ adjusted ß's of −10.2 [water] and −16.9 [urine]; CIs not reported); arsenic also present, but the association between log-transformed arsenic and IQ scores was smaller (full-scale IQ adjusted ß's of −6.15 [water] and −5.72 [urine]; CIs not reported)"

H.20:

[redacted] **Comments:** Please note when the result is not statistically significant and likely due to small sample sizes (e.g. discussion of the Green et al paper on page 37). Also for that paper, the results seem to be different by biological sex, an example of effect modification that would be expected for a neurotoxin.

Response: Agree (change made)

- We added the qualifier "not significant" for the results in girls. However, since the scope of this section is to present the observed IQ effects in children, we refrain from suggesting reasons for non-significance, such as sample size. In each study, there are a multitude of factors that could yield nonsignificant results, in addition to lack of power. The study-specific risk-of-bias evaluations describe study details (including sample size) and aspects that could impact the ability to detect an association. With respect to biological sex as an effect modifier, we consider our revised terminology in response to a previous comment to address [redacted] concern.

H.21:

[redacted] **Comments:** The results also need to be interpreted based on age of test administration. Some higher order functions do not develop until later ages and thus cannot be tested well in younger children. Also, as with other neurotoxins, deficits can occur at a variety of ages, and either persist or not. So the age at assessment becomes an important variable in the interpretation of findings and should be accounted for in the discussion.

Response: Disagree (edited for clarity)

- The available data are not provided in a way that allows for evaluating deficits occurring at a variety of ages and whether the deficits persist or not. Although some studies provide the results by specific ages, these are mainly high risk-of-bias studies conducted in areas with high fluorosis rates, and the tests were generally conducted in children 8–12 years old. The following text was added to the *Discussion* section as a limitation of the evidence base:

  "The database does not allow for comparison of ages and possible changes at different developmental stages in children to assess if there is a delay in development or if associations persist."

- We have already considered age at test administration in the risk-of-bias evaluation of individual studies in two different ways: (1) whether the test used to measure neurodevelopment or cognition was age-appropriate and (2) when a study included a range of ages, whether age was assessed as a potential confounder (for the reasons noted by [redacted]).

H.22:

[redacted] **Comments:** When discussing the variations in associations by genetic polymorphisms, it would be useful to discuss the function of the gene, especially the function related to neurodevelopment or the developing brain.

Response: Disagree (no change)

- Although information on the possible interaction of fluoride with genetic polymorphisms is an active area of investigation, only two studies were available as of the cutoff date for this systematic review. Our intent was to simply point this out as an emerging area of research rather than speculate about potential mechanisms of fluoride action, which would require much further study and a deeper understanding.

H.23:

[redacted] **Comments:** As indicated above, please be very careful in discussing dose response relationships, especially when these may be non-linear.

Response: Agree (no change)

- We agree that discussion of dose-response relationships should be done carefully, and we re-reviewed all of the dose-response text to address this concern. The Sup02_2022_Prepublication_NTP_Monograph summarizes the findings of the qualitative analysis of children's IQ studies that evaluated lower fluoride exposures without reporting on the evidence for dose response (available in full in the 2019 draft NTP Monograph). The Sup02_2022_Prepublication_NTP_Monograph refers the reader to the revised meta-analysis document as it provides a quantitative assessment of dose response to further inform this discussion.

b) How the overall set of confounders across the body of evidence from children's IQ studies was considered and presented.

H.24:

[redacted] **Comments:** Please see the discussion of confounding above. [redacted] do appreciate Figure 6 which describes the confounders measured in the low risk of bias studies, stratified by rating for confounding. In the three studies in which the RoB rating for confounding was high, however, it appears that such confounding may influence the results to some degree. It would be useful to have an assessment of the direction and magnitude of bias introduced by not clearly defining and controlling for key confounders, even if that discussion is somewhat speculative.

Response: Agree (no change)

- An assessment of the potential magnitude and direction of bias in the low risk-of-bias studies, as requested by [redacted], was included in *Appendix E* in the Sup02_2022_Prepublication_NTP_Monograph (previously *Appendix 4* of the Sup03_2021_draft_NTP_Monograph, the version of the monograph reviewed by [redacted]).

c) How the confidence rating in the body of evidence was developed and supported.

H.25:

[redacted] **Comments:** In general, the confidence rating in the body of evidence for this outcome is supported. However, several concerns necessitate a refinement of this confidence rating.

[redacted] agree with the prior review in that conclusions can only be made above the WHO drinking water limit for fluoride. It seems as though there is a lack of dose response curve estimation for lower levels of exposure, so an inference cannot be made over the entire range of exposure. Indeed, it is this lower dose range that is of interest for the US population.

Response: Agree (change made)

- As [redacted] notes, earlier versions of this monograph examined the evidence for dose response across the range of exposures represented in the human body of evidence, both from a qualitative and quantitative perspective. The current monograph intentionally does not dwell on this question, as the conclusions from individual included studies about dose response for cognitive neurodevelopmental associations at the lower fluoride exposure levels are somewhat conflicting. The uncertainty of the evidence at these lower levels is cited as one of the limitations of the evidence base. Given that the revised meta-analysis specifically addresses this question and incorporates newer literature, we have decided to revise these considerations in the Sup02_2022_Prepublication_NTP_Monograph to focus on the data on which we base our confidence statement, and to acknowledge the need for further studies at lower exposure levels. The following text has been added to the abstract and summary:

  "This review finds, with moderate confidence, that higher fluoride exposure (e.g., represented by populations whose total fluoride exposure approximates or exceeds the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride) is consistently associated with lower IQ in children. More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ."

H.26:

[redacted] **Comments:** Because the urine and serum biomarkers of fluoride represent relatively recent exposure, it is difficult to infer that the associations are from cumulative exposure without laying out the assumptions, i.e. long term residential history, similar habits of toothpaste use, etc.

Response: Agree (change made)

- Text was added to address the best measures for assessing long-term fluoride exposure (see quote below). Although urine and serum reflect recent exposures, they represent total fluoride exposure. The indicators and assumptions for long-term exposure in the cross-sectional studies are laid out in the *Overall Findings* section for IQ in children and the results are described in *Results by Study Design – Cross-sectional Studies* section where we address the assumptions for prior exposure, one of the factors that we considered in establishing the confidence level as moderate.

  "There is general consensus that the best measures of long-term fluoride exposure are bone and/or tooth measurements, and other than measures of dental fluorosis, these were not performed in any of the studies reviewed in this document."

H.27:

4. ***NTP concludes a rating of moderate confidence in the body of evidence for lower IQ in children associated with fluoride exposure***.

☐ Agree
X Agree in principle with the exception(s) listed below:
Please see point a above. The exception would be that there is low confidence of the association for levels of exposure in the lower dose range.
☐ Do not agree because:

Response: Agree (change made)

- We provided our response to this point above.

II. **Fluoride exposure and non-IQ neurodevelopmental or cognitive effects in children**

H.28:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on non-IQ neurodevelopmental or cognitive effects in children.

[redacted] **Comments:** Please see [redacted] comments in section III.1 above. The search terms used are encompassing of neurodevelopmental outcomes in children.

Response: No change requested

- No response necessary.

H.29:

2. Comment on whether the approach used to assess risk of bias for studies in children on non-IQ neurodevelopmental or cognitive effects was clearly described and appropriately applied.

[redacted] **Comments:** Please see [redacted] points in III 2 above. Further, for these outcomes, one would definitely need to stratify the results based on age of the child, as some of these skills develop differently. For example, children age 6-8 years are very different from neonates. Also, please note that the thinking regarding assessment of confounding would be outcome specific as some variables, e.g. SES, may not be applicable to some skills.

Response: Agree (change made)

- See our previous response to [redacted] comment on selection bias. In short, we have added clarifying language in the *Methods* section to indicate that selection bias was consider in determining the overall risk-of-bias status of each study (response above includes a quote of the revised monograph text).
- Furthermore, we agree that confounding is outcome-specific, but SES along with sex and age were identified as key covariates for all studies. This means that SES would need to be considered in any human study of fluoride and cognitive neurodevelopmental health effects; however, if there was reason to believe that SES (or age or sex) was not a potential confounder or risk-of-bias concern for a given study, then that would have been taken into consideration when determining the risk-of-bias rating for confounding. The risk-of-bias rating rationale would have described the reason that SES was not considered a concern for a particular study.

3. Comment on assessment of the human studies with regard to:

H.30:

a) How findings from individual "low risk of bias" studies were interpreted.

[redacted] **Comments:** Many of [redacted] comments in section III 3a are also applicable here.

As noted above, please note that the assessment of confounding needs to be outcome (and likely age) specific. For example, measures of socioeconomic status may not be confounders for outcomes measured in neonates (the Li study did not control for anything) but may be proxy measures for variables such as maternal smoking, that was not measured or controlled and which could be a confounder.

Response: Agree (change made)

- See section A1 where we addressed this comment when it was previously raised.

H.31:

[redacted] **Comments:** For the studies that measured multiple outcomes, there would need to be some adjustment for multiple testing, using either a conservative Bonferroni correction or some other method. This is particularly important here as the behavioral outcomes, for example, are correlated.

Response: Disagree (no change)

- *Appendix E* in the Sup02_2022_Prepublication_NTP_Monograph (previously *Appendix 4* of the Sup03_2021_draft_NTP_Monograph) includes considerations of adjustment methods (including use of the Benjamini–Hochberg false discovery rate) when information was provided by the study authors. We disagree that adjustment for multiple testing is necessary in our risk-of-bias assessment where studies are estimating an effect of exposure on an outcome. Adjustment for multiple comparisons is only necessary when a study is doing strict hypothesis testing (Rothman 1990).

H.32:

[redacted] **Comments:** (minor) Please note that often the GCI on the MSCA is considered a measure of IQ, so perhaps the study of Bashash et al (2017) could be considered in the IQ studies.

Response: Disagree (no change)

- The MSCA measures developmental abilities in children using tasks that assess verbal, quantitative, perceptual, memory and motor skills. Children can earn an IQ-like score (the General Cognitive Index; GCI) based on summed performance across tasks. We agree that the GCI can be considered as a measure of IQ; however, we considered it appropriate to categorize this test with other tests of cognitive function in the *Other Neurodevelopmental or Cognitive Effects in Children* section. Moreover, the *IQ in Children* section includes Bashash et al. (2017) for its results from the Wechsler Abbreviated Scale of Intelligence, which is typically considered an IQ test. Categorizing the MSCA results in the *Other Neurodevelopmental or Cognitive Effects in Children* section allowed us to avoid double-counting the Bashash et al. (2017) study in the *IQ in Children* section.

- Note that adding GCI to the *IQ in Children* section rather than the section on other neurodevelopmental outcomes may add to the strength of evidence, but it would not change the moderate confidence determination in the monograph. Furthermore, the revised meta-analysis includes sensitivity analyses with GCI scores from Bashash et al. (2017) and a second study that reported findings from the GCI portion of the MSCA.

H.33:

[redacted] **Comments:** Some of the associations are really quite large, e.g. adjusted betas of -19 in the study of Valedez Jimenez et al 2017, especially for the Bayley Scale. Such associations are either suspect or are not adjusted for the concentration of fluoride appropriately (maybe it is a log unit change). This needs to be clarified.

Response: Agree (change made)

- We have clarified in the tables and text that the associations are per log10-mg/L increase in fluoride exposure. The revised text in *Results in Infants* section of the Sup02_2022_Prepublication_NTP_Monograph reads as follows:

  "In infants 3 to 15 months of age, the Mental Development Index (MDI)—which measures functions including hand-eye coordination, manipulation, understanding of object relations, imitation, and early language development—was significantly inversely associated with maternal urinary fluoride in both the first and second trimesters (adjusted ß̂s per log10-mg/L increase = −19.05 with standard error of 8.9 for first trimester [p-value = 0.04] and −19.34 with standard error of 7.46 for second trimester [p-value = 0.013]) (Valdez Jimenez et al. 2017)."

H.34:

[redacted] **Comments:** Please clarify what a construction task is (page 56). Do you mean a fine motor copy task?

Response: Agree (change made)

- We revised the text to characterize the task more accurately as a visuoconstructional score from the Rey-Osterrieth Complex Figure Test. The revised sentence reads as follows:

  "Another study using visuoconstructional and memory scores from the Rey-Osterrieth Complex Figure Test in children 6–11 years old observed significantly lower scores with increasing concurrent child single spot urinary fluoride even after adjusting for age (partial correlation coefficients, per log-mg/L increase = −0.29 and −0.27 for copy [p-value <0.001] and immediate recall [p-value <0.001], respectively [CIs not reported]) (Rocha-Amador et al. 2009)."

H.35:

[redacted] **Comments:** Also on page 56 and highlighted in blue: this is unclear. Even though urinary arsenic is not associated with scores on these tasks, it could still very well be a confounder of the relationships between fluoride and the test scores.

Response: Agree (change made)

- As we discuss in *Appendix E* in the Sup02_2022_Prepublication_NTP_Monograph (previously *Appendix 4* of the Sup03_2021_draft_NTP_Monograph), although the model in Rocha-Amador et al. (2009) did not adjust for arsenic, arsenic in the F-As group was not associated with either outcome; therefore, arsenic as a co-exposure is not considered a major concern in this study. We revised text to mention the results adjusted for both fluoride and arsenic, as follows:

  "Another study using visuoconstructional and memory scores from the Rey-Osterrieth Complex Figure Test in children 6–11 years old observed significantly lower scores with increasing concurrent child single spot urinary fluoride even after adjusting for age (partial correlation coefficients, per log-mg/L increase = −0.29 and −0.27 for copy [p-value <0.001] and immediate recall [p-value <0.001], respectively [CIs not reported]) (Rocha-Amador et al. 2009). Although these children were also exposed to arsenic, the presence of arsenic could not explain the changes because, in the area with natural contamination by fluoride and arsenic (F–As), the test scores were not significantly associated with urinary arsenic levels (partial correlation coefficients, per log-mg/L increase = −0.05 and 0.02 for copy and immediate recall, respectively [CIs not reported]). The test scores were only marginally increased from fluoride alone when both fluoride and arsenic were included simultaneously in the model (partial correlation coefficients, per log-mg/L increase = −0.32 and −0.34 for copy and immediate recall, respectively [CIs not reported]) (Rocha-Amador et al. 2009)."

H.36:

[redacted] **Comments:** Also please address the issue that children with behavior problems may be more apt to, for example, drink excessive amounts of water or swallow toothpaste. This would be indicative of reverse causation.

Response: Disagree (no change)

- While polydipsia has been associated with clinical psychoses, we have failed to find reports of excessive consumption of water or toothpaste associated with the types of behaviors addressed in the studies examining fluoride exposure and other cognitive or neurodevelopmental conditions.

H.37:

b) How the confidence rating in the body of evidence was developed and supported.

[redacted] **Comments:** [redacted] fully agree with the low confidence rating for this body of evidence. The issues that [redacted] have highlighted above would only lend more support to the low confidence.

Response: No change requested

- No response necessary.

H.38:

4. ***The NTP concludes a rating of low confidence in the body of evidence for decreases in measures of other neurodevelopmental or cognitive effects in children associated with fluoride exposure***.

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

III. **Fluoride exposure and cognitive effects in adults**

H.39:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on cognitive effects in adults.

[redacted] **Comments:** Please see the comments above.

Response: No response necessary

- Comments were addressed where previously made by [redacted].

H.40:

2. Comment on whether the approach used to assess risk of bias for studies in adults on cognitive effects was clearly described and appropriately applied.

[redacted] **Comments:** Please see the comments above. [redacted] one additional comment here is that the results from China (Li et al 2016) perhaps indicate that the critical time of exposure is at earlier ages, since the exposure was residing in low and high fluorosis-endemic areas of China.

Response: Agree (no change)

- While we agree with [redacted] that earlier exposures could be an important factor in this study, there is insufficient information provided in the study to assess critical timing of exposure for cognitive impairments in adults.

3. Comment on assessment of the human studies with regard to:

H.41:

a) How findings from individual studies were interpreted.

[redacted] **Comments:** The studies were interpreted appropriately.

Response: No change requested

- No response necessary.

H.42:

b) How the confidence rating in the body of evidence was developed and supported.

[redacted] **Comments:** [redacted] fully support the confidence rating of low for this body of evidence.

Response: No change requested

- No response necessary.

H.43:

4. ***The NTP concludes a rating of low confidence in the body of evidence for changes in cognitive effects in adults with fluoride exposure.***

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

C. Studies in non-human animals

H.44:

The NTP agrees with the comments of the NASEM committee (NASEM 2020, 2021) concerning the overall poor quality of the experimental animal database on fluoride exposure and neurodevelopmental effects, with many studies suffering from major reporting deficiencies. As indicated above, the monograph focuses on the large human epidemiology database because it directly addresses the question of whether fluoride affects human neurodevelopment. Therefore, based on the recommendations of the NASEM committee, the experimental animal section and risk of bias details have been removed from this monograph and ***the NTP concludes that the scientific evidence from experimental animal data are inadequate to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans***.

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the inadequate designation.

References

Cowell WJ, Wright RJ. 2017. Sex-Specific Effects of Combined Exposure to Chemical and Non-chemical Stressors on Neuroendocrine Development: a Review of Recent Findings and Putative Mechanisms. Curr Envir Health Rpt. 4: 415–425. https://doi.org/10.1007/s40572-017-0165-9

D'Amico D, Parrott MD, Greenwood CE, Ferland G, Gaudreau P, Belleville S, Laurin D, Anderson ND, Kergoat MJ, Morais JA, Presse N, Fiocco AJ. 2020. Sex differences in the relationship between dietary pattern adherence and cognitive function among older adults: findings from the NuAge study. Nutr J. 19(58) . https://doi.org/10.1186/s12937-020-00575-3

Gochfeld M. 2017. Sex Differences in Human and Animal Toxicology. Toxicol Pathol. 45(1):172-189. doi: 10.1177/0192623316677327. PMID: 27895264; PMCID: PMC5371029.

Keller KL, Kling SMR, Fuchs B, Pearce AL, Reigh NA, Masterson T, Hickok K. 2019. A Biopsychosocial Model of Sex Differences in Children's Eating Behaviors. Nutrients. 11(3): 682. https://doi.org/10.3390/nu11030682

Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ. Modern Epidemiology, 4th edition. Wolters Kluwer, 2021.

Rothman KJ. 1990. No adjustments are needed for multiple comparisons. Epidemiology. 1(1):43-6. PMID: 2081237

Thomas DB, Basu N, Martinez-Mier EA, Sánchez BN, Zhang Z, Liu Y, Parajuli RP, Peterson K, Mercado-Garcia A, Bashash M, Hernández-Avila M, Hu H, Téllex-Rojo MM. 2016. Urinary and plasma fluoride levels in pregnant women from Mexico City. Env Res. 150: 489-495.

In November 2021, [redacted] received: 1) the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review,* 2) a copy of the NASEM Committee's comments on the 2020 draft NTP Monograph with NIEHS/DNTP responses (draft version of Sup01_Monograph), and 3) the [redacted] instructions. The instructions consisted of a preface, charge, instructions for the review, and a series of specific peer-review questions grouped by the following three topics: General Comments, Human Studies, and Studies in Non-Human Animals.

[redacted] were asked to provide their substantive scientific and technical comments and suggestions within the [redacted] form. In addition, they were asked whether they "Agree", "Agree in principle", or "Do not agree" with each NTP conclusion on confidence in a body of evidence.

The [redacted] instructions and specific peer-review questions are reproduced in the pages that follow in black text. [redacted] comments and responses to each question are also provided in black text starting with the words "[redacted] **comments**" in bold font. The NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] comments and [redacted] comments:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to this [redacted] would be marked "*see DocI_Monograph for detailed response*").
- [redacted] For comments DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

Preliminary comments on the draft NTP monograph prepared by the peer review [redacted] are noted below.
These preliminary comments are not binding and should not be construed to represent NTP determination or policy.

National Toxicology Program
NTP Monograph Letter Peer-Review Panel
Draft NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review

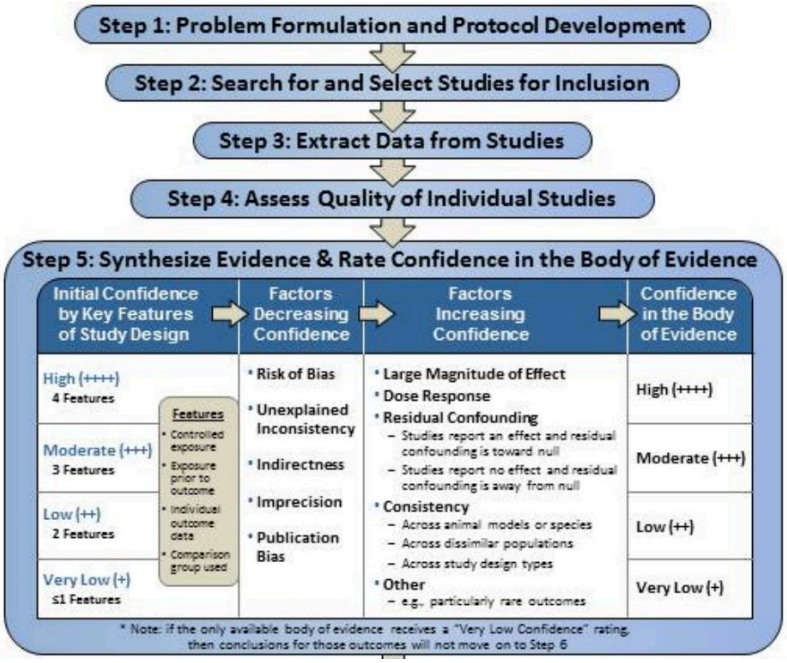
National Institute of Environmental Health Sciences
Research Triangle Park, NC

January 18, 2022

Fluoride State of the Science Document Review Form

[redacted]

Preface:
The objective of this evaluation was to conduct a systematic review of the published literature regarding the potential for exposure to fluoride to affect neurodevelopment and cognition in humans. The evaluation presented in the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* represents a comprehensive and current assessment. The methods used are from the Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration, which presents a seven-step framework for systematic review and evidence integration. Please note: this evaluation stops at step 5 of the systematic review process and does not proceed to step 6 to translate the confidence rating for the body of evidence into a level of evidence for health effects (see Figure 2 from the handbook).



Charge:

(1) Comment on the technical accuracy and whether the draft *NTP Monograph on State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* is clearly stated, and objectively presented.

(2) Determine whether the scientific evidence supports the NTP's confidence ratings for the bodies of evidence regarding neurodevelopmental and cognitive health effects associated with exposure to fluoride.

Instructions for Review:
All materials for this review are available in the Electronic Council Book (ECB). You will receive the specific URL and a password for accessing the ECB.

This evaluation identified 159 human studies relevant for assessing neurological health effects of exposure to fluoride; however, many studies included only secondary outcomes (e.g., 55 studies of thyroid hormones that were investigated as a potential mechanism). The scientific evidence in children and adults was evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood. Several studies evaluated learning and memory (n = 8 studies) or other cognitive developmental effects (e.g., total neurobehavioral scores and total mental capacity index in children, cognitive impairment in adults; n = 14 studies). Sixty-six human studies investigated IQ in children. Nineteen of the 66 IQ studies were determined to have low potential for bias and therefore, were categorized as "low risk of bias". Please give special attention to our assessment of these 19 studies.

- The 19 studies are available as PDFs and organized alphabetically in a folder on the ECB.
- All other studies are provided in the Health Assessment Workspace Collaborative, or HAWC database under the "studies list" tab, also organized alphabetically. You will also be provided a username and password for HAWC that will give you [redacted] permissions to access the PDFs in HAWC along with visualizations and other study information for this project at the following link (https://hawcproject.org/study/assessment/405/).

Please provide your substantive scientific and technical comments and suggestions within this [redacted] form. Identify and provide the rationale or scientific support for proposed changes or suggestions where possible.

If necessary, you can also provide additional editorial comments and recommendations for improving the report outside your specific charge questions (this form) within the draft report itself. Please note that only those comments included on the [redacted] form will be considered part of NTP's peer review report.

A. General Comments

I. Please comment on whether the scientific information presented in the draft monograph, including presentation of data in tables and figures, is technically correct, and clearly and objectively presented. Please suggest any improvements.

I.1: [redacted] **Comments:** The scientific information presented appears technically correct and objectively presented. A few suggestions are noted below to improve clarity. The background section could be reorganized for clarity and flow. It might be beneficial to begin the abstract and background with the pervasive use of fluoride in drinking water followed by a brief statement of the benefits. The benefits of fluoride in water has not been articulated. The benefits only need a sentence or two. The background appears to be more of a justification for the report rather than a true background of the evidence leading to the study/report.

Response: Agree (change made)

- We agree with the suggestion to reorganize the *Introduction* section. In response, we have moved text from the first paragraph of the *Introduction* closer to the end of the section. As such, the uses of and exposure sources to fluoride are now the first topics covered. We briefly discuss the benefits of fluoride but have not emphasized it or mentioned it in the *Objective* or *Specific Aims* as this topic is not the focus of the monograph. There is also no attempt in the monograph to compare hazards with benefits.

I.2: [redacted] **Comments:** Might consider beginning the background with the PHS recommendations.

Response: Disagree (no change)

- We have chosen not to highlight fluoridation of drinking water as the monograph focuses on the question of whether fluoride from all sources can affect neurodevelopmental outcomes and is written to avoid giving the mistaken impression that this systematic review is focused only on drinking water. While drinking water provides the majority of fluoride exposure in many of the studies, total exposure can vary widely even in optimally fluoridated areas based on personal habits in the use of dental products and consumption of beverages such as black tea that can contain fluoride.

I.3: [redacted] **Comments:** The abstract and background also need to be consistent in terms of presentation of human and animal studies. This consistent ordering of the studies (human, animal, mechanistic – for example) descriptions would improve flow and readability. Given the final conclusion of the animal studies section, is it possible to omit the non-human studies component?

Response: Disagree (no change)

- The ordering of topics in the various monograph sections has been determined after considering options and feedback from all reviewers. As a whole, we consider the current organization of topics in the monograph as appropriate to best support the ultimate rating of moderate confidence for effects of fluoride on children's IQ.

- With respect to the inclusion of the animal section, we consider it to be a valuable addition to the monograph even though the details have been largely relegated to earlier drafts that were reviewed by the NASEM Committee. The animal section provides an update to the 2016 NTP animal systematic review, identifies the studies that were conducted by the DNTP to address deficiencies in the 2016 NTP animal systematic review, and reiterates the lack of consistent evidence from this body of literature to support human findings.

I.4: [redacted] **Comments:** The term 'neurodevelopment' includes cognition, so if you would like to focus on cognition, you could simply state 'neurocognition.' Neurodevelopment is typically used as an umbrella term for all neurodevelopment, including cognition and motor function.

Response: Disagree (no change)

- We chose to use the terms "neurodevelopment" and "cognition" because the children's literature includes studies on both cognition and behavior.

I.5: [redacted] **Comments:** As currently written, the objective is not clearly stated. *Potential rewrite*: The objective of this report to assess the relationship between fluoride exposure and neurocognitive effects in humans and animals. To accomplish this objective, a systematic review of the literature was undertaken and mechanistic data was considered.

Response: Disagree (no change)

- We understand that the suggested refined objective may better reflect the ultimate emphasis of the monograph based on the data that were found; however, the systematic review was more comprehensive in scope and we consider it to be better represented by the current wording. Furthermore, the current wording is consistent with the published protocol.

I.6: [redacted] **Comments:** Why is the meta-analysis not included?

Response: No change requested

- The decision to pursue a narrative evidence synthesis rather than a meta-analysis was made while preparing the 2019 draft NTP Monograph because our goal of generating a document to support a hazard assessment did not require a quantitative estimate of hazard (e.g., numeric estimate of IQ points lost per mg F/L of drinking water or urine). However, as outlined in a new table that provides a timeline of draft monographs and important decision points (Table B-1 in *Appendix B* of the Sup02_2022_Prepublication_NTP_Monograph), comments received from the NASEM Committee that reviewed the 2019 draft NTP Monograph (NTP, 2019) recommended that we perform a meta-analysis and indicated that the outcome would be critical to reaching a hazard conclusion. We therefore performed a meta-analysis, which included a *dose-response meta-analysis*, and included the meta-analysis in the revised Sup04_2020_draft_NTP_Monograph (NTP, 2020). In its review of that Sup04_2020_draft_NTP_Monograph, the NASEM Committee again stated that the

document fell short of supporting our hazard call, and the Committee also had additional recommendations to improve the meta-analysis.

- After reflecting on the NASEM Committee comments on the Sup04_2020_draft_NTP_Monograph, we decided to remove the evidence integration step from the systematic review of the literature and instead issue the report (after further independent peer review) as a document outlining the state of the science on the association between fluoride exposure and deficits in neurodevelopment and cognition. We then decided to revise and submit the meta-analysis as a separate peer-reviewed publication because it was no longer required to support the "presumed" hazard call which was reached in the 2019 monograph and Sup04_2020_draft_NTP_Monograph. The meta-analysis, including the *dose-response meta-analysis*, was performed only on the studies addressing fluoride exposure in relation to deficits in children's IQ. The separate meta-analysis considers comments from the NASEM Committee in its revisions.

I.7: [redacted] **Comments:** Why limit to thyroid function as an effect/mechanism?

Response: No change requested

- Hypothyroidism and prematurity are among the few well-established risk factors for delayed or deficient neurodevelopment in children (for example, see review by Prezioso et al. [2018]). Many of the better-quality human studies controlled for gestational age at birth, and there is a growing body of literature on the interaction between fluoride exposure and low iodine levels in relation to children's IQ. This is why iodine was considered an important co-exposure in our risk-of-bias assessments (e.g., Goodman et al., 2022).

I.8: [redacted] **Comments:** Figure 1: [redacted] don't see where confounding or co-exposure is included.

Response: No change requested

- Confounding and co-exposures are part of the risk-of-bias assessment so are not individually listed in Figure 1. Details on confounding and co-exposures first appear in the *Quality Assessment of Individual Studies* section.

II. Please identify any information that should be added or deleted.

I.9: [redacted] **Comments:** Thyroid function isn't mentioned until the specific aims. It should be included in background along with other possible mechanisms, if known. It is unclear why thyroid function is being evaluated as the only mechanistic pathway. A figure or illustration depicting the theoretical pathway would be helpful.

Response: Agree (change made)

- We have added a footnote to the *Introduction* section of the Sup02_2022_Prepublication_NTP_Monograph to explain the focus on potential thyroid effects. The footnote reads:

  "The current review has evaluated the fluoride literature with an eye toward potential thyroid effects because a large literature base has accumulated

examining the interaction of fluoride with iodine uptake by the thyroid gland and consequential effects on synthesis of thyroid hormones, which are recognized to play significant roles in neurodevelopment in utero and during early childhood. This literature, along with a detailed proposed mechanism of action, was recently reviewed by Waugh (2019)."

I.10: [redacted] **Comments:** A brief discussion of serum fluoride needs to be included – similar to the urinary fluoride description (page 16).

Response: Agree (change made)

- We included a statement concerning serum fluoride in the *Exposure* section of the Sup02_2022_Prepublication_NTP_Monograph to explain why serum fluoride levels are considered a poor measure of long-term fluoride exposure. The new statement reads, *"Fluoride ion is rapidly absorbed from the gastrointestinal tract and is rapidly cleared from serum by distribution into calcified tissues and urinary excretion (IPCS 2002)."*

I.11: [redacted] **Comments:** Table 6 could include the following: 1) statistical methods; 2) confounders, particularly exposure to other known neurotoxicants, and how they were measured; 3) might rename 'Assessment timing' to age of participants or just combine the information with the location/subject's column

Response: Agree (edited for clarity)

- Although additional information could be added to Table 6, the information requested by [redacted] is already in *Appendix E* in the Sup02_2022_Prepublication_NTP_Monograph (previously *Appendix 4* of the Sup03_2021_draft_NTP_Monograph) for all the studies presented in Tables 6 and 7. Therefore, to address this comment, text has been added to the paragraphs that introduce Tables 6 and 7 to point to *Appendix E* for this additional information by study. We considered [redacted] suggestion to rename the 'assessment timing' column to 'age of participants'; however, we have retained the current column header as the information provided in this column is the age of participants at which outcome was assessed. The current header is the most concise way to communicate this.

B. Human studies

I. **Fluoride exposure and children's IQ**

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on measures of IQ in children.

I.12: [redacted] **Comments:** The approach described was appropriate. It is not clear when child and adult studies were separated from the main search or if each search was done independently (child and adult). It appears that it was only 'human studies.' [redacted] wonder how the search would change, if at all, if search terms for the target population was included? It should be clearly stated how and each population (child and adult) were separated.

Response: No change requested

- All life stages were relevant to the assessment according to our PECO statement (Population: "*Humans without restriction as to age or sex, geographic location, or life stage at exposure or outcome assessment*"). The same search was designed to identify child and adult studies, and the search did not include terms related to life stage (see response under B.III.1 for further explanation as to why this approach is thought to effectively capture relevant studies from all life stages). Although the process for deciding which bodies of evidence to synthesize and whether to separate groups of studies by health effects or age was described in the protocol, specific decisions were made based on the results of the literature search and selection. The groupings by age and the separation of child and adult studies were done after study selection and during the initial evaluation of the studies to determine what information was available. The initial evaluation sorted studies into children and adult studies to see if there was enough information to group the literature in a similar way as had been done for the 2016 NTP animal systematic review. As there was determined to be sufficient data, the decision was made to evaluate children separately from adults. The monograph explains that children and adults were evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood.

I.13: [redacted] **Comments:** The rationale for date selection needs to be more clearly articulated. The specific dates are included in the appendix, perhaps they could be included in the main text for clarity in the methods.

Response: Agree (change made)

- In an effort to provide further clarity on the progression of this multiyear assessment, we have developed a new table (Table B-1 in *Appendix B*) that provides a timeline of key activities contributing to the Sup02_2022_Prepublication_NTP_Monograph, including information relevant to the timing of the literature searches. For example, the expanded literature search to include non-English databases took place in May 2020 in response to the NASEM Committee's peer review report on the 2019 draft NTP Monograph.

2. Comment on whether the approach used to assess risk of bias was clearly described and appropriately applied to the set of studies designated as "low risk of bias."

I.14: [redacted] **Comments:** The approach to assess risk of bias was clearly described. A brief discussion is needed about critical confounders, including a biological exposure measure for tobacco use or exposure, such as serum cotinine, and parental IQ for the child studies If there are unique confounders for child and adult studies, this needs to be articulated. It currently appears that there are no unique confounders for child and adult.

Response: Agree (change made)

- [redacted] is correct that there are no unique confounders for children and adults. As noted in the monograph, the potential confounders that were considered important for all studies, populations, and outcomes were age, sex, and socioeconomic status regardless of whether the subjects were children or

adults. However, we realize that, as written in the Sup03_2021_draft_NTP_Monograph, it may be interpreted that age and sex confounders were only applied to children. Text has been updated in the Sup02_2022_Prepublication_NTP_Monograph to clarify that age and sex are important potential confounders regardless of life stage. For all other potential confounders considered in the evaluation, their importance was dependent on the study population and outcome being evaluated, and no specific potential confounder was unique to either children or adults.

- Smoking was considered an important confounder in adult studies evaluating Alzheimer's disease, but smoking was only considered a major concern if there were reasons to believe that it would be a source of bias.
- We agree with [redacted] that parental IQ is an important potential confounder in the studies of children. Because parental IQ, educational attainment, and other measures of socioeconomic status (SES) all likely share a common cause, the latter two covariates were considered to be potential proxy measures of parental IQ. Therefore, parental IQ was considered indirectly addressed if a study accounted for parental education and/or socioeconomic status. For clarification, we added a footnote to Figure 6 that lists all measures related to SES that were considered in the low risk-of-bias IQ-in-children studies.

3. Comment on assessment of the human studies with regard to:

a) How findings from individual studies designated as "low risk of bias" were interpreted.

I.15: [redacted] **Comments:** Findings from low-risk studies were interpreted well. They were interpreted objectively.

Response: No change requested

- No response necessary.

b) How the overall set of confounders across the body of evidence from children's IQ studies was considered and presented.

I.16: [redacted] **Comments:** The overall set of confounders has been thoughtfully considered and presented. Figure 6 is very comprehensive. Are there any unique confounders for the age groups (child and adult)?

Response: Agree (change made)

- This repeats a more extensive comment made previously on question B.I.2; see above for a more detailed response.

c) How the confidence rating in the body of evidence was developed and supported.

I.17:

[redacted] **Comments:** NTP used the GRADE system for rating confidence in the body of evidence. GRADE is a published method for reaching confidence. NTP also elaborated on factors they considered for potential downgrading and upgrading of research. Figure 1 outlines the process. It might be beneficial to include a 'scale' of factors that result in a score of high, moderate, low or very low in Figure 1, if applicable.

Response: Disagree (no change)

- As [redacted] points out, Figure 1 outlines the GRADE-based method, and the accompanying text elaborates on the factors considered for potential upgrading or downgrading of the confidence in the bodies of evidence. Given the complexity of the possible upgrade and downgrade decisions across the nine factors, we outline the process in Figure 1 rather than trying to predict all the combinations of factors that might result in different ratings of high, moderate, or low.

I.18:

4. ***NTP concludes a rating of moderate confidence in the body of evidence for lower IQ in children associated with fluoride exposure***.

☒ Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- Reviewer agreed with the moderate confidence rating.

II. **Fluoride exposure and non-IQ neurodevelopmental or cognitive effects in children**

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on non-IQ neurodevelopmental or cognitive effects in children.

I.19: [redacted] **Comments:** The approach described to search and select human studies on neurodevelopmental or cognitive function effects potentially associated with fluoride exposure was well-designed and executed. It should be stated if there were any literature review or data extraction methods for child and adult populations.

Response: Agree (no change)

- We agree that if literature review or data extraction methods had differed for child and adult populations, they would need to be clearly stated; however, in the case of this systematic review, the methods were not different. The systematic review protocol and monograph thoroughly describe the methods for screening (literature review) and data extraction and neither document indicates that methods would differ for children and adults. Study selection and data extraction methods were applied consistently across studies of both child and adult populations. Table 2 of the systematic review provides the inclusion and exclusion criteria used to determine study eligibility and states that there are no restrictions on age or life stage at exposure or outcome assessment, while not drawing any distinctions between child and adult studies. *Appendix 2* of the systematic review protocol lists data extraction elements and also does not draw any distinctions for studies in children versus adults.

I.20: [redacted] **Comments:** Page 13: Should consider adding team member initials to their roles in the review.

Response: Agree (change made)

- A front matter section titled *About This Review* has been added to the Sup02_2022_Prepublication_NTP_Monograph that lists the names of all team members along with a description of tasks to which they contributed (e.g., conducted literature screening, conducted data extraction).

I.21: [redacted] **Comments:** Page 13: there is a statement about studies 'evaluating only goiters or thyroid size were not extracted." If so, shouldn't they be part of the exclusion criteria? Similarly, data was not extracted from in vitro studies. This clarification is needed only because it appears that this report includes methods on data extraction for the meta-analysis that is in progress. For a reader, this description isn't necessary to understand the current report, but understand if these methods are needed.

Response: Disagree (edited for clarity)

- We have taken steps to increase clarity in the Sup02_2022_Prepublication_NTP_Monograph regarding the exclusion of topics for full evaluation. For example, details have been added to the *Data Extraction* methods discussion to further clarify why data on specific endpoints were not considered informative to the systematic review and did not undergo full data extraction or study quality evaluation (see below).

  "Data for primary and secondary outcomes, as well as thyroid hormone level data, were extracted from human studies. Studies evaluating only goiters or thyroid size were not extracted because they do not provide specific information on thyroid hormone levels that would inform whether a thyroid-mediated mechanism was involved in fluoride-associated changes in neurodevelopment."

  "Thyroid data were not extracted for animal studies due to inconsistency in the available data in humans."

  "In vitro studies were evaluated, although data were not extracted from these studies as none of the findings were considered informative with respect to biological plausibility."

- Note that the decision not to extract data on goiters was reached after studies went through the study selection process (where we apply inclusion and exclusion criteria to studies identified from the literature search). When this happens, it is standard practice to explain the reasoning in the systematic review methods, not to amend the protocol with this level of detail.

- The decision on thyroid data was reached by technical experts during the review because changes in thyroid size would not inform whether a thyroid-mediated mechanism was involved in fluoride-associated changes in neurodevelopment. The protocol did not include a level of detail on thyroid-related studies to specify preferred or less informative thyroid-related data. [redacted] makes a valid point that, in hindsight, the protocol could have specified that studies only reporting thyroid size or goiters would be excluded. Similarly, the consideration of mechanistic studies followed a tiered or phased approach to identify pockets of data that might support critical analysis with preference given to fluoride exposures of 20 ppm or less (deemed by technical experts to be most relevant to human exposures) and also to identify commonly reported thyroid-mediated mechanisms. The decision was also reached by technical experts during the

review that full data extraction of in vitro studies was not necessary to assess the biological plausibility of the human and animal results.

2. Comment on whether the approach used to assess risk of bias for studies in children on non-IQ neurodevelopmental or cognitive effects was clearly described and appropriately applied.

I.22: [redacted] **Comments:** Add a brief section on serum fluoride levels. Urinary fluoride levels is fully described, but serum has been omitted.

Response: Agree (change made)

- This repeats a more extensive comment made previously on question A.II.; see above for a more detailed response.

I.23: [redacted] **Comments:** One key feature for confidence rating is 'comparison group used.' This needs to be discussed further since fluoride exposure may be pervasive in water supplies. If so, in studies including a comparison group, include the comparison and how it was determined. Cross-sectional studies using biomarkers as continuous variables can be very strong.

Response: Disagree (edited for clarity)

- Tables 6, 7, and 8 already include data on exposure levels in comparison groups. Additionally, *Appendix E* in the Sup02_2022_Prepublication_NTP_Monograph (previously *Appendix 4* of the Sup03_2021_draft_NTP_Monograph) discusses in detail each low risk-of-bias study and indicates when biomarker measures were used.
- The comparisons in the epidemiological studies are between populations that had a range of fluoride exposures that could be compared with similar populations with lower or no fluoride exposures. To further distinguish between the comparison group and the group(s) exposed to higher levels of fluoride, we have added the word "higher" to specify "higher fluoride exposure," as appropriate, in several places throughout the Sup02_2022_Prepublication_NTP_Monograph. For example, we added the word "higher" to the sentence below from the *Results by Study Design – Cross-sectional Study Variations* section.

  "Overall, the cross-sectional studies consistently provide evidence that higher fluoride exposure is associated with lower IQ scores in children."

3. Comment on assessment of the human studies with regard to:

I.24:

a) How findings from individual "low risk of bias" studies were interpreted.

[redacted] **Comments:** Well done!

Response: No change requested

- No response necessary.

b) How the confidence rating in the body of evidence was developed and supported.

I.25: [redacted] **Comments:** Has the OHAT been published? If so, it should be referenced. Since it's a critical tool in this review, it needs to be further described. What other QA tools are available and why weren't they used? Were the Cochrane Review recommendations for assessment of the risk of bias in research studies followed?

Response: Agree (edited for clarity)

- We agree that the OHAT risk-of-bias tool should be referenced, and we have added this reference to both the protocol and the Sup02_2022_Prepublication_NTP_Monograph as https://ntp.niehs.nih.gov/go/riskbias. The risk-of-bias tool was reviewed by an expert panel as part of the development of the OHAT methods and is publicly posted on the NTP web pages.
- We disagree that the tool needs to be further described in the Sup02_2022_Prepublication_NTP_Monograph because it is described in detail in the protocol, which is appropriately referenced in the *Methods* section.
- The OHAT risk-of-bias tool was selected for this systematic review because it uses a parallel approach to assessing study quality across different study designs for both human and animal research, thus enabling synthesis and development of the confidence ratings to meet the objectives. It is the only tool that is designed to assess studies of environmental exposures, studies of varying study designs that were necessary for this systematic review, and studies in both humans and experimental animals. As described in the tool and the protocol for this systematic review, the OHAT risk-of-bias tool is based on Cochrane and AHRQ methods; therefore, the Cochrane Review recommendations for assessment of risk of bias of human studies were followed.

I.26:

4. ***The NTP concludes a rating of low confidence in the body of evidence for decreases in measures of other neurodevelopmental or cognitive effects in children associated with fluoride exposure***.

☒ Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

III. **Fluoride exposure and cognitive effects in adults**

I.27:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on cognitive effects in adults.

[redacted] **Comments:** Well described – since it appears to be the same for the child studies. Search terms does not include "child," "pediatrics," or "adult," or other terms to

separate out the child and adult studies. When were these terms added or were they added in the search?

Response: No change requested

- The search terms "child", "pediatrics", and "adult" were not included in the literature search. It was unnecessary to include these or other terms related to life stage because relevant studies of all life stages were captured with the existing search strategy. The search strategy included a set of exposure terms (e.g., "fluoride") and a set of health outcome terms (e.g., "neurodevelopment") as detailed in the appendices to the monograph. All life stages were relevant to the assessment according to our PECO statement (Population: "*Humans without restriction as to age or sex, geographic location, or life stage at exposure or outcome assessment*"), and we are confident that all relevant child and adult studies were identified by searching for relevant exposure and outcome terms only (i.e., all fluoride and neurodevelopmental studies would be identified across all life stages). Moreover, we are confident that the absence of search terms related to life stage would not result in missing studies with relevant exposures and relevant outcomes.

I.28:

2. Comment on whether the approach used to assess risk of bias for studies in adults on cognitive effects was clearly described and appropriately applied.

[redacted] **Comments:** Since it is similar to the methods used for child studies.

Response: No change requested

- No response necessary; [redacted] only notes that similar methods were used for studies in children.

3. Comment on assessment of the human studies with regard to:

I.29:

a) How findings from individual studies were interpreted.

[redacted] **Comments:** Not sure of this question – how is it different from the question in the 'child section'? Adult studies were interpreted well.

Response: No change requested

- No response necessary.

I.30:

b) How the confidence rating in the body of evidence was developed and supported.

[redacted] **Comments:** Similar response to the 'child section' above. The confidence in the adult studies was interpreted well.

Response: No change requested

- No response necessary.

I.31:

4. ***The NTP concludes a rating of low confidence in the body of evidence for changes in cognitive effects in adults with fluoride exposure***.

☒ Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

C. Studies in non-human animals

I.32:

The NTP agrees with the comments of the NASEM committee (NASEM 2020, 2021) concerning the overall poor quality of the experimental animal database on fluoride exposure and neurodevelopmental effects, with many studies suffering from major reporting deficiencies. As indicated above, the monograph focuses on the large human epidemiology database because it directly addresses the question of whether fluoride affects human neurodevelopment. Therefore, based on the recommendations of the NASEM committee, the experimental animal section and risk of bias details have been removed from this monograph and ***the NTP concludes that the scientific evidence from experimental animal data are inadequate to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans***.

☒ Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the inadequate designation.

I.33:

[redacted] **Comments:** If this is the conclusion of the review, [redacted] question the inclusion of non-human studies in this monograph.

Response: Disagree (no change)

- As discussed earlier, we contend that the animal studies section is a valuable part of the review because it provides a brief update to the 2016 NTP animal systematic review, identifies studies conducted by the DNTP to address deficiencies noted in the 2016 NTP animal systematic review, and reiterates the lack of consistent evidence from this body of literature to support human findings.

References:

Goodman CV, Hall M, Green R, Chevrier J, Ayotte P, Matinez-Mier EA, McGuckin T, Krzeczkowski J, Flora D, Hornung R, Lanphear B, Till C. (2022). Iodine Status Modifies the Association between

Fluoride Exposure in Pregnancy and Preschool Boys' Intelligence. Nutrients: 14(14):2920. https://doi.org/10.3390/nu14142920.

Prezioso G., Giannini C., and Chiarelli F. (2018). Effect of thyroid hormones on neurons and neurodevelopment. *Horm Res Paediatr*: 90:73-81. https://doi.org/10.1159/000492129.

In November 2021, [redacted] received: 1) the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review,* 2) a copy of the NASEM Committee's comments on the 2020 draft NTP Monograph with NIEHS/DNTP responses (draft version of Sup01_Monograph), and 3) the [redacted] instructions. The instructions consisted of a preface, charge, instructions for the review, and a series of specific peer-review questions grouped by the following three topics: General Comments, Human Studies, and Studies in Non-Human Animals.

[redacted] were asked to provide their substantive scientific and technical comments and suggestions within the [redacted] form. In addition, they were asked whether they "Agree", "Agree in principle", or "Do not agree" with each NTP conclusion on confidence in a body of evidence.

The [redacted] instructions and specific peer-review questions are reproduced in the pages that follow in black text. [redacted] comments and responses to each question are also provided in black text starting with the words "[redacted] **comments**" in bold font. The NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] comments and [redacted] comments:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to this [redacted] would be marked "*see DocJ_Monograph for detailed response*").
- [redacted] For comments DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

Preliminary comments on the draft NTP monograph prepared by the peer review [redacted] are noted below.
These preliminary comments are not binding and should not be construed to represent NTP determination or policy.

National Toxicology Program
NTP Monograph Letter Peer-Review Panel
Draft NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review

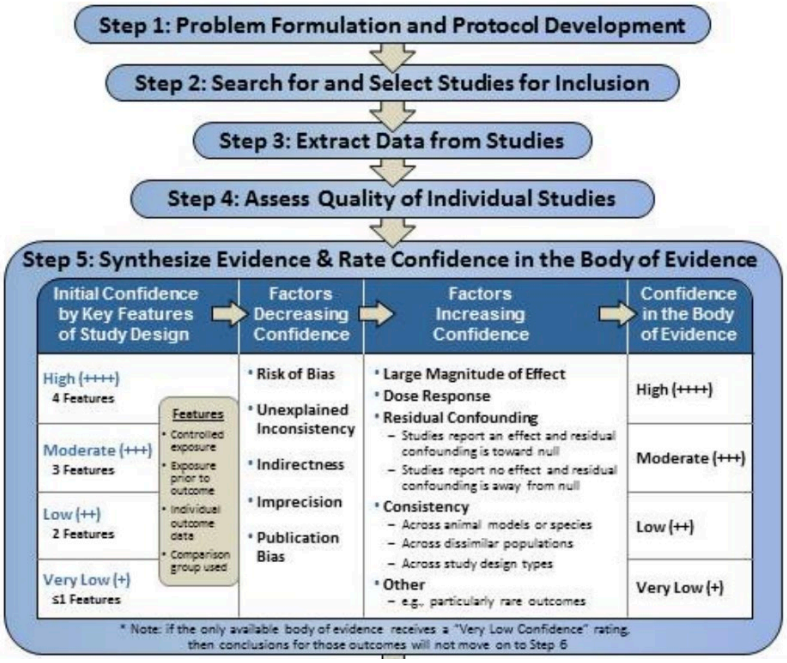
National Institute of Environmental Health Sciences
Research Triangle Park, NC

December 22, 2021

Fluoride State of the Science Document Review Form

[redacted]

Preface*:*
The objective of this evaluation was to conduct a systematic review of the published literature regarding the potential for exposure to fluoride to affect neurodevelopment and cognition in humans. The evaluation presented in the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* represents a comprehensive and current assessment. The methods used are from the Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration, which presents a seven-step framework for systematic review and evidence integration. Please note: this evaluation stops at step 5 of the systematic review process and does not proceed to step 6 to translate the confidence rating for the body of evidence into a level of evidence for health effects (see Figure 2 from the handbook).



Charge:

(1) Comment on the technical accuracy and whether the draft *NTP Monograph on State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* is clearly stated, and objectively presented.

(2) Determine whether the scientific evidence supports the NTP's confidence ratings for the bodies of evidence regarding neurodevelopmental and cognitive health effects associated with exposure to fluoride.

Instructions for Review:
All materials for this review are available in the Electronic Council Book (ECB). You will receive the specific URL and a password for accessing the ECB.

This evaluation identified 159 human studies relevant for assessing neurological health effects of exposure to fluoride; however, many studies included only secondary outcomes (e.g., 55 studies of thyroid hormones that were investigated as a potential mechanism). The scientific evidence in children and adults was evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood. Several studies evaluated learning and memory (n = 8 studies) or other cognitive developmental effects (e.g., total neurobehavioral scores and total mental capacity index in children, cognitive impairment in adults; n = 14 studies). Sixty-six human studies investigated IQ in children. Nineteen of the 66 IQ studies were determined to have low potential for bias and therefore, were categorized as "low risk of bias". Please give special attention to our assessment of these 19 studies.

- The 19 studies are available as PDFs and organized alphabetically in a folder on the ECB.
- All other studies are provided in the Health Assessment Workspace Collaborative, or HAWC database under the "studies list" tab, also organized alphabetically. You will also be provided a username and password for HAWC that will give you [redacted] permissions to access the PDFs in HAWC along with visualizations and other study information for this project at the following link (https://hawcproject.org/study/assessment/405/).

Please provide your substantive scientific and technical comments and suggestions within this [redacted] form. Identify and provide the rationale or scientific support for proposed changes or suggestions where possible.

If necessary, you can also provide additional editorial comments and recommendations for improving the report outside your specific charge questions (this form) within the draft report itself. Please note that only those comments included on the [redacted] form will be considered part of NTP's peer review report.

A. General Comments

1. Please comment on whether the scientific information presented in the draft monograph, including presentation of data in tables and figures, is technically correct, and clearly and objectively presented. Please suggest any improvements.
2. Please identify any information that should be added or deleted.

J.1: [redacted] **Comments:** Congratulations on a thorough and comprehensive systematic review – not only is the review itself impressive, but the HAWK system, your online portal, and all of your processes for assessing COI and training [redacted] were equally impressive.

Overall, the review is well organized, clearly written, and transparently documented. Below are a series of comments and questions, that if considered, may improve the review.

Response: provided below

- We appreciate [redacted] feedback and have provided responses to the series of comments in [redacted] table below where the issues were raised.

Section, page #	Comment
Objective and Specific Aims; page 5	**J.2:** Because this review has an extensive history that could be difficult for a reader to follow (i.e., the original 2016 review, and drafts from 2019, 2020, and the current draft), it would be helpful to develop a table or flowchart that documents that history. For example, you may consider noting the purpose/research question, findings, and noteworthy differences from previous/subsequent versions. See comments below, but the literature search section, in particular, was a little difficult to follow - and having the "big picture" of the review in a table or flowchart to refer to, would better allow the reader to follow all of the searches conducted, and how the differ, yet fit together to contribute to the present document. **Response: Agree (change made)** - In an effort to provide further clarity on the progression of this multiyear assessment, we have developed a new table (Table B-1 in *Appendix B* of the Sup02_2022_Prepublication_NTP_Monograph) that provides a timeline of key activities contributing to the Sup02_2022_Prepublication_NTP_Monograph, including literature searches that were utilized for the various drafts that underwent different peer reviews.
Objective and Specific Aims; page 5	**J.3:** It is not clear why the "hazard assessment step" was removed from the methodology. Is it because the authors deemed the step not possible based on available evidence? Or is it because the hazard assessment step will occur separately, taking into consideration both the review and the results of meta-analysis?

	Response: Agree (edited for clarity) ○ The *Preface* of the monograph clearly describes why the hazard assessment step was removed from the Sup02_2022_Prepublication_NTP_Monograph. Additionally, we developed a new table (Table B-1 in *Appendix B* of the Sup02_2022_Prepublication_NTP_Monograph) that provides a timeline of key activities contributing to the Sup02_2022_Prepublication_NTP_Monograph, including when the hazard assessment step was removed and that it was removed in response to the NASEM Committee's review report of the Sup04_2020_draft_NTP_Monograph. ○ In brief, the NASEM Committee's comments indicated they did not believe that the Sup04_2020_draft_NTP_Monograph presented a clear and convincing assessment to support its hazard conclusions. Although many of the comments offered by the NASEM Committee are addressed in the current document, we chose to delete the hazard assessment step and instead express our level of confidence in the evidence of an association between fluoride exposure and effects on cognitive neurodevelopment as our contribution to the larger ongoing discussion on the safe use of fluoride for oral health.
Methods, page 7	**J.4:** Would it be possible to define what is meant by "Categories focused on were those with more robust data at levels of fluoride more relevant to human exposure"? Should this information be documented as part of the PECO? Was this an inclusion criteria, or just used in prioritizing or weighting the evidence in drawing conclusions? **Response: Agree (edited for clarity)** ○ The sentence describing the use of categories with more robust data is a description of the approach used to collect, prioritize, and consider the available mechanistic data following the organization of the PECO statements. These were not inclusion criteria or an additional factor that could have been added to the PECO. Although the process for deciding which groupings of health effects to synthesize and whether to synthesize all groupings of health effects was described in the protocol, the specific decisions were made based on the results of the literature search and selection. This approach is specifically outlined in the *PECO Statements* section to describe how the in vitro/mechanistic data were evaluated and considered because it often differs from how human or animal data are assessed. We have edited the cited text for clarity and it now reads as follows: *"To prioritize and consider available mechanistic data, the categories focused on were those with more robust data at levels of fluoride more relevant to human exposure."*

Methods, page 8	**J.5:** The literature search section was somewhat confusing to follow, though, given the complexity of updating reviews, etc it is understandable why multiple searches were conducted. See previous comment regarding the various iterations of this review, historically, and how a table or flowchart may help the reader understand the progression of this review, and thus, better follow the searches that were carried out. For example, you may consider adding sub-headings within this section to distinguish which searches were run to capture which types of studies. **Response: Agree (edited for clarity)** ○ In response to [redacted] earlier comment on organization, and in an effort to provide further clarity on the progression of this multiyear assessment, we have developed a new table (Table B-1 in *Appendix B* of the Sup02_2022_Prepublication_NTP_Monograph) that provides a timeline of key activities contributing to the Sup02_2022_Prepublication_NTP_Monograph, including information relevant to the timing of the literature searches. For example, the expanded literature search to include non-English databases took place in May 2020 in response to the NASEM Committee's peer review report on the 2019 draft NTP Monograph.
Methods, page 9	**J.6:** Is there any plan to update the literature search run on May 1, 2020? Given that the search is now 1.5 years old, and this seems to be a topic with emerging evidence, it would be beneficial to update the search to ensure all relevant studies have been captured. **Response: Agree (No change)** ○ We performed an updated literature search in November 2021. There were a number of newer relevant publications identified, including several in Chinese journals. These newer publications (n = 7) are included as part of the meta-analysis, which is being prepared as a separate report for publication. We determined that, while the newer publications may slightly affect the quantitative results of the meta-analysis and dose-response meta-analysis, their findings are largely consistent with the literature reviewed in the current monograph and do not materially affect the level of confidence we have in the database. Because inclusion of these new studies in the monograph would necessitate further peer review, we have chosen not to include them.
Methods, page 10	**J.7:** This is the first time the "Flouride Action" website is mentioned (and the actual hyperlink appears in the subsequent section). It may be helpful to the reader to provide some rationale for why this website was specifically targeted. **Response: Agree (change made)** ○ We added new text to introduce the Fluoride Action Network and to clarify that the site was used as another resource because it is known to index fluoride publications. The new text appears as follows:

	"Fluoride Action Network website (http://fluoridealert.org/)—a site used as another resource to identify potentially relevant studies because it is known to index fluoride publications…" **J.8:** In addition, can it be assumed that any non-English paper that met criteria, regardless of outcome, would have been included? While it is understandable why the confirmatory search was done, it could be perceived as biased to only search for and include papers with null findings. **Response: Agree (change made)** ○ In terms of Chinese databases, we conducted the literature search independent of study findings, but we initially gave translation priority to studies that appeared to show no association. Although this was done to address potential publication bias, we agree that this was not appropriate and therefore have taken additional steps to translate and extract data from all relevant non-English studies identified from the Chinese database searches, including those that were not previously translated. Furthermore, the statements about null or no-effect studies have been deleted from the Sup02_2022_Prepublication_NTP_Monograph. ○ In addition, we updated the text in the *Literature Search* section to reflect that the search of Chinese databases was conducted to identify studies that may have been missed in previous searches because non-English language studies are not always indexed in the main databases used for this systematic review.
Methods, page 11	**J.9:** [redacted] can appreciate the use of machine-learning software to prioritize articles for screening. And the authors have done a nice job in describing and evaluating the algorithm employed when stopping at 98% - estimating that 2-4 studies may have been missed. However, given the high-profile nature of this review, and some level of uncertainty in the prediction algorithms of the tool, it may have been beneficial to manually screen the entire set of search results (2-4 studies is not an insignificant number when considering the total # of included articles). Use of machine-learning is helpful in that it can prioritize and identify sooner most included articles; however, when conducting systematic reviews used in large scale public health decision-making, it may be worth screening 100% of search results to ensure that all potentially relevant studies have been included. **Response: Disagree (no change)** ○ By using SWIFT Active Screener software to screen the initial literature search results, we avoided the need to manually screen over 13,000 abstracts. As outlined in the Sup02_2022_Prepublication_NTP_Monograph and systematic review protocol (https://ntp.niehs.nih.gov/go/785076), in addition to the screening of bibliographical databases, several additional methods to identify relevant literature were also employed. These included publicly

	posting the literature search results and asking peer reviewers at each stage whether they were aware of any additional relevant articles, screening the reference lists of reviews and included papers for possible articles, and conducting updated literature searches as outlined in response to a previous comment by the reviewer. The use of SWIFT Active Screener was estimated to result in the potential to miss one or two relevant human studies with primary neurodevelopmental or cognitive outcomes. The savings in time and impact were weighed against the potential impact of missing 1 or 2 studies relative to the nearly 100 human epidemiological studies identified with primary neurodevelopmental or cognitive outcomes, and this tradeoff was deemed to be acceptable.
Methods, page 13	**J.10:** If studies evaluating only goiters or thyroid size were not extracted, then why include them in the review altogether. Would it be more accurate to have amended the protocol to exclude these as outcomes of interest? **Response: Disagree (no change)** ○ The decision not to extract data on goiters was reached by technical experts during the review because changes in thyroid size would not inform whether a thyroid-mediated mechanism was involved in fluoride-associated changes in neurodevelopment. The protocol did not specify preferred or less informative thyroid-related data. [redacted] makes a valid point that, in hindsight, the protocol could have specified that studies only reporting thyroid size or goiters could have been excluded. Given that this decision was reached during the assessment, it is common practice to provide reasoning in the systematic review for these types of decisions, not to amend the protocol with these details.
Methods, page 15	**J.11:** Given that all included study designs were observational in nature, risk of bias due to confounding is a serious consideration. This [redacted] appreciates the thorough discussion of key confounders considered in risk of bias assessments but has concerns that even in studies rated as "low risk of bias," there remain serious concerns about the potential for confounding. This is especially important when considering an outcome like IQ, for which concerns are often raised about the specificity of the outcome, and its relationship with other constructs, such as SES, education, and race. **Response: Disagree (no change)** ○ We agree that risk of bias due to confounding is a serious consideration in any risk-of-bias assessment of observational studies but, due to both the comprehensive risk-of-bias assessment of each individual study and the assessment of potential confounding across studies, we disagree that there remain serious concerns about potential confounding among the low risk-of-bias studies. As described in the protocol, for a study to be considered low risk of bias for confounding, there had to be direct or indirect evidence that the key covariates (age, sex, and SES) and any other covariates considered important for the study's specific study

	population and/or outcome were sufficiently considered in terms of confounding. For example, studies of populations in China, India, and Mexico, where there is concern about exposures to high fluoride and high arsenic, were required to address arsenic. Figure 6 shows that 16 of 19 low risk-of-bias studies addressed each of the three key covariates and other important covariates, meeting the requirements for low risk of bias due to confounding. Looking across the body of literature, we observed considerable variation in covariates addressed across the 19 low risk-of-bias studies. When considering the impact of potential confounding on the consistency of results, no trends were discernable that would suggest that bias due to confounding has impacted or would explain the consistency in findings across the body of evidence that higher fluoride exposure is associated with lower IQ in children. ○ If a key covariate or other important covariate was not addressed in a study, we would also consider the most likely direction and magnitude of the potential bias. If the bias was likely to be toward the null, that may increase our confidence in the reported direction of the association. *Appendix E* (*Details for Low Risk-of-bias Studies*) includes detailed assessments of and justifications for each risk-of-bias rating, including considerations for the direction and magnitude of potential bias. ○ [redacted] identifies SES, education, and race. SES was a key covariate, education was considered as a measure or proxy of SES (see footnote to Figure 6), and race/ethnicity is listed in the protocol as a potentially important confounder; however, every study was conducted outside of the United States and there was no direct or indirect evidence to indicate that confounding by race/ethnicity was a concern. **J.12:** Is child sex a true confounder, in that its related both to the exposure and the outcome? (i.e., is there data to suggest that fluoride exposure differs based on sex?) **Response: No change requested** ○ We consider biological sex to be an important covariate and potential confounder for several reasons: (1) sex has historically been considered an important potential confounder in the literature (see Table 6 in Sup02_2022_Prepublication_NTP_Monograph) (Lash et al. 2021; Gochfeld 2017); (2) sex is an important risk factor for neurodevelopmental and cognitive outcomes (Cowell and Wright 2017); and (3) sex-related dietary ingestion and dietary differences are realistic in observational studies (D'Amico et al. 2020; Keller et al. 2019).
Methods, page 17	**J.13:** The paragraph describing RoB procedures could be moved up (prior to the PECO sections; as currently placed it gets a little lost and/or could be misperceived as relating only to outcome assessment).

	Response: Disagree (no change) ○ The systematic review process involves several steps and stages, and there is a general order by which these stages take place. The risk-of-bias discussion is located in an area of the *Methods* section that corresponds to the appropriate stage of the systematic review process, as is standard in publications of these types of reviews. Moving the risk-of-bias methods before the PECO (and therefore the literature search and screening methods) would create a misperception that the literature screening was influenced by study quality.
Methods, page 19	**J.14:** It is not clear why the meta-analysis portion of this review is being prepared as a separate report. **Response: Agree (edited for clarity)** ○ The decision to pursue a narrative evidence synthesis rather than a meta-analysis was made while preparing the 2019 draft NTP Monograph because our goal of generating a document to support a hazard assessment did not require a quantitative estimate of hazard (e.g., numeric estimate of IQ points lost per mg F/L of drinking water or urine). However, as outlined in a new table that provides a timeline of draft monographs and important decision points (Table B-1 in *Appendix B* of the Sup02_2022_Prepublication_NTP_Monograph), comments received from the NASEM Committee that reviewed the 2019 draft NTP Monograph (NTP, 2019) recommended that we perform a meta-analysis and indicated that the outcome would be critical to reaching a hazard conclusion. We therefore prepared a meta-analysis and included both the meta-analysis and dose-response meta-analysis in the revised Sup04_2020_draft_NTP_Monograph (NTP, 2020). In its review of that 2020 draft NTP Monograph, the NASEM Committee again stated that the document fell short of supporting our hazard call, and the Committee also had additional recommendations to improve the meta-analysis. After reflecting on the NASEM Committee comments on the Sup04_2020_draft_NTP_Monograph, we decided to remove the evidence integration step from the systematic review of the literature and instead issue the report (after further independent peer review) as a document outlining the state of the science on the association between fluoride exposure and deficits in neurodevelopment and cognition. This change is outlined in the *Preface* to the Sup02_2022_Prepublication_NTP_Monograph. Removing the evidence integration step from the systematic review precluded a determination of an overall hazard call. We then decided to revise and submit the meta-analysis as a separate peer-reviewed publication because it was no longer needed in an evaluation of confidence in the database of human evidence. An additional consideration was that the meta-analysis and dose-response analysis were performed only on the studies addressing fluoride exposure in relation to deficits in children's IQ, rather than on

	other neurological outcomes in children or cognition in adults. The separate meta-analysis considers comments from the NASEM Committee in its revisions.
Methods, page 20	**J.15:** [redacted] does not agree with the premise that all human studies are direct; it seems that certain measures of fluoride exposure have concerns with directness (i.e., endemic geographical region, job title). **Response: Disagree (no change)** ○ [redacted] cites two examples of fluoride exposure "endemic geographical region" and "job title" as potential concerns with directness. However, these examples are both direct evidence for this systematic review as defined in Table 1 the human PECO (Population, Exposure, Comparator and Outcome) Statement. Direct evidence comes from research that directly assesses exposures that are the focus of a given systematic review when described in populations that are also within the focus of a systematic review. As listed below, the PECO Statement in Table 1 specifies the population of interest as "humans without restriction" and exposure includes "job title" and "water levels" that cover groundwater exposure from endemic geographical regions. *"**Population**: Humans without restriction as to age or sex, geographic location, or life stage at exposure or outcome assessment* ***Exposure:** Exposure to fluoride based on administered dose or concentration, biomonitoring data (e.g., urine, blood, other specimens), environmental measures (e.g., air, water levels), or job title or residence…"* ○ [redacted] is using a definition of the population and exposures of interest that differ from the PECO statement for this review. [redacted] example would only apply when the specific question of a review is directed toward a narrow subpopulation and that would be stated in an alternate PECO. For example, if the review had been to evaluate the evidence on the association between occupational exposures to fluoride through mining and cognitive effects, then there would be direct and indirect human evidence. Direct evidence would include studies of miners with inhalation exposure or other occupational exposures determined by "job title" or other metrics); indirect evidence might include studies of oral exposures through water or dietary sources. The objective of this systematic review is to evaluate the evidence concerning the association between any fluoride exposure and neurodevelopmental and cognitive effects; therefore, all human studies are direct evidence.
Results, Figure 2, page 23	**J.16:** Identifying 15 references through other sources seems somewhat high. Was there a need to adjust the original search strategy to capture those references?

	Response: Agree (change made) ○ We have added text to clarify why the references identified by other sources were not captured in the database searches. In brief, 11 of the 15 references identified through other sources were not indexed in the bibliographic databases searched and therefore were not captured by the database searches. Many of the studies initially identified by other sources were non-English-language studies, and we recognized that additional targeted search strategies were required to identify non-English-language studies for this review. The supplemental search of Chinese databases was designed and conducted to address these challenges. Upon further review, we have clarified that four of the references in question were captured in the Chinese database searches, and we have made this correction to the text and study flow diagram. We were unable to identify the remaining 11 studies in any database searches. Regarding the impact of these 11 studies on the systematic review, only 1 of the 11 studies was a low risk-of-bias IQ study in children, and this study was included in the 19 low risk-of-bias studies on which the moderate confidence in the IQ-in-children body of evidence is based. The omission of this single study would not impact the moderate confidence rating. Of the remaining 10 studies, 7 were high risk-of-bias IQ-in-children studies and 1 was a high risk-of-bias adult study. The omission of the 7 (out of 53) high risk-of-bias IQ-in-children studies or the 1 (out of 8) high risk-of-bias adult studies would not impact any confidence conclusions in the monograph. Similarly, the two experimental animal studies would not impact the evaluation as the animal evidence was considered inadequate. The following new text appears as a footnote in the *Literature Search Results* section of the monograph: *"These 11 studies (9 human and 2 animal studies) were not identified through the electronic database searches, as they were not indexed in any of the electronic databases searched. Note that the supplemental search of non-English-language databases was designed in part to identify non-English-language studies that are not indexed in traditional bibliographic databases such as PubMed. It was successful in this goal, as multiple studies that were initially only identified through "other sources" were subsequently captured in the supplemental Chinese database search, leaving only 11 as identified through other sources."*
Results	**J.17:** In general, it would be helpful to the reader to describe the included study designs earlier in the respective results sections (i.e., along with the total #s of included articles, describing the # of cross-sectional, prospective cohort, etc is important). **Response: Agree (change made)** ○ We have added descriptions of the cohort, case-control, cross-sectional, and case report/case series study designs based on the NRC Report on

	Environmental Epidemiology (NRC 1997) as footnotes to Table 4 in the prepublication 2022 NTP Monograph, as follows: *"[c]Cohort studies are observational, studies in humans that examine a cohort prospectively or retrospectively over time.* *[d]Case-control studies are observational studies in humans that compare exposures of individuals who have a specific health effect or disease with exposures of controls who do not have the effect or disease. Controls generally come from the same population from which the cases were derived.* *[e]Cross-sectional studies are observational studies in humans that examine the relationship between exposures and outcomes or health effects assessed contemporaneously. Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).* *[f]A case report (or case study) is a descriptive study of a single individual or small group in which the study of an association between an observed effect and a specific environmental exposure is based on clinical evaluations and histories of the individual(s). A case series study in environmental epidemiology is designed to share health-related events on a collection of case reports on subjects with the same or similar health outcome(s) and environmental exposure(s)."* ○ We also added information on counts of studies per study design to the *Overview of Studies* subsections of the *IQ in Children*, *Other Neurodevelopmental or Cognitive Effects in Children*, and *Cognitive Effects in Adults* sections as indicated below. *"Nineteen studies (3 longitudinal prospective cohort and 16 cross-sectional studies) with low potential for bias evaluated the association between fluoride exposure and IQ in children (see Quality Assessment of Individual Studies section for methods on determining which studies pose low risk of bias)."* *"Nine low risk-of-bias studies (three prospective cohort and six cross-sectional studies) evaluated the association between fluoride exposure and cognitive neurodevelopmental effects other than IQ in children."* *"Two low risk-of-bias cross-sectional studies evaluated the association between fluoride exposure and cognitive effect in adults (Jacqmin et al. 1994; Li et al. 2016)."* **J.18:** It would also be of interest to expand Fig 3, or create a similar figure, to capture the ages at which fluoride exposure was measured. **Response: Disagree (no change)** ○ Although we agree that an expansion of Figure 3 could be interesting, the purpose of Figure 3 was to visualize the number of relevant studies identified in order to evaluate the outcome categories for pockets of data. For Figure 3, the studies were labeled as child or adult in order to

	evaluate if there were sufficient data to evaluate child and adult studies separately, as was done for the 2016 NTP animal evaluation.

B. Human studies

I. Fluoride exposure and children's IQ

J.19:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on measures of IQ in children.

[redacted] **Comments:** Yes, the approach used to search for and select studies was appropriate.

Response: No change requested

- No response necessary.

J.20:

2. Comment on whether the approach used to assess risk of bias was clearly described and appropriately applied to the set of studies designated as "low risk of bias."

[redacted] **Comments:** In general, it is difficult to understand how cross-sectional studies that adjusted for few, or no, confounders, employed somewhat indirect measures of fluoride exposure (or did not fully capture all sources of exposures to fluoride), or had concerns related to selection bias, were designated as "low risk of bias." If, for example, some confounders were accounted for in the design or analysis, other than statistical adjustment, it may be worth noting that on Table 6 (otherwise, it appears that many papers accounted for no confounders).

For example, Xiang, 2003a did not statistically adjust for any confounders. They did report some findings in relation to some of the confounders, but not to the extent that [redacted] would perceive them to have been fully accounted for.

For Xiang, 2011, the paper is published in a somewhat abbreviated format, appearing almost as a conference report or short correspondence. The journal does appear to be peer-reviewed currently, but it may be worth confirming that papers from 2011 were in fact peer-reviewed.

For Till, 2020, it does not appear that the authors applied inclusion/exclusion criteria related to the length of time subjects lived in the geographical areas tested. Therefore, it is difficult to know if exposure was accurately estimated. In addition, the authors did not confirm that formula preparation was done with tap or bottled water, but rather they used a proxy (maternal report of drinking tap/bottled); and it is unclear whether maternal drinking behaviors match formula preparation methods. Finally, the study measured exposure from 0-6mo, and did not account of fluoride exposure that occurred over the course of follow-up to age 3-4y (i.e., teeth brushing, supplemental intake, dietary intake). Therefore, it seems that there are some serious concerns related to potential exposure misclassification in this study.

Response: Disagree (edited for clarity)

- We appreciate [redacted] concern regarding cross-sectional studies; however, we disagree with the assertion that the low risk-of-bias cross-sectional IQ studies

have serious concerns related to confounding, exposure assessment, or selection bias that would preclude them from their designation as "low risk of bias." As described below and in detail in *Appendix E*, using the criteria in our protocol, we determined that these are well-conducted studies with minimal risk-of-bias concerns. The subsequent bullets in this response detail the strengths of these studies regarding their study design and low potential for bias due to confounding, exposure misclassification, and selection bias. In addition, we address the study-specific concerns raised by [redacted].

- **Confounding:** Due to both the comprehensive risk-of-bias assessment of each individual study and the assessment of potential confounding across studies, we disagree that there remain serious concerns about potential confounding among the low risk-of-bias studies. [redacted] is correct that Table 6 reports only the covariates that were adjusted for statistically. However, as is recommended by [redacted], Figure 6 does indicate when a covariate was adjusted for statistically and/or was not a concern for confounding in a particular study. As described in the protocol, for a study to be considered low risk of bias for confounding, there had to be direct or indirect evidence that the key covariates (age, sex, and SES), and any other covariates considered important for the specific study population and/or outcome, were sufficiently considered in terms of confounding. Examples of what it means for a covariate to be sufficiently considered in terms of confounding are described in a revised footnote to Figure 6 and include: it (the covariate) was statistically adjusted for in the final model, it was included in the model but not the final model because it did not substantially change the effect estimate, it was reported to have the same distribution in both the exposed and unexposed groups, and it was reported to not be associated with the exposure or outcome in that specific study population (thereby eliminating it as a potential confounder). Figure 6 shows that 14 of 16 low risk-of-bias cross-sectional studies addressed each of the three key covariates and other important covariates, meeting the requirements for low risk of bias due to confounding. Looking across the body of literature, we observed considerable variation in the covariates addressed across the 19 low risk of bias studies (16 cross-sectional and 3 prospective cohort studies). When considering the impact of potential confounding on the consistency of results, no trends were discernable that would suggest that bias due to confounding has impacted or would explain the consistency in findings across the body of evidence that higher fluoride exposure is associated with lower IQ in children.

  If a key covariate was not addressed in a study, we would also consider the most likely direction and magnitude of the potential bias. If the bias was likely to be toward the null, it may increase our confidence in the reported direction of the association (Xiang et al. [2003] is an example of this because it did not directly address potential co-exposure to arsenic; see further discussion of this study below). Detailed assessments of and justifications for risk-of-bias ratings for the low risk-of-bias studies, including considerations for likely direction and magnitude of bias, are provided in *Appendix E* (*Details for Low Risk-of-bias Studies*).

- **Exposure (characterization and considering potential misclassification):** Fifteen of the 16 low risk-of-bias cross-sectional studies that assessed the association

between fluoride exposure and IQ provide direct or indirect evidence that exposure was consistently assessed using acceptable methods and used individual, direct exposure data based on urine or water measures with appropriate analyses. For each study, a detailed summary of the exposure characterization, the risk-of-bias rating, and the basis for the rating for exposure characterization are provided in *Appendix E*, which includes discussion of any potential exposure misclassification and the potential impact on direction and magnitude of effect size. As we detail in *Appendix E* and summarize in the *Exposure Characterization in IQ Studies* section, there were few, if any, risk-of-bias concerns regarding exposure characterization in these studies. Thirteen of the 16 cross-sectional studies utilized an exposure measure (i.e., urine or serum) that would capture all sources of exposure to fluoride. Only one of the 16 cross-sectional studies had potential for bias due to exposure misclassification, which is discussed in detail in the *Exposure Characterization in IQ Studies* section and *Appendix E*. In this study (Seraj et al. 2012), a statistically significant association between water fluoride and IQ was reported. We determined that the potential exposure misclassification would bias the results toward the null, indicating that the true association may be greater than what was observed in this study.

- **Exposure (whether exposure preceded outcome):** Note that we acknowledge in the *Results by Study Design – Cross-sectional Studies* section that, as a general study design, cross-sectional studies often do not provide sufficient information to ensure that exposure preceded outcome. However, we do not judge studies simply by study type. Each study is assessed individually for multiple factors, including if the research design and conduct inform whether exposure preceded outcome assessment, as is the case for the low risk-of-bias cross-sectional IQ studies (see below):

  "In some cases, cross-sectional studies do provide indicators of prior exposure (e.g., prevalence of dental fluorosis, limiting study populations to subjects who lived in the same area for long periods of time). Evidence that exposure occurred prior to the outcome of interest increases the confidence in results and any potential association reported in these studies. Of the 16 low risk-of-bias cross-sectional studies, 12 established that exposure preceded the outcome assessment…"

- **Selection:** [redacted] also raised the concern about potential selection bias for cross-sectional studies. We agree with [redacted] that selection bias is an important consideration in risk-of-bias evaluations. As described previously, *Appendix E* includes a detailed summary of population selection for each low risk-of-bias studies and the basis for the ratings for selection bias and exposure characterization. All 16 low risk-of-bias cross-sectional IQ studies were rated either *definitely low risk of bias* or *probably low risk of bias* due to selection bias. In addition, we edited the following text in the *Methods* section to clarify that, in addition to the three key risk-of-bias questions, the answers to the other risk-of-bias questions were considered in assessing potential bias, including selection bias.

  "The other risk-of-bias questions, including selection of study participants, were also considered and were used to identify any other risk-of-bias concerns that may

indicate serious issues with a study that could cause it to be considered high risk of bias."

- **Individual studies cited by [redacted]:**
  - Xiang et al. (2003): This study was considered low risk of bias for confounding. [redacted] is correct that the study did not statistically adjust for the three key covariates; however, as described in *Appendix E*, the key covariates were considered not a concern for confounding because authors noted that these factors were similar between the two compared villages. We did note that there was potential co-exposure to arsenic, which would likely bias the observed association toward the null due to the reporting of higher arsenic levels in the control area.
  - Xiang et al. (2011): The journal *Fluoride* says it "contains peer-reviewed scientific reports on agricultural, analytical, biochemical, biological, chemical, clinical, dental, ecological, environmental, industrial, medical, metabolic, pharmacological, synergistic, toxicological, and veterinary aspects of inorganic and organic fluorides." Therefore, we do not have reason to believe that the manuscript was not peer-reviewed.
  - Till et al. (2020): [redacted] concerns for this study fall under potential exposure misclassification. We agree that this analysis was not designed to account for fluoride exposure postweaning. We also agree with the possibility of exposure misclassification. However, there is no evidence to suggest that the potential exposure misclassification is differential based on whether a participant lived in a fluoridated or non-fluoridated area. Therefore, as described in *Appendix E*, the possibility of exposure misclassification is non-differential and is likely similar in all participants, which would likely bias the association toward the null.

3. Comment on assessment of the human studies with regard to:
   a) How findings from individual studies designated as "low risk of bias" were interpreted.
   b) How the overall set of confounders across the body of evidence from children's IQ studies was considered and presented.
   c) How the confidence rating in the body of evidence was developed and supported.

J.21: [redacted] **Comments:** See above for comments on risk of bias ratings, and concerns related to confounding and/or residual confounding.
Page 39, last full paragraph includes the sentence, "Despite these few variations, the overall evidence of an effect on IQ is apparent." This [redacted] suggests editing the word "effect" to "association" or "correlation," given that the included studies are all observational.

Response: Agree (change made)

- Edits have been made throughout the Sup02_2022_Prepublication_NTP_Monograph to use the terms 'effect,' 'association,' and 'correlation' consistently and most appropriately. For example, the sentence referenced by [redacted] has been revised and reads as follows:

"Despite these few variations, the overall evidence of an association with lower IQ is apparent."

J.22: Page 40, "Gender considerations": Is there some biological plausibility that there would be sex differences in the relationship between fluoride exposure and neurocognitive outcomes. The term "susceptibility" is used several times, but it is unclear what that means. It seems to imply a biological reason, but it is unclear whether mechanistic evidence is supportive of that (or if gender differences actually represent some sort of residual confounding).

Response: Agree (change made)

- Note that this response refers to sex considerations because we updated the language from "gender" to "sex" in the monograph in response to a comment from [redacted]. There are several reasons we considered potential sex differences in this systematic review: (1) sex has historically been considered an important potential confounder in the literature (see Table 6 in Sup02_2022_Prepublication_NTP_Monograph) (Lash et al. 2021; Gochfeld 2017); (2) sex is an important risk factor for neurodevelopmental and cognitive outcomes (Cowell and Wright 2017); and (3) potential sex-related ingestion and dietary differences are realistic in cross-sectional studies (D'Amico et al. 2020; Keller et al. 2019).
- We have added the following text to the *Sex Considerations* section to address [redacted] comment, as follows:

  "Recent literature suggests that adverse neurodevelopmental effects of early-life exposure to fluoride may differ depending on timing of exposure and sex of the exposed. In a review of the human and animal literature, Green et al. (2020) concluded that, compared with females, male offspring appear to be more sensitive to prenatal but not postnatal exposure to fluoride, with several potential sex-specific mechanisms."

J.23: Page 48-49, **Assessment** of Risk of Bias: While the studies noted as "low risk of bias" are certainly lower risk than the studies noted as "high risk of bias," it appears that the evidence base is still subject to a number of important risks, particularly related to confounding and exposure classification (i.e., are they "low risk" or "lower risk"?).

Response: Agree (edited for clarity)

- [redacted] raises a valid point on clear terminology and word choice for the terms "low" and "lower" as well as "high" and "higher" to describe risk of bias. Word choice was carefully considered and reflects input from technical experts and [redacted]. In particular, use of the term "lower" risk of bias may raise the question, "lower than what?" Given this input, the decision was made to use a clear definition of "low risk-of-bias studies" and "high risk-of-bias studies" and to describe in detail how these terms are used early on in the document in the *Quality Assessment of Individual Studies* and *Risk-of-bias Considerations for Human Studies* sections. To clarify the definition of "high risk-of-bias studies," the quoted text below was added to the *Risk-of-bias Considerations for Human* Studies section. In addition, the detailed assessments of and justifications for risk-

of-bias ratings for the key studies are provided in *Appendix E* (*Details for Low Risk-of-bias Studies*). It is also important to note that the confidence rating of moderate for the association between higher fluoride exposures and lower children's IQ reflects assessment of risk of bias across the body of evidence as one of multiple specific factors evaluated in determining the confidence rating.

"Studies could also be considered high risk of bias if rated probably high risk of bias for one key risk-of-bias question along with other concerns, including potential for selection bias and concerns with statistical methods."

J.24: Page 48-49, Assessment of Unexplained Inconsistencies: While there is some consistency in findings suggesting that increased exposure to fluoride is associated with lower IQ, many studies reported mixed results (generally reporting a mix of inverse associations and null findings). How were these mixed findings taken into consideration when evaluating unexplained inconsistencies?

Response: Agree (change made)

1) We revised the text in the *Confidence Assessment of Findings on IQ in Children* section regarding unexplained inconsistencies. The 'Unexplained inconsistencies' bullet now reads as follows:

"Unexplained inconsistencies: The direction of the association is consistent in the majority of studies, and there was no downgrade for this factor. Eighteen of the 19 low risk-of-bias studies reported associations between higher fluoride levels and lower IQ scores in children. These studies were conducted in 5 different countries on more than 7,000 children from 15 different study populations. There is consistency in the direction of the association across prospective and cross-sectional study designs. There is also consistency in the direction of the association across studies using different fluoride exposure measures, including urinary and drinking water fluoride. The one study that did not observe an association did not provide results in a comparable manner and therefore this body of evidence is not considered to have unexplained inconsistencies."

- As we further explain in the *Summary of Key Findings for Low Risk-of-bias Children's IQ Studies* section of the Sup02_2022_Prepublication_NTP_Monograph, *"Although some studies that conducted multiple analyses observed within-study variations in results (e.g., differences between subsets of IQ tests), these variations were unique to individual studies and did not detract from the overall consistency in the findings that higher fluoride is associated with lower IQ scores."*

J.25:

4. ***NTP concludes a rating of moderate confidence in the body of evidence for lower IQ in children associated with fluoride exposure***.

☐ Agree:

X Agree in principle with the exception(s) listed below: [redacted] agree in principle with the direction of association concluded by NTP, but am uncertain that a rating of moderate is appropriate for a body of evidence comprised of mostly cross-sectional studies, that have not considered the full range of key confounders, or may have some concerns with exposure classification.

☐ Do not agree because:

Response: Disagree (no change)

- Study type (e.g., cross-sectional, cohort, case-control) should not serve as a proxy for assessing level of confidence in a body of evidence. Instead, NTP's framework for developing a confidence rating for a body of evidence starts with an initial confidence rating that is determined by the ability of the studies to address causality as reflected in the confidence that exposure preceded and was associated with the outcome (Rooney et al. 2014). This ability, in turn, is based on four key study design features (controlled exposure, exposure prior to outcome, individual outcome data, and comparison group) (https://ntp.niehs.nih.gov/go/ohathandbook). To meet the criteria for an initial confidence rating of moderate, studies must have three of the four key features. Among the 19 low risk-of-bias studies that form the basis of this body of evidence, 15 studies (3 prospective cohort studies and 12 cross-sectional studies) have 3 of the 4 features (individual outcome data, comparison group, and exposure prior to outcome) and so support an initial confidence rating of moderate. More specifically, the 12 cross-sectional studies provide sufficient details to establish that exposure preceded the outcome assessment (e.g., by providing prevalence of dental fluorosis, limiting study populations to subjects who lived in the same fluorosis area for long periods of time), in addition to having individual outcome data and a comparison group. Although cross-sectional studies can have limitations in ensuring that exposure preceded outcome, that is not the case with the cross-sectional studies that contributed to the determination of moderate confidence in an association between fluoride exposure and lower IQ in children. The three prospective cohort studies also provide individual outcome data, include a comparison group, and demonstrate that exposure preceded outcome, and so support initial confidence rating of moderate. Finally, the consistency of results across the body of evidence, including both study designs, and after consideration of all of the GRADE-based factors that may increase or decrease confidence, support the final confidence rating of moderate.
- Please see previous responses in Sections A and B.I.2.of this document that explain why we disagree that serious concerns remain about potential confounding and exposure misclassification among the low risk-of-bias studies that would impact our confidence in the literature.

II. Fluoride exposure and non-IQ neurodevelopmental or cognitive effects in children

J.26:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on non-IQ neurodevelopmental or cognitive effects in children.

[redacted] **Comments**: Yes, the approach used to search for and select studies was appropriate.

Response: No change requested

- No response necessary.

J.27:

2. Comment on whether the approach used to assess risk of bias for studies in children on non-IQ neurodevelopmental or cognitive effects was clearly described and appropriately applied.

[redacted] **Comments:** In general, this body of evidence has fewer apparent concerns with confounding or residual confounding. However, there are likely some concerns with exposure assessment/classification, given that some of the longitudinal studies assessed maternal fluoride status and neurocognitive outcomes later in childhood, without accounting for fluoride exposure of the child during the period of follow-up.

Response: Disagree (no change)

- We agree that, ideally, studies would account for a child's lifetime exposure; however, when considering risk of bias for exposure misclassification, in order for timing of exposure to impact the risk-of-bias rating, the exposure assessment would have to take place at a time that would not be appropriate for the outcome assessed (e.g., measurement of exposure after outcome). Evaluating exposure during a specific life stage prior to the outcome assessment does not indicate any misclassification for that specific life stage.

J.28:

3. Comment on assessment of the human studies with regard to:
   a) How findings from individual "low risk of bias" studies were interpreted.
   b) How the confidence rating in the body of evidence was developed and supported.

[redacted] **Comments:** The findings were interpreted correcting and a confidence rating of low seems an appropriate assessment.

Response: No change requested

- No response necessary.

J.29:

4. ***The NTP concludes a rating of low confidence in the body of evidence for decreases in measures of other neurodevelopmental or cognitive effects in children associated with fluoride exposure***.

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

III. Fluoride exposure and cognitive effects in adults

J.30:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on cognitive effects in adults.

[redacted] **Comments:** Yes, the approach used to search for and select studies was appropriate.

Response: No change requested

- No response necessary.

J.31:

2. Comment on whether the approach used to assess risk of bias for studies in adults on cognitive effects was clearly described and appropriately applied.

[redacted] **Comments:** Yes, the approach used to assess risk of bias was clearly described and generally appropriate.

Though, it is unclear whether these studies adequately captured a critical fluoride exposure window likely to impact neurocognitive health (i.e., does fluoride exposure in older adulthood impact neurocognitive health?) For example, lifelong fluoride exposure, and/or fluoride exposure at different lifestages that may be more critical to neurocognitive development, were not captured in these cross-sectional studies. Thus, it raises questions as to whether these cross-sectional studies are truly "low risk of bias," or are "lower" risk of bias than others.

Response: Agree (edited for clarity)

- We agree with [redacted] that none of the studies evaluated differential fluoride exposures in adults with adequate adjustment for earlier life exposures. The available body of evidence does not provide sufficient information to draw a conclusion on the critical period of exposure assessment/classification. While we agree that questions about critical periods of exposure and duration of exposure are important to understanding relative hazards from fluoride to neurocognitive health, these would not be addressed in the risk-of-bias evaluation of individual studies and are instead a limitation of the evidence base. If there had been more data and greater confidence in the body of evidence for studies in adults, the ability of the studies to address questions of lifelong exposure or critical exposure windows would have been added to the *Discussion* section. We consider the approach for the risk-of-bias evaluation to be appropriate for assessing the quality of the studies and conclude that an overall assessment of low confidence in an association between higher fluoride exposures and cognitive effects in adults is appropriate based on the body of evidence. In addition, the following text was added to the *Limitations of the Evidence Base* subsection of the *Discussion* of the Sup02_2022_Prepublication_NTP_Monograph to acknowledge the lack of studies to inform these questions.

  "No studies are available to evaluate lifelong exposure in adults, or fluoride exposure over a child's lifetime and neurodevelopmental or cognitive changes over time."

J.32:

3. Comment on assessment of the human studies with regard to:
   a) How findings from individual studies were interpreted.
   b) How the confidence rating in the body of evidence was developed and supported.

[redacted] **Comments:** The findings were interpreted correcting and a confidence rating of low and that the evidence is inadequate seems an appropriate assessment.

Response: No change requested

- No response necessary.

J.33:

4. ***The NTP concludes a rating of low confidence in the body of evidence for changes in cognitive effects in adults with fluoride exposure***.

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

C. Studies in non-human animals

J.34:

The NTP agrees with the comments of the NASEM committee (NASEM 2020, 2021) concerning the overall poor quality of the experimental animal database on fluoride exposure and neurodevelopmental effects, with many studies suffering from major reporting deficiencies. As indicated above, the monograph focuses on the large human epidemiology database because it directly addresses the question of whether fluoride affects human neurodevelopment. Therefore, based on the recommendations of the NASEM committee, the experimental animal section and risk of bias details have been removed from this monograph and ***the NTP concludes that the scientific evidence from experimental animal data are inadequate to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans***.

X Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the inadequate designation.

References

Cowell WJ, Wright RJ. 2017. Sex-Specific Effects of Combined Exposure to Chemical and Non-chemical Stressors on Neuroendocrine Development: a Review of Recent Findings and Putative Mechanisms. Curr Envir Health Rpt. 4: 415–425. https://doi.org/10.1007/s40572-017-0165-9

D'Amico D, Parrott MD, Greenwood CE, Ferland G, Gaudreau P, Belleville S, Laurin D, Anderson ND, Kergoat MJ, Morais JA, Presse N, Fiocco AJ. 2020. Sex differences in the relationship between dietary pattern adherence and cognitive function among older adults: findings from the NuAge study. Nutr J. 19(58) . https://doi.org/10.1186/s12937-020-00575-3

Gochfeld M. 2017. Sex Differences in Human and Animal Toxicology. Toxicol Pathol. 45(1):172-189. doi: 10.1177/0192623316677327. PMID: 27895264; PMCID: PMC5371029.

Keller KL, Kling SMR, Fuchs B, Pearce AL, Reigh NA, Masterson T, Hickok K. 2019. A Biopsychosocial Model of Sex Differences in Children's Eating Behaviors. *Nutrients*. 11(3): 682. https://doi.org/10.3390/nu11030682

Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ. Modern Epidemiology, 4th edition. Wolters Kluwer, 2021.

National Research Council (NRC). 1997. Environmental Epidemiology: Volume 2: Use of the Gray Literature and Other Data in Environmental Epidemiology. Washington (DC): National Academies Press (US); 1997. 2, Environmental-Epidemiology Studies: Their Design and Conduct. Available from: https://www.ncbi.nlm.nih.gov/books/NBK233644/

In November 2021, [redacted] received: 1) the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review,* 2) a copy of the NASEM Committee's comments on the 2020 draft NTP Monograph with NIEHS/DNTP responses (draft version of Sup01_Monograph), and 3) the [redacted] instructions. The instructions consisted of a preface, charge, instructions for the review, and a series of specific peer-review questions grouped by the following three topics: General Comments, Human Studies, and Studies in Non-Human Animals.

[redacted] were asked to provide their substantive scientific and technical comments and suggestions within the [redacted] form. In addition, they were asked whether they "Agree", "Agree in principle", or "Do not agree" with each NTP conclusion on confidence in a body of evidence.

The [redacted] instructions and specific peer-review questions are reproduced in the pages that follow in black text. [redacted] comments and responses to each question are also provided in black text starting with the words "[redacted] **comments**" in bold font. The NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

The prepublication 2022 NTP Monograph reflects changes made after consideration of the comments from the [redacted] along with all other input received through April of 2022. The prepublication 2022 NTP Monograph was subsequently sent to [redacted] for additional comments. A revised "track changes" version of the monograph was developed in September 2022 titled the "DocMon_Track_Changes_2022_NTP_Monograph." The following bullets describe how edits are documented in the track changes version of the monograph in response to [redacted] comments and [redacted] comments:

- [redacted] For comments related to DocG_Monograph, DocH_Monograph, DocI_Monograph, DocJ_Monograph, and DocK_Monograph:
    - Edits are marked with a comment bubble in the DocMon_Track_Changes_2022_NTP_Monograph that identifies the text in question and briefly describes any revisions.
    - The comment bubble contains the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response (e.g., comments made in response to this [redacted] would be marked "*see DocK_Monograph for detailed response*").
- [redacted] For comments DocA1_Monograph, DocA2_Monograph, DocB1_Monograph; DocB2_Monograph, and DocC_Monograph through DocF_Monograph:
    - Edits are marked in track changes format in the DocMon_Track_Changes_2022_NTP_Monograph.
    - A comment bubble has been added to the text in question containing the exact text of the [redacted] Comment.
    - The comment bubble also provides a reference to the specific response to comments document with the detailed NIEHS/DNTP response.

Preliminary comments on the draft NTP monograph prepared by the peer revie [redacted] are noted below. These preliminary comments are not binding and should not be construed to represent NTP determination or policy.

National Toxicology Program
NTP Monograph Letter Peer-Review Panel
Draft NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review

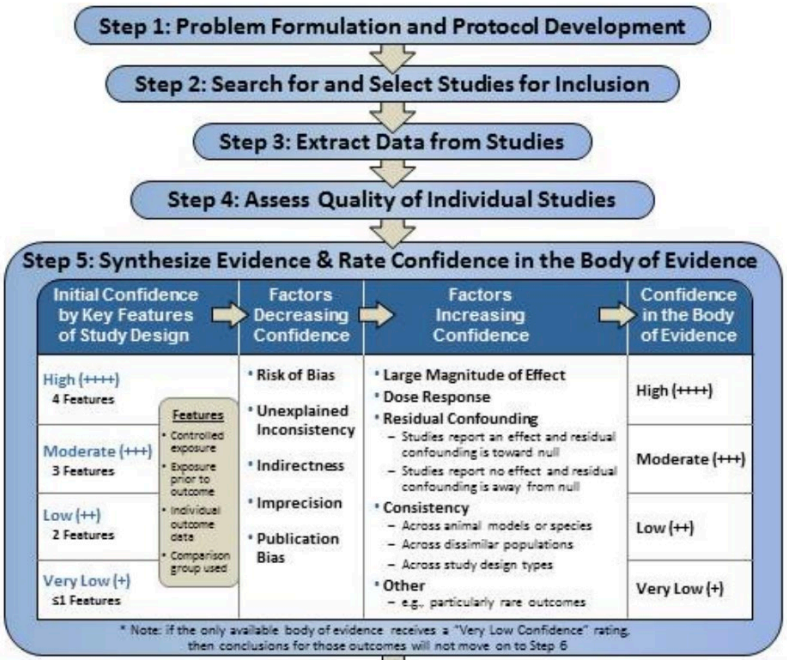
National Institute of Environmental Health Sciences
Research Triangle Park, NC

February 11, 2022

Fluoride State of the Science Document Review Form

[redacted]

Preface:
The objective of this evaluation was to conduct a systematic review of the published literature regarding the potential for exposure to fluoride to affect neurodevelopment and cognition in humans. The evaluation presented in the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* represents a comprehensive and current assessment. The methods used are from the Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration, which presents a seven-step framework for systematic review and evidence integration. Please note: this evaluation stops at step 5 of the systematic review process and does not proceed to step 6 to translate the confidence rating for the body of evidence into a level of evidence for health effects (see Figure 2 from the handbook).



Charge:

(1) Comment on the technical accuracy and whether the draft *NTP Monograph on State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* is clearly stated, and objectively presented.

(2) Determine whether the scientific evidence supports the NTP's confidence ratings for the bodies of evidence regarding neurodevelopmental and cognitive health effects associated with exposure to fluoride.

Instructions for Review:
All materials for this review are available in the Electronic Council Book (ECB). You will receive the specific URL and a password for accessing the ECB.

This evaluation identified 159 human studies relevant for assessing neurological health effects of exposure to fluoride; however, many studies included only secondary outcomes (e.g., 55 studies of thyroid hormones that were investigated as a potential mechanism). The scientific evidence in children and adults was evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood. Several studies evaluated learning and memory (n = 8 studies) or other cognitive developmental effects (e.g., total neurobehavioral scores and total mental capacity index in children, cognitive impairment in adults; n = 14 studies). Sixty-six human studies investigated IQ in children. Nineteen of the 66 IQ studies were determined to have low potential for bias and therefore, were categorized as "low risk of bias". Please give special attention to our assessment of these 19 studies.

- The 19 studies are available as PDFs and organized alphabetically in a folder on the ECB.
- All other studies are provided in the Health Assessment Workspace Collaborative, or HAWC database under the "studies list" tab, also organized alphabetically. You will also be provided a username and password for HAWC that will give you [redacted] permissions to access the PDFs in HAWC along with visualizations and other study information for this project at the following link (https://hawcproject.org/study/assessment/405/).

Please provide your substantive scientific and technical comments and suggestions within this [redacted] form. Identify and provide the rationale or scientific support for proposed changes or suggestions where possible.

If necessary, you can also provide additional editorial comments and recommendations for improving the report outside your specific charge questions (this form) within the draft report itself. Please note that only those comments included on the [redacted] form will be considered part of NTP's peer review report.

A. General Comments

K.1:

1. Please comment on whether the scientific information presented in the draft monograph, including presentation of data in tables and figures, is technically correct, and clearly and objectively presented. Please suggest any improvements.

   [redacted] **Comments:** The data was clearly presented and put together. In particular, the tables and figures are helpful. A few particular suggestions for the tables are mentioned in sections below.

   Response: No change requested

   - No response necessary because [redacted] feedback on tables is addressed where detailed suggestions are presented below.

2. Please identify any information that should be added or deleted.

   K.2: [redacted] **Comments:** It might be useful to have reminder, or reference back to the section in the text where the risk of bias information for human and animal studies is described in the methods (page 18), prior to presentation of the low risk of bias results for humans (page 28) and animals (page 67).

   Response: Agree (change made)

   - At the beginning of the *Low Risk-of-bias IQ Studies* section, we added the parenthetical text in the quote below to refer readers back to the *Methods* section that describes the risk-of-bias assessment for human studies; however, we determined that a similar reference back to risk-of-bias methods would be less helpful for the *Animal Learning and Memory Data* section, as the animal section does not discuss animal studies in terms of risk-of-bias status.

     "Nineteen studies (3 longitudinal prospective cohort and 16 cross-sectional studies) with low potential for bias evaluated the association between fluoride exposure and IQ in children (see Quality Assessment of Individual Studies section for methods on determining which studies pose low risk of bias)."

   K.3: [redacted] **Comments:** Additionally, it might be helpful to identify a limited set of confounder as required for evaluation. For example, those included in Figure 6 do not include all described in Table 6, and in fact [redacted] not present an important one: parental educational attainment.

   Response: Agree (edited for clarity)

   - Age, sex, and socioeconomic status (SES) are identified as the limited set of key covariates/potential confounders in the *Risk-of-bias Considerations for Human Studies* section. Each of these covariates had to be addressed in any human study of fluoride and cognitive neurodevelopmental health effects to be considered as low risk of bias for confounding. Other covariates may be considered important potential confounders depending on the specific study population and/or outcome assessed. We note that maternal education is listed in this section as a measure of SES. To provide further clarity that parental education is captured under SES in Figure 6, we added a footnote to Figure 6 that states, "*Covariates considered measures of SES include SES scaled scores, household/family income, child education, caretaker/parental education, and occupation/employment*."

B. Human studies

I. **Fluoride exposure and children's IQ**

K.4:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on measures of IQ in children.

[redacted] **Comments:** This monograph is clearly written and nicely uses tables and figures to display the search criteria and key information points. Furthermore, the level of detail in the methods provides an excellent path forward for understanding exact terms and criteria implemented.

Response: No change requested

- No response necessary.

K.5:

2. Comment on whether the approach used to assess risk of bias was clearly described and appropriately applied to the set of studies designated as "low risk of bias."

[redacted] **Comments:** As stated above, the methods section for this monograph is exemplary. Application of the criteria was clear and appropriate.

Response: No change requested

- No response necessary.

K.6:

3. Comment on assessment of the human studies with regard to:
How findings from individual studies designated as "low risk of bias" were interpreted.
How the overall set of confounders across the body of evidence from children's IQ studies was considered and presented.

[redacted] **Comments:** See comment below on parental IQ

Response: No change requested

- No response necessary because the question of parental IQ is addressed under [redacted] comment for question 4 below.

K.7:

How the confidence rating in the body of evidence was developed and supported.

4. ***NTP concludes a rating of moderate confidence in the body of evidence for lower IQ in children associated with fluoride exposure***.

☐ Agree
☒ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

[redacted] **Comments:** [redacted] am concerned that only one study in the low risk of bias category included parent IQ as a potential confounder. Given the known heritability of IQ, and established connections between socio-economic status (SES) and performance

testing, and SES and educational attainment, substantial confounding may be present. Of note, Figure 6 does not include parental educational attainment, which may be a proxy for an IQ related measure (or those via inherited variation) when a direct measure of IQ was not collected, though is mentioned in Table 6. Additionally, it is notable that many of the low risk of bias studies are cross-sectional and provide limited information regarding temporality and timing of exposure.

Response: Agree (edited for clarity)

- [redacted] made several comments related to this question, and for the first, we agree that parental IQ is important. We also agree that educational attainment (and SES) may be proxy measures of parental IQ. Therefore, parental IQ was considered indirectly addressed if a study accounted for parental educational attainment and/or SES. Figure 6 does not specifically include parental educational attainment because it was considered as a measure of SES. For clarification, we added a footnote to Figure 6 of the Sup02_2022_Prepublication_NTP_Monograph that lists the covariates identified in the studies included in Figure 6 that were considered measures of SES as follows:

  "Covariates considered measures of SES include SES scaled scores, household/family income, child education, caretaker/parental education, and occupation/employment."

- We disagree that many of the low risk-of-bias cross-sectional studies provide limited information regarding temporality and timing of exposure for determining the initial confidence rating. Most of the low risk-of-bias cross-sectional studies (12 of 16) did provide indicators of prior exposure (e.g., by providing prevalence of dental fluorosis, limiting study populations to subjects who lived in the same area for long periods of time). Evidence that exposure occurred prior to the outcome of interest increases the confidence in results (see Figure 1 in Sup02_2022_Prepublication_NTP_Monograph) and any potential association reported in these studies.

II. **Fluoride exposure and non-IQ neurodevelopmental or cognitive effects in children**

K.8:

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on non-IQ neurodevelopmental or cognitive effects in children.

[redacted] **Comments:** The methods were clearly described.

Response: No change requested

- No response necessary.

K.9:

2. Comment on whether the approach used to assess risk of bias for studies in children on non-IQ neurodevelopmental or cognitive effects was clearly described and appropriately applied.

[redacted] **Comments:** The approach for risk of bias was clearly described. The report might benefit from additional explanation of performance-based vs. reporter-based metrics of non-IQ outcomes, relative clinical importance, and interpretation.

Response: Agree (no change)

- *Appendix E* in the Sup02_2022_Prepublication_NTP_Monograph describes study-specific considerations for the risk-of-bias evaluation, including whether outcomes were assessed based on test performance or reporting, and the basis for the risk-of-bias rating. The following two excerpts from *Appendix E* illustrate how reporter-based and performance-based metrics of non-IQ outcomes were considered, respectively, in the risk-of-bias ratings and explanations.

  Excerpt 1

  "Outcome:

  Rating: Probably high risk of bias (–)

  Summary: The primary outcome variable, diagnosis of a learning disability by a health professional, was based on a single item from a household survey asked to all respondents: "Do you have a learning disability?" Answer options were: "yes," "no," "don't know," or the participant refused to answer. For Cycle 2, those who indicated having a learning disability were also asked what kind, with the answer options of: "ADD," "ADHD," "dyslexia," or "other." This question was omitted in Cycle 3, and the reason for omission was not described. Parents or guardians answered all questions for children aged 3–11 years, while children 12 years and older answered questions themselves. The self-reporting of a learning disability did not appear to have been confirmed by medical records or a health professional (– for methods based on self-report of diagnosis by a health care professional; also, in Cycle 3, no specific disabilities were described). Blinding was not a concern as spot urine samples were sent to a separate lab, and self-reports would not have knowledge of their urine or tap water exposure level (+ for blinding). Overall rating = –.

  Basis for rating: Probably high risk of bias based on indirect evidence that the outcome was measured using an insensitive method in the study population."

  Excerpt 2

  "Outcome:

  Rating: Definitely low risk of bias (++)

  Summary: Neurodevelopment was assessed with the Bayley Scales of Infant Development II (BSDI-II) that was noted to be reliable and valid for evaluating children from 3 months to 5 years of age. The average age of children assessed was 8 months, with a range of 3–15 months) (++ for methods). The study report stated that a trained psychologist who was blinded about the mother's fluoride exposure evaluated the infants at home (++ for blinding). Overall rating for methods and blinding = ++.

  Basis for rating: Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study

population, and that the outcome assessor was blind to participants' fluoride exposure."

3. Comment on assessment of the human studies with regard to:
How findings from individual "low risk of bias" studies were interpreted.

How the confidence rating in the body of evidence was developed and supported.

K.10:

4. ***The NTP concludes a rating of low confidence in the body of evidence for decreases in measures of other neurodevelopmental or cognitive effects in children associated with fluoride exposure***.

☒ Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

Response: No change requested

- [redacted] agreed with the low confidence rating.

Note: [redacted] only provided comments on the questions above. [redacted] indicated that they had reviewed the Sup03_2021_draft_NTP_Monograph and provided comments under Question A. "General Comments" and Sections I (Fluoride exposure and children's IQ) and II (Fluoride exposure and non-IQ neurodevelopmental or cognitive effects in children) of Question B. "Human Studies". However, they did not have time to provide comments on the remaining sections.

III. **Fluoride exposure and cognitive effects in adults**

1. Comment on whether the approach described in the methods to search for and select human studies on neurodevelopmental or cognitive function effects associated with fluoride exposure was appropriate for evaluating potential effects of fluoride exposure on cognitive effects in adults.

2. Comment on whether the approach used to assess risk of bias for studies in adults on cognitive effects was clearly described and appropriately applied.

3. Comment on assessment of the human studies with regard to:
How findings from individual studies were interpreted.

How the confidence rating in the body of evidence was developed and supported.

4. ***The NTP concludes a rating of low confidence in the body of evidence for changes in cognitive effects in adults with fluoride exposure***.

☐ Agree
☐ Agree in principle with the exception(s) listed below:
☐ Do not agree because:

C. Studies in non-human animals

The NTP agrees with the comments of the NASEM committee (NASEM 2020, 2021) concerning the overall poor quality of the experimental animal database on fluoride exposure and

neurodevelopmental effects, with many studies suffering from major reporting deficiencies. As indicated above, the monograph focuses on the large human epidemiology database because it directly addresses the question of whether fluoride affects human neurodevelopment. Therefore, based on the recommendations of the NASEM committee, the experimental animal section and risk of bias details have been removed from this monograph and ***the NTP concludes that the scientific evidence from experimental animal data are inadequate to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans***.

- ☐ Agree
- ☐ Agree in principle with the exception(s) listed below:
- ☐ Do not agree because:

Title: Association between fluoride exposure and children's intelligence: A systematic review and meta-analysis

Authors:

Kyla W. Taylor[1], PhD (corresponding author)

Sorina E. Eftim[2], PhD

Christopher A. Sibrizzi[2], MPH

Robyn B. Blain[2], PhD

Kristen Magnuson[2], MESM

Pamela A. Hartman[2], MEM

Andrew A. Rooney[1], PhD

John R. Bucher[1], PhD

Affiliations: [1]Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA; [2]ICF, Fairfax, VA, USA.

Corresponding author contact:

Kyla W. Taylor, PhD

530 Davis Drive, Morrisville NC 27719

Email: kyla.taylor@nih.gov

Phone: 919-530-0061

Abstract word count: 325

Manuscript word count: 3,751

Abstract

IMPORTANCE Water and water-based beverages are the main source of systemic fluoride intake; however, an individual's total exposure to fluoride also reflects contributions from other sources such as food, dental products, industrial emissions, and some pharmaceuticals. Previous meta-analyses suggest that exposure to fluoride adversely affects children's intelligence.

OBJECTIVE To perform a systematic review and meta-analysis to investigate associations between fluoride exposure and children's intelligence.

DATA SOURCES BIOSIS, EMBASE, PsychINFO, PubMed, Scopus, Web of Science, CNKI, and Wanfang databases were searched for relevant literature published up to November 2021.

STUDY SELECTION Inclusion criteria were assessment of cognitive outcomes, fluoride exposure, and statistical data on effect size.

DATA EXTRACTION AND SYNTHESIS Meta-analysis of Observational Studies in Epidemiology (MOOSE) reporting guidelines were followed for data extraction. The quality of individual studies was evaluated for risk of bias using a standardized tool. Pooled standardized mean differences (SMDs) and regression coefficients were estimated with random-effects models.

MAIN OUTCOMES AND MEASURES Children's intelligence levels reflected by intelligence quotient (IQ) scores.

RESULTS The meta-analysis of 55 studies (N = 18,845 children) with group-level exposures found that, when compared to children exposed to lower fluoride levels, children exposed to higher fluoride levels had lower mean IQ scores (pooled SMD: −0.46; 95% CI: −0.55, −0.37; p-value < 0.001). There was a dose-response relationship between group-level fluoride exposure measures and mean children's IQ. The meta-analysis of studies that reported individual-level measures of fluoride and children's IQ scores found a decrease of 1.81 points (95% CI: −2.80, −0.81; p-value < 0.001) per 1-mg/L increase in urinary

Commented [I1]: See Doc01_Meta-analysis, 1.B., page 1

Commented [I2]: See Doc02_Meta-analysis, 2.B., page 1

Commented [I3]: See Doc08_Meta-analysis, 8.G., page 5 and 6

Commented [I4]: See Doc01_Meta-analysis, 1.C., page 1

Commented [I5]: See Doc08_Meta-analysis, 8.K., page 7 and 8

Commented [I6]: See Doc01_Meta-analysis, 1.D., page 2

fluoride. Overall, the direction of the association was robust to stratification by study quality (high *vs*. low risk of bias), sex, age group, outcome assessment, study location, exposure timing, and exposure metric.

CONCLUSIONS AND RELEVANCE This meta-analysis confirms results of previous meta-analyses and extends them by including newer, more precise studies with individual-level exposure measures. The consistency of the data supports an inverse association between fluoride exposure and children's IQ.

Commented [I7]: See Doc08_Meta-analysis, 8.Q., page 10

Commented [I8]: See Doc06a_Meta-analysis, 6a.A., page 1

Introduction

Fluoride from natural sources occurs in some community water systems and, in the United States and some other countries, fluoride is added to public drinking water systems for the prevention of tooth decay. Water and water-based beverages are the main source of systemic fluoride intake; however, an individual's total exposure also reflects contributions from fluoride in other sources such as food, dental products, industrial emissions, and some pharmaceuticals.[1] Accumulating evidence suggests that fluoride exposure may affect brain development. A 2006 report from the National Research Council (NRC) concluded that high levels of naturally occurring fluoride in drinking water may be of concern for neurotoxic effects.[2] This report was largely based on studies from endemic fluorosis areas in China that had limitations in study design or methods (e.g., high risk of bias). Following the NRC review, more evidence has emerged in studies from India, Iran, Pakistan, New Zealand, Spain, and Canada (**Figure 1**). Two previous meta-analyses[3, 4] found an association between high fluoride exposure and lower children's IQ; however, many of the studies in these meta-analyses lacked the information necessary to evaluate study quality and all used group-level estimates of fluoride exposure. Since the most recent meta-analysis,[4] eleven new studies on exposure to fluoride and children's IQ have been published, including three prospective North American birth cohort studies[5-7] that used individual-level measures of maternal and children's urinary fluoride.

To incorporate this newer evidence, and to complement a larger systematic review[8] that concluded there is moderate confidence in the evidence of an inverse association between fluoride exposure and children's IQ, we conducted a meta-analysis of studies that provided group-and individual-level fluoride exposure measurements in relation to children's IQ scores.

Methods

The search, selection, extraction, and risk-of-bias evaluation of studies for this meta-analysis were part of a larger systematic review.[8] Brief methods are outlined below with detailed methods available in the protocol[9] and the **Supplemental Materials**.

Commented [I9]: See Doc03_Meta-analysis, 3.B. (page 1) and 3.C. (page 1, 2, and 3)

Commented [I10]: See Doc08_Meta-analysis, 8.F., page 5

Commented [I11]: See Doc01_Meta-analysis, 1.F and 1.G., page 3

Commented [I12]: See Doc01_Meta-analysis, 1.F., page 3

Commented [I13]: See Doc02_Meta-analysis, 2.C., page 1 and 2

Commented [I14]: See Doc02_Meta-analysis, 2.Q., page 5 and 6

Commented [I15]: See Doc05_Meta-analysis, 5.H., page 8 and 9

Commented [I16]: See Doc05_Meta-analysis, 5.H., page 8 and 9

Commented [I17]: See Doc02_Meta-analysis, 2.A. (page 1) and 2.Q. (page 5 and 6)

Systematic literature review

Literature searches were conducted in BIOSIS, EMBASE, PsychINFO, PubMed, Scopus, Web of Science, CNKI, and Wanfang databases through November 2021, without language restrictions. Search strategies are available in the protocol.[9]

Commented [I18]: See Doc01_Meta-analysis, 2.D., page 2

Study selection

To be eligible for inclusion, individual study publications had to satisfy review eligibility criteria outlined in the protocol.[9] References retrieved from the literature search were independently screened by two reviewers by title and abstract followed by full-text review. Studies that estimated the association between exposure to fluoride (based on environmental measures or biomonitoring data, reported as either individual-level or group-level measurements) and a quantitative measure of children's intelligence were included. Studies that did not report quantitative effect estimates (mean outcome measures or regression coefficients), measures of variability (95% confidence intervals [CIs], standard errors [SEs], or standard deviations [SDs]), or numbers of participants were excluded. Studies with missing measures of variability but with reported p-values for differences were included, and SDs were calculated using the approach in the Cochrane Handbook for Systematic Reviews.[10] To avoid sample overrepresentation, if the same cohort was followed at multiple timepoints resulting in multiple study publications,[11, 12] only the study publication that included the largest number of participants was included in this meta-analysis (see **eTable 1**).

Commented [I19]: See Doc02_Meta-analysis, 2.E., page 2

Commented [I20]: See Doc05_Meta-analysis, 5.L., page 10

Data extraction

Data were collected from included studies by one extractor and verified by a second extractor. Data were extracted in Health Assessment Workspace Collaborative (HAWC), an open source, web-based application for data extraction elements listed in the protocol. Data extraction results for included studies are publicly available and downloadable (https://hawcproject.org/assessment/405/).

Quality assessment: Risk-of-bias evaluation

Quality of individual studies, also called "risk of bias," was assessed using the National Toxicology Program's Office of Health Assessment and Translation approach.[13] Studies were independently evaluated by two trained assessors who answered risk-of-bias questions following prespecified criteria detailed in the protocol.[9] Risk-of-bias questions concerning confounding, exposure characterization, and outcome assessment were considered key. If not addressed appropriately, these questions were thought to have the greatest potential impact on the results.[9] The other risk-of-bias questions were used to identify other concerns that may indicate serious risk-of-bias issues (e.g., selection bias, statistical analysis). No study was excluded from the meta-analysis based on concerns for risk of bias; however, subgroup analyses were conducted with and without high risk-of-bias studies (i.e., studies rated "probably high" risk of bias for at least two key risk-of-bias questions or "definitely high" for any single question) to assess their impact on the results.

Commented [I21]: See Doc01_Meta-analysis, 1.H., page 3

Commented [I22]: See Doc02_Meta-analysis, 2.F., page 2 and 3

Commented [I23]: See Doc06a_Meta-analysis, 6a.B., page 2

Statistical analysis

We conducted the following analyses, planned *a priori* in the protocol: (1) *a mean-effects meta-analysis*, (2) *a dose-response mean-effects meta-analysis*, and (3) *a regression slopes meta-analysis.* We also conducted several subgroup and sensitivity analyses.

The *mean-effects meta-analysis* included studies that reported mean IQ scores and group-level exposures for at least one exposed and one reference group. The effect estimates in the primary *mean-effects meta-analysis* were the standardized mean differences (SMDs) for heteroscedastic population variances.[14-16] The SMDs were calculated from the difference in mean IQ scores between an exposed group and a reference group. If mean IQ scores were reported for multiple exposure groups within a single study, the highest exposure group was considered the exposed group and the lowest exposure group was considered the reference group. A sensitivity analysis was performed to evaluate the impact of all exposure groups combined compared to a reference group (see additional details on the approach, effect estimation, and study selection in the **Supplemental Materials**). Predefined subgroup analyses

Commented [I24]: See Doc06b_Meta-analysis, 6b.N., page 12 and 13

Commented [EAM25]: See Doc01_Meta-analysis, 1.J., page 4.
Note: Current language reflects revisions to the earlier version of document.

were stratified by risk of bias (high or low), study location (e.g., country), outcome assessment, exposure matrix (e.g., urinary fluoride or water fluoride concentrations), sex, and age group. To further evaluate potential sources of heterogeneity, we conducted meta-regression analyses using mean age in years (from the age range reported in each study) and year of publication in each study.

To determine whether the data support an exposure-response relationship, we conducted a *dose-response mean-effects meta-analysis.* This analysis included studies from the *mean-effects meta-analysis* that reported fluoride exposure levels and used a one-step approach as described in the protocol.[9, 17, 18] This approach uses linear mixed models to analyze all available mean effect estimates for the reference group and one or more exposure group and estimates a pooled dose-response curve using a restricted maximum likelihood estimation method. Model comparison was based on the maximum likelihood Akaike information criterion (AIC).[19] We also examined whether there was a dose-response relationship at lower exposure levels that corresponded with the U.S. Environmental Protection Agency drinking water standards[20] and World Health Organization drinking water guidelines[21] (details provided in the **Supplemental Materials**).

The *regression slopes meta-analysis* included studies that reported regression slopes to estimate associations between individual-level fluoride exposure and children's IQ. The primary regression slopes meta-analysis used regression slopes from models that adjusted for potential confounders. If results from multiple models were reported within a single study, either the most adjusted results or the main model results as presented by the study authors were selected. The study outcomes were evaluated with respect to a 1-mg/L unit increase in water or urinary fluoride, or 1-mg/day fluoride intake.

Data from individual studies were pooled using a random-effects model.[22] Heterogeneity was assessed by Cochran's Q test[23] and the I^2 statistic.[24] Forest plots were used to display results and to examine possible heterogeneity between studies. Potential publication bias was assessed by developing funnel plots and performing Egger regression on the estimates of effect size.[25-27] If publication bias was

Commented [I26]: See Doc02_Meta-analysis, 2.H., page 3

Commented [I27]: See Doc01_Meta-analysis, 1.E. (page 2) and 1.K. (page 4)

Commented [I28]: See Doc02_Meta-analysis, 2.I., page 3 and 4

Commented [I29]: See Doc06a_Meta-analysis, 6b.R., page 15 and 16

Commented [I30]: See Doc05_Meta-analysis, 5.J., page 9 and 10

present, trim-and-fill methods[28, 29] were used to estimate the number of missing studies and to predict the impact of the hypothetical "missing" studies on the pooled effect estimate. Subgroup analyses were performed to investigate sources of heterogeneity. Subgroup analyses were stratified by risk of bias (high or low), study location (e.g., country), outcome assessment, exposure matrix (e.g., urinary fluoride or water fluoride concentrations), pre- or post-natal exposure, and sex.

Statistical analyses were performed using the software STATA version 17.0[30] with the *combine, meta esize*, *meta set*, *meta summarize*, *drmeta*, *meta funnel*, *meta bias*, *meta trimfill* and *metareg* packages.[31]

Results

Study sample

Results of the study identification process are provided in **eFigure 1.** Characteristics of the 60 publications included in the meta-analysis are shown in **Table 1** (see **eTable 1** for list of excluded publications). A total of 55 publications reported mean IQ scores for group-level exposures. Eleven publications reported regression slopes for individual-level exposures based on urinary or water fluoride concentrations.[5-7, 11, 12, 32-37] Additional details on study characteristics are provided in the **Supplemental Materials**. Results from risk-of-bias evaluations are presented in **eFigure 2a** and **eFigure 2b**. Study-specific effect estimates used in the meta-analysis are presented in **eTable 2**.

Mean-effects meta-analysis

The meta-analysis of 55 studies (45 high risk-of-bias studies and 10 low risk-of-bias studies) that provided mean IQ scores shows that, when compared to children exposed to lower levels of fluoride, children exposed to higher fluoride levels had statistically significantly lower IQ scores (random-effects pooled SMD, −0.46; 95% CI: −0.55, −0.37; p-value < 0.001) (**Table 2, Figure 2**). There was evidence of high heterogeneity (I^2 = 87%, p-value < 0.001; **Table 2**) and publication bias (funnel plot and Egger's p-value < 0.001, Begg's p-value = 0.031; **eFigures 3 and 4**). Adjusting for possible publication bias

Commented [I31]: See Doc05_Meta-analysis, 5.I., page 9

Commented [I32]: See Doc06b_Meta-analysis_6b.V., page 18 and 19

Commented [I33]: See Doc02_Meta-analysis, 2.N., page 4

Commented [I34]: See Doc02_Meta-Analysis, 2.J. (page 4) and Doc08_Meta-analysis, 8.M. (page 8 and 9)
Note: Changes in study numbers from reviewer text reflects updated literature search.

Commented [I35]: See Doc02_Meta-analysis, 2.K., page 4

through trim-and-fill analysis suggested the imputation of seven additional studies to the right side, with an adjusted pooled SMD of – 0.36 (95% CI: −0.46, −0.26) (**eFigures 5 and 6**). The pattern of results across the 55 studies was consistent; 52 (95%) reported an inverse association with SMDs ranging from −5.34 (95% CI: −6.34, −4.34) to −0.04 (95% CI: −0.45, 0.36) (**Figure 2**). The(95% CI: −0.19, 0.21),[6] 0.01 (95% CI: −0.19, 0.22),[38] and 0.13 (95% CI: −0.16, 0.42).[5] Three studies[39, 40, 41 [translated in Li et al. 2008b]] lacked clear descriptions of their intelligence assessment methods; however, sensitivity analyses did not reveal substantial changes in the pooled SMD estimate when these studies were excluded or when a study[43] that reported the cognitive subset of evaluations using Bayley and McCarthy tests was included (**eTable 3**).

Among the low risk-of-bias studies (n = 10), [5, 6, 11, 32, 33, 36, 44-47] the random-effects pooled SMD was −0.22 (95% CI: −0.39, −0.05; p-value = 0.011) with high heterogeneity (I^2 = 83%) (**Table 2 and eFigure 7**). There was no evidence of publication bias (funnel plot and Egger's p-value = 0.93; **eFigures 8 and 9**). Among the high risk-of-bias studies (n = 45), the random-effects pooled SMD was −0.52 (95% CI: −0.63, −0.42; p-value < 0.001) with high heterogeneity (I^2 = 86%,) (**Table 2 and eFigure 7**). There was evidence of publication bias among the high risk-of-bias studies (funnel plot and Egger's p-value < 0.001; **eFigures 8 and 9**); adjusting for possible publication bias through trim-and-fill analysis supports the results with an adjusted pooled SMD estimate of −0.37 (95% CI: −0.48, −0.25) (**eFigures 10 and 11**). Subgroup analyses by sex, age group, study location, outcome assessment type, and exposure assessment type further support the consistent and robust pattern of an inverse association between fluoride exposure and children's IQ (**Table 2, eFigures 12-16**). The subgroup and meta-regression analyses did not explain a large amount of the overall heterogeneity; however, the degree of heterogeneity was lower We also examined whether there was a dose-response relationship at lower exposure levels that corresponded with the U.S. Environmental Protection Agency drinking water standards[20] and World restricted to Iran (I^2=56%), children ages 10 and older (I^2=68%), and girls (I^2=76%) (see **Supplemental Materials**).

Commented [I36]: See Doc06b_Meta-analysis, 6b.C., page 2 and 3

Commented [I37]: See Doc05_Meta-analysis, 5.J., page 9 and 10

Commented [I38]: See Doc01_Meta-analysis, 1.I., page 3

Commented [EAM39]: See Doc02_Meta-analysis, 2.L., page 4.
Note: Changes in study numbers from review text reflects updated literature search.

Commented [I40]: See Doc06b_Meta-analysis, 6b.G., page 7and 8

The sensitivity analysis to evaluate the impact of combining all exposed groups and comparing them to the reference group did not appreciably change the effect estimates (**eTable 3**). Sensitivity analyses that removed an outlier study[39] or a study with an unspecified IQ test[41 [translated in Li et al. 2008b]] also did not appreciably change the effect estimates (**eTable 3**).

Commented [I41]: See Doc08_Meta-analysis, 8.H., page 6

Dose-response mean-effects meta-analysis

The *dose-response mean-effects meta-analysis* combining data from 29 studies with group-level fluoride measurements in drinking water (23 high risk-of-bias and 6 low risk-of-bias studies) and 18 studies with group-level mean urinary fluoride levels (9 high risk-of-bias and 9 low risk-of-bias studies) show statistically significantly lower children's IQ scores with increasing fluoride exposures. Based on the linear models, the decrease in mean SMD between exposed and reference groups is −0.15 (95% CI: −0.20, −0.11; p-value < 0.001) for drinking water fluoride levels and −0.16 (95% CI: −0.24, −0.08; p-value < 0.001) for urinary fluoride levels (**eTable 4**). Based on the AIC and likelihood ratio tests, the best model fit was achieved when quadratic or restricted cubic spline exposure levels were added to the linear models for drinking water (**eFigure 17**); the linear model was the best fit for urinary fluoride (**eFigure 18**). Given the small difference in AICs between the different models, and for ease of interpretability, the linear model results were chosen for the purposes of discussion, although results from all models are presented (**eTable 4**). The direction of the associations did not change when the exposed groups were restricted to <4 mg/L or <2 mg/L fluoride in drinking water or fluoride in urine (**eTable 4** and **eTable 5**).

Commented [I42]: See Doc08_Meta-analysis, 8.I., page 6 and 7

Regression slopes meta-analysis

The *regression slopes meta-analysis* includes ten studies with individual-level exposure measures (1 high risk-of-bias and 9 low risk-of-bias studies) (**Table 1**). Each of these studies reported urinary fluoride levels,[5-7, 11, 12, 32-37] two reported fluoride intake,[6, 7] and two reported water fluoride levels.[6, 11] Two studies[7, 12] are not included in the primary meta-analysis they had overlapping populations with already-included studies[6, 11 respectively] (see **Supplemental Materials**). Similarly, three studies reporting scores

based on Bayley assessments[43, 48, 49] were only included in sensitivity analyses (see **Supplemental Materials**).

The overall pooled effect estimate from the nine studies with individual-level urinary fluoride measures shows that a 1-mg/L increase in urinary fluoride is associated with a statistically significant lower IQ score of 1.81 points (95% CI: −2.80, −0.81; p-value < 0.001) with evidence of heterogeneity (I^2 = 77%, p-value < 0.001; **Table 3, eFigure 19**) and no indications of publication bias (**eFigures 20 and 21**). When restricted to only low risk-of-bias studies, the decrease in IQ score was 1.33 points (95% CI: −2.09, −0.57; p-value < 0.001). There was evidence of moderate heterogeneity (I^2 = 46%, p-value < 0.072; **Table 3, eFigure 22**) and no indications of publication bias. The results for fluoride intake and water fluoride levels are available in **Supplemental Materials**.

Subgroup analyses by risk of bias, sex, country, exposure type, outcome assessment type, and pre- or post-natal exposure further support the consistent and robust pattern of an inverse association between fluoride exposure and children's IQ (**Table 3, eFigures 22–27**). The observed heterogeneity in the overall effect estimate was explained by the subgroup analyses, with no significant heterogeneity remaining in analyses of low-risk-of bias studies, by sex, by country, by assessment type, and by exposure timing (**Table 3**). The sensitivity analyses including reporting scores based on Bayley assessments[43, 48, 49] showed no substantial changes in the pooled effect estimates (**eTable 6**).

Discussion

The results of this meta-analysis support a statistically significant association between higher fluoride exposure and lower children's IQ. The direction of the association was robust to stratification by risk of bias, sex, age group, timing of exposure, study location, outcome assessment type, and exposure assessment type. There is also evidence of a dose-response relationship. Although the estimated decreases in IQ may seem small, research on other neurotoxicants has shown that subtle shifts in IQ at the population level can have a profound impact on the number of people who fall within the high and low

Commented [EAM43]: See Doc02_Meta-analysis, 2.M., page 4.
Note: Changes in study numbers from reviewer text reflects updated literature search.

Commented [I44]: See Doc02_Meta-analysis, 2.O., page 5

Commented [I45]: See Doc06b_Meta-analysis, 6b.G., page 7 and 8

Commented [I46]: See Doc02_Meta-analysis, 2.P., page 5

ranges of the population's IQ distribution.[50-54] For example, a 5-point decrease in a population's IQ would nearly double the number of people classified as intellectually disabled.[55]

The results of the *mean-effects meta-analysis* are consistent with two previous meta-analyses that, when comparing children exposed to lower fluoride levels, reported statistically significantly lower IQ scores in children exposed to higher fluoride levels ($p < 0.001$) (**Table 2**). However, this meta-analysis included more recently published studies that were considered low risk of bias and studies with different exposure assessment types. We also found a statistically significant dose-response between lower children's IQ with increasing fluoride exposures as measured in both drinking water (p-value < 0.001) and urine (p-value < 0.001). Associations appeared to be non-linear for drinking water and linear for urine. The Duan et al.[4] meta-analysis reported a significant non-linear dose-response relationship above 3 ppm [3 mg/L] in water. A more recent literature review[56] did not comment on the shape of the dose-response curve; however, based on the three publications from Mexico and Canada,[5-7] the author concluded that the association between maternal urinary fluoride and children's neurotoxicity appeared to be "dose dependent."

Whereas the previously published meta-analyses only included group-level exposures, the *regression slopes meta-analysis* included nine studies with individual urinary fluoride measures, a more precise exposure measure. It also included recent North American prospective cohort studies[5-7] with maternal urinary fluoride levels comparable to those found in the United States.[57] In contrast to urinary fluoride measures, drinking water measures capture only a portion of a person's total exposure to fluoride. Consequently, relying on drinking water levels alone likely underestimates an individual's total exposure to fluoride. For community water systems that add fluoride, the Public Health Service recommends a fluoride concentration of 0.7 mg/L; however, it is important to note that there are regions of the United States where public systems and private wells contain natural fluoride concentrations of more than 2 mg/L.[58] In April 2020, the Centers for Disease Control and Prevention (CDC) estimated that community water systems supplying water with ≥2 mg/L naturally occurring fluoride served 0.31% of the U.S.

Commented [I47]: See Doc08_Meta-analysis, 8.P., page 9 and 10

Commented [I48]: See Doc08_Meta-analysis, 8.N., page 9

Commented [I49]: See Doc01_Meta-analysis, 1.N. and 1.O., page 7

Commented [I50]: See Doc01_Meta-analysis, 1.M., page 6 and 7

population (~1 million people).[59] For the purposes of reducing dental fluorosis, the CDC recommends that parents use an alternative source of water for children aged 8 years and younger and for bottle-fed infants if their primary drinking water contains greater than 2 mg/L of fluoride.[60]

Commented [I51]: See Doc03_Meta-analysis, 3.B. (page 1 and 2) and 3.C. (page 2 and 3)

Commented [I52]: See Doc01_Meta-analysis, 1.Q., page 8

Commented [I53]: See Doc02_Meta-analysis, 2.T., page 6

Strengths and Limitations

Strengths of this meta-analysis include a large body of literature and predefined systematic search and screening process, a risk-of-bias assessment of individual studies, a variety of intelligence assessment methods and exposure matrices, varying exposure levels from multiple study locations, prespecified subgroup analyses, and use of both group-level and individual-level exposure data. The direction of the association is consistent across different analytical approaches and subgroup analyses.

There are also limitations to consider. Most of the studies included in the *mean-effects* and *dose-response mean effects meta-analyses* were considered to have study design and/or methodological limitations. For example, all but three studies were cross-sectional in design. However, among the low risk-of-bias cross-sectional studies, most provided information to suggest that exposure preceded the outcome (e.g., including only children who had lived in the area since birth, or children that had dental fluorosis). In addition, subgroup analyses suggest that the association between higher fluoride exposure and lower IQ was consistent even when restricted to low risk-of-bias studies (see **Table 2** and **eFigure 7** for additional details). Although we conducted subgroup analyses by sex, only 1 of the 14 studies that reported IQ scores separately for boys and girls analyzed fluoride exposure for each sex separately.[6] This is essential for evaluating whether a differential change in IQ by sex may be related to higher susceptibility or higher exposure in that sex. With a couple exceptions, the subgroup analyses in the *mean-effects meta-analysis* did not explain a large amount of the overall heterogeneity. However, the heterogeneity in the *regression slopes meta-analysis* was explained by subgroup analyses. This suggests that the aggregate nature of the *mean-effects meta-analysis* might not be sufficiently sensitive to capture potential sources of heterogeneity, as seen possible when using studies with individual-level data in the *regression slopes meta-analysis*. However, the large number of studies included in the *mean-effects meta-*

Commented [I54]: See Doc02_Meta-analysis, 2.U., page 7

Commented [I55]: See Doc08_Meta-analysis, 8.T., page 11

Commented [I56]: See Doc03_Meta-analysis, 3.E., page 4

Commented [I57]: See Doc06b_Meta-analysis, 6b.O., page 13 and 14

analysis and the consistency in the direction of the association across the analyses make this is less of a concern.

Another limitation of the *mean-effects meta-analyses* is that exposure values are assumed to be the same for each child in an exposure group, either because the study used a community-level water fluoride measure or a median, mean, or midpoint in water or urine as the exposure value. Fluoride exposure may vary considerably depending on individual behaviors and is best captured by individual-level measures of total exposure, such as urinary fluoride measures. Because drinking water measures capture only some of a person's total exposure to fluoride, it is reasonable to assume that some children in the meta-analysis had higher exposure to fluoride and those children may have skewed the mean IQ deficits of the entire group. Urinary fluoride levels include all ingested fluoride and are considered a valid measure to estimate total fluoride exposure.[61, 62] When compared with 24-hour urine samples, spot urine samples are more prone to the influence of timing of exposure (e.g., when water was last consumed, when teeth were last brushed) and can also be affected by differences in dilution. However, correlations between urinary fluoride concentrations from 24-hour samples and spot samples adjusted for urinary dilution have been described,[63] and with one exception[35] all studies in the *regression slopes meta-analysis*, accounted for dilution.

There is inconsistency in which model is the best fit at lower exposure levels (**eTable 4** and **eTable 5**) leading to uncertainty in the shape of the dose-response curve at these levels. More individual-level data would increase our certainty in the shape of the dose-response curve at these lower exposure levels. There are also several limitations to the existing approaches for evaluating potential for publication bias. The funnel plot asymmetry is a subjective assessment and is recommended only when at least 10 studies are included in the meta-analysis.[64] Furthermore, the Egger regression test and Begg's rank tests[25-27] may suffer from inflated type I power and limited power in certain situations.[65] Finally, the small number of studies reporting slopes for association with individual-level exposure data limits the power of the *regression slopes meta-analysis*.

Commented [I58]: See Doc06b_Meta-analysis, 6b.G., page 7 and 8

Commented [I59]: See Doc08_Meta-analysis, 8.E., page 4 and 5

Commented [I60]: See Doc02_Meta-analysis, 2.S., page 6

Commented [I61]: See Doc01_Meta-analysis, 1.M., page 6 and 7

Commented [I62]: See Doc06a_Meta-analysis, 6a.G., page 5

Commented [I63]: See Doc08_Meta-analysis, 8.D., page 3 and 4

Commented [I64]: See Doc03_Meta-analysis, 3.C., page 2 and 3

Commented [I65]: See Doc05_Meta-analysis, 5.I., page 9

This meta-analysis complements a larger systematic review[8] that concluded moderate confidence in the body of evidence that fluoride exposure is associated with lower IQ in children. Confidence would be increased with additional prospective cohort studies with individual urinary fluoride measures. Studies conducted in the United States, which as of the writing of this manuscript were not available, would also be valuable.

Commented [EAM66]: See Doc08_Meta-analysis, 8.B., page 2.

Conclusions

This meta-analysis extends the findings of our larger systematic review that concluded, with moderate confidence, that higher fluoride exposure is associated with lower children's IQ. These findings are consistent with prior meta-analyses and demonstrate that the direction of the association is robust to stratifications by risk of bias, sex, age group, outcome assessment, study location, exposure timing, and exposure measurement (including both drinking water and urinary fluoride). Therefore, the consistency of the data supports an inverse association between fluoride exposure and children's IQ.

Commented [I67]: See Doc06a_Meta-analysis, 6a.C., page 2

Acknowledgements

This work was supported by the Intramural Research Program (ES103316, ES103317) at the National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health, and was performed for NIEHS under contract GS00Q14OADU417 (Order No. HHSN273201600015U). Data sets and Stata code are available upon reasonable request from study authors. We appreciate the helpful comments and input on the draft manuscript provided by Suril Mehta, DrPH, MPH; Kelly Ferguson, PhD, MPH; Allen Wilcox, MD, PhD; and Alison Mogsinger-Reif, PhD. We would like to thank Jonathan Cohen, PhD, who performed independent verification of the dose-response meta-analysis; Cynthia Lin, PhD, and Alexandra Goldstone, MPH, for independent quality control of the data; and Jeremy Frye, MSLS, for conducting the literature search and reference management.

References

1. U.S. Environmental Protection Agency (US EPA). *Fluoride: Exposure and relative source contribution analysis*. U.S. Environmental Protection Agency; 2010. 820-R-10-015. Accessed 19 August 2019. http://www.epa.gov/dwstandardsregulations/fluoride-risk-assessment-and-relative-source-contribution

2. National Research Council (NRC). *Committee on fluoride in drinking water, board on environmental studies and toxicology. Fluoride in drinking water: A scientific review of EPA's standards*. National Research Council; 2006. Accessed 19 August 2019. http://www.nap.edu/catalog/11571/fluoride-in-drinking-water-a-scientific-review-of-epas-standards

3. Choi AL, Sun G, Zhang Y, Grandjean P. Developmental fluoride neurotoxicity: A systematic review and meta-analysis. *Environ Health Perspect*. 2012;120:1362-1368.

4. Duan Q, Jiao J, Chen X, Wang X. Association between water fluoride and the level of children's intelligence: A dose-response meta-analysis. *Public Health*. 2018;154:87-97. doi:https://doi.org/10.1016/j.puhe.2017.08.013

5. Bashash M, Thomas D, Hu H, et al. Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6-12 years of age in Mexico. *Environ Health Perspect*. 2017;125(9):1-12. doi:https://doi.org/10.1289/ehp655

6. Green R, Lanphear B, Hornung R, et al. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. *JAMA Pediatr*. 2019:E1-E9.

7. Till C, Green R, Flora D, et al. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. *Environ Int*. 2020;134:105315. doi:https://doi.org/10.1016/j.envint.2019.105315

8. National Toxicology Program (NTP). *Revised draft NTP Monograph on the systematic review of fluoride exposure and neurodevelopmental and cognitive health effects*. National Toxicology Program; 2020. https://www.nationalacademies.org/event/10-19-2020/docs/DDA97C9170D1A255D69C004CEB77B698E8D005011EFB

9. National Toxicology Program (NTP). *Protocol for systematic review of effects of fluoride exposure on neurodevelopment*. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health; 2020. https://ntp.niehs.nih.gov/ntp/ohat/fluoride/ntpprotocol_revised20200916_508.pdf

10. Higgins JT, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for systematic reviews of interventions version 6.0 (updated July 2019)*. Cochrane; 2019.

11. Yu X, Chen J, Li Y, et al. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. *Environ Int*. 2018;118:116-124. doi:https://doi.org/10.1016/j.envint.2018.05.042

12. Wang M, Liu L, Li H, et al. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. *Environ Int*. 2020b;134:105229. doi:https://doi.org/10.1016/j.envint.2019.105229

13. National Toxicology Program (NTP). *OHAT risk of bias rating tool for human and animal studies*. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health; 2015.

14. Bonett DG. Confidence intervals for standardized linear contrasts of means. *Psychol Methods*. 2008;13(2):99-109. doi:https://doi.org/10.1037/1082-989x.13.2.99

15. Hedges LV, Olkin I. *Statistical methods for meta-analysis*. Academic Press; 1985.

16. Rosenthal R. Parametric measures of effect size. In: Cooper H, Hedges LV, eds. *The Handbook of Research Synthesis*. Russell Sage Foundation; 1994.

17. Crippa A, Thomas I, Orsini N. A pointwise approach to dose-response meta-analysis of aggregated data. *Int J Stat Med Res*. 2018;7(2):25-32.

18. Crippa A, Discacciati A, Bottai M, Spiegelman D, Orsini N. One-stage dose–response meta-analysis for aggregated data. *Stat Methods Med Res*. 2019;28(5):1579-1596. doi:https://doi.org/10.1177/0962280218773122

19. Müller S, Scealy JL, Welsh AH. Model selection in linear mixed models. *Statist Sci*. 2013;28(2):135-167. doi:https://doi.org/10.1214/12-STS410

20. U.S. Environmental Protection Agency (US EPA). *New fluoride risk assessment and relative source contribution documents*. Office of Water; 2011. EPA-822-F-11-001. https://www.epa.gov/sites/default/files/2019-03/documents/fluoride-risk-assess-factsheet.pdf

21. World Health Organization (WHO). *Guidelines for drinking-water quality. Fourth edition*. 2011. https://apps.who.int/iris/bitstream/handle/10665/44584/9789241548151_eng.pdf?sequence=1&isAllowed=y&ua=1

22. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-88. doi:https://doi.org/10.1016/0197-2456(86)90046-2

23. Cochran WG. The combination of estimates from different experiments. *Biometrics*. 1954;10(1):101-129. doi:https://doi.org/10.2307/3001666

24. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-560. doi:https://doi.org/10.1136/bmj.327.7414.557

25. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics*. 1994;50(4):1088-1101. doi:https://doi.org/10.2307/2533446

26. Egger M, Smith G, Schneider M, Minder C, eds. *Systematic reviews in health care: Meta-analysis in context*. BMJ Publishing Group; 2008.

27. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-634. doi:https://doi.org/10.1136/bmj.315.7109.629

28. Duval S, Tweedie R. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *J Am Stat Assoc*. 2000;95(449):89-98. doi:https://doi.org/10.1080/01621459.2000.10473905

29. Duval S, Tweedie R. Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 2000;56(2):455-463. doi:https://doi.org/10.1111/j.0006-341X.2000.00455.x

30. StataCorp. *Stata Statistical Software: Release 17*. StataCorp LP; 2021.

31. Palmer TM, Sterne JAC, eds. *Meta-analysis in Stata: An updated collection from the Stata Journal*. 2nd ed. Stata Press; 2016.

32. Ding Y, Sun H, Han H, et al. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. *J Hazard Mater*. 2011;186:1942-1946.

33. Zhang S, Zhang X, Liu H, et al. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. *Toxicol Sci*. 2015b;144:238-245.

34. Cui Y, Zhang B, Ma J, et al. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. *Ecotoxicol Environ Saf*. 2018;165:270-277. doi:https://doi.org/10.1016/j.ecoenv.2018.09.018

35. Saeed M, Rehman MYA, Farooqi A, Malik RN. Arsenic and fluoride co-exposure through drinking water and their impacts on intelligence and oxidative stress among rural school-aged children of Lahore and Kasur districts, Pakistan. *Environ Geochem Health*. Nov 9 2021;doi:https://dx.doi.org/10.1007/s10653-021-01141-4

36. Xu K, An N, Huang H, et al. Fluoride exposure and intelligence in school-age children: evidence from different windows of exposure susceptibility. *BMC Public Health*. Nov 4 2020;20(1):1657. doi:https://dx.doi.org/10.1186/s12889-020-09765-4

37. Zhao L, Yu C, Lv J, et al. Fluoride exposure, dopamine relative gene polymorphism and intelligence: A cross-sectional study in China. *Ecotoxicol Environ Safety*. Feb 2021;209:111826. doi:https://dx.doi.org/10.1016/j.ecoenv.2020.111826

38. Broadbent JM, Thomson WM, Moffitt TE, Poulton R. Community water fluoridation and intelligence response. *Am J Public Health*. 2015;105:3-4.

39. Khan SA, Singh RK, Navit S, et al. Relationship between dental fluorosis and intelligence quotient of school going children in and around Lucknow District: A cross-sectional study. *J Clin Diagn Res*. 2015;9(11):10-15. doi:https://doi.org/10.7860/JCDR/2015/15518.6726

40. Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. [High fluoride and low iodine environment and subclinical cretinism in Xinjiang]. *Endem Dis Bull*. 1991;6(2):62-67.

41. Li Y, Li X, Wei S. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. *J West Chin University Med Sci*. 1994;25(2):188-191.

42. Li Y, Li X, Wei S. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. *Fluoride*. 2008b;41:331-335.

43. Ibarluzea J, Gallastegi M, Santa-Marina L, et al. Prenatal exposure to fluoride and neuropsychological development in early childhood: 1-to 4 years old children. *Environ Res*. Oct 8 2021:112181. doi:https://dx.doi.org/10.1016/j.envres.2021.112181

44. Xiang Q, Liang Y, Chen L, et al. Effect of fluoride in drinking water on children's intelligence. *Fluoride*. 2003a;36:84-94.

45. Seraj B, Shahrabi M, Shadfar M, et al. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. *J Dent*. 2012;9:221-229.

46. Trivedi M, Sangai N, Patel R, Payak M, Vyas S. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. *Fluoride*. 2012;45(4):377-383.

47. Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. *Neurosci Lett*. 2020;729:134981. doi:https://doi.org/10.1016/j.neulet.2020.134981

48. Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, et al. In utero exposure to fluoride and cognitive development delay in infants. *Neurotoxicology*. Mar 2017;59:65-70. doi:https://dx.doi.org/10.1016/j.neuro.2016.12.011

49. Cantoral A, Tellez-Rojo MM, Malin AJ, et al. Dietary fluoride intake during pregnancy and neurodevelopment in toddlers: A prospective study in the progress cohort. *Neurotoxicology*. Dec 2021;87:86-93. doi:https://dx.doi.org/10.1016/j.neuro.2021.08.015

50. Bellinger DC. Interpretation of small effect sizes in occupational and environmental neurotoxicology: individual versus population risk. *Neurotoxicology*. 2007;28(2):245-51. doi:https://dx.doi.org/10.1016/j.neuro.2006.05.009

51. Needleman HL. What can the study of lead teach us about other toxicants? *Environ Health Perspect*. 1990;86:183-9. doi:https://dx.doi.org/10.1289/ehp.9086183

52. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 1985;14(1):32-8. doi:https://dx.doi.org/10.1093/ije/14.1.32

53. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 2001;30(3):427-32; discussion 433-4. doi:https://dx.doi.org/10.1093/ije/30.3.427

54. Weiss B. Vulnerability of children and the developing brain to neurotoxic hazards. *Environ Health Perspect*. 2000;108 Suppl 3(Suppl 3):375-81. doi:https://dx.doi.org/10.1289/ehp.00108s3375

55. Braun JM. Early-life exposure to EDCs: Role in childhood obesity and neurodevelopment. *Nature Rev Endocrin*. 2016;13(3):161-173. doi:https://dx.doi.org/10.1038/nrendo.2016.186

56. Grandjean P. Developmental fluoride neurotoxicity: an updated review. *Environ Health*. 2019;18(1):110. doi:https://doi.org/10.1186/s12940-019-0551-x

57. Abduweli Uyghurturk D, Goin DE, Martinez-Mier EA, Woodruff TJ, DenBesten PK. Maternal and fetal exposures to fluoride during mid-gestation among pregnant women in northern California. *Environ Health*. 2020;19(1):38. doi:https://doi.org/10.1186/s12940-020-00581-2

58. McMahon PB, Brown CJ, Johnson TD, Belitz K, Lindsey BD. Fluoride occurrence in United States groundwater. *Sci Total Environ*. 2020;732:139217. doi:https://doi.org/10.1016/j.scitotenv.2020.139217

59. CDC Division of Oral Health. *Personal communication*. Centers for Disease Control and Prevention; 2020. September 3, 2020.

60. Centers for Disease Control and Prevention (CDC). *Community water fluoridation FAQs: Infant formula*. Centers for Disease Control and Prevention; 2015. Accessed 10 February 2022. https://www.cdc.gov/fluoridation/faqs/infant-formula.html

61. Villa A, Anabalon M, Zohouri V, Maguire A, Franco AM, Rugg-Gunn A. Relationships between fluoride intake, urinary fluoride excretion and fluoride retention in children and adults: An analysis of available data. *Caries Res*. 2010;44(1):60-8. doi:https://dx.doi.org/10.1159/000279325

62. Watanabe M, Kono K, Orita Y, Usuda K, Takahashi Y, Yoshida Y. Influence of dietary fluoride intake on urinary fluoride concentration and evaluation of corrected levels in spot urine. *Fluoride*. 1995;28(2):61-70.

63. Zohouri FV, Swinbank CM, Maguire A, Moynihan PJ. Is the fluoride/creatinine ratio of a spot urine sample indicative of 24-h urinary fluoride? *Community Dent Oxral Epidemiol*. Apr 2006;34(2):130-8. doi:https://dx.doi.org/10.1111/j.1600-0528.2006.00269.x

64. Higgins JPT, Chandler TJ, Cumpston M, Li T, Page MJ, Welch VA, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021)*. Cochrane; 2021.

65. Lin L, Chu H. Quantifying publication bias in meta-analysis. *Biometrics*. 2018;74(3):785-794. doi:https://doi.org/10.1111/biom.12817

66. Ren D, Li K, Liu D. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. *Chin J Control Endem Dis*. 1989;4(4):251.

67. Ren D, Li K, Liu D. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. *Fluoride*. 2008;41:319-320.

68. Chen YX, Han FL, Zhoua ZL, et al. [Research on the intellectual development of children in high fluoride areas]. *Chin J Control Endem Dis*. 1991;6(Suppl):99-100.

69. Chen YX, Han FL, Zhoua ZL, et al. Research on the intellectual development of children in high fluoride areas. *Fluoride*. 2008;41:120-124.

70. Guo XC, Wang RY, Cheng CF, et al. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. *Chin J Epidemiol*. 1991;10(2):98-100.

71. Guo XC, Wang RY, Cheng CF, et al. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. *Fluoride*. 2008a;41:125-128.

72. Sun M, Li S, Wang Y, Li F. [Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis]. *J Guiyang Med Coll*. 1991;16(3):204-206.

73. An J, Mei S, Liu A, Fu Y, Wang C. [Effect of high level of fluoride on children's intelligence]. *Chin J Control Endem Dis*. 1992;7(2):93-94.

74. Xu Y, Lu C, Zhang X. [The effect of fluorine on the level of intelligence in children]. *Endem Dis Bull*. 1994;9(2):83-84.

75. Li XS, Zhi JL, Gao RO. Effect of fluoride exposure on the intelligence of children. *Fluoride*. 1995;28:189-192.

76. Wang G, Yang D, Jia F, Wang H. [A study of the IQ levels of four- to seven-year-old children in high fluoride areas]. *Endem Dis Bull*. 1996;11(1):60-62.

77. Wang G, Yang D, Jia F, Wang H. A study of the IQ levels of four- to seven-year-old children in high fluoride areas. *Fluoride*. 2008b;41:340-343.

78. Yao L, Zhou J, Wang S, Cui K, Lin F. Analysis on TSH and intelligence level of children with dental Fluorosis in a high fluoride area. *Lit Inf Prev Med*. 1996;2(1):26-27.

79. Zhao LB, Liang GH, Zhang DN, Wu XR. Effect of a high fluoride water supply on children's intelligence. *Fluoride*. 1996;29:190-192.

80. Yao Y. Comparable analysis on the physical and mental development of children in endemic fluorosis area with water improvement and without water improvement. *Lit Inf Prev Med*. 1997;3(1):42-43.

81. Zhang J, Yao H, Chen Y. [The effect of high levels of arsenic and fluoride on the development of children's intelligence]. *Chin J Public Health*. 1998;17(2):119.

82. Lu Y, Sun ZR, Wu LN, Wang X, Lu W, Liu SS. Effect of high-fluoride water on intelligence in children. *Fluoride*. 2000;33:74-78.

83. Hong FG, Cao YX, Yang D, Wang H. [Research on the effects of fluoride on child intellectual development under different environmental conditions]. *Chin Prim Health Care*. 2001;15(3):56-57.

84. Hong FG, Cao YX, Yang D, Wang H. Research on the effects of fluoride on child intellectual development under different environmental conditions. *Fluoride*. 2008;41:156-160.

85. Hong F, Wang Hui Yang Dong Zhang Z. Investigation on the intelligence and metabolism of iodine and fluoride in children with high iodine and fluoride. *Chin J Endem Dis Control*. 2001b:12-14.

86. Wang X, Wang L, Hu P, Guo X, Luo X. [Effects of high iodine and high fluorine on children's intelligence and thyroid function]. *Chin J Endemiol*. 2001;20(4):288-290.

87. Li YP, Jing XY, Chen D, Lin L, Wang ZJ. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. *Chin J Public Health Manag*. 2003;19(4):337-338.

88. Li YP, Jing XY, Chen D, Lin L, Wang ZJ. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. *Fluoride*. 2008c;41:161-164.

89. Wang SX, Wang ZH, Cheng XT, et al. [Investigation and evaluation on intelligence and growth of children in endemic fluorosis and arsenism areas]. *Chin J Endemiol*. 2005;24:179-182.

90. Seraj B, Shahrabi M, Falahzade M, Falahzade F, Akhondi N. Effect of high fluoride concentration in drinking water on children's intelligence. *J Dent Med*. 2006;19(2):80-86.

91. Wang Z, Wang S, Zhang X, Li J, Zheng X, Hu C. Investigation of children's growth and development under long-term fluoride exposure. *Chin J Control Endem Dis*. 2006;21(4):239-241.

92. Fan Z, Dai H, Bai A, Li P, Li T, Li G. The effect of high fluoride exposure in children's intelligence. *J Environ Health*. 2007;24(10):802-803.

93. Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. Effect of high fluoride water on intelligence of school children in India. *Fluoride*. 2007;40:178-183.

94. Wang SX, Wang ZH, Cheng XT, et al. Arsenic and fluoride exposure in drinking water: Children's IQ and growth in Shanyin County, Shanxi Province, China. *Environ Health Perspect*. 2007;115:643-647.

95. Li F, Chen X, Huang R, Xie Y. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. *J Environ Health*. 2009;26(4):838-840.

96. Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. [Investigation and analysis of children's IQ and dental fluorosis in high fluoride area]. *Chin J Pest Control*. 2010;26(3):230-231.

97. Eswar P, Nagesh L, Devaraj CG. Intelligent quotients of 12-14 year old school children in a high and low fluoride village in India. *Fluoride*. 2011;44:168-172.

98. Kang J, Cheng Y, Wu K, Lin S, He G, Jin Y. [Effect of exposure to fluoride and arsenic in drinking water of Hangjinhouqi on children's intelligence]. *Chin School Health*. 2011:679-681.

99. Poureslami HR, Horri A, Garrusi B. A comparative study of the IQ of children age 7-9 in a high and a low fluoride water city in Iran. *Fluoride*. 2011;44:163-167.

100. Shivaprakash PK, Ohri K, Noorani H. Relation between dental fluorosis and intelligence quotient in school children of Bagalkot District. *J Indian Soc Pedod Prev Dent*. 2011;29:117-120.

101. Wang S, Zhu X. Investigation and analysis of intelligence level of children in high fluoride area. *Chin J Endem Dis Control*. 2012b;27(1):67-68.

102. Bai A, Li Y, Fan Z, Li X, Li P. [Intelligence and growth development of children in coal-burning-borne arsenism and fluorosis areas: An investigation study]. *Chin J Endemiol*. 2014;33(2):160-163.

103. Karimzade S, Aghaei M, Mahvi AH. Investigation of intelligence quotient in 9-12 year-old children exposed to high- and low-drinking water fluoride in West Azerbaijan Province, Iran. *Fluoride*. 2014;47:9-14.

104. Sebastian ST, Sunitha S. A cross-sectional study to assess the intelligence quotient (IQ) of school going children aged 10-12 years in villages of Mysore District, India with different fluoride levels. *J Indian Soc Pedod Prev Dent*. 2015;33:307-311.

105. Zhang P, Cheng L. Effect of coal-burning endemic fluorosis on children's physical development and intellectual level. *Chin J Endem Dis Control*. 2015c;30(6):458-459.

106. Das K, Mondal NK. Dental fluorosis and urinary fluoride concentration as a reflection of fluoride exposure and its impact on IQ level and BMI of children of Laxmisagar, Simlapal Block of Bankura District, W.B., India. *Environ Monit Assess*. 2016;188:218.

107. Mondal D, Dutta G, Gupta S. Inferring the fluoride hydrogeochemistry and effect of consuming fluoride-contaminated drinking water on human health in some endemic areas of Birbhum District, West Bengal. *Environ Geochem Health*. 2016;38:557-576.

108. Zhao Y, Cui Y, Yu J, et al. Study on the relationship between water-borne high iodine and thyroid hormone and children's intelligence level. *J Environ Health*. 2018:6-9.

109. Wang J, Yu M, Yang L, Yang X, Deng B. The effect of coal-burning fluoride exposure on children's intelligence and physical development. *J Environ Health*. 2020c;37(11):971-974. doi:https://dx.doi.org/10.16241/j.cnki.1001-5914.2020.11.007

110. Guo B, Yu J, Cui Y, et al. DBH gene polymorphism, iodine and fluoride and their interactions and their interaction with children's intelligence. *J Environ Hygiene*. 2021;11(02):134-140.

111. Lou D, Luo Y, Liu J, et al. Refinement impairments of verbal-performance intelligent quotient in children exposed to fluoride produced by coal burning. *Biol Trace Elem Res*. Feb 2021;199(2):482-489. doi:https://dx.doi.org/10.1007/s12011-020-02174-z

112. Wang R, He N, Wang Y, Hou G, Zhang P-J. Investigation and analysis of children's dental fluorosis and IQ level in high fluoride district of Hengshui City. *Med Anim Control*. 2021;37(8):796-800. doi:https://dx.doi.org/10.7629/yxdwfz202108023

Commented [I68]: See Doc07_Meta-analysis, 7.D., page 3

List of Tables and Figures

Tables and Figures

Table 1. Characteristics of Studies Included in the Meta-analysis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure: Assessment (Metric, Exposure Groups)	Fluoride Exposure: Levels	Intelligence Assessment	Overall RoB Rating	Confounders Considered
Ren et al. (1989)[66] [translated in Ren et al. 2008][me, o] *Cross-sectional*	China	8–14	No fluoride measurement Low iodine village/high fluoride and low iodine village	Not specified	Wechsler Intelligence Scale for Children	High	Sex; iodine
Chen et al. (1991)[68] [translated in Chen et al. 2008][me, w] *Cross-sectional*	China	7–14	Drinking water Nonendemic/endemic fluorosis village	0.89 mg/L (nonendemic) 4.55 mg/L (endemic)	Chinese Standardized Raven Test	High	Age; sex
Guo et al. (1991)[70] [translated in Guo et al. 2008a][me, o] *Cross-sectional*	China	7–13	Serum Reference area using wood/coal burning-related fluoride endemic area	0.1044 ± 0.0652 mg/L (reference) 0.1483 ± 0.0473 mg/L (endemic)	Chinese Binet Intelligence Test	High	Age; sex; SES
Lin et al. (1991)[40me, o] *Cross-sectional*	China	7–14	Urine, drinking water Reference area with iodine supplementation/high fluoride and low iodine village	Urine: 1.6 mg/L (reference area with iodine supplementation) 2.56 mg/L (high fluoride, low iodine village) Water: 0.34 mg/L (low iodine village) 0.88 mg/L (high fluoride, low iodine village)	Combined Raven's Test for Rural China	High	SES
Sun et al. (1991)[72me, o] *Cross-sectional*	China	6.5–12	No fluoride measurement Nonendemic/endemic (aluminum-fluoride endemic toxicosis)	Fluorosis: 98.36% (endemic)	Japan's Shigeo Kobayashi's 50-point scoring method	High	Age
An et al. (1992)[73me, w] *Cross-sectional*	China	7–16	Drinking water Nonhigh/high fluoride area	0.6−1.0 mg/L (nonhigh) 2.1−3.2 mg/L (secondary high) 5.2−7.6 mg/L (high) 2.1−7.6 mg/L (combined high)	Wechsler Intelligence Scale for Children-Revised	High	Age; race; SES

Commented [I69]: See Doc08_Meta-analysis, 8.K., page 7 and 8

Commented [I70]: See Doc03_Meta-analysis, 3.B., page 1 and 2

Commented [I71]: See Doc02_Meta-analysis, 2.V., page 7

Commented [I72]: See Doc01_Meta-analysis, 1.C., page 1

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Li et al. (1994)[41] [translated in Li et al. 2008b][me, o] *Cross-sectional*	China	12–13	Grain (cooked by burning high-fluoride coal) Reference group (no dental fluorosis)/high fluoride group I (no dental fluorosis)/high fluoride group II (dental fluorosis present)/high fluoride group III (dental fluorosis present)	0.5 mg/kg (reference group) 4.7 mg/kg (group I) 5.2 mg/kg (group II) 31.6 mg/kg (group III)	Proofing test	High	Age; sex; SES
Xu et al. (1994)[74me, w*] *Cross-sectional*	China	8–14	Drinking water Reference region/low- and high-fluoride regions[b]	0.8 mg/L (reference region) 0.38 mg/L (low fluoride) 1.8 mg/L (high fluoride)	Binet-Simon Scale	High	–
Li et al. (1995)[75me, o, u] *Cross-sectional*	China	8–13	Urine, dental fluorosis index (DFI) Nonfluorosis/fluorosis area due to soot from coal burning	1.02 mg/L; DFI: <0.4 (nonfluorosis) 1.81 mg/L; DFI: 0.8 (slight fluorosis) 2.01 mg/L; DFI: 2.5 (medium fluorosis) 2.69 mg/L; DFI: 3.2 (severe fluorosis)	China Rui Wen Scaler for Rural Areas	High	Sex
Wang et al. (1996)[76] [translated in Wang et al. 2008b][me, o, w] *Cross-sectional*	China	4–7	Drinking water (well) Low/high fluoride region Fluoride exposure from drinking water, contaminated food, and coal burning	0.58−1.0 mg/L (low) >1.0−8.6 mg/L (high)	Wechsler Preschool and Primary Scale of Intelligence	High	Age; sex
Yao et al. (1996)[78me, w] *Cross-sectional*	China	8–12	Drinking water Nonendemic/endemic fluorosis area	1 mg/L (nonendemic) 2 mg/L (slightly endemic) 11 mg/L (severely endemic)	Raven Test – Associative Atlas	High	Iodine; SES
Zhao et al. (1996)[79me, w] *Cross-sectional*	China	7–14	Drinking water Low fluoride village (Xinghua)/high fluoride village (Sima)	0.91 mg/L (low) 4.12 mg/L (high)	China Rui Wen Scaler for Rural Areas	High	Age; SES
Yao (1997)[80me, w*] *Cross-sectional*	China	7–12	Drinking water Nonfluorosis area/fluorosis area with water improvements/fluorosis area without water improvements	0.4 mg/L (nonfluorosis area) 0.33 mg/L (fluorosis area with water improvement) 2 mg/L (fluorosis area without water improvement)	Raven's Standard Progressive Matrices (China's Rural Version)	High	Iodine; SES
Zhang et al. (1998)[81me, o] *Cross-sectional*	China	4–10	Drinking water Reference/high fluoride group (all observation groups included arsenic exposure)	0.58 mg/L (reference) 0.8 mg/L (high fluoride)	Shigeo Kobayashi 50-pt. test	High	Age; arsenic

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Lu et al. (2000)[82me, w, u] *Cross-sectional*	China	10–12	Urine, drinking water Low/high fluoride area	Urine: 1.43 ± 0.64 mg/L (low) 4.99 ± 2.57 mg/L (high) Water: 0.37 ± 0.04 mg/L (low) 3.15 ± 0.61 mg/L (high)	Chinese Combined Raven Test-C2	High	SES
Hong et al. (2001)[83] [translated in Hong et al. 2008][me, w*] *Cross-sectional*	China	8–14	Drinking water Reference/high fluoride[b]	0.75 mg/L (reference) 2.90 mg/L (high fluoride)	Chinese Standardized Raven Test	High	Iodine; SES; demographics
Hong et al. (2001b)[85me, o] *Cross-sectional*	China	8–14	Urine, drinking water Nonendemic/endemic fluorosis areas (high fluoride, high iodine)	Urine: 0.796 ± 0.53 mg/L (nonendemic) 2.09 ± 1.03 mg/L (endemic) Water: 0.48 mg/L (nonendemic) 2.81 mg/L (endemic)	Combined Raven's Test for Rural China	High	–
Wang et al. (2001)[86me, o] *Cross-sectional*	China	8–12	Urine, drinking water Reference point (low fluoride, low iodine)/investigative point (high fluoride, high iodine)	Urine: 0.82 mg/L (low fluoride, low iodine) 3.08 mg/L (high fluoride, high iodine) Water: 0.5 mg/L (low fluoride, low iodine) 2.97 mg/L (high fluoride, high iodine)	Combined Raven's Test for Rural China	High	–
Li et al. (2003)[87] [translated in Li et al. 2008c][me] *Cross-sectional*	China	6–13	No fluoride measurement Reference/endemic fluorosis areas	Not specified	Chinese Standardized Raven Test	High	–

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Xiang et al. (2003a)[44me, w*, u] *Cross-sectional*	China	8–13	Urine, drinking water Nonendemic/endemic fluorosis areas	Urine: 1.11 ± 0.39 mg/L (reference) 3.47 ± 1.95 mg/L (high fluoride) Water: 0.36 ± 0.15 mg/L (nonendemic) 0.75 ± 0.14 mg/L (endemic fluorosis area group A) 1.53 ± 0.27 mg/L (endemic fluorosis area group B) 2.46 ± 0.3 mg/L (endemic fluorosis area group C) 3.28 ± 0.25 mg/L (endemic fluorosis area group D) 4.16 ± 0.22 mg/L (endemic fluorosis area group E) 2.47 ± 0.79 mg/L (high fluoride)	Combined Raven's Test for Rural China	Low	Age; sex; iodine; lead; SES
Wang et al. (2005)[89me, w, u] *Cross-sectional*	China	8–12	Urine, drinking water Reference/high fluoride group[c]	Urine: 1.51 mg/L(reference) 5.09 mg/L (high fluoride group) Water: 0.48 mg/L (reference) 8.31 mg/L (high fluoride group)	Chinese Combined Raven Test-C2	High	SES
Seraj et al. (2006)[90me, w *Cross-sectional*	Iran	7–11	Drinking water Low/high fluoride area	0.4 ppm (low) 2.5 ppm (high)	Raven Test	High	Sex
Wang et al. (2006)[91me, w, u] *Cross-sectional*	China	8–12	Urine, drinking water Reference/high (area severely affected by fluorosis)	Urine: 1.51 ± 1.66 mg/L (reference) 5.50 ± 2.40 mg/L (high) Water: 0.73 ± 0.28 mg/L (reference) 5.54 ± 3.88 mg/L (high)	Combined Raven's Test for Rural China	High	–
Fan et al. (2007)[92me, w, u] *Cross-sectional*	China	7–14	Urine, drinking water Low/high fluoride area	Urine: 1.78 ± 0.46 mg/L (low) 2.89 ± 1.97 mg/L (high) Water: 1.03 mg/L (low) 3.15 mg/L (high)	Chinese Combined Raven Test-C2	High	–
Trivedi et al. (2007)[93me, w, u] *Cross-sectional*	India	12–13	Urine, drinking water Low/high fluoride area	Urine: 2.30 ± 0.28 mg/L (low) 6.13 ± 0.67 mg/L (high) Water: 2.01 ± 0.009 mg/L (low) 5.55 ± 0.41 mg/L (high)	questionnaire prepared by Professor JH Shah	High	Age; sex

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Wang et al. (2007)[94me, o, w, u] *Cross-sectional*	China	8–12	Urine, drinking water Low fluoride, low arsenic/high fluoride, low arsenic area	Urine: 1.5 ± 1.6 mg/L (low fluoride, low arsenic) 5.1 ± 2.0 mg/L (high fluoride, low arsenic) Water: 0.5 ± 0.2 mg/L (low fluoride, low arsenic) 8.3 ± 1.9 mg/L (high fluoride, low arsenic)	Combined Raven's Test for Rural China	High	Age; sex; arsenic; SES
Li et al.(2009)[95me, o, u*] *Cross-sectional*	China	8–12	Urine Endemic fluorosis region caused by coal burning (reference/mild/medium/severe) Degree of dental fluorosis (normal/suspected/very mild/mild/medium/severe)	0.962 ± 0.517 mg/L (reference) 1.235 ± 0.426 mg/L (mild) 1.670 ± 0.663 mg/L (medium) 2.336 ± 1.128 mg/L (severe) 0.867 ± 0.233 mg/L (normal) 1.094 ± 0.355 mg/L (suspected) 1.173 ± 0.480 mg/L (very mild) 1.637 ± 0.682 mg/L (mild) 2.005 ± 0.796 mg/L (medium) 2.662 ± 1.093 mg/L (severe)	Combined Raven's Test for Rural China	High	Age; sex
Li et al. (2010)[96me] *Cross-sectional*	China	7–10	No fluoride measurement Nondental fluorosis children/dental fluorosis children	Not specified	Combined Raven's Test for Rural China	High	Sex
Ding et al. (2011)[32me, u*, rs] *Cross-sectional*	China	7–14	Dental fluorosis (normal/ questionable/very mild/mild/moderate) Urine Mean urinary fluoride levels (10 groups)	0.80 ± 0.55 mg/L (normal) 1.13 ± 0.73 mg/L (questionable) 1.11 ± 0.74 mg/L (very mild) 1.31 ± 0.78 mg/L (mild) 1.46 ± 0.79 mg/L (moderate) 0.26 mg/L (group 1) 0.45 mg/L (group 2) 0.56 mg/L (group 3) 0.66 mg/L (group 4) 0.75 mg/L (group 5) 0.89 mg/L (group 6) 1.08 mg/L (group 7) 1.33 mg/L (group 8) 1.74 mg/L (group 9) 2.96 mg/L (group 10) 0.10–3.55 mg/L	Combined Raven's Test for Rural China	Low	Age; arsenic; iodine; lead; SES; demographics

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Eswar et al.(2011)[97me, w] *Cross-sectional*	India	12–14	Drinking water Low/high fluoride villages	0.29 mg/L (low) 2.45 mg/L (high)	Standard Progressive Matrices	High	Age; sex
Kang et al. (2011)[98me, o] *Cross-sectional*	China	6–12	Drinking water Reference/high fluoride areas (both areas with high arsenic exposure)	1.24 ± 0.74 mg/L (all children) <1.2 mg/L (reference) ≥1.2 mg/L (high fluoride)	Chinese Combined Raven Test-C2	High	Age; sex
Poureslami et al. (2011)[99me, w] *Cross-sectional*	Iran	7–9	Drinking water Reference/endemic dental fluorosis city	0.41 mg/L (reference) 2.38 mg/L (endemic)	Persian version of Raven's Matrices Test	High	Sex
Shivaprakash et al. (2011)[100me, w] *Cross-sectional*	India	7–11	Drinking water No fluorosis/fluorosis severity groups (mild/moderate/severe)/all fluorosis	<0.5 ppm (no fluorosis) 2.5–3.5 ppm (mild) 2.5–3.5 ppm (moderate) 2.5–3.5 ppm (severe) 2.5–3.5 ppm (all)	Raven's Colored Progressive Matrices	High	Health factors; SES
Seraj et al. (2012)[45me, w] *Cross-sectional*	Iran	6–11	Drinking water Normal/medium/high fluoride levels	0.8 ± 0.3 mg/L (normal) 3.1 ± 0.9 mg/L (medium) 5.2 ± 1.1 mg/L (high)	Raven's Colored Progressive Matrices	Low	Age; sex; SES
Trivedi et al. (2012)[46me, w, u] *Cross-sectional*	India	12–13	Urine, ground water Low/high fluoride area	Urine: 0.42 ± 0.23 mg/L (low) 2.69 ± 0.92 mg/L (high) Water: 0.84 ± 0.38 mg/L (low) 2.3 ± 0.87 mg/L (high)	Questionnaire prepared by Professor JH Shah	Low	Sex; SES
Wang et al. (2012b)[101me] *Cross-sectional*	China	Primary school age	No fluoride measurement Reference/high fluoride areas	Not specified	Combined Raven's Test for Rural China	High	–
Bai et al. (2014)[102me, o] *Cross-sectional*	China	8–12	Urine Coal-burning-borne fluorosis areas (reference/lightly-affected/seriously-affected)	0.54 mg/L (reference) 0.81 mg/L (lightly-affected area) 1.96 mg/L (seriously-affected area)	Chinese Combined Raven Test-C2	High	SES
Karimzade et al. (2014)[103me, w] *Cross-sectional*	Iran	9–12	Drinking water Low/high fluoride area	0.25 mg/L (low) 3.94 mg/L (high)	Iranian version of the Raymond B Cattell test	High	Sex

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Broadbent et al. (2015)[38me, w*] *Prospective Cohort*	New Zealand	7–13	Drinking water Area without community water fluoridation (low)/area with community water fluoridation (high) Fluoride tablet use (never/ever) Fluoride toothpaste use (never/sometimes/always)	Water: 0.0–0.3 mg/L (low) 0.7–1.0 mg/L (high) Tablet use: 0 mg (never used) 0.5 mg (ever used) Range not specified for fluoride toothpaste use (always/sometimes/never)	Wechsler Intelligence Scale for Children-Revised	High	Sex; SES; low birth weight; breastfeeding
Khan et al. (2015)[39me] *Cross-sectional*	India	6–11	Drinking water Low fluoride areas (Tiwariganj)/high fluoride areas (Unnao) Fluorosis grades (normal/very mild/mild/moderate/severe)	0.19 mg/L (Tiwariganj) 2.41 mg/L (Unnao) Ranges not specified by fluorosis grades	Raven's Colored Progressive Matrices	High	Health factors; SES
Sebastian and Sunitha (2015)[104me, w*] *Cross-sectional*	India	10–12	Drinking water Low/normal/high fluoride villages	0.40 mg/L (low) 1.2 mg/L (normal) 2.0 mg/L (high)	Raven's Colored Progressive Matrices	High	Age; sex; SES
Zhang et al.(2015b)[33me, w*, u, rs] *Cross-sectional*	China	10–12	Urine, drinking water, serum Reference/high fluoride areas	Urine: 1.10 ± 0.67 mg/L (reference) 2.40 ± 1.01 mg/L (high) Water: 0.63 (0.58–0.68) mg/L (reference) 1.40 (1.23–1.57) mg/L (high) Serum: 0.06 ± 0.03 (reference) 0.18 ± 0.11 (high)	Combined Raven's Test for Rural China	Low	Age; sex; arsenic; iodine; drinking water fluoride; SES; thyroid hormone levels; COMT genotype
Zhang et al. (2015c)[105me, o] *Cross-sectional*	China	7−13	Urine Coal-burning endemic fluorosis area Reference (no dental fluorosis)/mild dental fluorosis/moderate dental fluorosis/critically ill dental fluorosis	0.83 ± 0.71 mg/L (reference) 1.54 ± 0.57 mg/L (mildly ill) 2.41 ± 0.76 mg/L (moderately ill) 3.32 ± 1.02 mg/L (critically ill)	Combined Raven's Test for Rural China	High	–

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Das and Mondal (2016)[106me, u] *Cross-sectional*	India	6–18	Urine, drinking water intake, dental fluorosis (normal/questionable/very mild/mild/moderate/severe)	Urine: 2.91 ± 1.76 mg/L (normal) 2.50 ± 2.39 mg/L (questionable) 2.58 ± 1.31 mg/L (very mild) 2.95 ± 1.44 mg/L (mild) 4.82 ± 3.57 mg/L (moderate) 3.81 ± 2.51 mg/L (severe) Water: 0.069 ± 0.021 mg/kg-d (normal) 0.064 ± 0.004 mg/kg-d (questionable) 0.060 ± 0.036 mg/kg-d (very mild) 0.060 ± 0.030 mg/kg-d (mild) 0.099 ± 0.063 mg/kg-d (moderate) 0.093 ± 0.040 mg/kg-d (severe)	Combined Raven's Test for Rural China	High	–
Mondal et al. (2016)[107me, w] *Cross-sectional*	India	10–14	Drinking water Low/high fluoride areas	Not reported (low) 0.33–18.08 mg/L (high)	Raven Standard Theoretical Intelligence Test	High	SES
Bashash et al. (2017)[5me, u, rs] *Prospective Cohort*	Mexico	6−12	Maternal urine Reference/high fluoride (based on children urinary fluoride)	<0.80 mg/L (reference) ≥0.80 mg/L (high)	Wechsler Abbreviated Scale of Intelligence	Low	Age; sex; weight at birth; parity; gestational age; maternal characteristics (smoking history, marital status, age at delivery, IQ, education, cohort)
Cui et al. (2018)[34rs] *Cross-sectional*	China	7–12	Urine	Boys: 1.3 (0.9−1.7)[d] mg/L Girls: 1.2 (0.9−1.6)[d] mg/L	Combined Raven's Test for Rural China	Low	Age; maternal education; smoking in family member; stress; anger; dopamine receptor-2 polymorphism

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Yu et al. (2018)[11me, w, u*, rs] *Cross-sectional*	China	7–13	Maternal urine Low/medium/high fluoride ranges Drinking water Normal/high fluoride	Urine: 0.01–1.60 mg/L (low) 1.60–2.50 mg/L (medium) 2.50–5.54 mg/L (high) Water: ≤1 mg/L (normal) >1 mg/L (high) Overall: 0.01−5.54 mg/L (urine) 0.20−3.90 mg/L (water)	Combined Raven's Test for Rural China	Low	Age; sex; health factors; SES
Zhao et al. (2018)[108me, o] *Cross-sectional*	China	7–12	Urine Reference/exposed areas All areas with iodine exposure	≤2.16 mg/L (reference) >2.16 mg/L (exposed)	Combined Raven's Test for Rural China	High	–
Green et al. (2019)[6me, w*, u*, rs] *Prospective Cohort*	Canada	3−4	Maternal urine, drinking water, maternal fluoride intake Nonfluoridated/fluoridated area	Urine: 0.40 ± 0.27 mg/L (nonfluoridated) 0.69 ± 0.42 mg/L (fluoridated) Water: 0.13 ± 0.06 mg/L (nonfluoridated) 0.59 ± 0.08 mg/L (fluoridated) Intake: 0.30 ± 0.26 mg/day (nonfluoridated) 0.93 ± 0.43 mg/day (fluoridated) Overall: 0.51 ± 0.36 mg/L (urine) 0.54 ± 0.44 mg/day (intake) 0.31 ± 0.23 mg/L (water)	Wechsler Primary and Preschool Scale of Intelligence-III	Low	Sex; city; maternal education; race/ethnicity; HOME score; prenatal secondhand smoke exposure
Cui et al. (2020)[47me, u] *Cross-sectional*	China	7–12	Urine Low/medium/high fluoride levels	<1.6 mg/L (low) 1.6–2.5 mg/L (medium) >=2.5 mg/L (high)	Combined Raven's Test	Low	Sex; arsenic; iodine

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Till et al. (2020)[7rs] *Prospective Cohort*	Canada	3−4	Residence, maternal urine, drinking water, infant fluoride intake from formula Nonfluoridated/fluoridated areas	Urine: 0.38−0.42 mg/L (nonfluoridated) 0.64−0.70 mg/L (fluoridated) Water: 0.13 mg/L (nonfluoridated) 0.58 mg/L (fluoridated) Intake: 0.02−0.08 mg/day (nonfluoridated) 0.12−0.34 mg/day (fluoridated)	Wechsler Primary and Preschool Scale of Intelligence-III	Low	Age; sex; maternal education; maternal race; HOME total score; secondhand smoke status in the child's house
Wang et al. (2020c)[109me, o] *Cross-sectional*	China	7−12	Urine Coal-burning endemic fluorosis area Nonendemic/endemic fluorosis regions	0.461 ± 0.210 mg/L (nonendemic) 0.689 ± 0.502 mg/L (endemic)	Combined Raven's Test for Rural China	High	Age; sex
Xu et al. (2020)[36me, u*, rs] *Cross-sectional*	China	7−13	Urine Reference/high prenatal exposure only/high childhood exposure only/both prenatal and childhood exposure group	0.82 ± 0.30 mg/L (reference) 0.98 ± 0.29 mg/L (high prenatal exposure only) 2.05 ± 0.58 mg/L (high childhood exposure only) 2.13 ± 0.59 mg/L (both prenatal and childhood exposure group)	Combined Raven's Test for Rural China	Low	Age; sex; gestational weeks; maternal education level; paternal education level; children's BMI
Guo et al., (2021)[110me] *Cross-sectional*	China	7−12	Urine Reference/exposed areas (also with iodine exposure)	1.16 mg/L (reference) 1.29 mg/L (iodine area 1) 2.01 mg/L (iodine area 2)	Combined Raven's Test for Rural China	High	–
Lou et al. (2021)[111me, o] *Cross-sectional*	China	8−12	Coal-burning endemic fluorosis area No fluoride measurement Nondental fluorosis children/dental fluorosis children	Not specified	Wechsler Intelligence Scale for Children-Revised in China (WISC-CR)	High	–
Saeed et al. (2021)[35me, o, rs] *Cross-sectional*	Pakistan	5−16	Urine, drinking water Reference/high fluoride areas Co-exposure with arsenic	Urine: 0.24 ± 0.15 mg/L (reference) 3.27 ± 2.60 mg/L (high) Water: 0.15 ± 0.13 mg/L (reference) 5.64 ± 3.52 mg/L (high)	Wechsler scale of intelligence (WISC-IV)	High	Age; sex; parental education; dental fluorosis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Wang et al. (2021)[112] me, w *Cross-sectional*	China	9−11	Drinking water Reference/high fluoride areas	1.0 ± 0.07 mg/L (reference) 2.8 ± 0.06 mg/L (high fluoride)	Combined Raven's Test	High	Age; sex
Zhao et al. (2021)[37rs] *Cross-sectional*	China	6−11	Urine Nonendemic/endemic fluorosis areas	1.03 (0.72, 1.47) mg/L	Combined Raven's Test for Rural China	Low	Age; sex; BMI; paternal educational level; maternal educational level; household income; abnormal birth; maternal age at delivery

Notes:
COMT = catechol-O-methyltransferase; RoB = risk of bias; SES = socioeconomic status; HOME = Home Observation for Measurement of the Environment
[a]An "me" superscript indicates that the studies included in the mean-effects meta-analysis; an "o" superscript indicates a study included in "other" exposures *mean-effects meta-analysis* (see **Table 2** footnote); a "w" superscript indicates studies included in the mean-effects dose-response meta-analysis using fluoride in water; a "u" superscript indicates studies included in the mean-effects dose-response meta-analysis using fluoride in urine; "*" indicates studies included in the mean-effects dose-response meta-analysis at levels < 1.5 mg/L; an "rs" superscript indicates studies included in the regression slopes meta-analysis.
[b]Additional exposure regions including iodine levels were not included in the analysis.
[c]Additional exposure regions including arsenic levels were not included in the analysis.
[d]Median (q1−q3).

Table 2. Pooled SMDs and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	SMD (95% CI)	Heterogeneity	
			p-value	I^2
Overall Effect	55	−0.46 (−0.55, −0.37)	<0.001	87%
Subgroup Analyses				
Risk of Bias				
Low	10	−0.22 (−0.39, −0.05)	<0.001	83%
High	45	−0.52 (−0.63, −0.42)	<0.001	86%
Sex				
Males	14	−0.62 (−0.81, −0.42)	<0.001	78%
Females	13	−0.53 (−0.72, −0.34)	<0.001	74%
Age Group				
<10 years[a]	13	−0.41 (−0.60, −0.22)	<0.001	80%
≥10 years	16	−0.55 (−0.70, −0.40)	<0.001	68%
Country				
China	39	−0.43 (−0.52, −0.34)	<0.001	85%
India	8	−0.99 (−1.55, −0.43)	<0.001	93%
Iran	4	−0.68 (−0.99, −0.38)	0.077	56%
Canada	1	0.01 (−0.19, 0.21)	NA	NA
Mexico	1	0.13 (−0.16, 0.42)	NA	NA
New Zealand	1	0.01 (−0.19, 0.22)	NA	NA
Pakistan	1	−0.25 (−0.65, 0.16)	NA	NA
Assessment Type				
CRT-RC tests	29	−0.36 (−0.46, −0.27)	<0.001	82%
Non-CRT-RC tests	26	−0.60 (−0.78, −0.42)	<0.001	89%
Raven's tests	10	−0.76 (−1.10, −0.43)	<0.001	91%
Other tests	16	−0.52 (−0.74, −0.29)	<0.001	89%
Exposure Type				
Water fluoride	32	−0.37 (−0.46, −0.27)	<0.001	82%
Dental fluorosis	7	−0.99 (−1.57, −0.41)	<0.001	96%
Other exposures[b]	16	−0.54 (−0.71, −0.37)	<0.001	81%
Previous Meta-analyses				
Duan et al. (2018)[4]	26	−0.52 (−0.62, −0.42)	<0.001	69%
Choi et al. (2012)[3]	27	−0.45 (−0.56, −0.34)	<0.001	80%

Notes: CI = confidence interval; CRT-RC = Combined Raven's Test–The Rural edition in China; NA = not applicable; SMD = standardized weighted mean difference

[a]An et al. (1992)[73] and Li et al. (2010)[96] include 10-year-old children in the <10 age group (7−10 years reported).

[b]Includes iodine [40, 66 [translated in Ren et al. 2008], 85, 86, 108]; arsenic[35, 81, 94]; aluminum[72]; and non-drinking water fluoride (i.e., fluoride from coal burning[41 [translated in Li et al. 2008b], 70 [translated in Guo et al. 2008a], 75, 76 [translated in Wang et al. 2008b], 89, 95, 102, 105, 109, 111]).

[c] p-value for differences between the estimates based on CRT-RC tests *vs.* non-CRT-RC tests.

[d] p-value for differences between the estimates based on CRT-RC tests, Raven's test and other tests. Note that non-CRT-RC test include Raven's tests and other tests.

Commented [l73]: See Doc06a_Meta-analysis, 6a.E., page 3 and 4

Commented [l74]: See Doc02_Meta-analysis, 2.X., page 7 and 8

Table 3. Pooled Regression Slopes and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Effect				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Subgroup Analyses				
Risk of Bias				
Low	8	−1.33 (−2.09, −0.57)	0.072	46%
High	1	−3.45 (−4.44, −2.46)	NA	NA
Sex				
Males	2	−2.23 (−5.45, 0.99)	0.092	65%
Females	2	−0.27 (-3.64, 3.10)	0.145	53%
Country				
Canada	1	−1.95 (−5.18, 1.28)	NA	NA
China	6	−1.06 (−1.70, −0.42)	0.191	33%
Mexico	1	−5.00 (−8.53, −1.47)	NA	NA
Pakistan	1	−3.45 (−4.44, −2.46)	NA	NA
Assessment Type				
CRT-RC tests	6	−1.06 (−1.70, −0.42)	0.191	33%
Non-CRT-RC tests	3	−3.43 (−4.35, −2.52)	0.457	0%
Exposure Type				
Urinary fluoride	9	−1.81 (−2.80, −0.81)	<0.001	77%
Intake	2	−3.87 (−7.15, −0.59)	0.737	0%
Water fluoride	2	−4.77 (−9.09, −0.45)	0.707	0%
Exposure timing				
Pre-natal exposure	3	−3.08 (−5.43, −0.72)	0.351	5%
Post-natal exposure	7	−1.84 (-2.97, -0.72)	<0.001	78%

Notes: CI = confidence interval; NA = not applicable

Commented [I75]: See Doc02_Meta-analysis, 2.H., page 3

Commented [I76]: See Doc06a_Meta-analysis, 6a.I., page 6

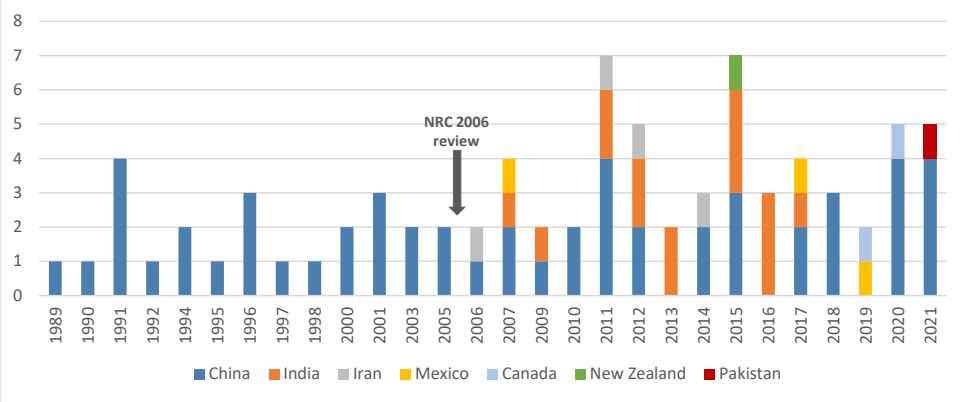


Figure 1. Number of Studies of Fluoride Exposure and IQ in Children by Country and Year of Publication

Note: Figure includes 80 epidemiological studies that were identified during the larger systematic review and the November 2021 literature search update that evaluated the effects of fluoride exposure on children's IQ.

Commented [EAM77]: See Doc02_Meta-analysis, 2.W, page 7

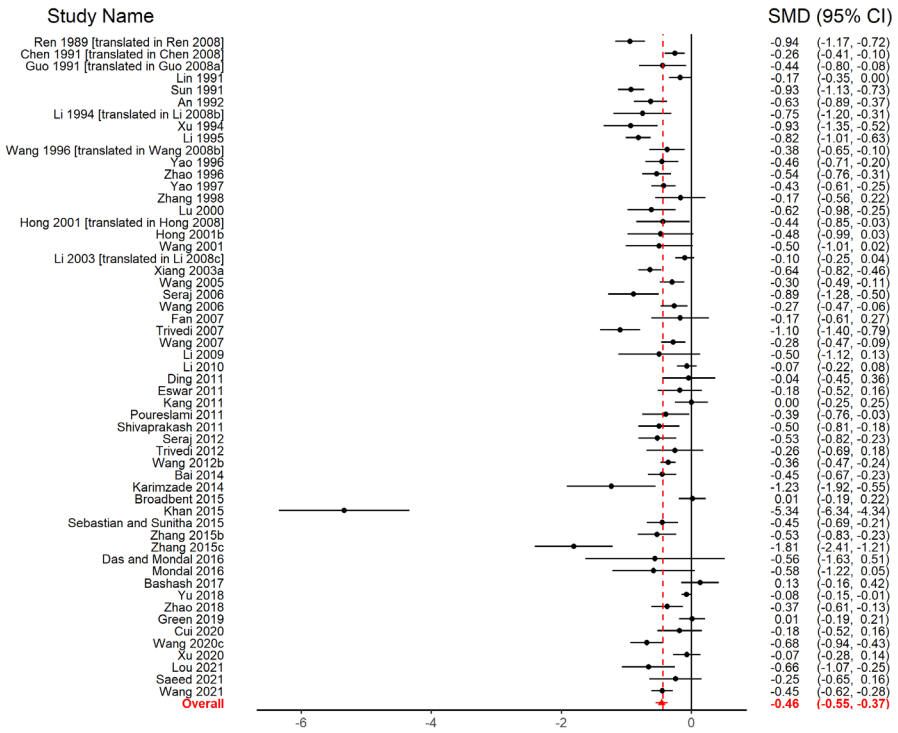


Figure 2. Association Between Fluoride Exposure and IQ Scores in Children

Note: Forest plot for random-effects meta-analysis of the association between fluoride exposure and child's IQ scores. Effect size is expressed as the standardized weighted mean difference for heteroscedastic population variances (SMD). The random-effects pooled SMD is shown as a solid triangle. Horizontal lines represent 95% CIs for the study-specific SMDs.

Commented [I78]: See Doc06b_Meta-analysis, 6b.R., page 15 and 16

SUPPLEMENTAL MATERIALS

List of Tables

List of Figures

Additional Detail on Methods

Systematic Literature Review

Literature searches were conducted in the following databases: BIOSIS, EMBASE, PsychINFO, PubMed, Scopus, Web of Science, CNKI, and Wanfang. Search strategies tailored for each database are available in the protocol (https://ntp.niehs.nih.gov/go/785076). The last search was performed on May 1, 2020. The identification of studies for the meta-analysis was part of a larger systematic review.[1]

Study Selection

In order to be eligible for inclusion in the systematic literature review, individual study publications (referred to in this paper as "studies") had to satisfy eligibility criteria outlined in the protocol (i.e., address PECO statement in Table 1 and specific exclusion criteria in Table 2, https://ntp.niehs.nih.gov/go/785076).

Commented [I1]: See Doc02_Meta-analysis, 2.E., page 2

The following exclusions were made:

(1) Case studies and case reports.

(2) Articles without original data (e.g., reviews, editorials, commentaries). Reference lists from these materials, however, were reviewed to identify potentially relevant studies not identified from the database searches. New studies identified were assessed for eligibility for inclusion.

(3) Conference abstracts or reports and dissertations.

References retrieved from the literature search were independently screened by two trained screeners at the title and abstract level to determine whether a reference met the evidence selection criteria. Studies that were not excluded during the title and abstract screening were further screened for inclusion with a full-text review by two independent reviewers. Translation assistance was obtained to assess the relevance of non-English studies. Following full-text review, the remaining studies were "included" and used for the evaluation.

Results of the study identification process are provided in **eFigure 1**.

Statistical Analysis

Mean-effects meta-analysis

A sensitivity analysis was performed to evaluate the impact of using any exposed group compared to the reference group. This was accomplished by using the approach outlined in the Cochrane Handbook for Systematic Reviews[2] which combines the data from all available exposure groups (n, mean, and standard deviation [SD]). Subgroup analyses were stratified by risk of bias (high or low), outcome assessment, exposure matrix (e.g., urine or water), pre- or post-natal exposures, outcome, gender, and age group. If results were not reported by gender or age-specific subgroups (<10, ≥10 years), they were calculated (if possible) by combining smaller subgroups. If SDs were not reported, but mean effects, sample sizes (n values), and p-values for differences between groups were available, SDs were calculated using the SE and t-statistic (assuming equal variances). To avoid sample overrepresentation, if the same cohort was followed at multiple timepoints resulting in multiple study publications (e.g., Yu et al.[3] and Wang et al.[4]), only the study publication that included the largest number of participants was included in the meta-analysis (see **eTable 1** for list of excluded studies and rationales). For studies with overlapping populations (i.e., multiple study publications that used the same cohort), results from one study publication were selected considering the following factors: most appropriate exposure metric, exposure

range, exposure period, number of subjects, and statistical adjustment for potential confounders (see **eTable 2** for study-specific effect estimates used in the meta-analysis).

Dose-response meta-analysis

To determine whether the data support an exposure-response relationship, we conducted a *dose-response mean-effects meta-analysis.* This analysis included studies from the *mean-effects meta-analysis* that reported fluoride exposure levels; we excluded studies for which there was evidence that co-exposures to arsenic or iodine might be differential (see **eTable 2**).

The *dose-response meta-analysis* was conducted using a one-step approach developed in the protocol (https://ntp.niehs.nih.gov/go/78500.76).[5, 6] The approach uses linear mixed models to analyze all available mean effect estimates for the reference group and one or more of the non-reference exposure groups. For each study, the median or mean fluoride level for each exposure group was assigned to its corresponding effect estimate. If median or mean levels by exposure group were not provided, the midpoint of the upper and lower boundaries in every exposure category was assigned as the average level. If the upper boundary for the highest exposure group was not reported, the boundary was assumed to have the same amplitude as the nearest exposure category. For each study, the SMDs and corresponding SEs were used to compare the differences in mean IQ between the exposed and reference groups. The corresponding SMD for the reference group was set to zero for this analysis. The SMDs and corresponding variances were used to estimate a pooled dose-response curve using a restricted maximum likelihood estimation method. To examine a potential nonlinear relationship between exposure to fluoride and children's IQ levels, quadratic terms and restricted cubic splines were created, and a potential departure from a linear trend was assessed by testing the coefficient of the quadratic term and a second spline equal to zero. Models were compared and the best model fit was determined based on the maximum likelihood Akaike information criterion (AIC).[7] The AIC is a goodness-of-fit measure that adjusts for the number of parameters in the model, and lower AIC values indicate better fitting models. Models using a pooled dose-response curve using a restricted maximum likelihood estimation method and a maximum likelihood method were also reported (**eTable4** and **eTable 5**, respectively).

Commented [I2]: See Doc06a_Meta-analysis, 6a.D., page 6 and 7

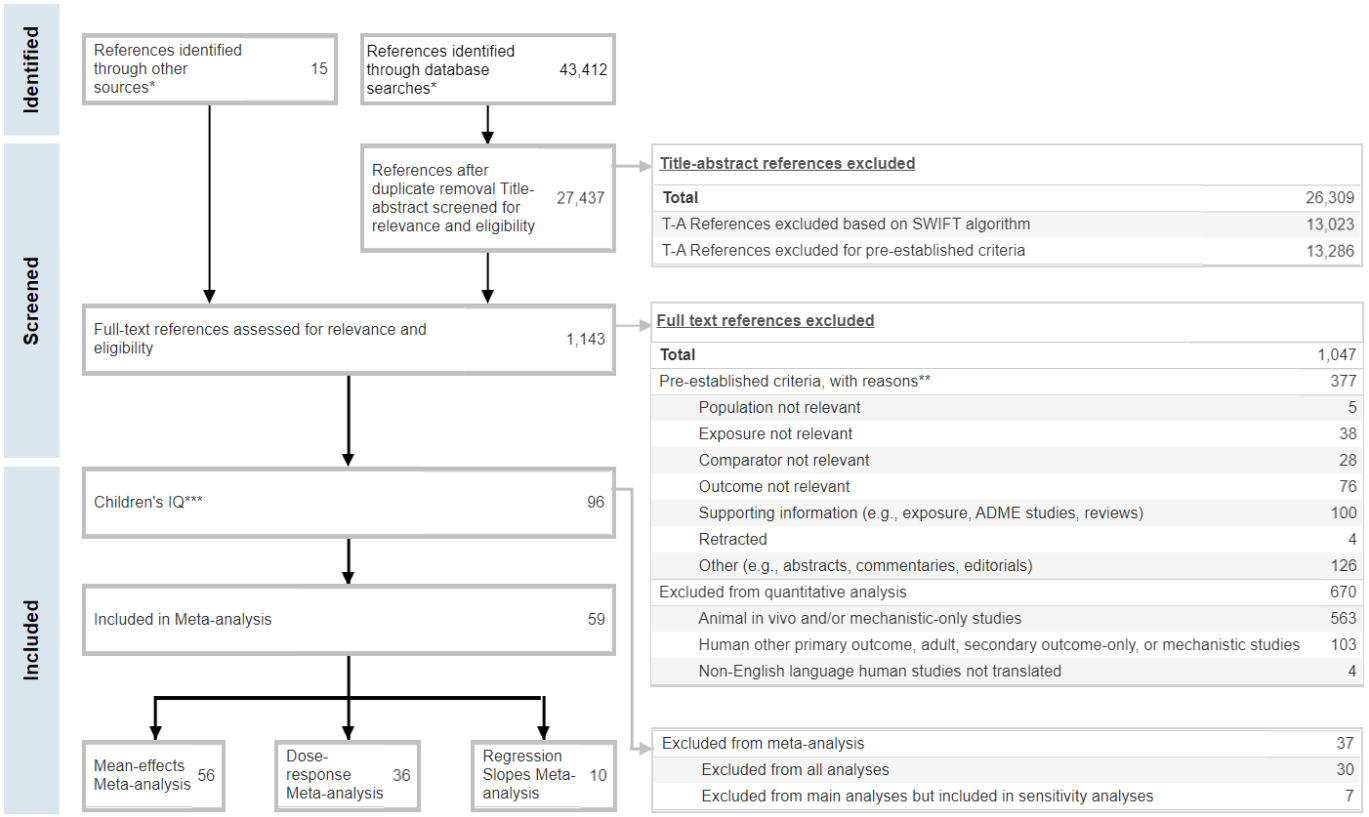
To examine whether there were effects at lower levels of exposure, we conducted sub-group analyses for both drinking water and urinary fluoride measures. Analyses were restricted to <4 mg/L, the EPA's current enforceable drinking water standard for fluoride; <2 mg/L, the EPA's non-enforceable secondary standard for fluoride in drinking water;[8] and <1.5 mg/L, the WHO's guideline for fluoride in drinking water.[9]

Results

Study Sample

Results of the study identification process are provided in **eFigure 1**. Characteristics of the 55 studies that compared mean IQ scores between groups of children with different levels of fluoride exposure are shown in **Table 1** of the main publication (see **eTable 1** for list of excluded publications). Study-specific effect estimates used in the meta-analyses are presented in **eTable 2**. One study per country was conducted in New Zealand, Mexico, Pakistan, and Canada; 4 studies were conducted in Iran, 8 studies were conducted in India, and the remaining 39 studies were performed in China (see **Table 1** of the main publication). Nine study populations were exposed to fluoride from coal burning[10 [translated in Guo et al. 2008a], 12 [translated in Li et al. 2008b], 14-16,17-19]; otherwise, it is assumed that study populations were exposed to fluoride primarily through drinking water. Measures of fluoride exposure included water fluoride (n = 32 studies), dental fluorosis (n = 7), and other non-drinking water sources of exposure to fluoride (e.g., fluoride exposure from coal burning [n = 16]). Fourteen studies presented results for boys and 13 studies reported results for girls; children < 10 years old and children ≥ 10 years old were examined in 13 and 16 studies, respectively (**Table 2**). The Combined Raven's Test for Rural China (CRT-RC) was used to measure

children's IQ in 29 studies. Other measures of IQ included the Wechsler intelligence tests,[20 [translated in Ren et al. 2008], 22 [translated in Wang et al. 2008b], 24, 25] Binet IQ test[10 [translated in Guo et al. 2008a], 26], Raven's Standard Progressive Matrices test,[27-36] Raymond B Cattell test,[37] Japan IQ test,[38, 39] Index of Mental Capacity,[12 [translated in Li et al. 2008b]] and other tests using a doctor-prepared questionnaire.[40, 41] There were 10 low risk-of-bias studies and 45 high risk-of-bias studies (https://hawcproject.org/summary/visual/assessment/405/Figure-X-Meta-analysis-RoB/).



eFigure 1. Prisma Flow Diagram of Study Inclusion

*This information was part of a larger systematic review effort resulting in many studies in the search strategy and PRISMA that were not considered for meta-analysis.

**Studies may have been excluded for more than one reason. The first one identified by the screener was recorded.

*** For the purpose of this PRISMA figure, the Children's IQ count includes three publications[42-44] based on subsamples (i.e., 50–60 children) of a larger Yu et al.[3] cohort. These three publications are not included in the meta-analysis and are not displayed in Figure 1 in the main publication.

eTable 1. List of Excluded Studies from Mean-effects Meta-analysis

Commented [l3]: See Doc05_Meta-analysis, 5.E. (pages 6 and 7), 5.G. (page 8), and 5.M. (page 11)

Reference, Country	Reason for Exclusion
Qin et al. (1990)[45] [translated in Qin et al. 2008], China	Missing mean or SD of outcome measure
Yang et al. (1994)[47] [translated in Yang et al. 2008], China	Overlapping population with Wang et al. (2001)[49]; Table 2 in Yang et al. (1994)[47] seemed incomplete
Wang et al. (2005b)[50] [translated in Wang et al. 2008a], China	Missing mean or SD of outcome measure
Rocha-Amador et al. (2007)[52], Mexico	Missing mean or SD of outcome measure
Liu et al. (2000)[53] [translated in Liu et al. 2008] , China	Overlapping population with Lu et al. (2000)[55]
Sudhir et al. (2009)[56], India	Missing mean or SD of outcome measure
He and Zhang (2010)[57], China	Missing mean or SD of outcome measure
Xiang et al. (2011)[58], China	Overlapping population with Xiang et al. (2003a)[59]
Saxena et al. (2012)[60], India	Missing mean or SD of outcome measure
Wang et al. (2012)[61], China	Overlapping population with Xiang et al. (2003a)[59]
Nagarajappa et al. (2013)[62], India	Seguin Foam Board test; due to the test measuring eye-hand coordination and cognitive-perceptual abilities
Pratap et al.(2013)[63], India	Missing mean or SD of outcome measure
Asawa et al. (2014)[64], India	Seguin Foam Board test; due to the test measuring eye-hand coordination and cognitive-perceptual abilities
Wei et al. (2014)[65], China	Missing mean or SD of outcome measure
Choi et al. (2015)[66], China	Cognitive functions other than IQ
Kundu et al. (2015)[67], India	Unusual IQ scores based on Raven's Standardized Progressive Matrices Test; used only for sensitivity analysis for the *mean-effects meta-analysis*
Aravind et al. (2016)[68], India	Unusually low IQ scores Raven's Standardized Progressive Matrices Test; used only for sensitivity analysis for the *mean-effects meta-analysis*
Jin et al.(2016)[69], China	Cognitive functions other than IQ; potential overlap with Zhang et al. (2015c)[70]
Kumar et al. (2016)[71], India	Seguin Foam Board test; due to the test measuring eye-hand coordination and cognitive-perceptual abilities
Jin et al.(2017)[72], China	Overlap with Jin et al. (2016)[69]; unusual IQ scores reported as percentiles
Razdan et al. (2017)[73], India	Unusually low IQ scores based on Raven's Standardized Progressive Matrices Test; used only for sensitivity analysis for the *mean-effects meta-analysis*
Valdez Jiménez et al. (2017)[74], Mexico	Bayley tests; used only for sensitivity analysis for the *regression slopes meta-analysis*
Wang et al. (2017)[75], China	Overlapping population with Xiang et al. (2003a)[59]

Reference, Country	Reason for Exclusion
Cui et al. (2018)[76], China	Missing mean or SD of outcome measure; used in *regression slopes meta-analysis*
Luo et al. (2018)[77], China	Overlapping population with Lou et al. (2021)[19]
Naik et al. (2018)[78], India	Missing sample sizes by exposure groups. Missing mean and SD for IQ scores
Sharma et al.(2018)[79], India	Missing mean and SD for IQ scores
Soto-Barreras et al. (2019)[80], Mexico	Missing mean or SD of outcome measure
Zhao et al. (2019)[43], China	Overlapping population with Yu et al. (2018)[3], but smaller sample size
Zhou et al. (2019)[44], China	Overlapping population with Yu et al. (2018)[3], but smaller sample size
Till et al.(2020)[81], Canada	Missing mean or SD of outcome measure; used in *regression slopes meta-analysis*
Wang et al. (2020b)[4], China	Missing mean or SD of outcome measure; used in sensitivity analysis for the *regression slopes meta-analysis*
Zhao et al. (2020)[42], China	Overlapping population with Yu et al. (2018)[3], but smaller sample size
Aggeborn and Öhman (2021)[82], Sweden	Cognitive functions other than IQ; cognitive test not specified
Cantoral et al. (2021)[83], Mexico	Bayley tests; used only for sensitivity analysis for the *regression slopes meta-analysis*
Farmus et al. (2021)[84], Canada	Same data as Till et al.(2020)[81]
Guo et al. (2021)[85], China	Overlapping population with Zhao et al. (2018),[86] but smaller sample size; excluded from overall *mean-effects meta-analysis* but used in mean-effects subgroup meta-analysis by age group
Ibarluzea et al. (2021)[87], Spain	Bayley and McCarthy tests; used only for sensitivity analysis for the *mean-effects meta-analysis*, *dose-response meta-analysis*, and *regression slopes meta-analysis*
Wang et al. (2021b)[88], China	Overlapping population with Wang et al. (2021)[89]; cognitive functions other than IQ
Yu et al. (2021)[90], China	Overlapping population with Yu et al. (2018)[3], but smaller sample size
Zhao et al. (2021)[91], China	Missing mean or SD of outcome measure; used in *regression slopes meta-analysis*
Zhou et al. (2021)[92], China	Overlapping population with Yu et al. (2018)[3], but smaller sample size

eTable 2. Study Characteristics and Study-specific Effect Estimates Included in the Meta-analyses and Sensitivity Analyses

Commented [I4]: See Doc05_Meta-analysis, 5.E. (pages 6 and 7) and 5.K. (page 10)

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Ren et al. (1989)[20] [translated in Ren et al. 2008][me, o] *Cross-sectional*	China	8–14	No fluoride measurement Low iodine village/high fluoride and low iodine village	Not specified	169, 85.00 (22.30) 160, 64.80 (20.40)			Subjects, Methods, Results section
Chen et al. (1991)[93] [translated in Chen et al. 2008][me, w] *Cross-sectional*	China	7–14	Drinking water Nonendemic/endemic fluorosis village	0.89 mg/L (nonendemic) 4.55 mg/L (endemic)	320, 104.03 (14.96) 320, 100.24 (14.52)	320, 104.03 (14.96) 320, 100.24 (14.52)		Results section, Table 1
Guo et al. (1991)[10] [translated in Guo et al. 2008a][me, o] *Cross-sectional*	China	7–13	Serum Reference area using wood/coal burning-related fluoride endemic area	0.1044 ± 0.0652 mg/L (reference) 0.1483 ± 0.0473 mg/L (endemic)	61, 81.39 (10.26) 60, 76.71 (10.85)			Calculated by ICF
Lin et al. (1991)[95me, o] *Cross-sectional*	China	7–14	Urine, drinking water Reference area with iodine supplementation/high fluoride and low iodine village	Urine: 1.6 mg/L (reference area with iodine supplementation) 2.56 mg/L (high fluoride, low iodine village) Water: 0.34 mg/L (low iodine village) 0.88 mg/L (high fluoride, low iodine village)	256, 78.00 (40.07) 250, 71.00 (40.07)			Calculated by ICF
Sun et al. (1991)[38me, o] *Cross-sectional*	China	6.5–12	No fluoride measurement Nonendemic area/endemic (aluminum-fluoride endemic toxicosis)	Fluorosis: 98.36% (endemic)	224, 82.68 (10.91) 196, 72.35 (11.36)			Calculated by ICF
An et al. (1992)[24me, w] *Cross-sectional*	China	7–16	Drinking water Nonhigh/high fluoride area	0.6−1.0 mg/L (nonhigh) 2.1−3.2 mg/L (secondary high) 5.2−7.6 mg/L (high) 2.1−7.6 mg/L (combined high)	121, 84.00 (12.10) 121, 75.90 (13.60)	121, 84.00 (12.10) 56, 76.10 (13.90) 65, 75.60 (13.30)		Table 1, Table 2

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Li et al. (1994)[12] [translated in Li et al. 2008b][me, o] *Cross-sectional*	China	12–13	Grain (cooked by burning high-fluoride coal) Reference group (no dental fluorosis)/high fluoride group I (no dental fluorosis)/high fluoride group II (dental fluorosis present)/high fluoride group III (dental fluorosis present)	0.5 mg/kg (reference group) 4.7 mg/kg (group I) 5.2 mg/kg (group II) 31.6 mg/kg (group III)	49, 267.20 (39.50) 36, 240.00 (30.80)			Table 1
Xu et al. (1994)[26me, w] *Cross-sectional*	China	8–14	Drinking water Reference region/low- and high-fluoride regions[b]	0.8 mg/L (reference region) 0.38 mg/L (low fluoride) 1.8 mg/L (high fluoride)	32, 83.83 (9.10) 97, 79.25 (2.25)	32, 83.83 (9.10) 21, 80.21 (8.27) 97, 79.25 (2.25)		Chart 1
Li et al. (1995)[14me, o, u] *Cross-sectional*	China	8–13	Urine, dental fluorosis index (DFI) Nonfluorosis/fluorosis area due to soot from coal burning	1.02 mg/L; DFI: <0.4 (nonfluorosis) 1.81 mg/L; DFI: 0.8 (slight fluorosis) 2.01 mg/L; DFI: 2.5 (medium fluorosis) 2.69 mg/L; DFI: 3.2 (severe fluorosis)	226, 89.90 (10.40) 230, 80.30 (12.90)	226, 89.90 (10.40) 227, 89.70 (12.70) 224, 79.70 (12.70) 230, 80.30 (12.90)		Table 2
Wang et al. (1996)[22] [translated in Wang et al. 2008b][me, o, w] *Cross-sectional*	China	4–7	Drinking water (well) Low/high fluoride regions Fluoride exposure from drinking water, contaminated food, and coal burning	0.58−1.0 mg/L (low) >1.0−8.6 mg/L (high)	83, 101.23 (15.84) 147, 95.64 (14.34)	83, 101.23 (15.84) 147, 95.64 (14.34)		Table 1
Yao et al. (1996)[28me, w] *Cross-sectional*	China	8–12	Drinking water Nonendemic/endemic fluorosis areas	1 mg/L (nonendemic) 2 mg/L (slightly endemic) 11 mg/L (severely endemic)	270, 98.46 (13.21) 78, 92.53 (12.34)	270, 98.46 (13.21) 188, 94.89 (11.15) 78, 92.53 (12.34)		Table 2
Zhao et al. (1996)[96me, w] *Cross-sectional*	China	7–14	Drinking water Low fluoride village (Xinghua)/high fluoride village (Sima)	0.91 mg/L (low) 4.12 mg/L (high)	160, 105.21 (14.99) 160, 97.69 (13.00)	160, 105.21 (14.99) 160, 97.69 (13.00)		Table 1

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Yao (1997)[27me, w*] *Cross-sectional*	China	7–12	Drinking water Nonfluorosis/fluorosis area with water improvements/fluorosis area without water improvements	0.4 mg/L (nonfluorosis area) 0.33 mg/L (fluorosis area with water improvement) 2 mg/L (fluorosis area without water improvement)	314, 99.98 (12.21) 183, 94.89 (11.15)	314, 99.98 (12.21) 326, 97.83 (11.27) 183, 94.89 (11.15)		Section 2.1 Intelligence Tests, page 2
Zhang et al. (1998)[39me, o] *Cross-sectional*	China	4–10	Drinking water Reference/high fluoride group (all observation groups included arsenic exposure)	0.58 mg/L (reference) 0.8 mg/L (high fluoride)	52, 87.69 (11.04) 51, 85.62 (13.23)			Table 1
Lu et al. (2000)[55me, w, u] *Cross-sectional*	China	10–12	Urine, drinking water Low/high fluoride area	Urine: 1.43 ± 0.64 mg/L (low) 4.99 ± 2.57 mg/L (high) Water: 0.37 ± 0.04 mg/L (low) 3.15 ± 0.61 mg/L (high)	58, 103.05 (13.86) 60, 92.27 (20.45)	58, 103.05 (13.86) 60, 92.27 (20.45)		Table 1
Hong et al. (2001)[97] [translated in Hong et al. 2008] [me, w] *Cross-sectional*	China	8–14	Drinking water Reference/high fluoride[b]	0.75 mg/L (reference) 2.90 mg/L (high fluoride)	32, 82.79 (8.98) 85, 80.58 (2.28)	32, 82.79 (8.98) 85, 80.58 (2.28)		Table 2
Hong et al. (2001b)[99me, o] *Cross-sectional*	China	8–14	Urine, drinking water Nonendemic/endemic fluorosis areas (high fluoride, high iodine)	Urine: 0.796 ± 0.53 mg/L (nonendemic) 2.09 ± 1.03 mg/L (endemic) Water: 0.48 mg/L (nonendemic) 2.81 mg/L (endemic)	30, 80.66 (11.93) 31, 75.89 (7.74)			Table 3, Table 4
Wang et al. (2001)[49,me, o] *Cross-sectional*	China	8–12	Urine, drinking water Reference point (low fluoride, low iodine)/investigative point (high fluoride, high iodine)	Urine: 0.82 mg/L (low fluoride, low iodine) 3.08 mg/L (high fluoride, high iodine) Water: 0.5 mg/L (low fluoride, low iodine) 2.97 mg/L (high fluoride, high iodine)	30, 81.67 (11.97) 30, 76.67 (7.75)			Table 2

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Li et al. (2003)[100] [translated in Li et al. 2008c][me] *Cross-sectional*	China	6–13	No fluoride measurement Reference/endemic fluorosis areas	Not specified	236, 93.78 (14.30) 720, 92.07 (17.12)			Table 1
Xiang et al. (2003a)[59,me, w*, u] *Cross-sectional*	China	8–13	Urine, drinking water Nonendemic/endemic fluorosis areas	Urine: 1.11 ± 0.39 mg/L (reference) 3.47 ± 1.95 mg/L (high fluoride) Water: 0.36 ± 0.15 mg/L (nonendemic) 0.75 ± 0.14 mg/L (endemic fluorosis area group A) 1.53 ± 0.27 mg/L (endemic fluorosis area group B) 2.46 ± 0.3 mg/L (endemic fluorosis area group C) 3.28 ± 0.25 mg/L (endemic fluorosis area group D) 4.16 ± 0.22 mg/L (endemic fluorosis area group E) 2.47 ± 0.79 mg/L (high fluoride)	290, 100.41 (13.21) 222, 92.02 (13.00)	290, 100.41 (13.21) 9, 99.56 (14.13) 42, 95.21 (12.22) 111, 92.19 (12.98) 52, 89.88 (11.98) 8, 78.38 (12.68)		Table 6, Table 8
Wang et al. (2005)[102,me, w, u] *Cross-sectional*	China	8–12	Urine, drinking water Reference/high fluoride group[c]	Urine: 1.51 mg/L(reference) 5.09 mg/L (high fluoride group) Water: 0.48 mg/L (reference) 8.31 mg/L (high fluoride group)	196, 112.36 (14.87) 253, 107.83 (15.45)	196, 112.36 (14.87) 253, 107.83 (15.45)		Table 1
Seraj et al. (2006)[29,me, w] *Cross-sectional*	Iran	7–11	Drinking water Low/high fluoride area	0.4 ppm (low) 2.5 ppm (high)	85, 98.90 (12.90) 41, 87.90 (11.00)	85, 98.90 (12.90) 41, 87.90 (11.00)		Methodology, Findings section (Text under Table 2)
Wang et al. (2006)[103,me, w, u] *Cross-sectional*	China	8–12	Urine, drinking water Reference/high (area severely affected by fluorosis)	Urine: 1.51 ± 1.66 mg/L (reference) 5.50 ± 2.40 mg/L (high) Water: 0.73 ± 0.28 mg/L (reference) 5.54 ± 3.88 mg/L (high)	166, 111.55 (15.19) 202, 107.46 (15.38)	166, 111.55 (15.19) 202, 107.46 (15.38)		Table 2

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Fan et al. (2007)[104,me, w, u] *Cross-sectional*	China	7–14	Urine, drinking water Low/high fluoride area	Urine: 1.78 ± 0.46 mg/L (low) 2.89 ± 1.97 mg/L (high) Water: 1.03 mg/L (low) 3.15 mg/L (high)	37, 98.41 (14.75) 42, 96.11 (12.00)	37, 98.41 (14.75) 42, 96.11 (12.00)		Table 1
Trivedi et al. (2007)[41,me, w, u] *Cross-sectional*	India	12–13	Urine, drinking water Low/high fluoride area	Urine: 2.30 ± 0.28 mg/L (low) 6.13 ± 0.67 mg/L (high) Water: 2.01 ± 0.009 mg/L (low) 5.55 ± 0.41 mg/L (high)	101, 104.44 (12.36) 89, 91.72 (10.66)	101, 104.44 (12.36) 89, 91.72 (10.66)		Table 2
Wang et al. (2007)[105,me, o, u, w] *Cross-sectional*	China	8–12	Urine, drinking water Low fluoride, low arsenic/high fluoride, low arsenic area	Urine: 1.5 ± 1.6 mg/L (low fluoride, low arsenic) 5.1 ± 2.0 mg/L (high fluoride, low arsenic) Water: 0.5 ± 0.2 mg/L (low fluoride, low arsenic) 8.3 ± 1.9 mg/L (high fluoride, low arsenic)	196, 104.80 (14.70) 253, 100.50 (15.80)	196, 104.80 (14.70) 253, 100.50 (15.80)		Table 2, Table 3
Li et al. (2009)[15,me, o, u*] *Cross-sectional*	China	8–12	Urine Endemic fluorosis region caused by coal burning (reference/mild/medium /severe) Degree of dental fluorosis (normal/suspected/ very mild/mild/medium/ severe)	0.962 ± 0.517 mg/L (reference) 1.235 ± 0.426 mg/L (mild) 1.670 ± 0.663 mg/L (medium) 2.336 ± 1.128 mg/L (severe) 0.867 ± 0.233 mg/L (normal) 1.094 ± 0.355 mg/L (suspected) 1.173 ± 0.480 mg/L (very mild) 1.637 ± 0.682 mg/L (mild) 2.005 ± 0.796 mg/L (medium) 2.662 ± 1.093 mg/L (severe)	20, 102.70 (17.61) 20, 93.85 (18.11)	20, 102.70 (17.61) 20, 97.30 (18.56) 20, 93.90 (17.60) 20, 93.85 (18.11)		Table 1
Li et al. (2010)[106,me] *Cross-sectional*	China	7–10	No fluoride measurement Nondental fluorosis children/dental fluorosis children	Not specified	329, 97.36 (18.24) 347, 98.73 (21.07)			Table 3

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Ding et al. (2011)[107,me, u*, rs] *Cross-sectional*	China	7–14	Dental fluorosis (normal/ questionable/very mild/mild/ moderate) Urine Mean urinary fluoride levels (10 groups)	0.80 ± 0.55 mg/L (normal) 1.13 ± 0.73 mg/L (questionable) 1.11 ± 0.74 mg/L (very mild) 1.31 ± 0.78 mg/L (mild) 1.46 ± 0.79 mg/L (moderate) 0.26 mg/L (group 1) 0.45 mg/L (group 2) 0.56 mg/L (group 3) 0.66 mg/L (group 4) 0.75 mg/L (group 5) 0.89 mg/L (group 6) 1.08 mg/L (group 7) 1.33 mg/L (group 8) 1.74 mg/L (group 9) 2.96 mg/L (group 10) Range: 0.10−3.55 mg/L	136, 104.07 (12.30) 28, 103.54 (13.59)	136, 104.07 (12.30) 54, 103.00 (16.10) 74, 102.11 (15.05) 39, 106.03 (12.33) 28, 103.54 (13.59)	−0.59 (−1.09, −0.08) per 1 mg/L urinary F	Table 2, Section 3 Results and discussion (under Fig. 2)
Eswar et al. (2011)[31,me, w] *Cross-sectional*	India	12–14	Drinking water Low/high fluoride villages	0.29 mg/L (low) 2.45 mg/L (high)	65, 88.80 (15.30) 68, 86.30 (12.80)	65, 88.80 (15.30) 68, 86.30 (12.80)		Table 1
Kang et al. (2011)[108me, o] *Cross-sectional*	China	6–12	Drinking water Reference/high fluoride areas (both areas high arsenic exposure)	1.24 ± 0.74 mg/L (all children) <1.2 mg/L (reference) ≥1.2 mg/L (high fluoride)	90, 96.8 (12.7) 178, 96.8 (16.3)			Table 1. Section 2.1
Poureslami et al. (2011)[32,me, w] *Cross-sectional*	Iran	7–9	Drinking water Reference/endemic dental fluorosis city	0.41 mg/L (reference) 2.38 mg/L (endemic)	60, 97.80 (15.95) 59, 91.37 (16.63)	60, 97.80 (15.95) 59, 91.37 (16.63)		Table 3, Results section (under Table 3)
Shivaprakash et al. (2011)[33,me, w] *Cross-sectional*	India	7–11	Drinking water No fluorosis/fluorosis severity groups (mild/moderate/ severe)/all fluorosis	<0.5 ppm (no fluorosis) 2.5–3.5 ppm (mild) 2.5–3.5 ppm (moderate) 2.5–3.5 ppm (severe) 2.5–3.5 ppm (all)	80, 76.36 (20.84) 80, 66.63 (18.09)	80, 76.36 (20.84) 80, 66.63 (18.09)		Table 1

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Seraj et al. (2012)[30,me, w] *Cross-sectional*	Iran	6–11	Drinking water Normal/medium/high fluoride levels	0.8 ± 0.3 mg/L (normal) 3.1 ± 0.9 mg/L (medium) 5.2 ± 1.1 mg/L (high)	91, 97.77 (18.91) 96, 88.58 (16.01)	91, 97.77 (18.91) 106, 89.03 (12.99) 96, 88.58 (16.01)		Table 2
Trivedi et al. (2012)[40,me, w, u] *Cross-sectional*	India	12–13	Urine, ground water Low/high fluoride area	Urine: 0.42 ± 0.23 mg/L (low) 2.69 ± 0.92 mg/L (high) Water: 0.84 ± 0.38 mg/L (low) 2.3 ± 0.87 mg/L (high)	50, 97.17 (17.96) 34, 92.58 (18.25)	50, 97.17 (17.96) 34, 92.58 (18.25)		Table 3, Results section (above Table 3)
Wang et al. (2012b)[109me] *Cross-sectional*	China	Primary school age	No fluoride measurement Reference/high fluoride areas	Not specified	455, 98.36 (14.56) 800, 92.21 (18.45)			Table 1
Bai et al. (2014)[16,me, o] *Cross-sectional*	China	8–12	Urine Coal-burning-borne fluorosis areas (reference/lightly-affected/seriously-affected)	0.54 mg/L (reference) 0.81 mg/L (lightly-affected area) 1.96 mg/L (seriously-affected area)	164, 107.92 (13.62) 162, 101.22 (15.97)			Table 2
Karimzade et al. (2014)[37,me, w] *Cross-sectional*	Iran	9–12	Drinking water Low/high fluoride area	0.25 mg/L (low) 3.94 mg/L (high)	20, 104.25 (20.75) 19, 81.21 (16.17)	20, 104.25 (20.75) 19, 81.21 (16.17)		Table 1

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Broadbent *et al.* (2015)[25,me, w*] *Prospective Cohort*	New Zealand	7–13	Drinking water Area without community water fluoridation (low)/area with community water fluoridation (high) Fluoride tablet use (never/ever) Fluoride toothpaste use (never/sometimes/always)	Water: 0.0–0.3 mg/L (low) 0.7–1.0 mg/L (high) Tablet use: 0 mg (never used) 0.5 mg (ever used) Range not specified for fluoride toothpaste use (always/sometimes/never)	99, 99.80 (14.50) 891, 100.00 (15.10)	99, 99.80 (14.50) 891, 100.00 (15.10)		Table 1
Khan et al. (2015)[34,me] *Cross-sectional*	India	6–11	Drinking water Low fluoride areas (Tiwariganj)/high fluoride areas (Unnao) Fluorosis grades (normal/very mild/mild/moderate/severe)	0.19 mg/L (Tiwariganj) 2.41 mg/L (Unnao) Ranges not specified by fluorosis grades	241, 110.10 (9.00) 5, 62.40 (2.40)			Table/Fig-5
Kundu et al. (2015)[67,sa] *Cross-sectional*	India	8−12	Drinking water Low fluoride areas/high fluoride areas	Not specified	100, 85.80 (18.85) 100, 76.20 (19.10)			Table 2
Sebastian and Sunitha (2015)[35,me, w*] *Cross-sectional*	India	10–12	Drinking water Low/normal/high fluoride villages	0.40 mg/L (low) 1.2 mg/L (normal) 2.0 mg/L (high)	135, 86.37 (13.58) 135, 80.49 (12.67)	135, 86.37 (13.58) 135, 88.60 (14.01) 135, 80.49 (12.67)		Table 1, Table 2
Zhang et al. (2015b)[110,me, w*, u, rs] *Cross-sectional*	China	10–12	Urine, drinking water, serum Reference/high fluoride areas	Urine: 1.10 ± 0.67 mg/L (reference) 2.40 ± 1.01 mg/L (high) Water: 0.63 (0.58–0.68) mg/L (reference) 1.40 (1.23–1.57) mg/L (high) Serum: 0.06 ± 0.03 (reference) 0.18 ± 0.11 serum (high)	96, 109.42 (13.30) 84, 102.33 (13.46)	96, 109.42 (13.30) 84, 102.33 (13.46)	−2.42 (−4.59, −0.24) per 1 mg/L urinary F	Table 1, Table 3

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Zhang et al. (2015c)[70me, o] *Cross-sectional*	China	7−13	Urine Coal-burning endemic fluorosis area Reference (no dental fluorosis)/mild dental fluorosis/middle dental fluorosis/critically ill dental fluorosis	0.83 ± 0.71 mg/L (reference) 1.54 ± 0.57 mg/L (mildly ill) 2.41 ± 0.76 mg/L (moderately ill) 3.32 ± 1.02 mg/L (critically ill)	30, 110.34 (11.52) (reference) 30, 90.52 (10.37) (critically ill)			Table 1, Table 3
Aravind et al. (2016)[68,sa] *Cross-sectional*	India	10−12	Drinking water Low/high fluoride levels	<1.2 ppm (low) >2 ppm (high)	96, 41.03 (16.36) 96, 31.59 (16.81)			Table 1
Das and Mondal (2016)[111,me, u] *Cross-sectional*	India	6–18	Urine, drinking water intake Dental fluorosis (normal/questionable/very mild/ mild/ moderate/severe)	Urine: 2.91 ± 1.76 mg/L (normal) 2.50 ± 2.39 mg/L (questionable) 2.58 ± 1.31 mg/L (very mild) 2.95 ± 1.44 mg/L (mild) 4.82 ± 3.57 mg/L (moderate) 3.81 ± 2.51 mg/L (severe) Water: 0.069 ± 0.021 mg/kg-d (normal) 0.064 ± 0.004 mg/kg-d (questionable) 0.060 ± 0.036 mg/kg-d (very mild) 0.060 ± 0.030 mg/kg-d (mild) 0.099 ± 0.063 mg/kg-d (moderate) 0.093 ± 0.040 mg/kg-d (severe)	4, 108.30 (53.20) 23, 85.91 (37.68)	4, 108.30 (53.20) 17, 103.18 (33.35) 27, 107.70 (27.92) 35, 92.83 (26.90) 43, 84.51 (35.16) 23, 85.91 (37.68)		Table 3
Mondal et al. (2016)[36,me, w] *Cross-sectional*	India	10–14	Drinking water Low/high fluoride areas	Not reported (low) 0.33–18.08 mg/L (high)	22, 26.41(10.46) 18, 21.17 (6.77)	22, 26.41 (10.46) 18, 21.17 (6.77)		Table 9

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Bashash et al. (2017)[112,me, u, rs] *Prospective Cohort*	Mexico	6−12	Maternal urine Reference/high fluoride levels (based on children urinary fluoride)	<0.80 mg/L (reference) ≥0.80 mg/L (high)	77, 95.37 (10.31) 112, 96.80 (11.16)	77, 95.37 (10.31) 112, 96.80 (11.16)	−2.50 (−4.12, −0.59) per 0.5 mg/L maternal urinary F	Abstract, Table 3
Razdan et al. (2017)[73,sa] *Cross-sectional*	India	12–14	Drinking water Low/high fluoride levels	0.6 ppm (low) 4.99 ppm (high)	69, 38.61 (6.34) 75, 13.95 (5.14)			Table 2
Valdez Jiménez et al. (2017)[74sa] *Prospective Cohort*	Mexico	Infancy	Maternal urine, drinking water	Urine: 1.9 ± 1.0 mg/L (1st trimester) 2.0 ± 1.1 mg/L (2nd trimester) 2.7 ± 1.1 mg/L (3rd trimester) Water: 2.6 ± 1.1 mg/L (1st trimester) 3.1 ± 1.1 mg/L (2nd trimester) 3.7 ± 1.0 mg/L (3rd trimester)			Bayley MDI: −19.05 (8.9) per 1 log10 mg/L maternal urinary F (1st trimester) −19.34 (7.46) per 1 log10 mg/L maternal urinary F (2nd trimester)	Table 2, Table 4
Cui et al. (2018)[76,rs] *Cross-sectional*	China	7–12	Urine	Boys: 1.3 (0.9−1.7)[d] mg/L Girls: 1.2 (0.9−1.6)[d] mg/L			−2.47 (−4.93, −0.01) per 1 log urinary F	Table 2
Yu et al. (2018)[3,me, w, u*, rs] *Cross-sectional*	China	7–13	Maternal urine Low/medium/high fluoride ranges Drinking water Normal/high fluoride	Urine: 0.01–1.60 mg/L (low) 1.60–2.50 mg/L (medium) 2.50–5.54 mg/L (high) Water: ≤1 mg/L (normal) >1 mg/L (high) Overall: 0.01−5.54 mg/L (urine) 0.20−3.90 mg/L (water)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	0.36 (−0.29, 1.01) per 0.5 mg/L maternal urinary F	Table 1, Table 3
Zhao et al. (2018)[86me, o] *Cross-sectional*	China	7–12	Urine Reference/exposed areas All areas with iodine exposure	≤2.16 mg/L (reference) >2.16 mg/L (exposed)	199, 114.52 (12.72) 100, 109.59 (14.24)			Table 4

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Green et al. (2019)[113,me, w*, u*, rs] *Prospective Cohort*	Canada	3−4	Maternal urine, drinking water, maternal fluoride intake Nonfluoridated/ fluoridated area	Urine: 0.40 ± 0.27 mg/L (nonfluoridated) 0.69 ± 0.42 mg/L (fluoridated) Water: 0.13 ± 0.06 mg/L (nonfluoridated) 0.59 ± 0.08 mg/L (fluoridated) Intake: 0.30 ± 0.26 mg/day (nonfluoridated) 0.93 ± 0.43 mg/day (fluoridated) Overall: 0.51 ± 0.36 mg/L (urine) 0.54 ± 0.44 mg/day (intake) 0.31 ± 0.23 mg/L (water)	238, 108.07 (13.31) 162, 108.21 (13.72)	238, 108.07 (13.31) 162, 108.21 (13.72)	−1.95 (−5.19, 1.28) per 1 mg/L maternal urinary F −5.29 (−10.39, −0.19) per 1 mg/L water F −3.66 (−7.16, 0.15) per 1 mg maternal F intake	Table 2, text page 945, eTable 4
Cui et al. (2020)[114,me, u] *Cross-sectional*	China	7–12	Urine Low/medium/high fluoride levels	<1.6 mg/L (low) 1.6–2.5 mg/L (medium) >=2.5 mg/L (high)	396, 112.16 (11.50) 36, 110.00 (14.92)	396, 112.16 (11.50) 66, 112.05 (12.01) 36, 110.00 (14.92)		Table 1

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Till et al. (2020)[81,rs] *Prospective Cohort*	Canada	3−4	Residence, maternal urine, drinking water, infant fluoride intake from formula Nonfluoridated areas/fluoridated	Urine: 0.38−0.42 mg/L (nonfluoridated) 0.64−0.70 mg/L (fluoridated) Water: 0.13 mg/L (nonfluoridated) 0.58 mg/L (fluoridated) Intake: 0.02−0.08 mg/day (nonfluoridated) 0.12−0.34 mg/day (fluoridated)			−2.69 (−7.38, 2.01) per 0.5 mg/day infant F intake (formula)	Table 2
Wang et al. (2020b)[4,sa] *Cross-sectional*	China	7−13	Urine, drinking water	Urine: 0.01−5.54 mg/L Water: 0.20−3.90 mg/L			−1.214 (−1.987, −0.442) per 1 mg/L urinary F −1.037 (−2.040, −0.035) per 1 mg/L urinary F (males) −1.379 (−2.628, −0.129) per 1 mg/L urinary F (females); −1.587 (−2.607, −0.568) per 1 mg/L water F −1.422 (−2.792, −0.053) per 1 mg/L water F (males) −1.649 (−3.201, −0.097) per 1 mg/L water F (females)	Table 4
Wang et al. (2020c)[18me, o] *Cross-sectional*	China	7−12	Urine Coal-burning endemic fluorosis area Nonendemic/endemic fluorosis regions	0.461 ± 0.210 mg/L (nonendemic) 0.689 ± 0.502 mg/L (endemic)	100, 97 (20.3) 170, 82.5 (21.7)			Section 2.1, Table 2

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Xu et al. (2020)[115] me, u*, rs *Cross-sectional*	China	7−13	Urine Reference/high prenatal exposure only/high childhood exposure only/both prenatal and childhood exposure group	0.82 ± 0.30 mg/L (reference) 0.98 ± 0.29 mg/L (high prenatal exposure only) 2.05 ± 0.58 mg/L (high childhood exposure only) 2.13 ± 0.59 mg/L (both prenatal and childhood exposure group)	228, 123.92 (12.50) 141, 123.04 (11.24)	228, 123.92 (12.50) 107, 119.76 (11.28) 157, 124.65 (10.88) 141, 123.04 (11.24)	−0.055 (−1.626, 1.517) per 1 mg/L urinary F 2.785 (−0.832, 6.403) per 1 mg/L urinary F (<1.7 mg/L) −4.965 (−9.198, −0.732) per 1 mg/L urinary F (≥1.7 mg/L) 4.054 (−3.169, 11.277) per 1 mg/L prenatal urinary F (<1.7 mg/L) −3.929 (−9.396, 1.538) per 1 mg/L prenatal urinary F (≥1.7 mg/L) 3.146 (−1.138, 7.430) per 1 mg/L postnatal urinary F (<1.7 mg/L) −6.595 (−13.323, 0.133) per 1 mg/L postnatal urinary F (≥1.7 mg/L)	Table 1, Table 3, author correspondence
Cantoral et al. (2021)[83sa] *Prospective Cohort*	Mexico	1−2	Maternal fluoride intake	1.12 ± 0.54 mg/day			Bayley III cognitive scores: −1.14 (−3.26, 0.99) per 0.5 mg/L maternal F intake 0.07 (−2.37, 2.51) per 0.5 mg/L maternal F intake (females) −3.50 (−6.58, −0.42) per 0.5 mg/L maternal F intake (males)	Table 3, Table 4

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Guo et al., (2021)[85me] *Cross-sectional*	China	7−12	Urine Reference/exposed areas (all areas with iodine exposure)	1.16 mg/L (reference) 1.29 mg/L (iodine area 1) 2.01 mg/L (iodine area 2)	7−9 years: 71, 116.71 (12.16) 35, 118.11 (12.8) 22, 113.95 (12.26) 10−12 years: 79, 109.86 (12.05) 48, 110.83 (10.58) 44, 105.39 (13.6)			Table 2, Table 3
Ibarluzea et al. (2021)[87sa] *Prospective Cohort*	Spain	1, 4	Maternal urine Nonfluorinated/ fluoridated communities	Urine: 0.38 ± 0.27 mg/L (nonfluorinated) 0.70 ± 0.41 mg/L (fluoridated) Water: <0.1 mg/L (nonfluorinated) 0.81 ± 0.15 mg/L (fluoridated)	Bayley MDI scores: 153, 97.696 (14.91) 160, 100.395 (15.411) McCarthy GCI scores: 123, 98.666 (15.531) 124, 101.473 (15.423)	Bayley MDI scores: 153, 97.696 (14.91) 160, 100.395 (15.411) McCarthy GCI scores: 123, 98.666 (15.531) 124, 101.473 (15.423)	Bayley MDI scores: 4.67 (−1.78, 11.13) per 1 mg/L maternal urinary F 7.86 (−1.68, 17.40) per 1 mg/L maternal urinary F (males) 1.77 (−7.32, 10.87) per 1 mg/L maternal urinary F (females) McCarthy GCI scores: −2.16 (−8.56, 4.23) per 1 mg/L maternal urinary F −1.79 (−11.85, 8.27) per 1 mg/L maternal urinary F (males) −3.60 (−12.07, 4.86) per 1 mg/L maternal urinary F (females)	Section 2.2, author correspondence
Lou et al. (2021)[19me, o] *Cross-sectional*	China	8−12	Coal-burning endemic fluorosis area No fluoride measurement Nondental fluorosis children/dental fluorosis children	Not specified	44, 96.64 (11.70) 55, 88.51 (12.77)			Table 4

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Saeed et al. (2021)[116me, o, rs] *Cross-sectional*	Pakistan	5−16	Urine, drinking water Reference/high fluoride areas Co-exposure with arsenic	Urine: 0.24 ± 0.15 mg/L (reference) 3.27 ± 2.60 mg/L (high fluoride) Water: 0.15 ± 0.13 mg/L (reference) 5.64 ± 3.52 mg/L (high fluoride)	30, 100.93 (13.10) 118, 97.26 (15.39)		−3.54 (0.50) per 1 mg/L urinary F	Table 1, Table 3
Wang et al. (2021)[89me, w] *Cross-sectional*	China	9−11	Drinking water Reference/high fluoride areas	1.0 ± 0.07 mg/L (reference) 2.8 ± 0.06 mg/L (high fluoride)	303, 109.0 (14.4) 275, 102.1 (16.3)	303, 109.0 (14.4) 275, 102.1 (16.3)		Section 2.1, Table 2
Zhao et al. (2021)[91rs] *Cross-sectional*	China	6−11	Urine Nonendemic/endemic fluorosis areas	1.03 (0.72, 1.47) mg/L			−5.957 (−9.712, −2.202) per 1 log urinary F	Section 3.1, Table 3

Notes:

SD = standard deviation; SE = standard error; MDI = Mental Development Index; GCI = General Cognitive Index

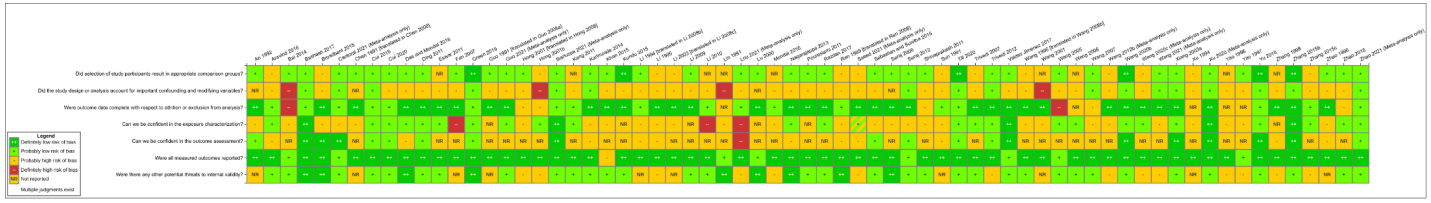
[a]An "me" superscript indicates that the studies included in the *mean-effects meta-analysis*; an "o" superscript indicates a study included in "other" exposures mean-effects analysis (see Table 2 footnote in the main publication); a "w" superscript indicates studies included in the *mean-effects dose-response meta-analysis* using fluoride in water; a "u" superscript indicates studies included in the *mean-effects dose-response meta-analysis* using fluoride in urine; "*" indicates studies included in the *mean-effects dose-response meta-analysis* at levels < 1.5 mg/L; an "rs" superscript indicates studies included in the *regression slopes meta-analysis*.

[b]Additional exposure regions including iodine levels were not included in the analysis.

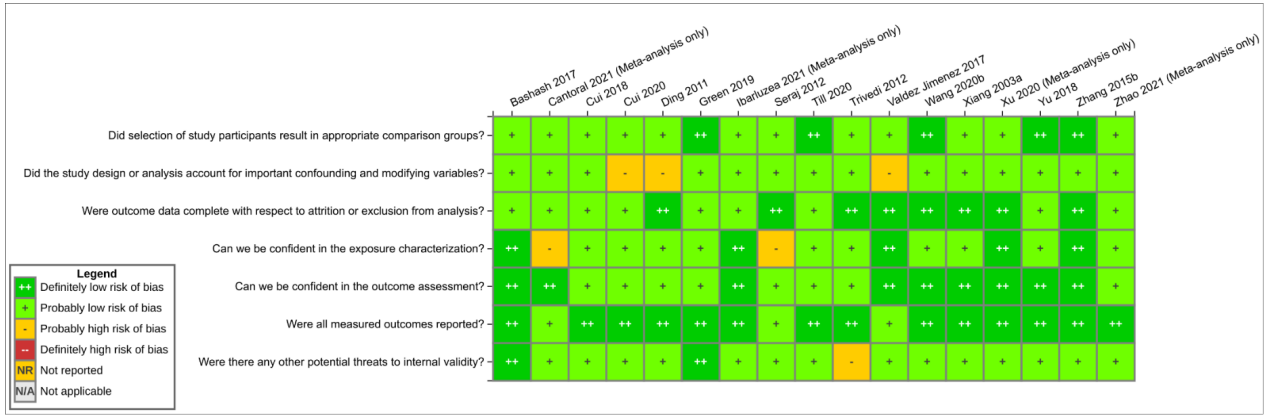
[c]Additional exposure regions including arsenic levels were not included in the analysis.

[d]Median (q1−q3).

(a) All Studies



(b) Low Risk-of-bias Studies



eFigure 2. Results from Risk-of-bias Evaluations for Studies Included in the Meta-analyses and Sensitivity Analyses[a]

Panel (a) presents risk-of-bias results for all studies. An interactive version of eFigure 2(a) is available here: https://hawcproject.org/summary/visual/assessment/405/eFigure-2-Meta-analysis-RoB/. Panel (b) presents risk-of-bias results for low risk-of-bias studies only. An interactive version of eFigure 2(b) is available here: https://hawcproject.org/summary/visual/assessment/405/eFigure-2b-Meta-analysis-RoB-low-RoB-studies/.

The following studies are included in the *mean-effects meta-analysis* and *mean-effects dose-response meta-analysis*: Bashash et al. (2017),[112] Cui et al. (2020),[114] Ding et al. (2011),[107] Green et al. (2019),[113] Seraj et al. (2012),[30] Trivedi et al. (2012),[40] Xiang et al. (2003a),[59] Xu et al. (2020),[115] Yu et al. (2018),[3] and Zhang et al.(2015b).[110]

The following studies are included in the *regression slopes meta-analysis*: Bashash et al. (2017),[112] Cui et al. (2018),[76] Ding et al. (2011),[107] Green et al. (2019),[113] Till et al. (2020),[81] Xu et al. (2020),[115] Yu et al. (2018),[3] Zhang et al.(2015b),[110] and Zhao et al. (2021).[91]

Four studies are only included in sensitivity analyses. All four of these studies are included in sensitivity analyses for the *regression slopes meta-analysis* and include Cantoral et al. (2021),[83] Ibarluzea et al. (2021),[87] Valdez Jiménez et al. (2017),[74] and Wang et al. (2020b).[4] Ibarluzea et al. (2021)[87] is also included in sensitivity analyses for the *mean-effects meta-analysis* and *mean-effects dose-response meta-analysis*.

Mean-effects Meta-analysis

in fluoridated *vs*. non-fluoridated areas in Canada,[113] or in New Zealand.[25] No other studies included in the main *mean-effects meta-analysis* made comparisons between fluoridated *vs*. non-fluoridated areas. In both studies, levels of fluoride in water were low, even in communities with fluoridated drinking water, likely limiting the power to detect an effect.

In Bashash et al.,[112] the SMD compares mean IQ scores in children with urinary fluoride levels below *vs*. above 0.80 mg/L in Mexico.[112] Unlike other studies in the *mean-effects meta-analysis* which compared mean IQ scores between fluoridated *vs*. non-fluoridated areas, or areas with high *vs.* low fluoride exposures (see **eTable 2**), the Bashash et al.[112] study was not designed to measure fluoride exposure by geographical area. However, since the mean IQ scores were provided in the manuscript for children with urinary fluoride levels below *vs*. above 0.80 mg/L, we included them in this analysis. It's worth noting that there was no significant difference when comparing MUF levels between the groups of children with urinary fluoride levels above or below 0.80 mg/L, however when children's IQs were regressed against MUF, a statistically significant inverse association was found.

Commented [I5]: See Doc06a_Meta-analysis, 6a.M., page 8 and 9

Commented [I6]: See Doc06b_Meta-analysis, 6b.C., page 2 and 3

Meta-regression results

The results of the meta-regression models indicate that year of publication and mean age of study children did not explain a large degree of heterogeneity as neither were significant predictors of the relationship between fluoride and children's intelligence, and the residual I^2 remained high (85% and 87%, respectively). Year of publication (SMD = 0.01, 95% CI: -0.01, 0.02) and mean age (SMD = -0.04, 95% CI: -0.13, 0.04) explained relatively little between-study variance (adjusted R^2 of 12% and 5%, respectively). When both year of publication and mean age were included in the model, there were no notable improvements to the amount of between-study variance explained (adjusted R^2 = 13%) or percent residual variation due to heterogeneity (residual I^2 = 85%).

Excluding the outlier study[34] resulted in a slightly lower heterogeneity for the overall effect estimate (I^2=84%) and for the India-specific effect estimate (I^2=69%). The meta-regression indicates that mean age is a significant predictor of the effect (SMD = -0.06, 95% CI: −0.12, −0.01, p-value =0.025), explaining 9% of the between-study variance. Year of publication (SMD = 0.01, 95% CI: 0.001, 0.02, p-value=0.028) explained a larger degree of between-study variance (R^2 = 19 %).

Commented [I7]: See Doc06b_Meta-analysis, 6b.G., page 7 and 8

Commented [I8]: See Doc08_Meta-analysis, 8.E., page 4 and 5

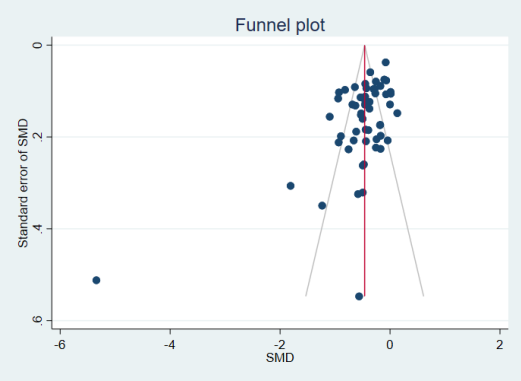
Mean-effects meta-analysis sensitivity analyses

eTable 3. Sensitivity Analyses for Mean-effects Meta-analysis: Pooled SMDs and 95% CIs for Children's IQ Score and Exposures to Fluoride

Commented [l9]: See Doc05_Meta-analysis, 5.D., page 4 and 5

Analysis	Number of Studies	SMD (95% CI)	Heterogeneity	
			p-value	I^2
Excluding Khan et al. (2015)[34]				
	54	−0.43 (−0.51, −0.34)	<0.001	84%
Excluding Lin et al. (1991)[95]				
	54	−0.47 (−0.56, −0.37)	<0.001	87%
Excluding Li et al. (1994)[12] [translated in Li et al. 2008b]				
	54	−0.46 (−0.55, −0.36)	<0.001	87%
Excluding Trivedi et al. (2012)[40]				
	54	−0.46 (−0.56, −0.37)	<0.001	87%
Low risk of bias studies, excluding Trivedi et al. (2012)[40]				
	9	−0.22 (−0.40, −0.04)	<0.001	85%
Including Ibarluzea et al. (2021),[87] Bayley MDI score				
	56	−0.45 (−0.54, −0.36)	<0.001	88%
Including Ibarluzea et al. (2021),[87] McCarthy GCI score				
	56	−0.45 (−0.54, −0.36)	<0.001	87%
Including Aravind et al. (2016),[68] Kundu et al. (2015),[67] Razdan et al. (2017)[73]				
	58	−0.52 (−0.62, −0.42)	<0.001	93%
Including Aravind et al. (2016),[68] Kundu et al. (2015),[67] Razdan et al. (2017)[73], Ibarluzea et al. (2021),[87] Bayley MDI score				
	59	−0.51 (−0.61, −0.40)	<0.001	91%
Including Aravind et al. (2016),[68] Kundu et al. (2015),[67] Razdan et al. (2017)[73], Ibarluzea et al. (2021),[87] McCarthy GCI score				
	59	−0.51 (−0.61, −0.40)	<0.001	91%
Any exposure group				
	55	−0.44 (−0.54, −0.34)	<0.001	91%

Notes:
CI = confidence interval; SMD = standardized weighted mean difference; MDI = Mental Development Index; GCI = General Cognitive Index.



eFigure 3. Funnel Plot of Included Studies

This funnel plot shows individual studies included in the analysis according to random-effect standardized weighted mean difference (SMD) estimates (x-axis) and the standard error (SE) of each study-specific SMD (y-axis). The solid vertical line indicates the pooled SMD estimate for all studies combined and the dashed lines indicated pseudo 95% confidence limits around the pooled SMD estimate.

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
           beta1 =     -3.20
      SE of beta1 =     0.576
                z =     -5.55
       Prob > |z| =    0.0000

Begg's test for small-study effects

Kendall's score =   -299.00
    SE of score =   137.750
                z =     -2.18
     Prob > |z| =    0.0305
```

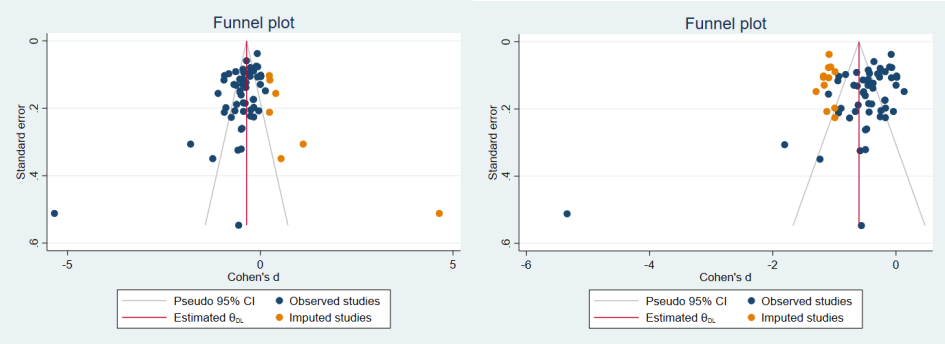
eFigure 4. Test for Publication Bias

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                           Number of studies =     62
  Model: Random-effects                    observed =     55
 Method: DerSimonian-Laird                 imputed =      7

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------
             Studies |        Cohen's d    [95% conf. interval]
---------------------+--------------------------------------
            Observed |           -0.461      -0.554      -0.368
  Observed + Imputed |           -0.357      -0.459      -0.255
------------------------------------------------------------
```

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration                           Number of studies =     67
  Model: Random-effects                    observed =     55
 Method: DerSimonian-Laird                 imputed =      12

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------
             Studies |        Cohen's d    [95% conf. interval]
---------------------+--------------------------------------
            Observed |           -0.461      -0.554      -0.368
  Observed + Imputed |           -0.601      -0.713      -0.489
------------------------------------------------------------
```

eFigure 5. Trim-and-fill Analysis

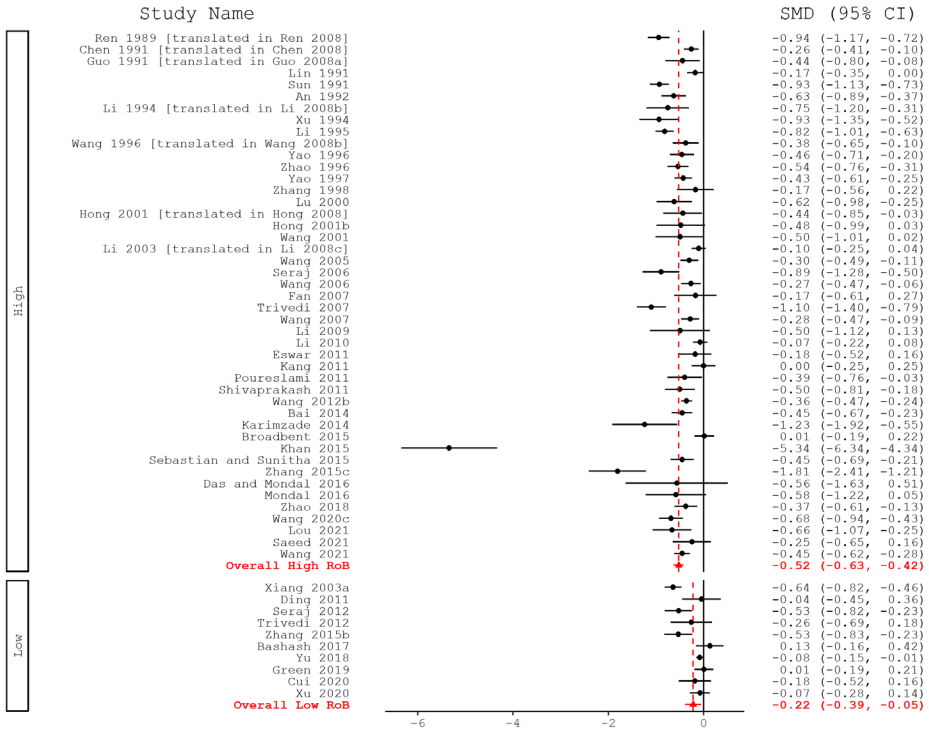
Left panel shows the random-effects pooled SMD after filling in to the right using a run estimator (the linear estimator to the right showed no change in pooled SMD); right panel shows random-effects pooled SMD after filling in to the left using a linear estimator (the run estimator to the left showed no change in pooled SMD).



eFigure 6. Filled-in Funnel Plots to Eliminate Publication Bias

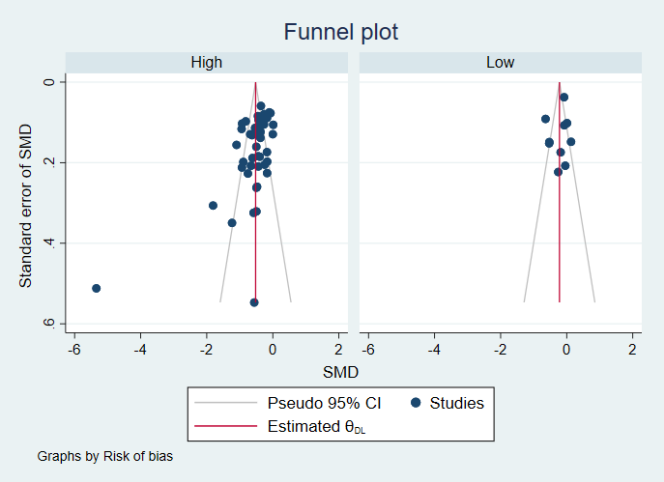
Left panel shows the funnel plot filled in to the right using a run estimator (the linear estimator to the right showed no change in pooled SMD); right panel shows the funnel plot filled in to the left using a linear estimator (the run estimator to the left showed no change in pooled SMD).

Risk-of-bias Subgroup Analysis



eFigure 7. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Risk of Bias

Commented [EAM10]: See Doc06b_Meta-analysis, 6b.W., page 19 through 21.



eFigure 8. Funnel Plot by Risk-of-bias Evaluation

Commented [EAM11]: See Doc06b_Meta-analysis, 6b.W., page 19 through 21.

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird                    High RoB

H0: beta1 = 0; no small-study effects
            beta1 =     -3.41
      SE of beta1 =     0.618
                z =     -5.52
       Prob > |z| =    0.0000

. *meta bias if rob==1, begg rob==1
. meta bias if rob==2, egger random(dl) nometashow

Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird                    Low RoB

H0: beta1 = 0; no small-study effects
            beta1 =     -0.17
      SE of beta1 =     1.835
                z =     -0.09
       Prob > |z| =    0.9275
```

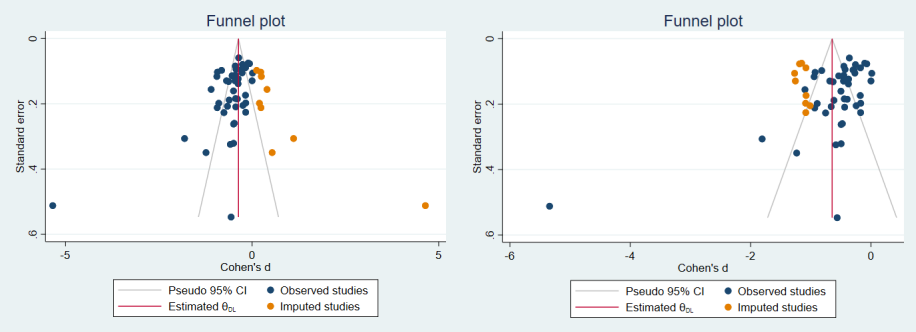
eFigure 9. Test for Publication Bias by Risk of Bias

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                           Number of studies =     54
  Model: Random-effects                     observed =     45
 Method: DerSimonian-Laird                  imputed =      9

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------
             Studies |        Cohen's d    [95% conf. interval]
---------------------+--------------------------------------
            Observed |           -0.521      -0.625      -0.416
  Observed + Imputed |           -0.365      -0.484      -0.246
------------------------------------------------------------
```

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration                           Number of studies =     54
  Model: Random-effects                     observed =     45
 Method: DerSimonian-Laird                  imputed =      9

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------
             Studies |        Cohen's d    [95% conf. interval]
---------------------+--------------------------------------
            Observed |           -0.521      -0.625      -0.416
  Observed + Imputed |           -0.646      -0.765      -0.526
------------------------------------------------------------
```

eFigure 10. Trim-and-fill Analysis for High Risk-of-bias Studies

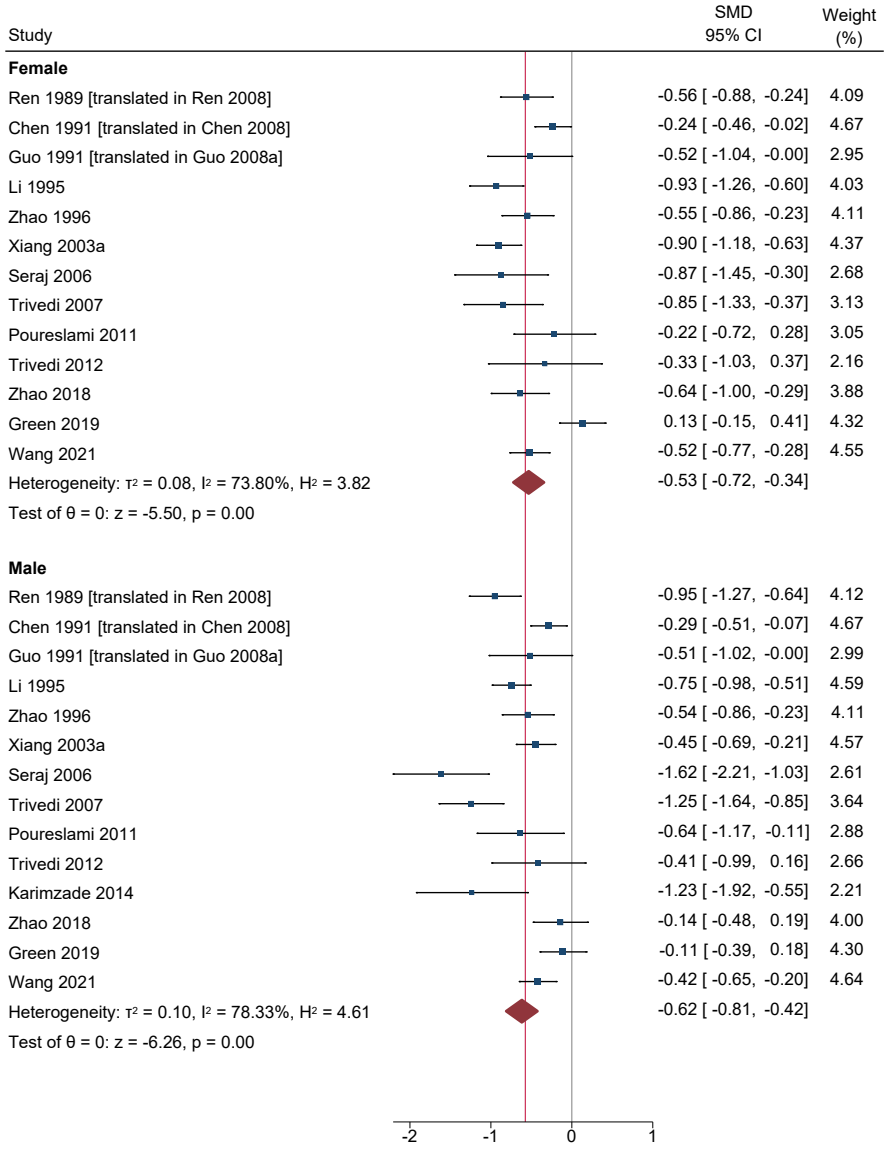
Filling in to the right using a linear estimator or to the left using a run estimator showed no change in the pooled SMD.



eFigure 11. Filled-in Funnel Plots for High Risk-of-bias Studies

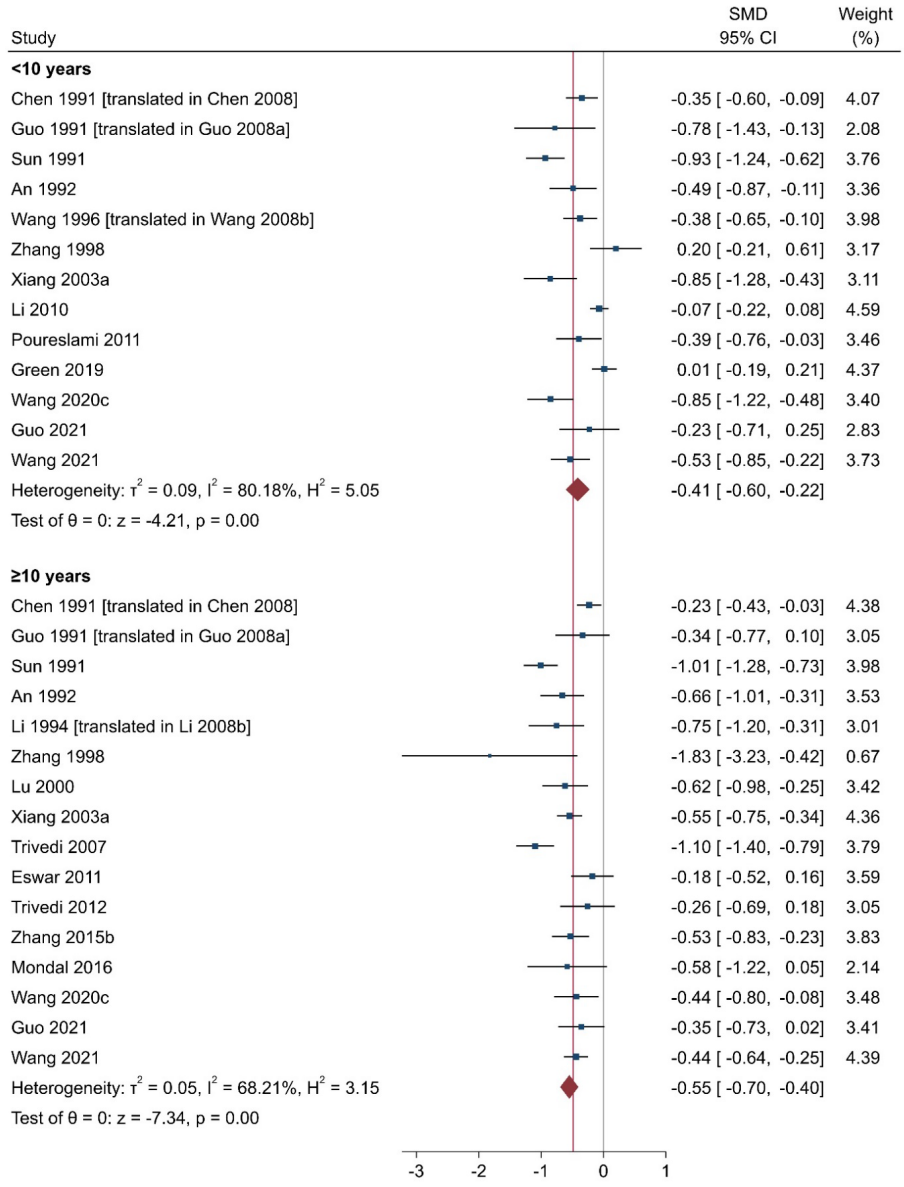
Left panel shows the random-effects pooled SMD after filling in to the right using a run estimator (the linear estimator to the right showed no change in the pooled SMD); right panel shows random-effects pooled SMD after filling in to the left using a linear estimator (the run estimator to the left showed no change in the pooled SMD).

Sex Subgroup Analysis



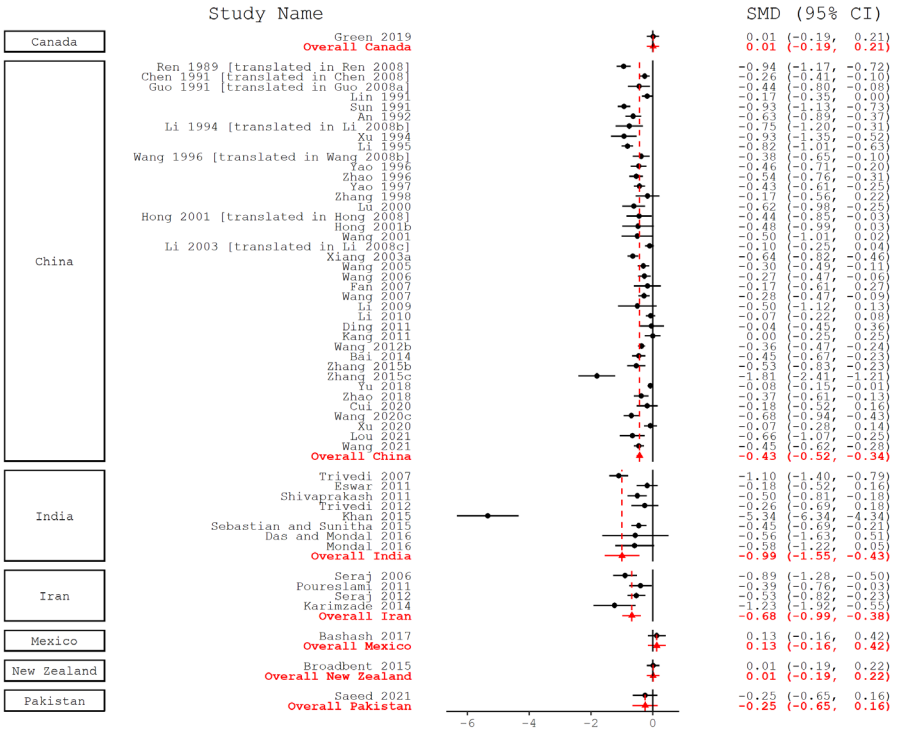
eFigure 12. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Sex

Age Group Subgroup Analysis



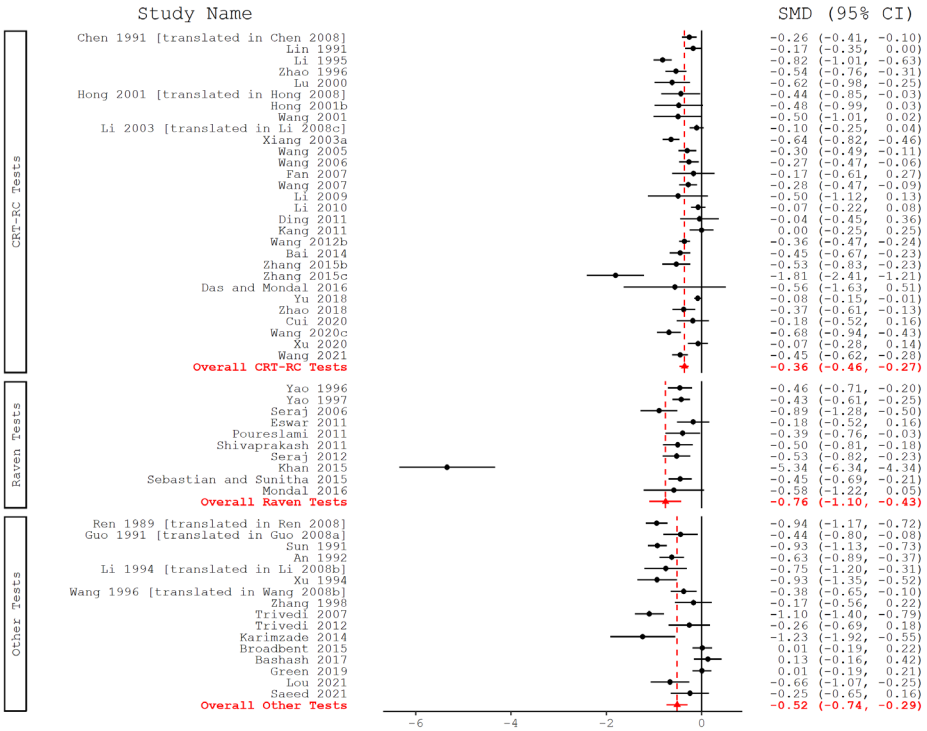
eFigure 13. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Age Group

Country Subgroup Analysis



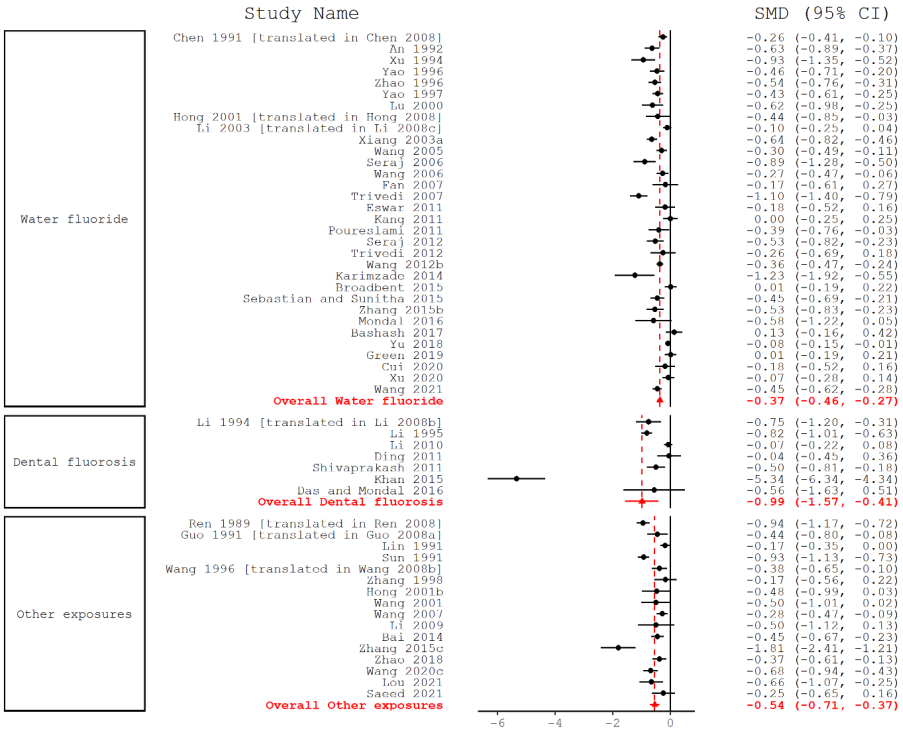
eFigure 14. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Country

Assessment Type Subgroup Analysis



eFigure 15. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Assessment Type

Exposure Type Subgroup Analysis



eFigure 16. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Exposure Type

Exposure types include water, dental fluorosis, and other exposures (iodine, arsenic, aluminum, and fluoride from coal burning).

Dose-Response Meta-analysis Using Mean Effect Estimates

When analyses were restricted to exposed groups with <4 mg/L (i.e., 0 to <4 mg/L) fluoride in drinking water (n = 21 publications [6 low and 15 high risk-of-bias studies]), there was a statistically significant inverse association between fluoride exposure and children's IQ (SMD: −0.22; 95% CI: −0.27, −0.17; p-value < 0.001) (**eTable 4**). When restricted to <2 mg/L (i.e., 0 to <2 mg/L) in drinking water (n = 7 publications [3 low and 4 high risk-of-bias studies]), the magnitude of the effect estimate did not substantially change (SMD: −0.15; 95% CI: −0.41, 0.12; p-value = 0.274). However, when restricted to exposed groups with <1.5 mg/L (i.e., 0 to <1.5 mg/L) in drinking water (n = 7 publications [3 low and 4 high risk-of-bias studies]), there was no longer an association between fluoride in drinking water and children's IQ (SMD: 0.05; 95% CI: −0.36, 0.45; p-value = 0.816). When analyses were further restricted to low risk-of-bias publications at <4 mg/L, <2 mg/L, and <1.5 mg/L, the associations remained in the same direction and were larger in magnitude compared to when data from both low and high risk-of-bias studies were combined (**eTable 4** and **eTable 5**).

When analyses were restricted to exposed groups with <4 mg/L urinary fluoride (n = 13 publications [9 low and 4 high risk-of-bias studies]), there was a statistically significant inverse association between children's urinary fluoride exposure and IQ (SMD: −0.17; 95% CI: −0.30, −0.05; p-value = 0.005) (**eTable 4**). When restricted to <2 mg/L urinary fluoride (n = 7 publications [5 low and 2 high risk-of-bias studies]), there was an inverse association (SMD: −0.06; 95% CI: −0.14, 0.01; p-value = 0.094). When restricted to exposed groups with <1.5 mg/L urinary fluoride (n = 5 publications [4 low and 1 high risk-of-bias studies]), there was an inverse association (SMD: −0.09; 95% CI: −0.16, −0.01; p-value = 0.026). When analyses were further restricted to low risk-of-bias publications, the associations at <2 mg/L and <1.5 mg/L became smaller in magnitude and were statistically significant at <1.5 mg/L (p-value = 0.472 and p-value = 0.028, respectively) (**eTable 4**). Similar results were observed when the maximum likelihood estimation method was used (**eTable 5**).

Commented [I12]: See Doc06a_Meta-analysis, 6a.L., page 7 and 8

Commented [I13]: See Doc06a_Meta-analysis, 6a.L., page 7 and 8

Commented [I14]: See Doc06a_Meta-analysis, 6a.L., page 7 and 8

Commented [I15]: See Doc01_Meta-analysis, 1.P. (page 8) and Doc06b_Meta-analysis, 6b.EE. (page 24 and 25)

eTable 4. Dose-Response Meta-analysis Using Mean Effects—Model Selection[a]

Commented [I16]: See Doc01_Meta-analysis, 1.K., Page 4 and 5

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Water Fluoride – All Studies					
No. Studies/No. Observations Number of Children		29/39 11,656	21/27 8,723	7/9 2,971	7/7 2,832
Linear Model[b]	Beta (95% CI) p-value AIC	−0.15 (−0.20, −0.11) p < 0.001 AIC = 53.8	−0.22 (−0.27, −0.17) p < 0.001 AIC = 16.1	−0.15 (−0.41, 0.12) p = 0.274 AIC = 11.8	0.05 (−0.36, 0.45) p = 0.816 AIC = 8.2
Quadratic Model[c]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.27 (−0.34, −0.21); p < 0.001 0.02 (0.01, 0.03); p < 0.001 AIC = 48.8 p* < 0.001	−0.12 (−0.35, 0.11); p = 0.318 −0.04 (−0.10, 0.03); p = 0.280 AIC = 21.2 p* = 0.012	0.79 (−0.01, 1.58); p = 0.052 −0.56 (−0.97, −0.16); p = 0.006 AIC = 12.5 p* = 0.007	0.30 (−0.53, 1.14); p = 0.477 −0.23 (−1.01, 0.55); p = 0.561 AIC = 11.3 p* = 0.04
Restricted Cubic Splines Model[d]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.29 (−0.39, −0.20); p < 0.001 0.48 (0.18, 0.78); p = 0.002 AIC = 42.3 p* < 0.001	−0.14 (−0.34, 0.06), p = 0.162 −0.23 (−0.66, 0.20), p = 0.295 AIC = 16.9 p* = 0.009	1.15 (0.07, 2.22) p = 0.037 −1.20 (−2.03, −0.36) p = 0.005 AIC = 10.5 p* = 0.010	0.49 (−0.50, 1.47) p = 0.334 −0.69 (−2.40, 1.02) p = 0.428 AIC = 10.2 p* = 0.05
Water Fluoride – Low Risk-of-bias Studies					
No. Studies/No. Observations Number of Children		6/11 4,355	6/9 4,251	3/4 921	3/3 879
Linear model	Beta (95% CI) p-value AIC	−0.19 (−0.34, −0.05) p = 0.009 AIC = 10.3	−0.22 (−0.36, −0.07) p = 0.003 AIC = 3.9	−0.34 (−0.72, 0.03) p = 0.070 AIC = 4.5	−0.32 (−0.91, 0.26) p = 0.276 AIC = 4.1

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Urinary Fluoride – All Studies					
No. Studies/No. Observations Number of Children		18/32 8,502	13/26 6,885	7/11 4,654	5/8 3,992
Linear Model[b]	Beta (95% CI) p-value AIC	−0.16 (−0.24, −0.08) p < 0.001 AIC = 73.8	−0.17 (−0.30, −0.05) p = 0.005 AIC = 68.0	−0.06 (−0.14, 0.01) p = 0.094 AIC = 1.2	−0.09 (−0.16, −0.01) p = 0.026 AIC= 2.8
Quadratic Model[c]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.10 (−0.31, 0.11); p = 0.360 −0.01 (−0.05, 0.02); p = 0.496 AIC = 84.3 p* = 0.14	0.07 (−0.23, 0.38); p = 0.645 −0.07 (−0.16, 0.01); p = 0.071 AIC = 75.8 p* = 0.08	−0.22 (−0.65, 0.20); p = 0.303 0.08 (−0.13, 0.30); p = 0.456 AIC = 9.2 p* = 0.42	0.65 (−1.46, 2.76); p = 0.548 −0.66 (−2.11, 0.80); p = 0.379 AIC = 8.3 p* = 0.10
Restricted Cubic Splines Model[d]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.12 (−0.28, 0.04); p = 0.150 −0.10 (−0.43, 0.23); p = 0.545 AIC = 79.6 p* = 0.13	−0.03 (−0.22, 0.16); p = 0.741 −0.24 (−0.47, −0.002); p = 0.048 AIC = 73.3 p* = 0.07	−0.14 (−0.32, 0.04); p = 0.130 0.13 (−0.17, 0.43); p = 0.395 AIC = 8.5 p* = 0.37	−0.52 (−1.65, 0.62); p = 0.371 0.63 (−1.32, 2.59); p = 0.524 AIC = 6.7 p* = 0.07
Urinary Fluoride – Sensitivity analysis including Ibarluzea et al. **(2021)[87] Bayley MDI scores**					
No. Studies/No. Observations Number of Children		19/33 8,815	14/27 7,445	8/12 4,967	6/9 4,305
Linear model	Beta (95% CI) p-value AIC	−0.15 (−0.23, −0.07) p < 0.001 AIC = 75.0	−0.15 (−0.28, −0.03) p = 0.015 AIC = 69.0	−0.04 (−0.14, 0.05) p = 0.371 AIC = 1.7	−0.08 (−0.15, −0.003) p = 0.043 AIC = 3.6
Urinary Fluoride – Sensitivity analysis including Ibarluzea et al. **(2021)[87] McCarthy GCI scores**					
No. Studies/No. Observations Number of Children		19/33 8,749	14/27 7,445	8/12 4,901	6/9 4,239

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Linear model	Beta (95% CI) p-value AIC	−0.15 (−0.23, −0.07) p < 0.001 AIC = 74.5	−0.16 (−0.28, −0.04) p = 0.011 AIC = 68.6	−0.05 (−0.14, 0.04) p = 0.259 AIC = 1.3	−0.08 (−0.16, −0.01) p = 0.036 AIC = 3.0
Urinary Fluoride – Low Risk-of-bias Studies					
No. Studies/No. Observations Number of Children		9/15 5,713	9/15 5,713	5/8 4,141	4/7 3,952
Linear model	Beta (95% CI) p-value AIC	−0.10 (−0.21, 0.01) p = 0.082 AIC = 5.9	−0.10 (−0.21, −0.01) p = 0.082 AIC = 5.9	−0.05 (−0.17, 0.08) p = 0.472 AIC = 2.8	−0.08 (−0.16, −0.01) p = 0.028 AIC = 2.5

Notes:

AIC = Akaike information criterion; SMD = standardized mean difference; p = p-value for effect estimate; p* = p-value for likelihood ratio tests; MDI = Mental Development Index; GCI = General Cognitive Index

[a]Parameter estimates are changes in SMDs (beta [95% CI]) based on the restricted maximum likelihood models; model fit is represented by the maximum likelihood AIC.

[b]The estimates represent change in SMD for the linear model and AIC, respectively.

[c]The estimates represent change in SMD for the linear term, change in SMD for quadratic term, AIC, and p-values for likelihood ratio test versus linear model, respectively. Potential departure from a linear trend was assessed by testing the coefficient of the quadratic term equal to zero.

[d]The estimates represent change in SMD for the first spline term, change in SMD for the second spline term, AIC, and p-value for likelihood ratio test vs linear model, respectively. Potential departure from a linear trend was assessed by testing the coefficient of the second spline equal to zero.

eTable 5. Dose-response Meta-analysis Using Mean Effects: Maximum Likelihood Models[a]

Commented [I17]: See Doc01_Meta-analysis, 1.K., page 4 and 5

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Water Fluoride – All Studies					
No. Studies/No. Observations Number of Children		29/39 11,656	21/27 8,723	7/9 2,971	7/7 2,832
Linear Model[b]	Beta (95% CI) p-value AIC	−0.15 (−0.20, −0.11) p < 0.001 AIC = 47.9	−0.22 (−0.27, −0.17) p < 0.001 AIC = 10.5	−0.15 (−0.39, 0.08) p = 0.202 AIC = 9.6	0.02 (−0.33, 0.36) p = 0.928 AIC = 6.7
Quadratic Model[c]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.26 (−0.32, −0.20); p < 0.001 0.02 (0.01, 0.03); p < 0.001 AIC = 33.0 p* < 0.001	−0.11 (−0.33, 0.11); p = 0.332 −0.04 (−0.10, 0.02); p = 0.229 AIC= 10.2 p* = 0.012	0.64 (0.04, 1.24); p = 0.036 −0.49 (−0.81, −0.16); p = 0.003 AIC = 8.2 p* = 0.007	0.34 (−0.37, 1.04); p = 0.349 −0.26 (−0.88, 0.35); p = 0.405 AIC = 8.5 p* = 0.04
Restricted Cubic Splines Model[d]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.29 (−0.38, −0.21); p < 0.001 0.48 (0.20, 0.78); p = 0.001 AIC = 33.9 p* < 0.001	−0.13 (−0.32, 0.05); p = 0.162 −0.24 (−0.65, 0.16); p = 0.233 AIC= 9.7 p* = 0.009	0.27 (−0.09, 0.62); p = 0.140 −0.44 (−0.83, −0.04); p = 0.029 AIC = 8.9 p* = 0.010	0.26 (−0.26, 0.79); p = 0.321 −0.49 (−1.54, 0.56); p = 0.363 AIC = 8.7 p*= 0.05
Water Fluoride – Low Risk-of-bias Studies					
No. Studies/No. Observations Number of Children		6/11 4,355	6/9 4,251	3/4 921	3/3 879
Linear model	Beta (95% CI) p-value AIC	−0.19 (−0.31, −0.06) p = 0.003 AIC = 6.7	−0.21 (−0.33, −0.09) p = 0.001 AIC = 0.3	−0.35 (−0.63, −0.07) p = 0.015 AIC = 2.7	−0.34 (−0.80, 0.12) p = 0.153 AIC = 3.3

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Urinary Fluoride – All Studies					
No. Studies/No. Observations Number of Children		18/32 8,502	13/26 6,885	7/11 4,654	5/8 3,992
Linear Model[b]	Beta (95% CI) p-value AIC	−0.16 (−0.23, −0.08) p < 0.001 AIC = 69.2	−0.17 (−0.29, −0.06) p = 0.004 AIC = 64.2	−0.07 (−0.13, 0.003) p = 0.060 AIC = −3.7	−0.12 (−0.36, 0.12) p = 0.325 AIC = 0.8
Quadratic Model[c]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.19 (−0.44, 0.06); p = 0.131 0.01 (−0.02, 0.05); p = 0.462 AIC = 73.0 p*= 0.14	0.08 (−0.21, 0.37); p = 0.587 −0.08 (−0.16, 0.0004); p = 0.051 AIC = 67.2 p* = 0.08	−0.23 (−0.62, 0.17); p = 0.267 0.08 (−0.12, 0.29); p = 0.423 AIC = 1.7 p* = 0.42	−0.11 (−1.45, 1.23); p = 0.868 0.02 (−0.74, 0.77); p = 0.967 AIC = 4.1 p* = 0.10
Restricted Cubic Splines Model[d]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.12 (−0.28, 0.04); p = 0.138 −0.10 (−0.41, 0.21); p = 0.524 AIC = 72.9 p*= 0.13	−0.03 (−0.21, 0.15); p = 0.775 −0.24 (−0.47, −0.02); p = 0.034 AIC = 66.8 p* = 0.07	−0.13 (−0.29, 0.03); p = 0.107 0.12 (−0.14, 0.38); p = 0.366 AIC = 1.5 p* = 0.37	−0.26 (−0.72, 0.20); p = 0.270 0.36 (−0.58, 1.29); p = 0.453 AIC = 3.5 p* = 0.07
Urinary Fluoride – Sensitivity analysis including Ibarluzea et al. **(2021)[87] Bayley MDI scores**					
No. Studies/No. Observations Number of Children		19/33 8,815	14/27 7,445	8/12 4,967	6/9 4,305
Linear model	Beta (95% CI) p-value AIC	−0.15 (−0.23, −0.07) p < 0.001 AIC = 70.3	−0.16 (−0.28, −0.04) p = 0.010 AIC = 65.2	−0.06 (−0.13, 0.01) p = 0.086 AIC = −3.2	−0.08 (−0.15 −0.003) p = 0.043 AIC = −1.2
Urinary Fluoride – Sensitivity analysis including Ibarluzea et al. **(2021)[87] GCI scores**					
No. Studies/No. Observations Number of Children		19/33 8,749	14/27 7,445	8/12 4,901	6/9 4,239

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	**<4 mg/L**	**<2 mg/L**	**<1.5 mg/L**
Linear model	Beta (95% CI) p-value AIC	−0.15 (−0.23, −0.07) p < 0.001 AIC = 69.8	−0.16 (−0.28, −0.04) p = 0.008 AIC = 64.9	−0.04 (−0.20, 0.13) p = 0.653 AIC = −0.9	−0.08 (−0.16, −0.01) p = 0.036 AIC = −1.7
Urinary Fluoride – Low Risk-of-bias Studies					
No. Studies/No. Observations Number of Children		9/15 5,713	9/15 5,713	5/8 4,141	4/7 3,952
Linear model	Beta (95% CI) p-value AIC	−0.10 (−0.20, 0.004) p = 0.059 AIC = 2.0	−0.10 (−0.20, 0.004) p = 0.059 AIC = 2.0	−0.07 (−0.14, 0.01) p = 0.073 AIC = −1.8	−0.08 (−0.16, −0.01) p = 0.028 AIC = −2.2

Notes:

AIC = Akaike information criterion; SMD = standardized mean difference; p = p-value for effect estimate; p* = p-value for likelihood ratio tests; MDI = Mental Development Index; GCI = General Cognitive Index

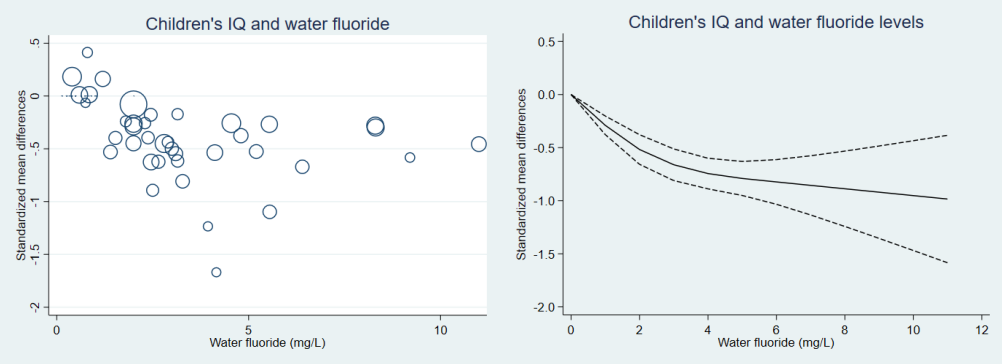
[a]Parameter estimates are changes in SMDs (beta [95% CI]) based on the maximum likelihood models; model fit is represented by the maximum likelihood AIC.

[b]The estimates represent change in SMD for the linear model and AIC, respectively.

[c]The estimates represent change in SMD for the linear term, change in SMD for quadratic term, AIC, and p-values for likelihood ratio test versus linear model, respectively. Potential departure from a linear trend was assessed by testing the coefficient of the quadratic term equal to zero

[d]The estimates represent change in SMD for the first spline term, change in SMD for the second spline term, AIC, and p-value for likelihood ratio test vs linear model, respectively. Potential departure from a linear trend was assessed by testing the coefficient of the second spline equal to zero.

Water Fluoride Exposure



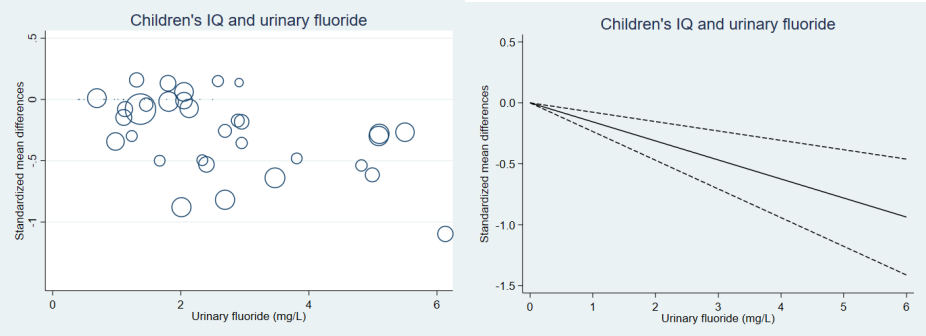
eFigure 17. Pooled Dose-Response Association Between Fluoride in Water and Standardized Mean Differences in Children's IQ

Left panel: circles indicate standardized weighted mean differences (SMDs) in individual studies; size of bubbles is proportional to precision (inverse of variance) of the standardized mean differences. Right panel: Water fluoride levels were modeled with restricted cubic splines terms in a random-effects model (solid line). Dashed lines represent the 95 % confidence intervals for the spline model. Please see **eTable 2** for characteristics of the studies included in the *dose-response meta-analysis* (studies with water fluoride exposure and at least two exposure levels).

Commented [I18]: See Doc08_Meta-analysis., 8.R., page 10 and 11

Commented [I19]: See Doc03_Meta-analysis, 3.F., page 4

Urinary Fluoride Exposure



eFigure 18. Pooled Dose-Response Association Between Fluoride in Urine and Standardized Mean Differences in Children's IQ

Left panel: Circles indicate standardized weighted mean differences in individual studies; size of bubbles is proportional to precision (inverse of variance) of the standardized mean differences. Right panel: Urinary fluoride levels were modeled with a linear random-effects model (solid line). Dashed lines represent the 95 % confidence intervals for the linear model. Please see **eTable 2** for characteristics of the studies included in the *dose-response meta-analysis* (studies with urinary fluoride exposure and at least two exposure levels).

Commented [I20]: See Doc03_Meta-analysis, 3.F., page 4

Regression Slopes Meta-analysis

Studies with overlapping populations

Yu et al.[3] and Wang et al.[4] used the same study cohort of children recruited in 2015 from the rural areas of Tianjin City, China. Since Wang et al.[4] (n = 571) used a subset of the original study sample from Yu et al.[3] (n = 2,886), only results from Yu et al.[3] were included in the meta-analysis. A sensitivity analysis was performed to evaluate the impact of using the effect estimate from Wang et al.[4] rather than the pooled effect estimate from Yu et al.[3]. Green et al.[113] and Till et al.[81] used the same Maternal-Infant Research on Environmental Chemicals (MIREC) cohort that reported drinking tap water in 10 Canadian cities, with the studies overlapping for 398 mother-child pairs. Both studies reported effect estimates for maternal urinary fluoride (MUF) and water fluoride concentrations. In the Green et al.[113] study, 512 mother-child pairs had MUF data compared to 398 pairs in Till et al.[81]. Water fluoride levels were available for 420 pairs in Green et al.[113] compared to 398 pairs in Till et al.[81]. Both studies reported effect estimates adjusted for maternal education, maternal race, child's sex, HOME total score, and secondhand smoke status in the child's home. In addition, Till et al.[81] adjusted for child's age at IQ testing (the age range for all children was 3–4 years old). Because of the larger sample size and because covariate adjustments were similar, results from Green et al.[113] were included in the main analysis. However, because of the more adjusted estimates from Till et al.[81] compared to Green et al.[113], a sensitivity analysis was performed using the water fluoride result for formula-fed children and the MUF result from Till et al.[81]. For fluoride from intake, the estimates from both studies were used since they represent total fluoride intake from Green et al.[113] and infant fluoride intake from formula Till et al.[81].

Three studies were excluded with reported slopes because the exposure was measured at the community level.[25, 30, 35] Only one study[116] included in this meta-analysis was considered high risk of bias. For Bashash et al.[112], Yu et al.[3] and Till et al.[81], units of exposure were transformed from 0.5 mg/L to 1 mg/L. Cui et al.[76], and Zhao et al. (2021)[91] reported associations between IQ and log-transformed exposure, and units of exposure were transformed from 1 log mg/L to 1 mg/L[117]. Yu et al.[3] reported estimates from piecewise linear regression models and provided three ranges for urinary fluoride exposure (low 0.01–1.60 mg/L, medium 1.60–2.50 mg/L, high 2.50–5.54 mg/L) and two ranges for water fluoride (low 0.20–3.40 mg/L and high 3.40–3.90 mg/L). Since these piecewise effect estimates are likely correlated, the study-specific pooled effect estimates were used for urine and water fluoride exposures for the overall effect meta-analysis. A sensitivity analysis was performed to evaluate the impact of using pooled estimates rather than piecewise estimates from Yu et al.[3].

For studies reporting multiple measures of fluoride exposure, the results associated with measured or estimated individual-level exposures, biomarker levels (such as urinary fluoride), or fluoride intake levels were prioritized over water fluoride concentrations (see protocol; https://ntp.niehs.nih.gov/go/785076); however, subgroup analyses by exposure metric (urinary fluoride, fluoride intake, and water fluoride) were also performed.

Regression slopes meta-analysis sensitivity analyses

Information about demographic variables was not always accessible, making it difficult to study the impact of potential confounders on effect estimates. Sensitivity analyses for the regression slopes explored the impact of using unadjusted estimates, and results were not significantly impacted (**eTable 6**). Also, most of the estimates used in the *mean-effects meta-analyses* come from studies that used fluoride concentrations at the community level to represent exposure. Therefore, unless community-level clustering is accounted for in the analysis, the standard errors of the difference in means between exposed and reference groups are likely to be biased. This is less of an issue in studies using individual-level exposures (e.g., the *regression slopes meta-analysis*). However, most studies lacked adjustment for clustering,[3, 76, 110] or for complex sampling strategies.[3, 110] Therefore, we performed sensitivity analyses to

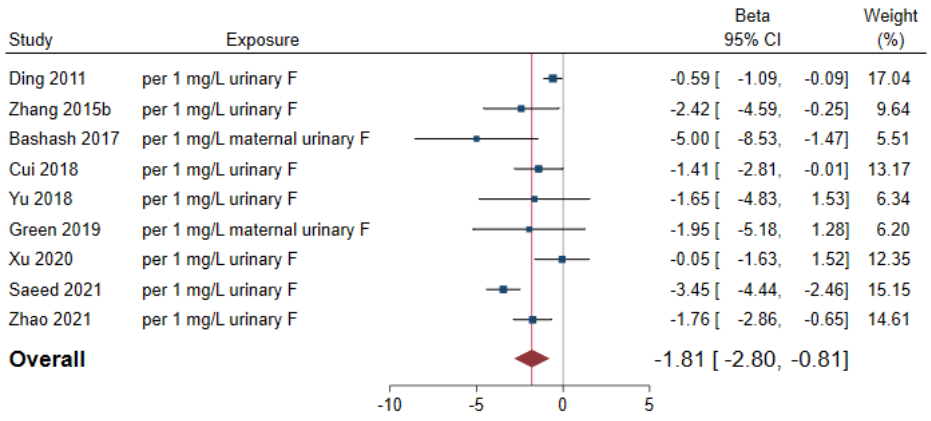
assess the impact of such issues and there were minimal changes in the pooled slopes (**eTable 6**). In the *regression slopes meta-analysis*, from the Green et al.[113] and Bashash et al.[112] studies, we used the estimates reported from the models using the clustering variable (city or cohort, respectively) as a fixed effect. However, the sensitivity analysis using the regression slopes from the corresponding models with random effects from the Green et al.[113] and Bashash et al.[112] studies,[118, 119] showed that a 1-mg/L increase in urinary fluoride was associated with a statistically significant lower IQ score of 1.80 points (95% CI: −2.80, −0.81). This suggests that clustering is not a significant issue in the results of our *regression slopes meta-analysis*.

eTable 6. Regression Slopes Meta-analysis

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimate				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Sensitivity Analyses				
Using the piecewise estimates from Yu et al. (2018)[3]				
Full-scale IQ	11	−1.68 (−2.65, −0.71)	<0.001	79%
Using effect estimates from Wang et al. (2020b)[4] rather than Yu et al. (2018)[3]				
Full-scale IQ	9	−1.70 (−2.55, −0.85)	<0.001	77%
Using Till et al. *(2020)[81] rather than Green et al. (2019)[113] estimates*				
Full-scale IQ	9	−1.83 (−2.80, −0.86)	<0.001	77%
Using estimates from random effect models for Green et al. (2019)[113] and Bashash et al. (2017)[112]				
Full-scale IQ	9	−1.80 (−2.80, −0.80)	<0.001	76%
Males	2	−2.39 (−5.89, 1.10)	0.070	69%
Females	2	−0.53 (−3.43, 2.37)	0.186	43%
Excluding Cui et al.[76]				
Full-scale IQ	8	−1.89 (−3.03, −0.74)	<0.001	80%
Excluding Yu et al. (2018)[3] and Zhang et al. (2015b)[110]				
Full-scale IQ	7	−1.76 (−2.90, −0.62)	<0.001	82%
Using unadjusted estimates from Bashash et al. (2017),[112] Cui et al. (2018),[76] Green et al. (2019)[113], *Yu et al. (2018)[3]*				
Full-scale IQ	9	−1.82 (−2.81, −0.83)	<0.001	76%
Using Verbal or Performance IQ scores from Green et al. (2019)[113]				
Verbal IQ	9	−1.78 (−2.78, −0.79)	<0.001	77%
Performance IQ	9	−1.77 (−2.77, −0.77)	<0.001	77%
Using Bashash et al. (2017)[112] McCarthy GCI scores, Valdez Jimenez et al. (2017)[74] (Bayley MDI scores), Cantoral et al. (2021)[83] (Bayley III cognitive scores), Ibarluzea et al. (2021)[87] (Bayley MDI scores).				
Urinary fluoride	11	−1.78 (−2.78, −0.78)	<0.001	75%
Intake	3	−3.28 (−5.87, −0.68)	0.799	0%
Water fluoride	2	−4.77 (−9.09, −0.45)	0.707	0%
Using Bashash et al. (2017)[112] McCarthy GCI scores, Valdez Jimenez et al. (2017)[74] (Bayley MDI scores), Cantoral et al. (2021)[83] (Bayley III cognitive scores), Ibarluzea et al. (2021)[87] (McCarthy GCI scores).				
Urinary fluoride	11	−1.90 (−2.86, −0.94)	<0.001	73%
Intake	3	−3.28 (−5.87, −0.68)	0.799	0%
Water fluoride	2	−4.77 (−9.09, −0.45)	0.707	0%

Notes:
CI = confidence interval; GCI = General Cognitive Index; MDI = Mental Development Index.

Commented [l21]: See Doc05_Meta-analysis, 5.C. (page 3 and 4), 5.D. (page 4 through 6) and 5.F. (page 7 and 8). See Doc06b_Meta-analysis, 6b.W., page 19 through 21.

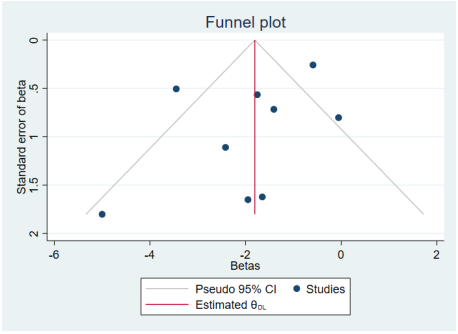


eFigure 19. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Overall Analysis

Estimates (betas) for individual studies are shown with solid boxes representing the weight, and the pooled estimate is shown as a solid diamond. Horizontal lines represent 95% Cis for the study-specific betas.

Commented [l22]: See Doc06b_Meta-analysis, 6b.C., page 2 and 3

Commented [l23]: See Doc06a_Meta-analysis, 6a.M., page 8 and 9



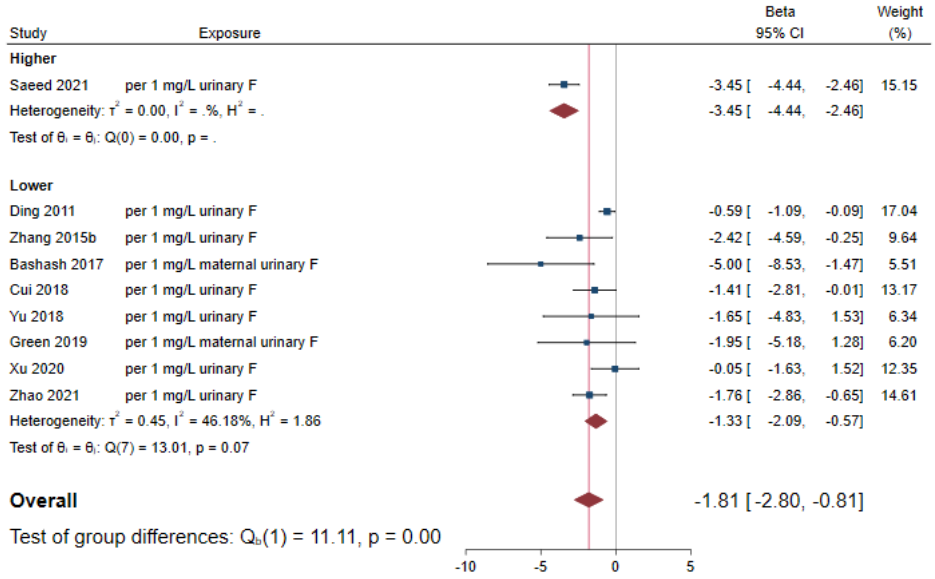
eFigure 20. Funnel Plot for Studies with Individual-level Exposures

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -1.06
      SE of beta1 =     1.066
                z =     -1.00
       Prob > |z| =    0.3192
```

eFigure 21. Test for Publication Bias for Studies with Individual-level Exposures

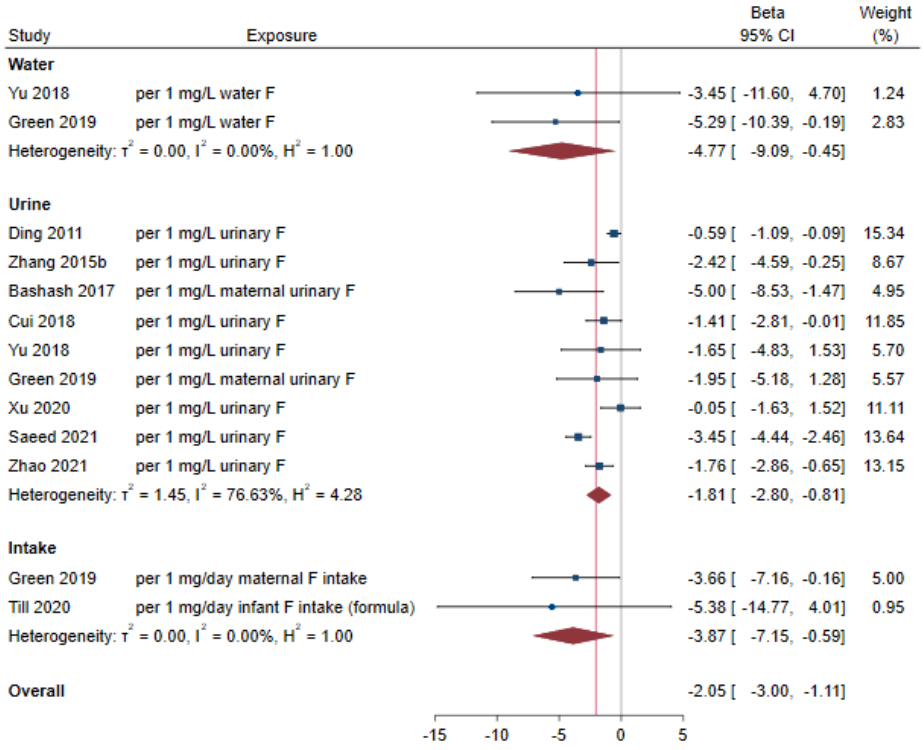
Subgroup Analyses

Risk-of-bias Subgroup Analysis



eFigure 22. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Risk of Bias

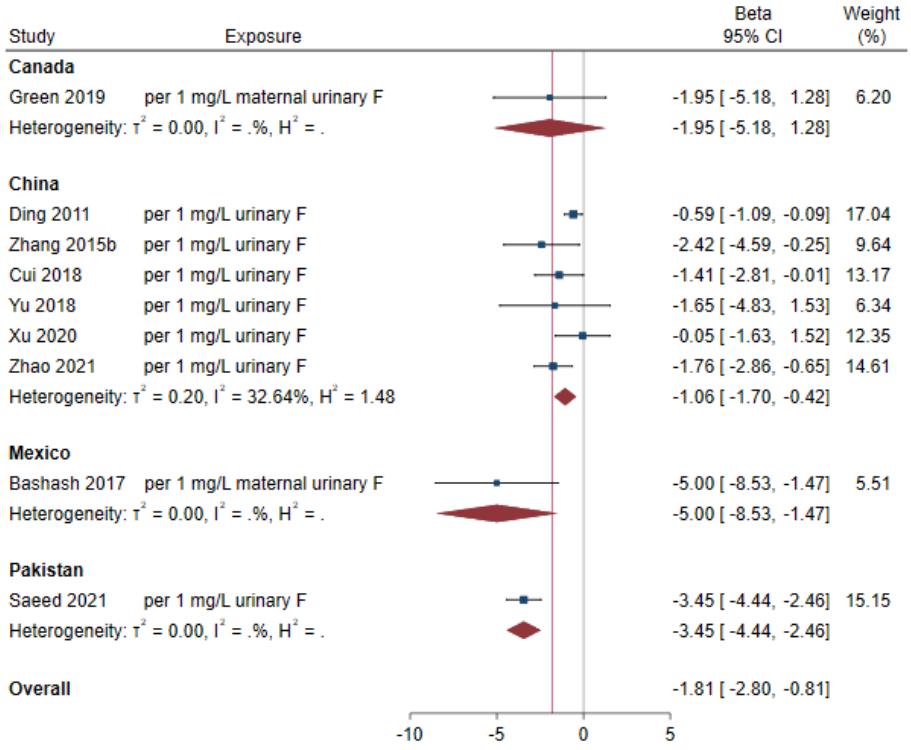
Exposure Type Subgroup Analysis



eFigure 23. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Exposure Type

Commented [l24]: See Doc06b_Meta-analysis, 6b.C., page 2 and 3

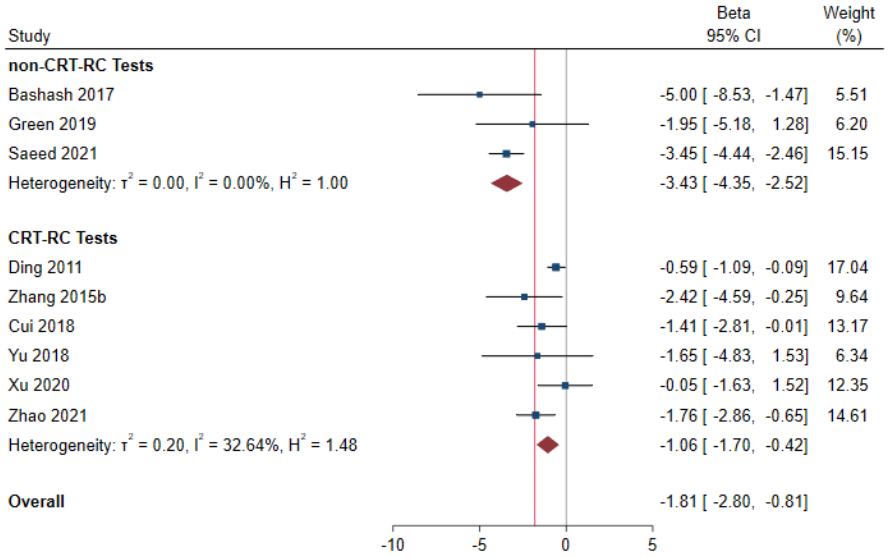
Country Subgroup Analysis



eFigure 24. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Country

Note: The analyses for publication bias for studies from China, Canada, and Mexico rely on a very small number of studies each and are not shown.

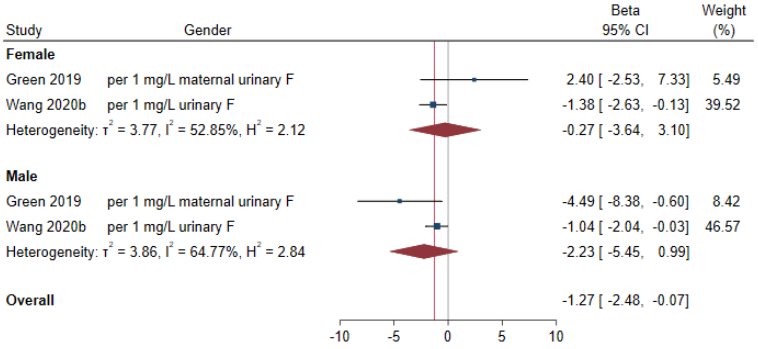
Assessment Type Subgroup Analysis



eFigure 25. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Assessment Type

Note: The analyses for publication bias for CRT-RC studies and non-CRT-RC studies include only six and three studies, respectively, and are not shown.

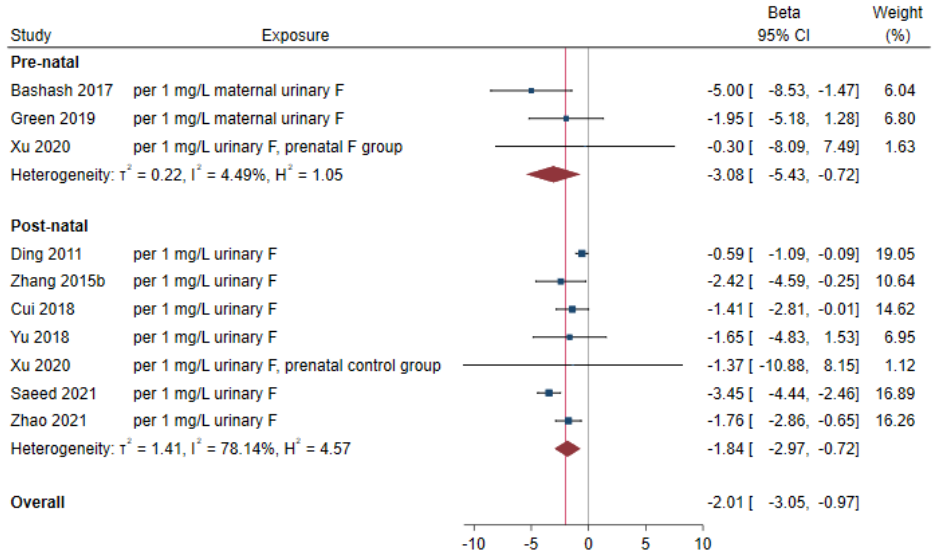
Sex Subgroup Analysis



eFigure 26. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Sex

Note: The analysis for publication bias by gender relies on two studies each and are not shown.

Pre-natal vs post-natal exposure Subgroup Analysis



eFigure 27. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Prenatal vs. Postnatal Exposure

References

1. National Toxicology Program (NTP). *Revised draft NTP Monograph on the systematic review of fluoride exposure and neurodevelopmental and cognitive health effects*. National Toxicology Program; 2020. https://www.nationalacademies.org/event/10-19-2020/docs/DDA97C9170D1A255D69C004CEB77B698E8D005011EFB
2. Higgins JT, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for systematic reviews of interventions version 6.0 (updated July 2019)*. Cochrane; 2019.
3. Yu X, Chen J, Li Y, et al. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. *Environ Int*. 2018;118:116-124. doi:https://doi.org/10.1016/j.envint.2018.05.042
4. Wang M, Liu L, Li H, et al. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. *Environ Int*. 2020b;134:105229. doi:https://doi.org/10.1016/j.envint.2019.105229
5. Crippa A, Thomas I, Orsini N. A pointwise approach to dose-response meta-analysis of aggregated data. *Int J Stat Med Res*. 2018;7(2):25-32.
6. Crippa A, Discacciati A, Bottai M, Spiegelman D, Orsini N. One-stage dose–response meta-analysis for aggregated data. *Stat Methods Med Res*. 2019;28(5):1579-1596. doi:https://doi.org/10.1177/0962280218773122
7. Müller S, Scealy JL, Welsh AH. Model selection in linear mixed models. *Statist Sci*. 2013;28(2):135-167. doi:https://doi.org/10.1214/12-STS410
8. U.S. Environmental Protection Agency (US EPA). *New fluoride risk assessment and relative source contribution documents*. Office of Water; 2011. EPA-822-F-11-001. https://www.epa.gov/sites/default/files/2019-03/documents/fluoride-risk-assess-factsheet.pdf
9. World Health Organization (WHO). *Guidelines for drinking-water quality. Fourth edition*. 2011. https://apps.who.int/iris/bitstream/handle/10665/44584/9789241548151_eng.pdf?sequence=1&isAllowed=y&ua=1
10. Guo XC, Wang RY, Cheng CF, et al. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. *Chin J Epidemiol*. 1991;10(2):98-100.
11. Guo XC, Wang RY, Cheng CF, et al. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. *Fluoride*. 2008a;41:125-128.
12. Li Y, Li X, Wei S. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. *J West Chin University Med Sci*. 1994;25(2):188-191.
13. Li Y, Li X, Wei S. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. *Fluoride*. 2008b;41:331-335.
14. Li XS, Zhi JL, Gao RO. Effect of fluoride exposure on the intelligence of children. *Fluoride*. 1995;28:189-192.
15. Li F, Chen X, Huang R, Xie Y. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. *J Environ Health*. 2009;26(4):838-840.
16. Bai A, Li Y, Fan Z, Li X, Li P. [Intelligence and growth development of children in coal-burning-borne arsenism and fluorosis areas: An investigation study]. *Chin J Endemiol*. 2014;33(2):160-163.

17. Zhang P, Cheng L. Effects of coal-burning endemic fluorosis on children's physical development and intelligence. *Chin J Endem Dis Control*. 2015:458-459.

18. Wang J, Yu M, Yang L, Yang X, Deng B. The effect of coal-burning fluoride exposure on children's intelligence and physical development. *J Environ Health*. 2020c;37(11):971-974. doi:https://dx.doi.org/10.16241/j.cnki.1001-5914.2020.11.007

19. Lou D, Luo Y, Liu J, et al. Refinement impairments of verbal-performance intelligent quotient in children exposed to fluoride produced by coal burning. *Biol Trace Elem Res*. Feb 2021;199(2):482-489. doi:https://dx.doi.org/10.1007/s12011-020-02174-z

20. Ren D, Li K, Liu D. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. *Chin J Control Endem Dis*. 1989;4(4):251.

21. Ren D, Li K, Liu D. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. *Fluoride*. 2008;41:319-320.

22. Wang G, Yang D, Jia F, Wang H. [A study of the IQ levels of four- to seven-year-old children in high fluoride areas]. *Endem Dis Bull*. 1996;11(1):60-62.

23. Wang G, Yang D, Jia F, Wang H. A study of the IQ levels of four- to seven-year-old children in high fluoride areas. *Fluoride*. 2008b;41:340-343.

24. An J, Mei S, Liu A, Fu Y, Wang C. [Effect of high level of fluoride on children's intelligence]. *Chin J Control Endem Dis*. 1992;7(2):93-94.

25. Broadbent JM, Thomson WM, Moffitt TE, Poulton R. Community water fluoridation and intelligence response. *Am J Public Health*. 2015;105:3-4.

26. Xu Y, Lu C, Zhang X. [The effect of fluorine on the level of intelligence in children]. *Endem Dis Bull*. 1994;9(2):83-84.

27. Yao Y. Comparable analysis on the physical and mental development of children in endemic fluorosis area with water improvement and without water improvement. *Lit Inf Prev Med*. 1997;3(1):42-43.

28. Yao L, Zhou J, Wang S, Cui K, Lin F. Analysis on TSH and intelligence level of children with dental Fluorosis in a high fluoride area. *Lit Inf Prev Med*. 1996;2(1):26-27.

29. Seraj B, Shahrabi M, Falahzade M, Falahzade F, Akhondi N. Effect of high fluoride concentration in drinking water on children's intelligence. *J Dent Med*. 2006;19(2):80-86.

30. Seraj B, Shahrabi M, Shadfar M, et al. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. *J Dent*. 2012;9:221-229.

31. Eswar P, Nagesh L, Devaraj CG. Intelligent quotients of 12-14 year old school children in a high and low fluoride village in India. *Fluoride*. 2011;44:168-172.

32. Poureslami HR, Horri A, Garrusi B. A comparative study of the IQ of children age 7-9 in a high and a low fluoride water city in Iran. *Fluoride*. 2011;44:163-167.

33. Shivaprakash PK, Ohri K, Noorani H. Relation between dental fluorosis and intelligence quotient in school children of Bagalkot District. *J Indian Soc Pedod Prev Dent*. 2011;29:117-120.

34. Khan SA, Singh RK, Navit S, et al. Relationship between dental fluorosis and intelligence quotient of school going children in and around Lucknow District: A cross-sectional study. *J Clin Diagn Res*. 2015;9(11):10-15. doi:https://doi.org/10.7860/JCDR/2015/15518.6726

35. Sebastian ST, Sunitha S. A cross-sectional study to assess the intelligence quotient (IQ) of school going children aged 10-12 years in villages of Mysore District, India with different fluoride levels. *J Indian Soc Pedod Prev Dent*. 2015;33:307-311.

36. Mondal D, Dutta G, Gupta S. Inferring the fluoride hydrogeochemistry and effect of consuming fluoride-contaminated drinking water on human health in some endemic areas of Birbhum District, West Bengal. *Environ Geochem Health*. 2016;38:557-576.

37. Karimzade S, Aghaei M, Mahvi AH. Investigation of intelligence quotient in 9-12 year-old children exposed to high- and low-drinking water fluoride in West Azerbaijan Province, Iran. *Fluoride*. 2014;47:9-14.

38. Sun M, Li S, Wang Y, Li F. [Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis]. *J Guiyang Med Coll*. 1991;16(3):204-206.

39. Zhang J, Yao H, Chen Y. [The effect of high levels of arsenic and fluoride on the development of children's intelligence]. *Chin J Public Health*. 1998;17(2):119.

40. Trivedi M, Sangai N, Patel R, Payak M, Vyas S. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. *Fluoride*. 2012;45(4):377-383.

41. Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. Effect of high fluoride water on intelligence of school children in India. *Fluoride*. 2007;40:178-183.

42. Zhao Q, Tian Z, Zhou G, et al. SIRT1-dependent mitochondrial biogenesis supports therapeutic effects of resveratrol against neurodevelopment damage by fluoride. *Theranostics*. 2020;10:4822-4838. doi:https://doi.org/10.7150/thno.42387

43. Zhao Q, Niu Q, Chen J, et al. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol*. Jan 18 2019;93(3):709-726. doi:https://doi.org/10.1007/s00204-019-02390-0

44. Zhou G, Tang S, Yang L, et al. Effects of long-term fluoride exposure on cognitive ability and the underlying mechanisms: Role of autophagy and its association with apoptosis. *Toxicol Appl Pharmacol*. 2019;378:114608. doi:https://doi.org/10.1016/j.taap.2019.114608

45. Qin L, Huo S, Chen R, Chang Y, Zhao M. [Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children]. *Chin J Endem Dis*. 1990;5(4):203-204.

46. Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children. *Fluoride*. 2008;41:115-119.

47. Yang Y, Wang X, X. G, Hu P. [The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine]. *Chin J Epidemiol*. 1994;15(4):296-298.

48. Yang Y, Wang X, Guo X, Hu P. The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine. *Fluoride*. 2008;41:336-339.

49. Wang X, Wang L, Hu P, Guo X, Luo X. [Effects of high iodine and high fluorine on children's intelligence and thyroid function]. *Chin J Endemiol*. 2001;20(4):288-290.

50. Wang SY, Zhang HX, Fan W, et al. [The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children]. *J Appl Clin Ped*. 2005b;20(9):897-899.

51. Wang SY, Zhang HX, Fan W, et al. The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children. *Fluoride*. 2008a;41:344-348.

52. Rocha-Amador D, Navarro ME, Carrizales L, Morales R, Calderon J. Decreased intelligence in children and exposure to fluoride and arsenic in drinking water. *Cad Saude Publica*. 2007;23(Suppl 4):S579-587.

53. Liu SL, Lu Y, Sun ZR, et al. [Report on the intellectual ability of children living in high-fluoride water areas]. *Chin J Control Endem Dis*. 2000;15(4):231-232.

54. Liu SL, Lu Y, Sun ZR, et al. Report on the intellectual ability of children living in high-fluoride water areas. *Fluoride*. 2008;41:144-147.

55. Lu Y, Sun ZR, Wu LN, Wang X, Lu W, Liu SS. Effect of high-fluoride water on intelligence in children. *Fluoride*. 2000;33:74-78.

56. Sudhir KM, Chandu GN, Prashant GM, Subba Reddy VV. Effect of fluoride exposure on intelligence quotient (IQ) among 13-15 year old school children of known endemic area of fluorosis, Nalgonda District, Andhra Pradesh. *J Indian Assoc Public Health Dent*. 2009;2009(13):88-94.

57. He MX, Zhang CN. [Investigation of children's intelligence quotient and dental fluorosis in drinking water-type of endemic fluorosis area in Pucheng County, Shaanxi Province before and after drinking water change]. *Chin J Endemiol*. 2010;29:547-548.

58. Xiang Q, Liang Y, Chen B, Chen L. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. *Fluoride*. 2011;44:191-194.

59. Xiang Q, Liang Y, Chen L, et al. Effect of fluoride in drinking water on children's intelligence. *Fluoride*. 2003a;36:84-94.

60. Saxena S, Sahay A, Goel P. Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India. *J Neurosci Rural Pract*. 2012;3:144-149.

61. Wang G, Gao M, Zhang M, Yang M, Xiang Q. Correlation between total fluoride intake and children's IQ. *J Southeast Univ Med Ed*. 2012:743-746.

62. Nagarajappa R, Pujara P, Sharda AJ, et al. Comparative assessment of intelligence quotient among children living in high and low fluoride areas of Kutch, India: A pilot study. *Iran J Public Health*. 2013;42:813-818.

63. Pratap SV, Singh CD, Sandeep T, et al. A correlation between serum vitamin, acetylcholinesterase activity and IQ in children with excessive endemic fluoride exposure in Rajasthan, India. *Int Res J Medical Sci*. 2013;1(3):12-16.

64. Asawa K, Pujara P, Thakkar JP, et al. Assessment of intelligence quotient among schoolchildren of fishermen community of Kutch, Gujarat, India. *Int Marit Health*. 2014;65(2):73-8. doi:https://doi.org/10.5603/imh.2014.0017

65. Wei N, Li Y, Deng J, Xu S, Guan Z. [The effects of comprehensive control measures on intelligence of school-age children in coal-burning-borne endemic fluorosis areas]. *Chin J Endemiol*. 2014;33(3):320-322.

66. Choi AL, Zhang Y, Sun G, et al. Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study. *Neurotoxicol Teratol*. 2015;47:96-101.

67. Kundu H, Basavaraj P, Singla A, Gupta R, Singh K, Jain S. Effect of fluoride in drinking water on children's intelligence in high and low fluoride areas of Delhi. *J Indian Assoc Public Health Dent*. Apr 2015;13(2):116-121. doi:https://doi.org/10.4103/2319-5932.159043

68. Aravind A, Dhanya RS, Narayan A, Sam G, Adarsh VJ, Kiran M. Effect of fluoridated water on intelligence in 10-12-year-old school children. *J Int Soc Prev Community Dent*. Dec 2016;6(Suppl 3):S237-S242. doi:https://doi.org/10.4103/2231-0762.197204

69. Jin T, Han T, Wei Y, Wu Y, Wang Z, Zhang H. Investigation on working memory level of children aged 8-12 years in coal-burning fluorosis area. *J Environ Health*. 2016:409-411.

70. Zhang P, Cheng L. Effect of coal-burning endemic fluorosis on children's physical development and intellectual level. *Chin J Endem Dis Control*. 2015c;30(6):458-459.

71. Kumar R, Gupta VK, Kumar A, Shetty R, Singh M, Pandey V. Evaluation of correlation of dental fluorosis and cognitive status. *Ann Int Med Dental Res*. 2016;2(4):155.

72. Jin T, Wang Z, Wei Y, Wu Y, Han T, Zhang H. Investigation on intelligence level of children aged 8-12 years old in coal-burning fluorosis area. *J Environ Health*. 2017:229-231.

73. Razdan P, Patthi B, Kumar JK, Agnihotri N, Chaudhari P, Prasad M. Effect of fluoride concentration in drinking water on intelligence quotient of 12-14-year-old children in Mathura District: A cross-sectional study. *J Int Soc Prev Community Dent*. 2017;7:252-258. doi:https://doi.org/10.4103/jispcd.JISPCD_201_17

74. Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, et al. In utero exposure to fluoride and cognitive development delay in infants. *Neurotoxicology*. Mar 2017;59:65-70. doi:https://dx.doi.org/10.1016/j.neuro.2016.12.011

75. Wang G, Zhang M, Wang Q, et al. Investigation on the relationship between serum fluoride content and IQ of children before and after reducing fluoride to water. *Capital Pub Health*. 2017:274-277.

76. Cui Y, Zhang B, Ma J, et al. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. *Ecotoxicol Environ Saf*. 2018;165:270-277. doi:https://doi.org/10.1016/j.ecoenv.2018.09.018

77. Luo Y, Ma R, Liu Z, Guan Z, Lou D, Zheng D. Intelligence investigation and forensic significance of children in coal-burning fluorosis area. *Chin J Forensic Med*. 2018:590-593.

78. Naik SP, Bankur PK, Sathe S, Haris PMM, Kadar N, Satyanarayan A. Impact of fluoridated water on intelligence quotient levels of school children: A exploratory study. *Int J Oral Care Res*. 2018;6(1):63-66.

79. Sharma P, Bhardwaj AK, Singh M, Kumar D, Sharma A, Grover A. Does fluorosis affect the intelligence profile of children? A cross sectional analysis of school children of district Una, Himachal Pradesh, India. *Int J Community Med Public Health*. 2018;5(3):1047-53. doi:http://dx.doi.org/10.18203/2394-6040.ijcmph20180759

80. Soto-Barreras U, Escalante-Villalobos KY, Holguin-Loya B, et al. Effect of fluoride in drinking water on dental caries and IQ in children. *Fluoride*. 2019;52:474-482.

81. Till C, Green R, Flora D, et al. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. *Environ Int*. 2020;134:105315. doi:https://doi.org/10.1016/j.envint.2019.105315

82. Aggeborn L, Öhman M. The effects of fluoride in drinking water. *J Polit Econ*. 2020;129(2):465-491. doi:https://dx.doi.org/10.1086/711915

83. Cantoral A, Tellez-Rojo MM, Malin AJ, et al. Dietary fluoride intake during pregnancy and neurodevelopment in toddlers: A prospective study in the progress cohort. *Neurotoxicology*. Dec 2021;87:86-93. doi:https://dx.doi.org/10.1016/j.neuro.2021.08.015

84. Farmus L, Till C, Green R, et al. Critical windows of fluoride neurotoxicity in Canadian children. *Environ Res*. Sep 2021;200:111315. doi:https://doi.org/10.1016/j.envres.2021.111315

85. Guo B, Yu J, Cui Y, et al. DBH gene polymorphism, iodine and fluoride and their interactions and their interaction with children's intelligence. *J Environ Hygiene*. 2021;11(02):134-140.

86. Zhao Y, Cui Y, Yu J, et al. Study on the relationship between water-borne high iodine and thyroid hormone and children's intelligence level. *J Environ Health*. 2018:6-9.

87. Ibarluzea J, Gallastegi M, Santa-Marina L, et al. Prenatal exposure to fluoride and neuropsychological development in early childhood: 1-to 4 years old children. *Environ Res*. Oct 8 2021:112181. doi:https://dx.doi.org/10.1016/j.envres.2021.112181

88. Wang R, He N, Hou G, Zhang P, Jin W, Wang Y. Hengshui Survey and analysis of the mental health status of children aged 8-12 years in the high-fluoride area of Shanghai. *Med Anim Control*. 2021b;37(06):583-586.

89. Wang R, He N, Wang Y, Hou G, Zhang P-J. Investigation and analysis of children's dental fluorosis and IQ level in high fluoride district of Hengshui City. *Med Anim Control*. 2021;37(8):796-800. doi:https://dx.doi.org/10.7629/yxdwfz202108023

90. Yu X, Xia L, Zhang S, et al. Fluoride exposure and children's intelligence: Gene-environment interaction based on SNP-set, gene and pathway analysis, using a case-control design based on a cross-sectional study. *Environ Int*. 2021;155:106681. doi:https://doi.org/10.1016/j.envint.2021.106681

91. Zhao L, Yu C, Lv J, et al. Fluoride exposure, dopamine relative gene polymorphism and intelligence: A cross-sectional study in China. *Ecotoxicol Environ Safety*. Feb 2021;209:111826. doi:https://dx.doi.org/10.1016/j.ecoenv.2020.111826

92. Zhou G, Zhao Q, Luo C, et al. Low-moderate fluoride exposure and intelligence among Chinese school-aged children: Role of circulating mtDNA content. *Sci Total Environ*. Sep 10 2021;786:147330. doi:https://dx.doi.org/10.1016/j.scitotenv.2021.147330

93. Chen YX, Han FL, Zhoua ZL, et al. [Research on the intellectual development of children in high fluoride areas]. *Chin J Control Endem Dis*. 1991;6(Suppl):99-100.

94. Chen YX, Han FL, Zhoua ZL, et al. Research on the intellectual development of children in high fluoride areas. *Fluoride*. 2008;41:120-124.

95. Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. [High fluoride and low iodine environment and subclinical cretinism in Xinjiang]. *Endem Dis Bull*. 1991;6(2):62-67.

96. Zhao LB, Liang GH, Zhang DN, Wu XR. Effect of a high fluoride water supply on children's intelligence. *Fluoride*. 1996;29:190-192.

97. Hong FG, Cao YX, Yang D, Wang H. [Research on the effects of fluoride on child intellectual development under different environmental conditions]. *Chin Prim Health Care*. 2001;15(3):56-57.

98. Hong FG, Cao YX, Yang D, Wang H. Research on the effects of fluoride on child intellectual development under different environmental conditions. *Fluoride*. 2008;41:156-160.

99. Hong F, Wang Hui Yang Dong Zhang Z. Investigation on the intelligence and metabolism of iodine and fluoride in children with high iodine and fluoride. *Chin J Endem Dis Control*. 2001b:12-14.

100. Li YP, Jing XY, Chen D, Lin L, Wang ZJ. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. *Chin J Public Health Manag*. 2003;19(4):337-338.

101. Li YP, Jing XY, Chen D, Lin L, Wang ZJ. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. *Fluoride*. 2008c;41:161-164.

102. Wang SX, Wang ZH, Cheng XT, et al. [Investigation and evaluation on intelligence and growth of children in endemic fluorosis and arsenism areas]. *Chin J Endemiol*. 2005;24:179-182.

103. Wang Z, Wang S, Zhang X, Li J, Zheng X, Hu C. Investigation of children's growth and development under long-term fluoride exposure. *Chin J Control Endem Dis*. 2006;21(4):239-241.

104. Fan Z, Dai H, Bai A, Li P, Li T, Li G. The effect of high fluoride exposure in children's intelligence. *J Environ Health*. 2007;24(10):802-803.

105. Wang SX, Wang ZH, Cheng XT, et al. Arsenic and fluoride exposure in drinking water: Children's IQ and growth in Shanyin County, Shanxi Province, China. *Environ Health Perspect*. 2007;115:643-647.

106. Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. [Investigation and analysis of children's IQ and dental fluorosis in high fluoride area]. *Chin J Pest Control*. 2010;26(3):230-231.

107. Ding Y, Sun H, Han H, et al. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. *J Hazard Mater*. 2011;186:1942-1946.

108. Kang J, Cheng Y, Wu K, Lin S, He G, Jin Y. [Effect of exposure to fluoride and arsenic in drinking water of Hangjinhouqi on children's intelligence]. *Chin School Health*. 2011:679-681.

109. Wang S, Zhu X. Investigation and analysis of intelligence level of children in high fluoride area. *Chin J Endem Dis Control*. 2012b;27(1):67-68.

110. Zhang S, Zhang X, Liu H, et al. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. *Toxicol Sci*. 2015b;144:238-245.

111. Das K, Mondal NK. Dental fluorosis and urinary fluoride concentration as a reflection of fluoride exposure and its impact on IQ level and BMI of children of Laxmisagar, Simlapal Block of Bankura District, W.B., India. *Environ Monit Assess*. 2016;188:218.

112. Bashash M, Thomas D, Hu H, et al. Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6-12 years of age in Mexico. *Environ Health Perspect*. 2017;125(9):1-12. doi:https://doi.org/10.1289/ehp655

113. Green R, Lanphear B, Hornung R, et al. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. *JAMA Pediatr*. 2019:E1-E9.

114. Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. *Neurosci Lett*. 2020;729:134981. doi:https://doi.org/10.1016/j.neulet.2020.134981

115. Xu K, An N, Huang H, et al. Fluoride exposure and intelligence in school-age children: evidence from different windows of exposure susceptibility. *BMC Public Health*. Nov 4 2020;20(1):1657. doi:https://dx.doi.org/10.1186/s12889-020-09765-4

116. Saeed M, Rehman MYA, Farooqi A, Malik RN. Arsenic and fluoride co-exposure through drinking water and their impacts on intelligence and oxidative stress among rural school-aged children of Lahore and Kasur districts, Pakistan. *Environ Geochem Health*. Nov 9 2021;doi:https://dx.doi.org/10.1007/s10653-021-01141-4

117. Rodriguez-Barranco M, Tobias A, Redondo D, Molina-Portillo E, Sanchez M-J. Standardizing effect size from linear regression models with log-transformed variables for meta-analysis. *BMC Med Res Methodol*. 2017;17(44)doi:https://doi.org/10.1186/s12874-017-0322-8

118. Bashash M. Personal communication from Morteza Bashash to Sorina Eftim regarding estimates from models with cohort effect considered using a random effect model. 2021.

119. Till C. Personal communication from Christine Till to Sorina Eftim regarding estimates from multi-level models predicting child FSIQ from MUF using city as a random effect. 2021.

Peer Review of the draft Meta-analysis Manuscript to Evaluate the Association between Fluoride Exposure and Children's Intelligence

[redacted] received a draft version of the manuscript as well as a copy of the NASEM Committee comments on the meta-analysis and the NIEHS/DNTP responses (draft version of Sup01_Meta-analysis). [redacted] provided comments in track changes on the draft manuscript in Microsoft Word. The full comments have been reproduced below verbatim along with the specific text referred to by as quotes under a heading for the specific section of the document (e.g., "Abstract section"). Note that the yellow highlighting was in the document as provided. Formatting has been applied to aid in reading. Responses have been added in blue text following each of the comments beginning with the word "**Response**" in bold font.

[redacted]
Date: July 27, 2021

1.A: ***Overall Comments:*** *This is such impressive work and glad you have put it into what [redacted] sure will be a high impact paper. [redacted] attaching some comments. [redacted] tried to highlight a couple of places where [redacted] thought your tox language needed to be modified to be more easily palatable to the clinical audience you're targeting, particularly the dose-response results. Those tended to be pretty confusing. [redacted] also had a lot of questions reading the abstract, prior to reading the paper. These are all editorial and [redacted] think your analysis is robust and your conclusions are in line with what the data are showing. Congratulations on great work.*

Response: Agree (no change requested)

- We appreciate [redacted] comments about this work, especially that the analyses are robust and the conclusions are in line with what the data are showing.

1.B: ***Abstract section:*** *"To perform a systematic review and meta-analysis to investigate associations between fluoride exposure and children's intelligence."*

[redacted] **Comment:** Maybe it's because [redacted] know what's coming but [redacted] find myself wanting to know in this sentence when fluoride exposure occurred for the studies included.

Response: Disagree (no change)

- The additional detail is not necessary in the abstract to convey the scope of the manuscript and there is a word limit for the abstract. The full description of fluoride exposure is included in the *Methods* section wherein the timing of exposure is described as "*pre- or post-natal exposure.*"

1.C: ***Abstract section:*** *"Inclusion criteria were assessment of cognitive outcomes, fluoride exposure, and statistical data on effect size."*

[redacted] **Comment:** Again, reading abstract before the article, but [redacted] am wondering if the exposure was all drinking water, or you combined studies with info on exposure biomarkers and drinking water levels?

Response: Disagree (no change)

- The exposures to fluoride included, but were not limited to, drinking water. There were biomonitoring measures (e.g., urinary fluoride) and other environmental measures (e.g.,

coal burning, areas endemic for dental fluorosis). We do not consider the additional detail in the abstract necessary to convey the scope of the manuscript, and due to the strict word limit for the abstract, the full description of fluoride exposures considered is included in the *Methods* section and in *Table 1. Characteristics of Studies Included in the Meta-analysis* (excerpt provided below).

Excerpt of Table 1. Characteristics of Studies Included in the Meta-analysis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Ren et al. (1989)[66] [translated in Ren et al. 2008][me, o] *Cross-sectional*	China	8–14	No fluoride measurement Low iodine village/high fluoride and low iodine village	Not specified	Wechsler Intelligence Scale for Children	High	Sex; iodine
Chen et al. (1991)[68] [translated in Chen et al. 2008][me, w] *Cross-sectional*	China	7–14	Drinking water Nonendemic/endemic fluorosis village	0.89 mg/L (nonendemic) 4.55 mg/L (endemic)	Chinese Standardized Raven Test	High	Age; sex
Guo et al. (1991)[70] [translated in Guo et al. 2008a][me, o] *Cross-sectional*	China	7–13	Serum Reference area using wood/coal burning-related fluoride endemic area	0.1044 ± 0.0652 mg/L (reference) 0.1483 ± 0.0473 mg/L (endemic)	Chinese Binet Intelligence Test	High	Age; sex; SES

1.D: ***Abstract section:*** *"The meta-analysis of forty-six studies with group-level exposures found that children exposed to higher fluoride levels had lower IQ scores (SMD, −0.49; 95%CI, −0.60, −0.38; p-value<0.001)."*

[redacted] **Comment:** What does the SMD correspond to for this effect estimate?

Response: Agree (no change requested)

- This SMD compares groups of children living in areas with "high" fluoride exposure with children living in areas with "low" fluoride exposure. After updating the literature search, we revised the sentence to say:

  "The meta-analysis of 55 studies (N = 18,845 children) with group-level exposures found that, when compared to children exposed to lower fluoride levels, children exposed to higher fluoride levels had lower mean IQ scores (pooled SMD: −0.46; 95% CI: −0.55, −0.37; p-value < 0.001)."

1.E: ***Abstract section:*** *"When analyses were restricted to studies with groups exposed to ≤2 mg/L fluoride in drinking water, the mean SMD in children's IQ scores remained lower (SMD, -0.27; 95% CI: -0.36, -0.17; p-value<0.001)."*

[redacted] **Comment:** Can you briefly state why the cutoffs were chosen for this analysis?

Response: Agree (change made)

- We have removed this sentence from the abstract but have added the following rationale for the cutoffs to the *Methods* section:

  "We also examined whether there was a dose-response relationship at lower exposure levels that corresponded with the U.S. Environmental Protection Agency drinking water standards[20] and World Health Organization drinking water guidelines[21] (details provided in the Supplemental Materials)."

1.F: ***Introduction section:*** *(the yellow text highlighting was added by [redacted]) "However, many of the studies lacked details and key information necessary to evaluate study quality (e.g., measurement of covariates and other neurotoxic co-exposures)."*

[redacted] **Comment:** Both of the meta-analyses?

Response: Agree (change made)

- Yes, this statement refers to both previous meta-analyses (Choi et al. 2012 and Duan et al. 2018). We have revised the text in the *Introduction* section as follows:

  "Two previous meta-analyses[3, 4] found an association between high fluoride exposure and lower children's IQ; however, many of the studies in these meta-analyses lacked the information necessary to evaluate study quality and all used group-level estimates of fluoride exposure."2

1.G: [redacted] ***Comment:*** *You have references formatted in a few different ways so just double check throughout before submitting*

Response: Agree (change made)

- We have reformatted the references using superscript notation to align with the journal requirements.
- The changes are reflected throughout the document and an example from the previous comment is repeated below to show the formatting:

  "Two previous meta-analyses[3, 4] found an association between high fluoride exposure and lower children's IQ…"

1.H: ***Methods section:*** *(the yellow text highlighting was added by [redacted]) "Quality of individual studies, also called "risk of bias", was assessed using NTP's HAT approach."*

[redacted] **Comment:** Spell out?

Response: Agree (change made)

- This sentence has been revised in the *Methods* section as follows:

  "Quality of individual studies, also called "risk of bias," was assessed using the National Toxicology Program's Office of Health Assessment and Translation approach."

1.I: ***Results section:*** *(the yellow text highlighting was added by [redacted]) "Three studies[OBJ] unclear descriptions of their intelligence assessment methods, and sensitivity analyses did not reveal substantial changes in the pooled SMD estimate with or without these studies (−0.57; 95% CI: −0.69, −0.45)."*

[redacted] **Comment:** Something funny going on here

Response: Agree (change made)

- This formatting issue was a glitch with the citations and a broken link to the references which has been fixed.
- This sentence has been revised in the *Results* section as follows:

"Three studies[39, 40, 41 [translated in Li et al. 2008b]] lacked clear descriptions of their intelligence assessment methods; however, sensitivity analyses did not reveal substantial changes in the pooled SMD estimate when these studies were excluded or when a study[43] that reported the cognitive subset of evaluations using Bayley and McCarthy tests was included (eTable 3)."

1.J: ***Results section:*** *"For studies that had more than one exposed group (n = 17), a sensitivity analysis was performed to evaluate the impact of using all exposed groups combined compared to the reference group."*

[redacted] **Comment:** Made this a new paragraph.

Response: Agree (change made)

- We have taken [redacted] suggestion to make the text a new paragraph.

1.K: ***Results section:*** *"When the analyses were restricted to studies with the exposed group <1.5 mg/L fluoride in drinking water (n = 9; 2 lower risk-of-bias and 7 higher risk-of-bias studies) there was a non-significant positive association between fluoride exposure and children's IQ (SMD, 0.32; 95% CI: −0.57, 1.20). When restricted to studies with the exposed group <1.5 mg/L urinary fluoride (n = 4; 2 lower risk-of-bias and 2 higher risk-of-bias studies), the association was negative (SMD, −0.13; 95% CI: −0.29, 0.03; p-value=0.111]. When including groups exposed to < 2 mg/L urinary fluoride (n = 6; 3 lower risk-of-bias studies and 3 higher risk-of-bias studies), the association did not change substantially (−0.09; 95% CI: −0.22, 0.03; p-value=0.143). However, when including groups exposed to <2 mg/L fluoride in drinking water (n = 9; 2 lower risk-of-bias and 7 higher risk-of-bias studies), the association remained significant (SMD, −0.27; 95% CI: −0.36, −0.17; p-value<0.001) (eTable 4)."*

[redacted] **Comment:** This paragraph was tough for me to understand. First, [redacted] uncertain about why the cutoffs were chosen. Second, ordering should maybe start with <2? Going from all exposure levels to slightly lower exposure levels, to lowest exposure levels (<1.5). [redacted] also suggest grouping together the drinking water and urinary results in two separate paragraphs (including the overall linear results at the beginning of each paragraph). Last, [redacted] might be interested in comparing the number of low/high risk of bias studies across each grouping but it's too hard to follow, so maybe you could have a table of this info and summarize in a sentence or two here?

Response: Agree (change made)

- As noted in a previous comment, we have added the following rationale for the cutoffs to the *Methods* section:

  "We also examined whether there was a dose-response relationship at lower exposure levels that corresponded with the U.S. Environmental Protection Agency drinking water standards[20] and World Health Organization drinking water guidelines[21] (details provided in the Supplemental Materials)."

- Due to word count restrictions, we have limited the discussion of the results in the main manuscript to the linear model results. Therefore, the need for two separate paragraphs describing the drinking water and urinary results became unnecessary. We included additional results in the *Results* section of the supplemental materials. New tables

suggested by [redacted] have been added to the supplemental materials (eTable 4 and eTable 5; excerpts provided below). As suggested by [redacted], these tables were reordered to go from the least restrictive (all data) to most restrictive (<1.5 mg/L) exposure levels. These tables provide the overall linear results separately for drinking water and urinary results, the numbers of low and high risk-of-bias studies across each group, and the results when restricted to only the low risk-of-bias studies.

Excerpt of eTable 4. Dose-Response Meta-analysis Using Mean Effects—Model Selection

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Water Fluoride – All Studies					
No. Studies/No. Observations Number of Children		29/39 11,656	21/27 8,723	7/9 2,971	7/7 2,832
Linear Model[b]	Beta (95% CI) p-value AIC	−0.15 (−0.20, −0.11) p < 0.001 AIC = 53.8	−0.22 (−0.27, −0.17) p < 0.001 AIC = 16.1	−0.15 (−0.41, 0.12) p = 0.274 AIC = 11.8	0.05 (−0.36, 0.45) p = 0.816 AIC = 8.2
Quadratic Model[c]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.27 (−0.34, −0.21); p < 0.001 0.02 (0.01, 0.03); p < 0.001 AIC = 48.8 p* < 0.001	−0.12 (−0.35, 0.11); p = 0.318 −0.04 (−0.10, 0.03); p = 0.280 AIC = 21.2 p* = 0.012	0.79 (−0.01, 1.58); p = 0.052 −0.56 (−0.97, −0.16); p = 0.006 AIC = 12.5 p* = 0.007	0.30 (−0.53, 1.14); p = 0.477 −0.23 (−1.01, 0.55); p = 0.561 AIC = 11.3 p* = 0.04
Restricted Cubic Splines Model[d]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.29 (−0.39, −0.20); p < 0.001 0.48 (0.18, 0.78); p = 0.002 AIC = 42.3 p* < 0.001	−0.14 (−0.34, 0.06), p = 0.162 −0.23 (−0.66, 0.20), p = 0.295 AIC = 16.9 p* = 0.009	1.15 (0.07, 2.22) p = 0.037 −1.20 (−2.03, −0.36) p = 0.005 AIC = 10.5 p* = 0.010	0.49 (−0.50, 1.47) p = 0.334 −0.69 (−2.40, 1.02) p = 0.428 AIC = 10.2 p* = 0.05
Water Fluoride – Low Risk-of-bias Studies					
No. Studies/No. Observations Number of Children		6/11 4,355	6/9 4,251	3/4 921	3/3 879
Linear model	Beta (95% CI) p-value AIC	−0.19 (−0.34, −0.05) p = 0.009 AIC = 10.3	−0.22 (−0.36, −0.07) p = 0.003 AIC = 3.9	−0.34 (−0.72, 0.03) p = 0.070 AIC = 4.5	−0.32 (−0.91, 0.26) p = 0.276 AIC = 4.1

Excerpt of eTable 5. Dose-response Meta-analysis Using Mean Effects: Maximum Likelihood Models

Exposure Analysis	Parameters	Fluoride Exposure			
		All data	<4 mg/L	<2 mg/L	<1.5 mg/L
Urinary Fluoride – All Studies					
No. Studies/No. Observations Number of Children		18/32 8,502	13/26 6,885	7/11 4,654	5/8 3,992
Linear Model[b]	Beta (95% CI) p-value AIC	−0.16 (−0.23, −0.08) p < 0.001 AIC = 69.2	−0.17 (−0.29, −0.06) p = 0.004 AIC = 64.2	−0.07 (−0.13, 0.003) p = 0.060 AIC = −3.7	−0.12 (−0.36, 0.12) p = 0.325 AIC = 0.8
Quadratic Model[c]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.19 (−0.44, 0.06); p = 0.131 0.01 (−0.02, 0.05); p = 0.462 AIC = 73.0 p*= 0.14	0.08 (−0.21, 0.37); p = 0.587 −0.08 (−0.16, 0.0004); p = 0.051 AIC = 67.2 p* = 0.08	−0.23 (−0.62, 0.17); p = 0.267 0.08 (−0.12, 0.29); p = 0.423 AIC = 1.7 p* = 0.42	−0.11 (−1.45, 1.23); p = 0.868 0.02 (−0.74, 0.77); p = 0.967 AIC = 4.1 p* = 0.10
Restricted Cubic Splines Model[d]	Beta (95% CI); p-value Beta (95% CI); p-value AIC p-value*	−0.12 (−0.28, 0.04); p = 0.138 −0.10 (−0.41, 0.21); p = 0.524 AIC = 72.9 p*= 0.13	−0.03 (−0.21, 0.15); p = 0.775 −0.24 (−0.47, −0.02); p = 0.034 AIC = 66.8 p* = 0.07	−0.13 (−0.29, 0.03); p = 0.107 0.12 (−0.14, 0.38); p = 0.366 AIC = 1.5 p* = 0.37	−0.26 (−0.72, 0.20); p = 0.270 0.36 (−0.58, 1.29); p = 0.453 AIC = 3.5 p* = 0.07
Urinary Fluoride – Low Risk-of-bias Studies					
No. Studies/No. Observations Number of Children		9/15 5,713	9/15 5,713	5/8 4,141	4/7 3,952
Linear model	Beta (95% CI) p-value AIC	−0.10 (−0.20, 0.004) p = 0.059 AIC = 2.0	−0.10 (−0.20, 0.004) p = 0.059 AIC = 2.0	−0.07 (−0.14, 0.01) p = 0.073 AIC = −1.8	−0.08 (−0.16, −0.01) p = 0.028 AIC = −2.2

1.L: ***Discussion section:*** *"However, the associations did not remain significant when exposure was restricted to <1.5 mg/L, the current WHO safe water guideline."*

[redacted] **Comment:** There it is! Was looking for that info in the abstract and the methods/results. [redacted] question though—why use these cutoffs for the urinary levels as well? Do they directly translate?

Response: Agree (no change requested)

- These cutoffs are useful for comparison across different exposure measures. Drinking water levels roughly translate to urinary levels, but there is variation depending on the level of fluoride in the drinking water as well as individual behaviors. There is literature suggesting that among people living in areas with high levels of fluoride in drinking water, 1 mg/L in drinking water fluoride approximates 1 mg/L in urinary fluoride (e.g., Kumar et al. 2017); however, there is also literature suggesting that, at lower drinking water fluoride levels, drinking water only represents a portion of a person's total exposure to fluoride (EPA 2010), which includes exposure from other sources like dental products, foods, and beverages. Therefore, relying on drinking water levels may underestimate exposure.

  References

  S Kumar, S Lata, J Yadav, and JP Yadav (2017) Relationship between water, urine and serum fluoride and fluorosis in school children of Jhajjar District, Haryana, India. Appl Water Sci 7, 3377–3384. https://doi.org/10.1007/s13201-016-0492-2

  EPA (2010) Fluoride: Exposure and Relative Source Contribution Analysis. 890-R-10-015. US Environmental Protection Agency. Office of Water. Washington, D.C. Available at https://www.epa.gov/sites/default/files/2019-03/documents/fluoride-exposure-relative-report.pdf

1.M: ***Discussion section:*** *"While the results of our meta-analyses were consistent with two previous meta-analyses (Table 2), they differed in several ways. Our meta-analyses included more recently published studies that are lower risk of bias, and studies with different exposure assessment types. …If children with higher exposures had a greater IQ deficiency than children with lower exposures, the highly exposed children may have driven the mean IQ deficits of the entire group. Therefore, it is important to keep in mind that fluoride concentrations in drinking water alone do not reflect the magnitude of fluoride exposures to children who consume excessive amounts of fluoridated toothpaste or to formula-fed babies who consume powdered formula that is reconstituted with fluoridated water."*

[redacted] **Comment:** Not following the point you're trying to make in these sentences. It seems like first you were trying to make a point about how this meta-analysis is better than the previous, but then you're commenting on a problem with exposure assessment more generally. Seems like these should be separate discussion paragraphs.

Response: Agree (change made)

- We agree that the original paragraph sounds disjointed. Therefore, we revised the text and separated these topics into different paragraphs in the *Discussion* section as shown below:

  "Whereas the previously published meta-analyses only included group-level exposures, the regression slopes meta-analysis included nine studies with individual urinary fluoride measures, a more precise exposure measure. It also included recent North American

prospective cohort studies[5-7] with maternal urinary fluoride levels comparable to those found in the United States.[57]"

In a later paragraph in the Discussion section, we say:

"Fluoride exposure may vary considerably depending on individual behaviors and is best captured by individual-level measures of total exposure, such as urinary fluoride measures. Because drinking water measures capture only some of a person's total exposure to fluoride, it is reasonable to assume that some children in the meta-analysis had higher exposure to fluoride and those children may have skewed the mean IQ deficits of the entire group. Urinary fluoride levels include all ingested fluoride and are considered a valid measure to estimate total fluoride exposure.[61, 62]"

1.N: ***Discussion section:*** *"Consistent with previous literature, our dose-response meta-analysis shows statistically significantly lower children's IQ with increasing fluoride exposure."*

[redacted] **Comment:** Individual studies or the meta-analyses?

Response: Agree (change made)

- This was referring to one meta-analysis and another literature review, but we have since removed "consistent with previous literature" and directly cite the literature we are referring to in the *Discussion* section as follows:

  "The Duan et al.[4] meta-analysis reported a significant non-linear dose-response relationship above 3 ppm [3 mg/L] in water. A more recent literature review[56] did not comment on the shape of the dose-response curve; however, based on the three publications from Mexico and Canada,[5-7] the author concluded that the association between maternal urinary fluoride and children's neurotoxicity appeared to be "dose dependent."

1.O: ***Discussion section:*** *"Consistent with previous literature, our dose-response meta-analysis shows statistically significantly lower children's IQ with increasing fluoride exposure. Duan et al (2018) suggested a significant non-linear dose-response relationship above 3ppm [3 mg/L] in water [5]."*

[redacted] **Comment:** [redacted] having a hard time interpreting what this means. Relationship was non-linear, but there was an effect when levels were above 3 mg/L in the water (but not when levels were lower than that)? [redacted] not sure [redacted] would say this is "consistent with the literature" since it is just one other study? Maybe [redacted] just am not used to this dose-response language but consider that your other readers may not be either.

Response: Agree (change made)

- We have revised this part of the *Discussion* section to no longer say "consistent with the literature."

1.P: ***Discussion section****: "Our dose-response meta-analysis also revealed a significant dose-response relationship at <2mg/L fluoride in drinking water, levels that occur naturally in some U.S. drinking water systems."*

[redacted] **Comment:** [redacted] think you could consider language like "when levels were restricted to those below <2mg/L". [redacted] find it confusing because it's almost like you're comparing <2mg/L to some other level, which isn't what you're doing

Response: Agree (change made)

- We have removed this sentence from the manuscript and revised text in the supplemental materials to clarify that the dose-response analyses were restricted to (1) <4 mg/L, (2) <2 mg/L, and (3) <1.5 mg/L as follows:

  "When analyses were restricted to exposed groups with <4 mg/L (i.e., 0 to <4 mg/L) fluoride in drinking water (n = 21 publications [6 low and 15 high risk-of-bias studies]), there was a statistically significant inverse association between fluoride exposure and children's IQ (SMD: −0.22; 95% CI: −0.27, −0.17; p-value < 0.001) (eTable 4). When restricted to <2 mg/L (i.e., 0 to <2 mg/L) in drinking water (n = 7 publications [3 low and 4 high risk-of-bias studies]), the magnitude of the effect estimate did not substantially change (SMD: −0.15; 95% CI: −0.41, 0.12; p-value = 0.274). However, when restricted to exposed groups with <1.5 mg/L (i.e., 0 to <1.5 mg/L) in drinking water (n = 7 publications [3 low and 4 high risk-of-bias studies]), there was no longer an association between fluoride in drinking water and children's IQ (SMD: 0.05; 95% CI: −0.36, 0.45; p-value = 0.816)."

1.Q: ***Discussion section:*** *"Our dose-response meta-analysis also revealed a significant dose-response relationship at <2mg/L fluoride in drinking water, levels that occur naturally in some U.S. drinking water systems. As of April 2020, the CDC estimated that 0.59% of persons living in the United States (~ 1.9 million people) were served by community water systems (CWS) containing ≥ 1.5 mg/L naturally occurring fluoride and 0.31% (~1 million people) were served by CWS containing ≥ 2 mg/L (https://www.cdc.gov/fluoridation/data-tools/reporting-system.html)."*

[redacted] **Comment:** Not sure the point you're trying to make here. Shouldn't you be comparing CDC estimates to where levels are <2mg/L? [redacted] think these statistics are really important for contextualizing results but not sure this is the best result to compare them to?

Response: Agree (change made)

- We agree that our point was not clear. The point we wanted to make was that, even though the recommended level of artificially fluoridated water in the United States is 0.7 mg/L, some people may still be exposed to higher levels of naturally occurring fluoride in their drinking water. The revised text reads as follows:

  "For community water systems that add fluoride, the Public Health Service recommends a fluoride concentration of 0.7 mg/L; however, it is important to note that there are regions of the United States where public systems and private wells contain natural fluoride concentrations of more than 2 mg/L.[58] In April 2020, the Centers for Disease Control and Prevention (CDC) estimated that community water systems supplying water with ≥2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people).[59] For the purposes of reducing dental fluorosis, the CDC recommends that parents use an alternative source of water for children aged 8 years and younger and for bottle-fed infants if their primary drinking water contains greater than 2 mg/L of fluoride.[60]"

1.R: ***Discussion section:*** *"However, because all studies were considered lower risk of bias along with the moderate statistical heterogeneity and robustness our findings suggest that the small number of studies is unlikely to have influenced the meta-analysis findings."*

[redacted] **Comment:** Reword this sentence kind of run on

Response: Agree (change made)

- We agree with [redacted] comment; however, in the process of revising the manuscript, we have removed that sentence from the *Discussion* section and the issue no longer applies.

1.S: ***Conclusions section:*** *"The association remained statistically significant when restricted to <2 mg/L fluoride in drinking water (p-value<0.001), levels that occur naturally in some U.S. community water systems."*

[redacted] **Comment:** For more impact you could clarify this here with some statistic

Response: Agree (change made)

- We have removed this sentence from the *Conclusions* section because it was no longer accurate after the literature update.

Peer Review of the draft Meta-analysis Manuscript to Evaluate the Association between Fluoride Exposure and Children's Intelligence

[redacted] received a draft version of the manuscript as well as a copy of the NASEM Committee comments on the meta-analysis and the NIEHS/DNTP responses (draft version of Sup01_Meta-analysis). [redacted] provided comments in track changes on the draft manuscript in Microsoft Word. The full comments have been reproduced below verbatim along with the specific text referred to by [redacted] as quotes under a heading for the specific section of the document (e.g., "Abstract section"). Note that the red formatted text was in the document as provided. Formatting has been applied to aid in reading. Responses have been added in blue text following each of the comments beginning with the word "**Response**" in bold font.

[redacted]
Date: July 1, 2021

2.A: [redacted] **Comment:** Is the paper being submitted alongside a companion SR to discuss non-meta issues with fluoride and IQ?

Response: Agree (no change requested)

- Yes, the NTP Monograph on the systematic review of fluoride exposure and cognitive neurodevelopmental health effects is being published first and is referred to and cited in the *Methods* section as follows:

  "The search, selection, extraction, and risk-of-bias evaluation of studies for this meta-analysis were part of a larger systematic review.[8]"

2.B: Abstract section

[redacted] **suggested text:** [redacted] inserted text (shown here in red font) as follows: "*To perform a systematic review and meta-analysis to investigate epidemiological? associations between fluoride exposure and children's intelligence*."

[redacted] **Comment:** Even though this is for [NIEHS/DNTP removed journal name], [redacted] still think somewhere you should specify these are only epi studies

Response: Disagree (no change)

- We consider the objective in the *Abstract* to already imply that the meta-analysis only includes epidemiological studies with the word "children's" (i.e., "to investigate associations between fluoride exposure and children's intelligence"). In addition, details such as study eligibility are provided in the meta-analysis protocol (see Appendix 6 to the systematic review protocol located here: https://ntp.niehs.nih.gov/go/785076) and the *Methods* section of the manuscript.

2.C: Introduction section

[redacted] **suggested text:** [redacted] inserted text (shown here in red font) adding one sentence as follows**:** *This analysis was used to inform a larger systematic review on fluoride exposure and neurodevelopment*.

[redacted] **Comment:** [redacted] added this (to the Introduction) because doing a meta in isolation is bound to raise flags with reviewers, so best to mention this before the methods.

Response: Agree (change made)

- New text was added to the *Introduction* section to say that the meta-analysis complements "a larger systematic review" as follows:

  "To incorporate this newer evidence, and to complement a larger systematic review[8] that concluded there is moderate confidence in the evidence of an inverse association between fluoride exposure and children's IQ, we conducted a meta-analysis of studies that provided group-and individual-level fluoride exposure measurements in relation to children's IQ scores."

2.D: **Methods section:** "Literature searches were conducted in BIOSAS, EMBASE, PsychINFO, PubMed, Scopus, Web of Science, CNKI, and Wanfang databases through May 1, 2020 without language restrictions. Search strategies are available in the protocol.[8]"

[redacted] **Comment:** This also may raise flags if you are submitting 1+ year later. Worthwhile to be explicit for the early cutoff date (i.e., "the cutoff date chosen as part of our larger SR").

Response: Agree (change made)

- The literature search was updated in November 2021; therefore, the *Methods* section contains revised text:

  "Literature searches were conducted in BIOSIS, EMBASE, PsychINFO, PubMed, Scopus, Web of Science, CNKI, and Wanfang databases through November 2021, without language restrictions."

2.E: [redacted] **Comment:** You may benefit from upfront defining your inclusion criteria in a Supp table to stave off queries of age ranges, neurocognitive tests, etc.

Response: Disagree (no change)

- We define inclusion criteria in the protocol, which is referenced in the *Methods* sections of the manuscript and the supplemental materials, respectively, as follows:

  "To be eligible for inclusion, individual study publications had to satisfy review eligibility criteria outlined in the protocol.[9]"

  "In order to be eligible for inclusion in the systematic literature review, individual study publications (referred to in this paper as "studies") had to satisfy eligibility criteria outlined in the protocol (i.e., address PECO statement in Table 1 and specific exclusion criteria in Table 2, https://ntp.niehs.nih.gov/go/785076*)."*

2.F: **Methods section:** "The other risk-of-bias questions were also considered and were used to identify any other concerns that may indicate serious risk-of-bias issues (e.g., statistical analysis)."

[redacted] **Comment:** Maybe also mention as another example "selection bias", another critical domain in these studies

Response: Agree (change made)

- We have revised this sentence in the *Methods* section as follows:

"The other risk-of-bias questions were used to identify other concerns that may indicate serious risk-of-bias issues (e.g., selection bias, statistical analysis)."

2.G: **Methods section:** "No study was excluded from the meta-analysis based on concerns for risk of bias; however, subgroup analyses were conducted with and without higher risk-of-bias studies (i.e., studies rated probably high risk of bias for at least two key risk-of-bias questions or definitely high for any single question) to assess their impact on the results."

[redacted] **Comment:** Did you consider the magnitude and direction of the risk of bias? If your high RoB studies are high risk of bias for non-diff exp and outcome misclassification and all of your results bias toward the null, it may impact the interpretation of your results

Response: Agree (no change requested)

- Yes, this information has been considered and is available in Appendix E to the prepublication 2022 NTP Monograph.

2.H: **Methods section:** "Subgroup analyses were stratified by risk of bias (higher or lower), study location (e.g., country), outcome assessment, exposure matrix (e.g., urine or water), pre- or post-natal exposure, gender-specific groups, and age-specific subgroups."

[redacted] **Comment:** Are there others apart from urinary biomarkers and water exposure? Also, [redacted] maybe categorize as "urinary F, water F concentrations" or "biomonitoring, environmental sampling"

Response: Agree (change made)

- We have revised this sentence in the *Methods* section as follows:

  "Predefined subgroup analyses were stratified by risk of bias (high or low), study location (e.g., country), outcome assessment, exposure matrix (e.g., urinary fluoride or water fluoride concentrations), sex, and age group."

- Other exposure types were also considered, such as fluoride intake (see excerpt of Table 3 below).

  Excerpt of Table 3. Pooled Regression Slopes and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Effect				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Subgroup Analyses				
Exposure Type				
Urinary fluoride	9	−1.81 (−2.80, −0.81)	<0.001	77%
Intake	2	−3.87 (−7.15, −0.59)	0.737	0%
Water fluoride	2	−4.77 (−9.09, −0.45)	0.707	0%

2.I: **Methods section:** "The study outcomes were evaluated with respect to a 1-mg/L unit increase in exposure."

[redacted] **Comment:** Is a 1 mg/L unit increase the same for urinary F (biomarker) and water measures (envir monitoring)?

Response: Agree (change made)

- We have revised this sentence in the *Methods* section as follows:

  "The study outcomes were evaluated with respect to a 1-mg/L unit increase in water or urinary fluoride, or 1-mg/day fluoride intake."

2.J: **Results section:** "The meta-analysis of 46 studies (37 lower risk of bias studies and 9 higher risk of bias studies) that provided mean IQ scores showed that children exposed to higher fluoride levels had statistically significantly lower IQ scores (random-effects pooled SMD, −0.49; 95% CI: −0.60, −0.38; p-value<0.001) (Figure 2)."

[redacted] **Comment:** That is ½ an IQ point. That's a big deal.

Response: Agree (no change requested)

2.K: **Results section:** "There was evidence of high heterogeneity (I^2 = 89%, p-value < 0.001, Table 2) and publication bias (funnel plot and Egger's p-value < 0.001, Begg's p = 0.04, eFigures 2 and 3)."

[redacted] **Comment:** Given the high level of heterogeneity, should you mention that this is further justification to conduct subgroup analyses (in addition to your justification for only reporting RE models)? Just a suggestion.

Response: Disagree (no change)

- We do not consider this to be necessary given that the protocol and *Methods* section describe that prespecified subgroup analyses were performed to investigate sources of heterogeneity.

2.L: **Results section:** "Among the higher risk-of-bias studies (n = 37), the random-effects pooled SMD was −0.55 (95% CI: −0.68, −0.43) with high heterogeneity (I^2 = 84%, p-value < 0.001) (Table 2 and eFigure 6)."

[redacted] **Comment:** Just a note that seeing consistency in your metas between your high and low RoB studies adds to your justification of an association

Response: Agree (no change requested)

2.M: **Results section:** "The overall pooled effect estimate from the six studies with individual-level urinary fluoride measures shows that a 1-mg/L increase in urinary fluoride is associated with a statistically significant lower IQ score of 1.58 points (95% CI: −2.63, −0.53; p-value=0.003)."

[redacted] **Comment:** Wow; is this after accounting for potential confounders? If so, that is substantial

Response: Agree (no change requested)

- Correct, this represents the pooled effect estimate using each study's adjusted regression coefficient.

2.N: **Results section**: "Adjusting for possible publication bias through trim-and-fill analysis supports the conclusion that a 1-mg/L increase in urinary fluoride was associated with lower IQ, with an adjusted pooled effect estimate of −0.87 (95% CI: −1.93, 0.19) (eFigure 19)."

[redacted] **Comment:** Report your p-value

Response: Agree (change made)

- We added p-values throughout the *Results* section.

2.O: **Results section:** "A 1-mg/L increase in fluoride intake and water fluoride are also significantly associated with a lower IQ score of 3.87 points (95% CI: −7.15, −0.59; p-value=0.021) and 4.77 points (95% CI: −9.10, −0.45; p-value=0.031), respectively (Table 3); however, the results for both metrics are based on a small sample of studies (n=2 for each measure) and should be interpreted with caution."

[redacted] **Comment:** Is an N of 2 even worth reporting, or including as a main result?

Response: Agree (change made)

- We have replaced the above sentence with the following:

  "The results for fluoride intake and water fluoride levels are available in Supplemental Materials."

2.P: [redacted] **Comment:** This section (Discussion) is great, but it is missing a robust discussion of the biological plausibility or proposed mechanism of action. A ton of implications that [redacted] think deserves its own pub.

Response: Disagree (no change)

- Potential biological mechanisms are covered in the state of the science report. However, currently, the data on mechanisms are too limited and heterogeneous to make a determination of biological plausibility and therefore we do not think it is appropriate to include this in the *Discussion*. However, we do agree this is an important area for continuing study and deserves a separate analysis and publication expanding on the potential limitations and promising research on mechanisms.

2.Q: **Discussion section:** "The results of our three meta-analyses support an inverse association between fluoride exposure and children's IQ. Results were robust to stratification by risk of bias, gender, age group, timing of exposure, study location, outcome assessment type, and exposure assessment type. The association remained statistically significant when the exposed group was restricted to <2 mg/L fluoride in drinking water (p-value<0.001), levels that occur naturally in some U.S. community water systems. However, the associations did not remain significant when exposure was restricted to <1.5 mg/L, the current WHO safe water guideline."

[redacted] **Comment:** Somewhere in this paragraph consider adding a sentence that this meta is used to inform the larger SR, and the meta is a piece of the larger equation

Response: Disagree (no change)

- We have included that this meta-analysis is part of the larger systematic review in the *Introduction* and *Methods* sections, respectively, as follows:

"To incorporate this newer evidence, and to complement a larger systematic review[8] that concluded there is moderate confidence in the evidence of an inverse association between fluoride exposure and children's IQ, we conducted a meta-analysis of studies that provided group-and individual-level fluoride exposure measurements in relation to children's IQ scores."

"The search, selection, extraction, and risk-of-bias evaluation of studies for this meta-analysis were part of a larger systematic review.[8]"

2.R: **Discussion section:** "Individual levels are a more precise measure of exposure compared to group-level measures; however, drinking water levels comprise only a portion of a person's total exposure to fluoride."

[redacted] **Comment:** Do you mean "household concentrations" or something else?

Response: Agree (change made)

- "Drinking water levels" in the above sentence referred to individual exposures to drinking water. However, during our revisions, this sentence was removed.

2.S: **Discussion section**

[redacted] **suggested text:** [redacted] inserted text (shown here in red font) at the end of the following sentence: "*Consequently, it is reasonable to assume that some children in our mean effects meta-analyses had higher exposure to fluoride from other common sources (e.g., dental products, foods and beverages)*; *though these are generally considered negligible relative to water*."

[redacted] **Comment:** [redacted] almost positive the relative source contribution of water compared to other sources is really disparately large

Response: Agree (change made)

- It has been estimated that other sources make up about 30% of total fluoride exposure.
- We have revised this sentence in the *Discussion* section as follows:

  "Because drinking water measures capture only some of a person's total exposure to fluoride, it is reasonable to assume that some children in the meta-analysis had higher exposure to fluoride and those children may have skewed the mean IQ deficits of the entire group."

2.T: [redacted] **Comment:** [redacted] know a major concern for EPA and other groups is infants whose sole source of consumption is formula from reconstituted tap water.

Response: Agree (change made)

- We have revised the *Discussion* section to include the following:

  "For the purposes of reducing dental fluorosis, the CDC recommends that parents use an alternative source of water for children aged 8 years and younger and for bottle-fed infants if their primary drinking water contains greater than 2 mg/L of fluoride."

2.U: **Discussion section:** "There are also several limitations to consider. Studies included in our meta-analyses also had various intrinsic limitations."

[redacted] **Comment**: Such as?

Response: Agree (change made)

- We have revised this sentence in the *Discussion* section as follows:

  "Most of the studies included in the mean-effects and dose-response mean effects meta-analyses were considered to have study design and/or methodological limitations. For example, all but three studies were cross-sectional in design."

2.V: [redacted] **Comment:** Table 1: You need to include study design in table 1

Response: Agree (change made)

- We have revised Table 1 to include study design in the first column (excerpt provided below).

  Excerpt of Table 1. Characteristics of Studies Included in the Meta-analysis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Ren et al. (1989)[66] [translated in Ren et al. 2008][me, o] *Cross-sectional*	China	8–14	No fluoride measurement Low iodine village/high fluoride and low iodine village	Not specified	Wechsler Intelligence Scale for Children	High	Sex; iodine
Chen et al. (1991)[68] [translated in Chen et al. 2008][me, w] *Cross-sectional*	China	7–14	Drinking water Nonendemic/endemic fluorosis village	0.89 mg/L (nonendemic) 4.55 mg/L (endemic)	Chinese Standardized Raven Test	High	Age; sex
Guo et al. (1991)[70] [translated in Guo et al. 2008a][me, o] *Cross-sectional*	China	7–13	Serum Reference area using wood/coal burning-related fluoride endemic area	0.1044 ± 0.0652 mg/L (reference) 0.1483 ± 0.0473 mg/L (endemic)	Chinese Binet Intelligence Test	High	Age; sex; SES

2.W: [redacted] **Comment:** Figure 1: Add a Y axis (even if it is in the title)

Response: Agree (change made)

- We have revised Figure 1 to include the y-axis description in the title as follows:

  "Figure 1. Number of Studies of Fluoride Exposure and IQ in Children by Country and Year of Publication".

2.X: [redacted] **Comment:** Figure 2: Do you have space to add additional columns to increase the informativeness of this forest plot? Adding in the Ns and study designs would be helpful, but most importantly the exposure assessment used.

Response: Disagree (no change)

- We have kept Figure 2 as is for readability, but the subgroup analysis stratified by exposure assessment type is included in Table 2 (excerpt provided below).

Excerpt of Table 2. Pooled SMDs and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	SMD (95% CI)	Heterogeneity	
			p-value	I^2
Overall Effect	55	−0.46 (−0.55, −0.37)	<0.001	87%
Subgroup Analyses				
Risk of Bias				
Low	10	−0.22 (−0.39, −0.05)	<0.001	83%
High	45	−0.52 (−0.63, −0.42)	<0.001	86%
Sex				
Males	14	−0.62 (−0.81, −0.42)	<0.001	78%
Females	13	−0.53 (−0.72, −0.34)	<0.001	74%
Age Group				
<10 years[a]	13	−0.41 (−0.60, −0.22)	<0.001	80%
≥10 years	16	−0.55 (−0.70, −0.40)	<0.001	68%
Country				
China	39	−0.43 (−0.52, −0.34)	<0.001	85%
India	8	−0.99 (−1.55, −0.43)	<0.001	93%
Iran	4	−0.68 (−0.99, −0.38)	0.077	56%
Canada	1	0.01 (−0.19, 0.21)	NA	NA
Mexico	1	0.13 (−0.16, 0.42)	NA	NA
New Zealand	1	0.01 (−0.19, 0.22)	NA	NA
Pakistan	1	−0.25 (−0.65, 0.16)	NA	NA
Assessment Type				
CRT-RC tests	29	−0.36 (−0.46, −0.27)	<0.001	82%
Non-CRT-RC tests	26	−0.60 (−0.78, −0.42)	<0.001	89%
Raven's tests	10	−0.76 (−1.10, −0.43)	<0.001	91%
Other tests	16	−0.52 (−0.74, −0.29)	<0.001	89%
Exposure Type				
Water fluoride	32	−0.37 (−0.46, −0.27)	<0.001	82%
Dental fluorosis	7	−0.99 (−1.57, −0.41)	<0.001	96%
Other exposures[b]	16	−0.54 (−0.71, −0.37)	<0.001	81%
Previous Meta-analyses				
Duan et al. (2018)[4]	26	−0.52 (−0.62, −0.42)	<0.001	69%
Choi et al. (2012)[3]	27	−0.45 (−0.56, −0.34)	<0.001	80%

Peer Review of the draft Meta-analysis Manuscript to Evaluate the Association between Fluoride Exposure and Children's Intelligence

[redacted] received a draft version of the manuscript as well as a copy of the NASEM Committee comments on the meta-analysis and the NIEHS/DNTP responses (draft version of Sup01_Meta-analysis). The full [redacted] comments have been reproduced below verbatim. Formatting has been applied to aid in reading. Responses have been added in blue text following each of the comments beginning with the word "**Response**" in bold font.

[redacted]
Date: September 14, 2021

3.A: [redacted] **Comment:** [redacted] this review represents an enormous effort. Meta-analyses are not my specialty but, by all evidence, what you have done is state-of-the-art. [redacted] do have some thoughts on how the paper might be framed for a clinical journal.

Response: Agree (no change requested)

- We appreciate the comment that this meta-analysis is state of the art.

3.B: [redacted] **Comment:** [redacted] realize the NTP defines its role in its reports strictly and narrowly–but if this paper is intended for a medical journal, then readers will expect some context. The paper could benefit from a brief section in the introduction on things like the sources of fluoride (ground water vs water supplement vs diet vs dental treatment), the drinking-water levels considered "optimum," the levels associated with toxicity (dental fluorosis), and an idea of the range of levels found in human populations. (This is probably not a complete list – [redacted] not an expert in this area.)

Response: Agree (change made)

- We agree that context for the findings is important and have added (1) information on the sources of fluoride to the *Introduction* section; (2) drinking water levels considered optimal as recommended by the U.S. Public Health Service to the *Discussion* section (this also addresses the levels associated with dental fluorosis because the optimal level is meant to provide enough fluoride to prevent tooth decay in children and adults while limiting the risk of dental fluorosis); and (3) the degree of exposures to high levels of naturally occurring fluoride in the United States to the *Discussion* section. As for the range of levels found in human populations, this varies widely depending on the geographic location of the population, the source of the exposure, and individual behaviors. Therefore, we considered it best to provide the exposure levels for each individual study population as reported by the study authors, which are available in Table 1 (excerpt provided below).
- We added text to the *Introduction* section as follows:

  "Fluoride from natural sources occurs in some community water systems and, in the United States and some other countries, fluoride is added to public drinking water systems for the prevention of tooth decay. Water and water-based beverages are the main source of systemic fluoride intake; however, an individual's total exposure also reflects contributions from fluoride in other sources such as food, dental products, industrial emissions, and some pharmaceuticals."

- We added text to the *Discussion* section as follows:

  "For community water systems that add fluoride, the Public Health Service recommends a fluoride concentration of 0.7 mg/L; however, it is important to note that there are regions of the United States where public systems and private wells contain natural fluoride concentrations of more than 2 mg/L.[58] In April 2020, the Centers for Disease Control and Prevention (CDC) estimated that community water systems supplying water with ≥2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people).[59]"

  Excerpt of Table 1. Characteristics of Studies Included in the Meta-analysis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure: Assessment (Metric, Exposure Groups)	Fluoride Exposure: Levels	Intelligence Assessment	Overall RoB Rating	Confounders Considered
Broadbent et al. (2015)[38me, w*] *Prospective Cohort*	New Zealand	7–13	Drinking water Area without community water fluoridation (low)/area with community water fluoridation (high) Fluoride tablet use (never/ever) Fluoride toothpaste use (never/sometimes/always)	Water: 0.0–0.3 mg/L (low) 0.7–1.0 mg/L (high) Tablet use: 0 mg (never used) 0.5 mg (ever used) Range not specified for fluoride toothpaste use (always/sometimes/never)	Wechsler Intelligence Scale for Children-Revised	High	Sex; SES; low birth weight; breastfeeding
Cui et al. (2018)[34rs] *Cross-sectional*	China	7–12	Urine	Boys: 1.3 (0.9–1.7)[d] mg/L Girls: 1.2 (0.9–1.6)[d] mg/L	Combined Raven's Test for Rural China	Low	Age; maternal education; smoking in family member; stress; anger; dopamine receptor-2 polymorphism
Green et al. (2019)[6me, w*, u*, rs] *Prospective Cohort*	Canada	3–4	Maternal urine, drinking water, maternal fluoride intake Nonfluoridated/fluoridated area	Urine: 0.40 ± 0.27 mg/L (nonfluoridated) 0.69 ± 0.42 mg/L (fluoridated) Water: 0.13 ± 0.06 mg/L (nonfluoridated) 0.59 ± 0.08 mg/L (fluoridated) Intake: 0.30 ± 0.26 mg/day (nonfluoridated) 0.93 ± 0.43 mg/day (fluoridated) Overall: 0.51 ± 0.36 mg/L (urine) 0.54 ± 0.44 mg/day (intake) 0.31 ± 0.23 mg/L (water)	Wechsler Primary and Preschool Scale of Intelligence-III	Low	Sex; city; maternal education; race/ethnicity; HOME score; prenatal secondhand smoke exposure

3.C: [redacted] **Comment:** Is there a threshold effect? You report that "when restricted to exposed groups with <1.5 mg/L in drinking water…, there was no longer an association between fluoride in drinking water and children's IQ (SMD: 0.01; 95% CI: -0.37, 0.39; p-value=0.972)." This seems crucially important. The US Public Health Service recommends an optimum drinking-water concentration of 0.7 mg/L. According to an NHANES paper, 95% of US kids are exposed to drinking water with less than 1 mg/L fluoride. If all of this is correct, then the range at which the effects you find might actually occur are rare in the US, and perhaps other places as well. That needs to be emphasized.

Response: Agree (change made)

- Although [redacted] refers to a quote, the text has been paraphrased by [redacted], and the actual text in the version of the manuscript reviewed by [redacted] is as follows:

  "When the analyses were restricted to studies with the exposed group <1.5 mg/L fluoride in drinking water … there was a non-significant positive association between fluoride exposure and children's IQ (SMD, 0.32; 95% CI: −0.57, 1.20)."

- We agree that readers may consider the question of threshold and shape of the dose-response curve at low doses based on the results of the meta-analysis. We revised our discussion of the shape of the dose-response curve at low doses in the *Discussion* section of the manuscript as follows:

"There is inconsistency in which model is the best fit at lower exposure levels (eTable 4 and eTable 5) leading to uncertainty in the shape of the dose-response curve at these levels. More individual-level data would increase our certainty in the shape of the dose-response curve at these lower exposure levels."

- [redacted] links uncertainty in the shape of the dose-response curve at lower doses to a potential threshold and then focuses on drinking water concentrations only and the number of U.S. people with drinking water below 1 mg/L fluoride. However, as we stated in response to the previous question, there are multiple sources of fluoride that contribute to total exposure. We added a sentence to the *Introduction* section of the manuscript to emphasize the importance of total fluoride exposure and additional context to the *Discussion* section on the number of people in the United States served by water systems >2 mg/L fluoride as described below.

  Text was added to the *Introduction* section as follows:

  "Water and water-based beverages are the main source of systemic fluoride intake; however, an individual's total exposure also reflects contributions from fluoride in other sources such as food, dental products, industrial emissions, and some pharmaceuticals."

  Text was added to the *Discussion* section as follows:

  "For community water systems that add fluoride, the Public Health Service recommends a fluoride concentration of 0.7 mg/L; however, it is important to note that there are regions of the United States where public systems and private wells contain natural fluoride concentrations of more than 2 mg/L.[58] In April 2020, the Centers for Disease Control and Prevention (CDC) estimated that community water systems supplying water with ≥2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people).[59]"

3.D: [redacted] **Comment:** You know that people in certain quarters fear the government is "poisoning" water with fluoride. You are wading into that territory when you publish this in a medical journal, and you should provide as clear a picture of the practical implications as you can. The paper might benefit from adding a coauthor who is an expert in the clinical and public health context of fluoride research, and who could help connect this intensive statistical analysis to its public health setting.

Response: Disagree (no change)

- Yes, we are fully aware of the controversial nature of this topic. We are primarily interested in providing an accurate as possible analysis of the relevant literature with a transparent listing of strengths and limitations of the database. The issue of water fluoridation is not emphasized as we found no studies in the literature that were specifically designed or powered to examine this practice in relation to children's neurological development. In addition, we fully agree that the practical implications of this research are potentially wide ranging. However, given the additional analyses and scope of considerations involved, we consider the implications in the public health setting to be deserving of a more comprehensive risk-benefit analysis that is beyond the scope of this effort.

3.E: [redacted] **Comment:** The effect size appears markedly stronger in boys. [redacted] could not tell if this result might depend on the studies based on group measures rather than individual measures. If group measures are contributing, the result could be due to boys in hot climates drinking more water and thus having higher exposure. If the result is not persistent in studies that rest on individual exposure measures, then the interpretation could lean towards a biological vulnerability of boys. This seems like an important distinction to explore.

Response: Agree (change made)

- We have added text to the *Discussion* section to acknowledge this topic as a limitation as follows:

  "Although we conducted subgroup analyses by sex, only 1 of the 14 studies that reported IQ scores separately for boys and girls analyzed fluoride exposure for each sex separately.[6] This is essential for evaluating whether a differential change in IQ by sex may be related to higher susceptibility or higher exposure in that sex."

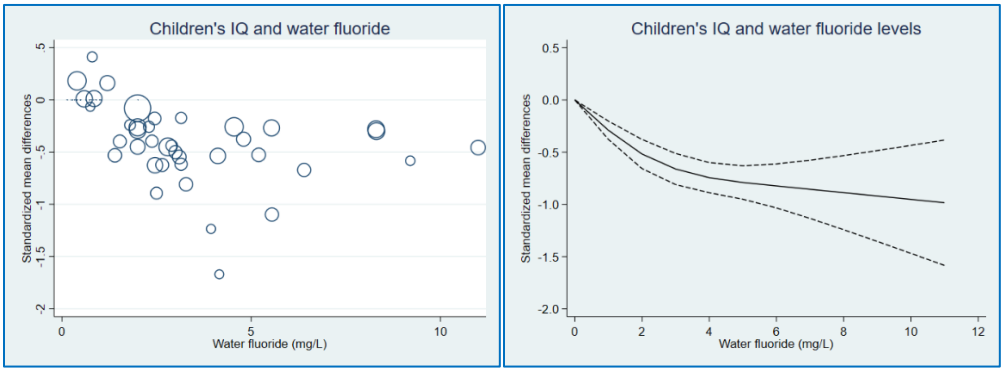
- This topic is also addressed more fully in the prepublication 2022 NTP Monograph which, now that it has undergone exhaustive peer review, will be cited in the manuscript.

3.F: [redacted] **Comment:** Some minor comments.

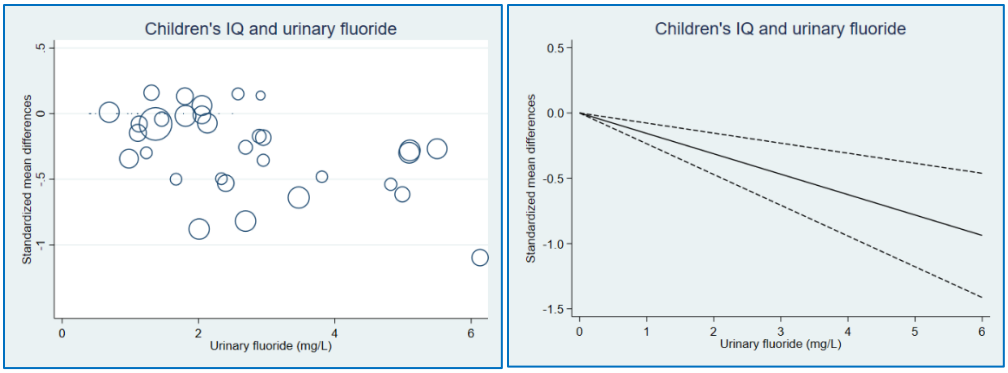
[redacted] **Comment:** [redacted] would suggest that the Y axis on the two panels in Supplementary Figure 16 be made on the same scale, so that it is easier to move between the two panels.

Response: Agree (change made)

- [redacted] is referring to eFigure 26. We agree with the commenter's suggestion on scale and also found this change could be applied to eFigure 25. Therefore, we have updated eFigure 25 and eFigure 26 (which are eFigure 17 and eFigure 18 in the current draft supplemental materials) so that the y-axes on the two panels use the same scale.



eFigure 17. Pooled Dose-Response Association Between Fluoride in Water and Standardized Mean Differences in Children's IQ



eFigure 18. Pooled Dose-Response Association Between Fluoride in Urine and Standardized Mean Differences in Children's IQ

3.G: [redacted] **Comment:** With regard to outcome, there should be some mention of the usual standard deviation of IQ in the population, to give a better idea of how large a one-point difference might be.

Response: Disagree (no change)

- The standard deviations of measured IQs are specific to study population. Since the meta-analyses we perform pool the results of many different study populations together, and range between mean-effects, dose-response, and regression slopes, we consider it to be misleading to provide a "usual standard deviation IQ."

3.H: [redacted] **Comment:** Medical journals typically do not allow footnotes.

Response: Agree (change made)

- All footnotes have been removed from the manuscript.

3.I: [redacted] **Comment:** [redacted] assume the words IQ "score" and IQ "point" are equivalent, but as [redacted] first read the abstract, [redacted] couldn't be sure. Consistent use might avoid any possible confusion.

Response: Agree (change made)

- When compared against appropriate population norms, the IQ score has a point value. Thus, decreases in IQ score can be expressed as numeric points. However, not all studies report scores in this manner, with some reporting only raw IQ test scores. We have reviewed text in the manuscript to make sure that all references to changes in IQ scores or points correctly reflect the underlying information.

3.J: [redacted] **Comment:** Finally, [redacted] found the constant "thanks" to [redacted] in each draft response to be distracting. Better to get to the point.

Response: Agree (change made)

- We have toned down the "thanks" in our responses to the NASEM Committee comments, which is reflected in Sup01_Meta-analysis and Sup01_Monograph.

3.K: [redacted] **Comment:** [redacted] hope this is useful. Thanks for the opportunity to have a look at this important piece of work. [[redacted]]

Response: Agree (no change requested)

Peer Review of the draft Meta-analysis Manuscript to Evaluate the Association between Fluoride Exposure and Children's Intelligence

[redacted] received a draft version of the manuscript as well as a copy of the NASEM Committee comments on the meta-analysis and the NIEHS/DNTP responses (draft version of Sup01_Meta-analysis). The full [redacted] comments have been reproduced below verbatim. Formatting has been applied to aid in reading. A response has been added in blue text following the comments beginning with the word "**Response**" in bold font.

[redacted]
Date: September 17, 2021

4.A: [redacted] **Comment**: [redacted] gone over the paper in detail, and this is excellent work. [redacted] genuinely don't have any concerns or suggestions. [redacted] think the analysis itself is excellent, and you thoroughly addressed comments.

Response: Agree (no change requested)

- We appreciate [redacted] comments that [redacted] does not have any concerns or suggestions and that we have thoroughly addressed the NASEM Committee comments.

In February 2022, the [redacted] provided comments to NIEHS/DNTP on the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* and a draft manuscript on a meta-analysis of fluoride exposure and IQ in children. This document contains a subset of the overall [redacted] comments related to the meta-analysis manuscript along with the NIEHS/DNTP responses. The meta-analysis-related comments from the [redacted] are reproduced here in black text, and the NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font.

February 1, 2022

[redacted]

Feedback to NTP/NIEHS regarding:

1. Fluoride state of the science document
2. Fluoride and IQ Meta-analysis manuscript

5.A: Issue: Keeping findings in context

As NASEM noted in their review of the 2019 Draft Monograph, "the context into which the monograph falls calls for much more carefully developed and articulated communication on this issue." [redacted] fully concurs with this recommendation and with NASEM's 2019 assessment that "NTP needs to state clearly that the monograph is not designed to be informative with respect to decisions about the concentrations of fluoride that are used for water fluoridation."

NTP stated in the revised draft of the monograph that the evidence of "effects on cognitive neurodevelopment are inconsistent, and therefore unclear" at the levels typically found in drinking water in the US. NASEM agreed with this assessment, stating that "[m]uch of the evidence presented in the report comes from studies that involve relatively high fluoride concentrations. Little or no conclusive information can be garnered from the revised monograph about the effects of fluoride at low exposure concentrations (less than 1.5 mg/L)."

[redacted] is extremely concerned that the revised 2021 NTP report and the meta-analysis omit this important context that was previously included. Without clarification, readers may interpret that exposure to fluoride at any concentration is associated with lower IQ, a conclusion that is not borne out by the available science or the findings of the systematic review.

Recommendation:

- [redacted] requests NTP include a statement in the systematic review abstract and fulltext, as well as the meta-analysis, like that found in the 2020 draft monograph: "When focusing on findings from studies with exposures in ranges typically found in drinking water in the United States (0.7 mg/L for optimally fluoridated community water systems) that can be evaluated for dose response, effects on cognitive neurodevelopment are inconsistent, and therefore unclear."

  Response: Disagree (no change)

  - We remain sensitive to the need to provide context concerning fluoride exposures in the United Stated from fluoridated water, and we have included the PHS recommendations for optimal water fluoridation in the meta-analysis manuscript. However, we also stress that the subject of our fluoride monograph and meta-analysis is total fluoride exposures from all sources. The 2022 update of the meta-analysis includes a number of new non-U.S. studies that further inform the relationship between IQ deficits in children and exposures to fluoride that were not available for inclusion in the 2020 draft NTP Monograph. These studies provide additional information to sharpen the dose-response mean-effects estimates and improve the *regression slopes meta-analysis*. Although the clarity of effects at lower fluoride exposures is improving, there are no studies on the potential association between fluoride exposures and IQ in children in the United States, and no nationally representative urinary fluoride levels are available, making it difficult to make more specific

statements about the relevance of our meta-analysis findings to the U.S. population.

Note: [redacted] comments on the animal studies for the prepublication 2022 NTP Monograph are not reproduced here as they are not relevant to the meta-analysis. See DocA1_Monograph for the monograph-relevant comments and responses.

5.B: Issue: Limitations section

In its response letter, NASEM requested adding clarifying information in the manuscript. NTP itemized items in the state-of-the-science manuscript on limitations of the evidence based and the systematic review. However, these limitations do not address the following issues comprehensively:

Note: [redacted] comments on the protocol and literature search (numbered as "1" and "2" in the original comments) for the prepublication 2022 NTP Monograph are not reproduced here as they are not directly relevant to the meta-analysis. To avoid confusion, the number "3" was removed from following comment. See DocA1_Monograph for the monograph-relevant comments and responses.

5.C: Some included studies with complex sample designs did not report if they used population weights to generate estimates.

Recommendation: In addition to listing this as a limitation, NTP should identify these studies in the body of the report.

Response: Agree (change made)

- We have addressed these issues in the meta-analysis. We specifically mentioned the studies in the *Results* section of supplemental materials. In addition, we performed a new sensitivity analysis excluding results from the studies that did not account for complex sampling strategies (Yu et al. (2018), Zhang et al. (2015b)). Based on this analysis, the pooled effect estimate did not change appreciably (see excerpt of eTable 6 below).

Excerpt of eTable 6. Regression Slopes Meta-analysis

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimate				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Sensitivity Analysis				
Excluding Yu et al. (2018)[3] and Zhang et al. (2015b)[110]				
Full-scale IQ	7	−1.76 (−2.90, −0.62)	<0.001	82%

- Additionally, our risk-of-bias assessment carefully considered accounting for sampling strategy or clustering in determining study-specific potential for bias. Our analyses stratify results by risk-of-bias status to evaluate the potential impact on the overall effect estimates from studies that have high potential for bias versus studies that have low potential for bias.

5.D: Clustering: NASEM identified that in some population studies, participants living in the same communities were assigned the same measure of fluoride exposure without considering the effect in the data analysis. These correlation may artificially increase the statistical power.

Recommendation: Limitations should note the studies where clustering was a potential threat and specifically whether the investigators addressed this.

Response: Agree (change made)

- Based on the NASEM Committee's comment, we revised text in Appendix E of the prepublication NTP 2022 Monograph (previously Appendix 4 in the 2020 draft NTP Monograph) to clearly specify which low risk-of-bias studies addressed clustering as a feature of the study design or statistical analysis. When clustering was not accounted for, we describe the expected impact that this may have on the study's results.

  We have performed several additional sensitivity analyses to address the NASEM Committee's comments on clustering (further described below). The new results are presented in eTable 3 and eTable 6 of the supplemental materials (excerpts provided below).

  For example, we added a sensitivity analysis excluding Trivedi et al. (2012) from the *mean-effects meta-analysis* (both the overall effect analysis and the low risk-of-bias subgroup analysis) to assess the impact of clustering. Excluding Trivedi et al. (2012) did not change the results appreciably. The results of this new sensitivity analysis compared to the main overall effect estimate are shown below in the excerpt of eTable 3.

Excerpt of eTable 3. Sensitivity Analyses for Mean-effects Meta-analysis: Pooled SMDs and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimates				
Overall effect	55	−0.46 (−0.55, −0.37)	<0.001	87%
Low risk of bias	10	−0.22 (−0.39, −0.05)	<0.001	83%
Sensitivity Analyses excluding Trivedi et al. (2012)[40]				
Overall effect	54	−0.46 (−0.56, −0.37)	<0.001	87%
Low risk of bias	9	−0.22 (−0.40, −0.04)	<0.001	85%

- As suggested by the NASEM Committee, lack of accounting for clustering has little impact on studies with individual-level exposure measures (e.g., urinary fluoride levels) that also account for important confounders capturing the cluster (city) effect. For example, the minimal impact of clustering is illustrated by Bashash et al. (2017) who accounted for clustering at the cohort level by using cohort as a fixed effect in the models. In addition, the models accounted for many important confounders, which are also likely to reflect the cohort effect. The unadjusted and adjusted effect estimates were similar (β [95% CI] = −2.37 [−4.45, −0.29] and −2.50 [−4.12, −0.59], respectively).
- In the case of Green et al. (2019), we contacted the study authors and received the results from models using city as a random intercept. The overall adjusted effect estimates with city as a fixed effect and with city as a random effect were not significantly different from each other (β [95% CI] = −1.95 [−5.19, 1.28] and −2.20 [−5.39, 0.98], respectively).
- To address the NASEM Committee's concerns about clustering, we performed two new sensitivity analyses—one using the unadjusted effect estimates from Bashash et al. (2017), Cui et al. (2018), Green et al. (2019), and Yu et al. (2018) and another using the estimates from the random effect models from Bashash et al. (2017) and Green et al. (2019). The additional sensitivity analyses had minimal impact on the overall results of the meta-analysis (see excerpt of eTable 6 below).

Excerpt of eTable 6. Regression Slopes Meta-analysis

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimate				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Sensitivity Analyses				
Using estimates from random effect models for Green et al (2019)[113] and Bashash et al. (2017)[112]				
Full-scale IQ	9	−1.80 (−2.80, −0.80)	<0.001	76%
Using unadjusted estimates from Bashash et al. (2017),[112] Cui et al. (2018),[76] Green et al. (2019)[113], Yu et al. (2018)[3]				
Full-scale IQ	9	−1.82 (−2.81, −0.83)	<0.001	76%

Note: [redacted] comment on contacting authors of studies considered in the prepublication 2022 NTP Monograph with reporting quality questions as part of the risk-of-bias assessment are not reproduced here as they are not directly relevant to the meta-analysis. See DocA1_Monograph for the monograph-relevant comments and responses.

Meta-Analysis: The meta-analysis, originally requested by NASEM to obtain measures of association and sensitivity analysis across selected studies was removed to be published separately.

Note: [redacted] comment on the association between the prepublication 2022 NTP Monograph and meta-analysis is not reproduced here as it is not directly relevant to the meta-analysis itself. See DocA1_Monograph for the monograph-relevant comments and responses.

[redacted] concluded their comment with the statement that:
[NTP] should address NASEM's critiques of the September 2020 draft (abstracted below):

5.E:

a. Provide sufficient information about each study to allow the reader to understand why particular outcomes/results were selected (data transparency)

Response: Agree (change made)

- The NASEM Committee suggested the addition of a table providing more information on each study that "would allow readers to identify which result from each study was used and support a better understanding of why NTP selected the results that it did for inclusion in the meta-analysis." We found the suggestion helpful and have newly included eTable 2 (Study Characteristics and Study-specific Effect Estimates Included in the Meta-analyses and Sensitivity Analyses; excerpt below) to clarify study details including selected effect estimates used from each study (i.e., means, standard deviations, sample sizes, regression slopes with 95% confidence intervals, and

exposure levels). The source of the results (e.g., table, figure) from each study publication is also listed. eTable 1 (excerpt below) provides details on excluded studies and studies with overlapping populations.

Excerpt of eTable 2. Study Characteristics and Study-specific Effect Estimates Included in the Meta-analyses and Sensitivity Analyses

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Bashash et al. (2017)[112,me, u, rs] *Prospective Cohort*	Mexico	6–12	Maternal urine Reference/high fluoride levels (based on children urinary fluoride)	<0.80 mg/L (reference) ≥0.80 mg/L (high)	77, 95.37 (10.31) 112, 96.80 (11.16)	77, 95.37 (10.31) 112, 96.80 (11.16)	−2.50 (−4.12, −0.59) per 0.5 mg/L maternal urinary F	Abstract, Table 3
Razdan et al. (2017)[73,sa] *Cross-sectional*	India	12–14	Drinking water Low/high fluoride levels	0.6 ppm (low) 4.99 ppm (high)	69, 38.61 (6.34) 75, 13.95 (5.14)			Table 2
Valdez Jiménez et al. (2017)[74sa] *Prospective Cohort*	Mexico	Infancy	Maternal urine, drinking water	Urine: 1.9 ± 1.0 mg/L (1st trimester) 2.0 ± 1.1 mg/L (2nd trimester) 2.7 ± 1.1 mg/L (3rd trimester) Water: 2.6 ± 1.1 mg/L (1st trimester) 3.1 ± 1.1 mg/L (2nd trimester) 3.7 ± 1.0 mg/L (3rd trimester)			Bayley MDI: −19.05 (8.9) per 1 log10 mg/L maternal urinary F (1st trimester) −19.34 (7.46) per 1 log10 mg/L maternal urinary F (2nd trimester)	Table 2, Table 4
Cui et al. (2018)[76,rs] *Cross-sectional*	China	7–12	Urine	Boys: 1.3 (0.9–1.7)[d] mg/L Girls: 1.2 (0.9–1.6)[d] mg/L			−2.47 (−4.93, −0.01) per 1 log urinary F	Table 2
Yu et al. (2018)[3,me, w, u*, rs] *Cross-sectional*	China	7–13	Maternal urine Low/medium/high fluoride ranges Drinking water Normal/high fluoride	Urine: 0.01–1.60 mg/L (low) 1.60–2.50 mg/L (medium) 2.50–5.54 mg/L (high) Water: ≤1 mg/L (normal) >1 mg/L (high) Overall: 0.01–5.54 mg/L (urine) 0.20–3.90 mg/L (water)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	0.36 (−0.29, 1.01) per 0.5 mg/L maternal urinary F	Table 1, Table 3

Excerpt of eTable 1. List of Excluded Studies from Mean-effects Meta-analysis

Reference, Country	Reason for Exclusion
Qin et al. (1990)[45] [translated in Qin et al. 2008], China	Missing mean or SD of outcome measure
Yang et al. (1994)[47] [translated in Yang et al. 2008], China	Overlapping population with Wang et al. (2001)[49]; Table 2 in Yang et al. (1994)[47] seemed incomplete
Wang et al. (2005b)[50] [translated in Wang et al. 2008a], China	Missing mean or SD of outcome measure
Rocha-Amador et al. (2007)[52], Mexico	Missing mean or SD of outcome measure
Liu et al. (2000)[53] [translated in Liu et al. 2008] , China	Overlapping population with Lu et al. (2000)[55]
Sudhir et al. (2009)[56], India	Missing mean or SD of outcome measure
He and Zhang (2010)[57], China	Missing mean or SD of outcome measure
Xiang et al. (2011)[58], China	Overlapping population with Xiang et al. (2003a)[59]
Saxena et al. (2012)[60], India	Missing mean or SD of outcome measure
Wang et al. (2012)[61], China	Overlapping population with Xiang et al. (2003a)[59]

5.F:

b. Conduct additional sub-group analyses (study design, attention to concerns about blinding, complex sampling designs, and statistical analyses that account for clustered sampling designs)

Response: Agree (change made)

- We have performed several additional sensitivity analyses to address the NASEM Committee's comments on blinding, complex sampling designs, and clustering. The results are presented in eTable 6 (excerpt

below). One analysis excluded Cui et al. (2018) to respond to the Committee's concerns about blinding. To address the NASEM Committee's comments about complex sampling designs, we conducted a sensitivity analysis excluding Yu et al. (2018) and Zhang et al. (2015b). To address the Committee's concerns about clustering, we performed two sensitivity analyses—one using the unadjusted effect estimates and one using the estimates from the random effect models from Bashash et al. (2017) and Green et al. (2019). The additional sensitivity analyses had minimal impact on the overall results of the meta-analysis (see excerpt of eTable 6 below).

Excerpt of eTable 6. Regression Slopes Meta-analysis

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimate				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Sensitivity Analyses				
Using estimates from random effect models for Green et al. (2019)[113] and Bashash et al. (2017)[112]				
Full-scale IQ	9	−1.80 (−2.80, −0.80)	<0.001	76%
Males	2	−2.39 (−5.89, 1.10)	0.070	69%
Females	2	−0.53 (−3.43, 2.37)	0.186	43%
Excluding Yu et al. (2018)[3] and Zhang et al. (2015b)[110]				
Full-scale IQ	7	−1.76 (−2.90, −0.62)	<0.001	82%
Using unadjusted estimates from Bashash et al. (2017),[112] Cui et al. (2018),[76] Green et al. (2019)[113], Yu et al. (2018)[3]				
Full-scale IQ	9	−1.82 (−2.81, −0.83)	<0.001	76%

5.G:

c. Revisit the inclusion of data from overlapping studies

Response: Agree (change made)

- The NASEM Committee identified one set of overlapping populations—Xiang et al. (2003) and Xiang et al. (2011)—and suggested review of all of the analyses to ensure that overlapping publications are not included in any meta-analyses. We have removed the Xiang et al. (2011) assessment of IQ associated with serum fluoride levels from the meta-analyses. We have also confirmed that there are no overlapping publications used in the same meta-analysis. As stated previously, eTable 1 (excerpt above) provides details on studies with overlapping populations.

5.H:

d. Describe the meta-analysis methods in a single location for ease of reading

Response: Agree (change made)

- The separation of the meta-analysis from the NTP Monograph supports greater clarity in the presentation of methods for the meta-analysis versus the overall systematic review methods for the NTP Monograph. The peer-reviewed protocol contains the complete methodological details for the meta-analysis in one location. The *Methods* section of the meta-analysis manuscript also has improved clarity as it is now solely focused on the meta-analysis.

5.I:

e. Acknowledge weaknesses in the subjective way publication bias was assessed

Response: Agree (change made)

- We agree with the NASEM Committee's overall comment that, "NTP provides a reasonably thorough and appropriate evaluation of publication bias." The NASEM Committee recommended NTP consider "adjusting for possible publication bias" rather than "eliminating publication bias" when referring to results of fill-and-trim analyses. We accepted the recommendation, addressed Committee comments, and to provide additional clarity, we have added a brief discussion of the existing approaches for evaluating potential for publication bias to the *Methods* section of the meta-analysis manuscript, as follows:

  "Potential publication bias was assessed by developing funnel plots and performing Egger regression on the estimates of effect size.[25-27] If publication bias was present, trim-and-fill methods[28, 29] were used to estimate the number of missing studies and to predict the impact of the hypothetical "missing" studies on the pooled effect estimate."

- We agree that the limitations of the tests used to evaluate publication bias should be mentioned, and we have added the following to the *Discussion* section as follows:

  "There are also several limitations to the existing approaches for evaluating potential for publication bias. The funnel plot asymmetry is a subjective assessment and is recommended only when at least 10 studies are included in the meta-analysis.[64] Furthermore, the Egger regression test and Begg's rank tests[25-27] may suffer from inflated type I power and limited power in certain situations.[65]"

5.J:

f. Assess heterogeneity multiple ways

Response: Agree (change made)

- The NASEM Committee had several comments on heterogeneity and noted that NTP primarily used the Cochran's Q test to assess

heterogeneity. The Committee did not suggest assessing heterogeneity in multiple ways but noted that "heterogeneity can also be assessed by providing a count or percentage of the number of studies to the right or left of the null value. Some would consider that a much simpler, more intuitive, and perhaps more useful way of assessing heterogeneity, especially in light of the marked differences between the studies in design, study populations, exposure and outcome assessment methods, and statistical analyses. Although that approach should not be used as the sole basis of conclusions, it can be a useful first step in exploring why heterogeneity might exist."

- The meta-analysis manuscript now includes clear references to the studies with effect estimates to the right of the null in the *Results* section of the manuscript as follows:

  "The three studies with a non-negative association reported SMD estimates of 0.01 (95% CI: −0.19, 0.21),[6] 0.01 (95% CI: −0.19, 0.22),[38] and 0.13 (95% CI: −0.16, 0.42).[5]"

- In the *Methods* section, we provide details on how heterogeneity was assessed as follows:

  "Heterogeneity was assessed by Cochran's Q test[23] and the I2 statistic.[24] Forest plots were used to display results and to examine possible heterogeneity between studies."

5.K:

g. Provide the rationale for selecting individual outcomes from a single study when multiple outcomes were present

Response: Agree (change made)

- We reviewed the analyses to ensure that a consistent approach matching the data criteria outlined in the meta-analysis protocol was applied to all studies. Results were selected considering the most appropriate exposure metric, exposure range, exposure period, number of subjects, and statistical adjustment for potential confounders. See excerpt of eTable 2 referenced in our response to comment "a" above for study-specific effect estimates used in the meta-analysis.

5.L:

h. Revisit decisions made to exclude particular study results

Response: Agree (change made)

- The NASEM Committee recommended that NTP review the process to exclude study results from the meta-analysis. In response, we reviewed the analyses to ensure that a consistent approach matching the data criteria outlined in the meta-analysis protocol was applied to all studies.

For reasons why particular outcomes/results were selected, see our responses to comments "a" and "g" above.

- We also revised the meta-analysis to include standardized mean differences (SMDs) from Green et al. (2019). We agree with the Committee that Ding et al. (2011) and Zhang et al. (2015) were correctly included in both the *mean-effects* and *regression slopes meta-analyses*.

5.M: Issue: New evidence

Two studies (Ibarluzea et al., 2021 and Aggeborn & Ohman, 2021) published in 2021 were not included in the systematic review or meta-analysis. These studies have comparable methods to other included studies.

Recommendation: The Ibarluzea and Aggeborn & Oehman studies should be evaluated and included when assessing the evidence, similar to the 15 additional studies from the Chinese databases. [redacted] also recommends NTP include a comparison between Ibarluzea et al., 2021, and Green et al., 2019, because both studies investigate fluoride exposures at levels used for water fluoridation.

Response: Agree (change made)

- We have updated the literature search for the meta-analysis through November 1, 2021, using appropriate methods to identify critically assessed and relevant new publications. After integrating the results, the conclusions of the meta-analysis were essentially unchanged.
- In updating the literature search for the meta-analysis, 10 new studies (including Ibarluzea et al. 2021) were added to the evidence database. These new studies were published in the past 2 years and their addition left the findings of the analysis essentially unchanged. Our meta-analysis now includes 60 studies of children's cognition and fluoride exposure, 13 of which are high quality.
- Aggeborn and Ohman (2021) had been previously reviewed when it was a 2017 non-peer-reviewed white paper but was excluded because it was not peer-reviewed. The study was excluded from the meta-analysis because it assessed cognitive functions other than IQ and the cognitive tests were not specified (see supplemental materials, eTable 1).

This document contains the complete first set of comments provided by the [redacted] in January 2022 in its original format and the NIEHS/DNTP responses to those comments. Note that the yellow highlighting as well as the purple and red formatted text were in the document as provided. The NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

6a.A: From Abstract

RESULTS The meta-analysis of 46 studies (N = 15,538 children) with group-level exposures found that children exposed to higher fluoride levels had lower mean IQ scores (pooled SMD: −0.49; 95% CI: −0.60, −0.38; p-value < 0.001). Results were robust to stratification by study quality (high *vs*. low risk of bias), gender, age group, outcome assessment, study location, exposure timing, and exposure metric. There was a dose-response relationship between mean children's IQ and group-level fluoride exposure measures. [1] The meta-analysis of the association between individual-level measures of fluoride and children's IQ found a decrease of 1.58 IQ points (95% CI: −2.63, −0.53; p-value = 0.003) per 1-mg/L increase in urinary fluoride. **CONCLUSIONS AND RELEVANCE** Our meta-analysis confirms results of previous meta-analyses and extends them by including newer, more precise studies with individual-level exposure measures. The data support a consistent inverse association between fluoride exposure and children's IQ.

[1] This dose-response statement is not consistent with the level of fidelity of the data presented/available and infers there are negative health effects attributable to fluoride. This is a critical concern that applies to the highlighted statements below.

Response: Disagree (no change)

- The highlighted text accurately describes the available data, analysis, and results. The language in the abstract and throughout the manuscript objectively and fairly describes the data, including strengths and limitations.
- In its 2021 report on the 2020 draft NTP Monograph, the NASEM Committee agreed with our statements on consistency: "As noted in the revised monograph, 44 of the 46 studies represented in that figure had effect estimates to the left of zero—results that indicate an association between higher fluoride exposures and lower IQ. Those results highlight the marked consistency in the current epidemiologic literature on fluoride and childhood IQ."
- Please note that the last sentence that was highlighted was subsequently changed as follows:

  "The consistency of the data supports an inverse association between fluoride exposure and children's IQ."

6a.B: FROM Manuscript

No study was excluded from the meta-analysis based on concerns for risk of bias; however, subgroup analyses were conducted with and without high risk-of-bias studies (i.e., studies rated "probably high" risk of bias for at least two key risk-of-bias questions or "definitely high" for any single key question) to assess their impact on the results.

Response: No change requested

- This text was highlighted but was not accompanied by a comment or request for revision. We assume that the text was highlighted to imply that this approach is a flaw. Excluding studies from systematic reviews or systematic reviews with meta-analyses is not considered a best practice in the systematic review community (Higgins et al. 2021). As a well-documented systematic review and meta-analysis, this evaluation follows a protocol where inclusion and exclusion criteria were defined a priori. As the non-highlighted text above clearly states, a subgroup analysis was conducted with and without the high risk-of-bias studies.

  Reference: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). 2021. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane. Available from www.training.cochrane.org/handbook

6a.C: Conclusions

Our meta-analysis confirms and extends prior meta-analyses that reported associations between higher fluoride exposures and lower IQ levels of children. The results were robust to stratifications by risk of bias, gender, age group, outcome assessment, study location, exposure timing, and exposure type (including both drinking water and urinary fluoride). Therefore, the data support a consistent inverse association between fluoride exposure and children's IQ.

Response: No change requested

- This text was highlighted but was not accompanied by a comment or request for revision. We are unaware of the reviewer's thoughts on this highlighted text but will note that the sentences in the conclusion are factual statements describing the data.

6a.D: From Supplemental Documentation

If median or mean levels by exposure group were not provided, the midpoint of the upper and lower boundaries in every exposure category was assigned as the average level. If the upper boundary for the highest exposure group was not reported, the boundary was assumed to have the same amplitude as the nearest exposure category

Response: No change requested

- Again, these sentences were highlighted but were not accompanied by a comment or request for revision. We are unaware of the reviewer's thoughts on this highlighted text but will note that this method is common practice in dose-response analyses in determining exposure levels for each data point (Boffetta et al. 2020) and is described in our peer-reviewed protocol.

Reference: Boffetta, P., Zunarelli, C., & Borron, C. (2020). Dose-Response Analysis of Exposure to Arsenic in Drinking Water and Risk of Skin Lesions: A Systematic Review of the Literature. Dose-response: a publication of International Hormesis Society, 18(4), 1559325820957823. https://doi.org/10.1177/1559325820957823

From NTP 2020 revision NTP Protocol: Systematic Review of Effects of Fluoride Exposure on Neurodevelopment (nih.gov)

Note: A comment related to the protocol for the NTP Monograph (see https://ntp.niehs.nih.gov/go/785076) is not reproduced here as it is not directly relevant to the meta-analysis.

6a.E: Additional concerns

- Measure assessment of "intelligence" was different in different studies (the scores/scales for different countries, different tools and the interpretation of the "mean" of disparate classification systems. Examples: Wechsler Abbreviated Scale of Intelligence vs. "Wechsler Intelligence Scale for Children-Revised" vs. "Combined Raven's Test for Rural China" or "Wechsler Primary and Preschool Scale of Intelligence-III"

  Response: Disagree (no change)
  - We view the use of these different tests in studies of different study populations as the proper approach and consider whether the test is appropriate for a given population as part of risk-of-bias assessment. As per our protocol, for a "definitely" or "probably low risk-of-bias" rating for outcome assessment, it is required that studies use an intelligence test that is appropriate to the population being studied. The consistency of the direction of the association across a diverse range of tests supports the conclusions of our meta-analysis.
  - The difference in tests is also a reason we used the standardized mean difference (SMD) as the unit of measure in our meta-analysis. The SMD is commonly used in meta-analysis when the studies all assess the same outcome (e.g., intelligence) but measure it in a variety of ways (e.g., WISC-R, Combined Raven's Test for Rural China, etc.). It is necessary to standardize the results of the studies to a uniform scale before they can be combined (Higgins et al. 2021).
  - In addition, this comment fails to acknowledge that we also conducted a subgroup analysis stratified by assessment type. The results of this subgroup analysis show that the direction of the association is robust to stratification by assessment type and that assessment type does not explain the observed heterogeneity. The results of this subgroup analysis compared to the main overall effect estimate are shown below.

    Excerpt of Table 2. Pooled SMDs and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	SMD (95% CI)	Heterogeneity	
			p-value	I^2
Overall Effect				
Overall Effect	55	−0.46 (−0.55, −0.37)	<0.001	87%
Subgroup Analyses				
Assessment Type				
CRT-RC tests	29	−0.36 (−0.46, −0.27)	<0.001	82%
Non-CRT-RC tests	26	−0.60 (−0.78, −0.42)	<0.001	89%
Raven's tests	10	−0.76 (−1.10, −0.43)	<0.001	91%
Other tests	16	−0.52 (−0.74, −0.29)	<0.001	89%

Table 2 Notes: CI = confidence interval; CRT-RC = Combined Raven's Test–The Rural edition in China; NA = not applicable; SMD = standardized weighted mean difference

Reference: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). 2021. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane. Available from www.training.cochrane.org/handbook

6a.F:

- Definition of "high"/"low" fluoride levels were different across studies and not defined by the author. One newer study Bashash et al. (2017) defined "Low" as <0.80 mg/L and "High" ≥0.80 mg/L but without upper limit and the difference between Low/High in this example could be as small as 1/100th

Response: Disagree (no change)

- It would be inappropriate for us to define high and low fluoride levels for the purpose of this meta-analysis**.** Our approach is consistent with all previous fluoride meta-analyses (Choi et al. 2015, Duan et al. 2018, Miranda et al. 2021). Table 1 transparently reports the high and low fluoride levels as presented in each individual study.
- In its peer review of the first draft of the meta-analysis, which was included in the 2020 draft NTP Monograph, the NASEM Committee agreed with this method for the *mean-effects meta-analysis*: "The overall approach appears to be sound in comparing mean IQ scores for the most and least highly exposed to fluoride even though the absolute fluoride concentrations are not comparable among studies.", and "The committee found the meta-analysis to be a valuable addition to the monograph and acknowledges the tremendous amount of work that was required. The meta-analysis applied standard, broadly accepted methods, and the data shown in Figure A5-1 and the related evaluations are especially informative (NTP 2020a, p. 235)."
- In addition, this comment fails to acknowledge that we also conducted a *regression slopes meta-analysis* that used studies reporting continuous data estimating associations between individual-level fluoride exposure and children's

IQ. In this analysis, differences across studies with respect to what study authors might consider high or low fluoride levels are irrelevant.

6a.G:

- Quantified dose of exposure not presented. Urinary spot testing not good surrogate and no correlation to quantifiable exposure given relatively rapid clearance of Fl from the body.

  Response: Disagree (no change)

  - We understand the concerns regarding urinary fluoride levels. However, fluoride levels measured during pregnancy and in children include all ingested fluoride and are considered a valid measure to estimate total fluoride exposure (Villa *et al.* 2010, Watanabe *et al.* 1995).
  - We acknowledge that the type and timing of urinary sample collection is important to consider, and we have considered these factors in our analysis as described in the prepublication 2022 NTP Monograph. When compared to 24-hour urine samples, spot urine samples are more prone to the influence of timing of exposure and can also be affected by differences in dilution; however, many studies attempted to account for dilution using either urinary creatinine or specific gravity. Good correlations between 24-hour samples and urinary fluoride concentrations from spot samples adjusted for urinary dilution have been described in the literature (e.g., Zohouri et al. 2006). Both 24-hour samples and spot urine samples adjusted for dilution are considered acceptable, with 24-hour samples considered the more accurate measure of fluoride. If authors made appropriate efforts to reduce the concern for bias (e.g., accounting for dilution), studies that used this metric were generally considered to have probably low risk of bias for exposure.
  - However, we have added the following sentence to the *Strengths and Limitations* section of the meta-analysis to acknowledge this concern:

    "When compared with 24-hour urine samples, spot urine samples are more prone to the influence of timing of exposure (e.g., when water was last consumed, when teeth were last brushed) and can also be affected by differences in dilution."

6a.H:

- The lack of a direct measure of dose and thus, exposure is a significant design limitation and the strength and specificity of the conclusions are out of proportion given the limitations; the statements/conclusions of the manuscript overstate what can be fairly concluded from the studies.

  Response: Disagree (no change)

  - The conclusions of our meta-analysis are consistent with two prior meta-analyses of studies using group-level exposures and extend these analyses with a

confirmatory *regression slopes meta-analysis* that uses individual-level exposure and outcome assessments. In the *Discussion* section, we clearly address the limitations of a *mean-effects meta-analysis,* including the way in which exposure is measured.

6a.I:

- The results could be used to recommend improvements to future studies but the lack of an individual fluoride exposure variable and dose measurement precludes the conclusions asserted in this paper. This weakness could be responsible for complete misclassification of many of the data points.

  Response: Disagree (no change)

  - This meta-analysis does not lack individual fluoride exposure variables. Our *regression slopes meta-analysis* includes 11 studies with individual-level exposure measures (with 10 high quality publications) from 6 different study populations. Each of these studies reported individual urinary fluoride levels, with two also reporting fluoride intake and two also reporting water fluoride levels.
  - As we mentioned in our response to a previous comment, urinary fluoride in children is a valid measure to estimate total fluoride exposure. In addition, the consistency of the results from the *regression slopes meta-analysis* stratified by exposure type (Table 3 excerpt provided below) suggest that the results cannot be explained by a "complete misclassification of many of the data points."

    Excerpt of Table 1. Pooled Regression Slopes and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Effect				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Subgroup Analyses				
Exposure Type				
Urinary fluoride	9	−1.81 (−2.80, −0.81)	<0.001	77%
Intake	2	−3.87 (−7.15, −0.59)	0.737	0%
Water fluoride	2	−4.77 (−9.09, −0.45)	0.707	0%

6a.J:

- The strength and specificity of the conclusions are out of proportion and overstated given the significant limitations of the available data from these studies

  Response: Disagree (no change)

  - The statements made in the meta-analysis are measured and representative of the data.

6a.K: Additional Background:

From Dose Response Assessment - an overview | ScienceDirect Topics,

Dose–Response Assessment
Dose–response assessment characterizes the quantitative relationship between exposure (usually determined in toxicity studies) and the occurrence of adverse health effects. Typically applied or administered dose, rather than effective tissue dose, is used to develop the dose–response relationship.

- These are important points that support the premise that there really is no measure or attempt to measure "dose" of exposure. An environmental, naturally occurring metal (Fl^-) merely being in the environment does not constitute an exposure of any particular magnitude. This is missing.

  Response: Disagree (no change)
  - We are somewhat unclear on the points being raised. Concerning a "typical" dose- -response relationship, the comment above is correct that most dose-response relationships are based on estimates of applied or administered dose; however, this is also commonly considered a practical limitation of the method. We explain in the manuscript that drinking water measures are indirect measures of exposure and that internal measures such as those reflected by urinary fluoride data are preferred. We also disagree that fluoride is classified as a metal.

6a.L: Also, [redacted] read of eTable 4. Dose-Response Meta-analysis Using Mean Effects – Model Selection for Water Fluoride does *not* represent dose response, as [redacted] see it. For example: Linear Model: the P value for <2 mg/L and <1.5 mg/L are not significant and it makes sense then, that if the "All data" p= <0.001 is disproportionately influenced by the <4 mg/L exposure. It is also not clear from this table whether these numbers represent <4 but >= 2mg/L and <2 but >=1.5, etc. Are these mutually exclusive categories? Needs clarification

Response: Disagree (edited for clarity)
- We disagree that statistical significance is necessary to indicate a dose-response relationship. Data should be evaluated in their full context for epidemiological studies, and statistical significance is only one consideration (EPA 2020). We report p-values and consider them as an important, but not exclusive contribution to the overall data interpretation.
- However, we have taken the suggestion to clarify that the exposure categories are not mutually exclusive and have added the range of exposure for each group when they are first mentioned in the supplemental materials as follows:

  <4 mg/L (i.e., 0 to <4 mg/L) fluoride in drinking water

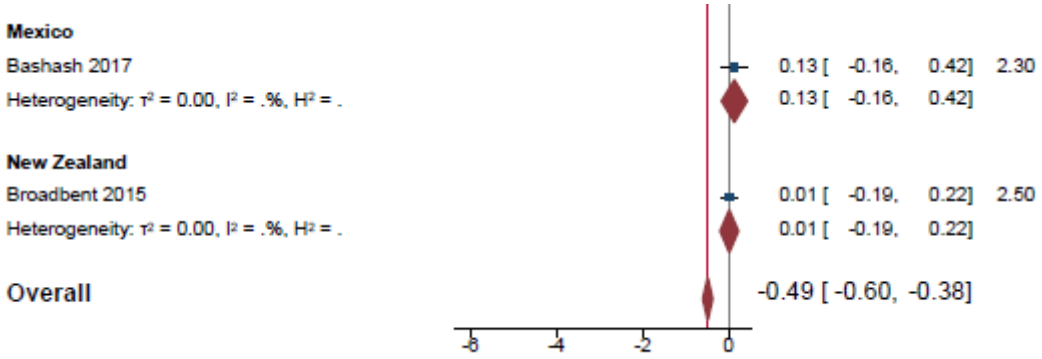
<2 mg/L (i.e., 0 to <2 mg/L) fluoride in drinking water

<1.5 mg/L (i.e., 0 to <1.5 mg/L) fluoride in drinking water

Reference: U.S. EPA. ORD Staff Handbook for Developing IRIS Assessments (Public Comment Draft, Nov 2020). U.S. EPA Office of Research and Development, Washington, DC, EPA/600/R-20/137, 2020.

6a.M: Children's Urinary Fluoride – All Studies

Also, worth noting: the newer studies included in the analysis… Mexico and New Zealand Country subgroup analysis are both at **Zero** or **above zero**, eFigure 13. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Age Group



Response: No change requested

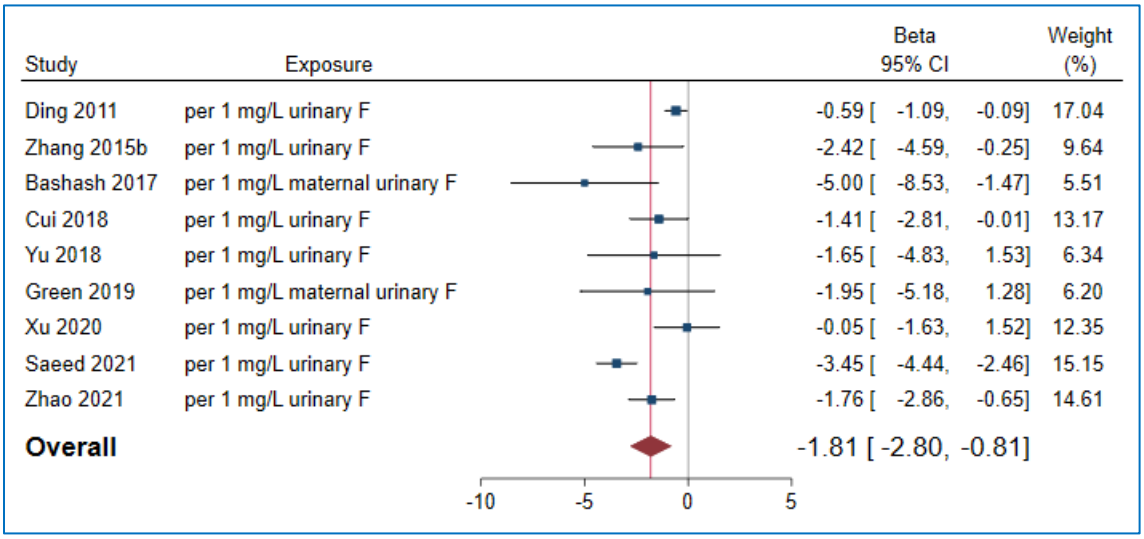
- This is an excellent example of DNTP considering all data irrespective of direction of effect. In fact, we point out these non-negative effect estimates in the *Results* section of the supplemental materials:

  "The three studies with non-negative associations reported SMD estimates of 0.01 (95% CI: −0.19, 0.21),[113] 0.01 (95% CI: −0.19, 0.22),[25] and 0.13 (95% CI: −0.16, 0.42).[112] Two of the three studies with non-negative SMDs compare mean IQs in children living in fluoridated vs. non-fluoridated areas in Canada,[113] or in New Zealand.[25] No other studies included in the main mean-effects meta-analysis made comparisons between fluoridated vs. non-fluoridated areas. In both studies, levels of fluoride in water were low, even in communities with fluoridated drinking water, likely limiting the power to detect an effect.

  In Bashash et al.,[112] the SMD compares mean IQ scores in children with urinary fluoride levels below vs. above 0.80 mg/L in Mexico.[112] Unlike other studies in the mean-effects meta-analysis which compared mean IQ scores between fluoridated vs. non-fluoridated areas, or areas with high vs. low fluoride exposures (see eTable 2), the Bashash et al.[112] study was not designed to measure fluoride exposure by geographical area. However, since the mean IQ scores were provided in the manuscript for children with urinary fluoride levels below vs. above 0.80 mg/L, we included them in this analysis. It's worth

noting that there was no significant difference when comparing MUF levels between the groups of children with urinary fluoride levels above or below 0.80 mg/L, however when children's IQs were regressed against MUF, a statistically significant inverse association was found."

- Note: In our November 2021 update of the literature, we also included the INMA cohort study (Ibarluzea et al., 2021) that found positive associations between fluoride exposure and cognitive effects in boys.



eFigure 19. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Overall Analysis

eFigure 19 note: Estimates (betas) for individual studies are shown with solid boxes representing the weight, and the pooled estimate is shown as a solid diamond. Horizontal lines represent 95% CIs for the study-specific betas.

This document contains the complete second set of comments provided by [redacted] in February 2022 in its original format and the NIEHS/DNTP responses to those comments. The NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

6b.A: [redacted] **Critique of the NTP meta-analysis manuscript**

Summary:

The group of studies included in this meta-analysis had three significant issues identified by the manuscript authors that weaken its results: publication bias, high heterogeneity, and lack of uniformity in measuring and reporting the primary studies' outcome or IQ measure. Overshadowing these problems is the inappropriate use of a meta-analysis for observational studies when randomized clinical trials are not available. The intervention and control arms among a group of similar randomized trials are comparable because randomization tends to balance the arms with respect to both known and unknown confounders, but this is not true of observational studies. In the present meta-analysis, those categorized as consuming higher levels of fluoride are compared to those consuming lower levels. The fluoride exposure is not randomized and may be dictated by national or regional governments. Two potential consequences are spurious associations between fluoride and IQ and differential results by country. [redacted] begin with these two consequences.

Response: Disagree (no change)

- First, it is important to make clear that this meta-analysis was conducted at the strong recommendation of the NASEM Committee's peer review of the 2019 draft NTP Monograph. The 2019 draft NTP Monograph evaluated a large number of human observational studies but did not include a meta-analysis. The NASEM Committee's peer review report stated that the "committee strongly recommends that NTP reconsider its decision not to perform a meta-analysis."
- Second, the NASEM Committee agreed with the methods used in the meta-analysis. In its peer review of the first draft of the meta-analysis, which was included in the 2020 draft NTP Monograph, the NASEM Committee stated: "The critical information regarding comparison of study results comes from the new meta-analysis, which seeks to extract and integrate comparable findings from selected studies as discussed further below. The overall approach appears to be sound in comparing mean IQ scores for the most and least highly exposed to fluoride even though the absolute fluoride concentrations are not comparable among studies.", and "The committee found the meta-analysis to be a valuable addition to the monograph and acknowledges the tremendous amount of work that was required. The meta-analysis applied standard, broadly accepted methods, and the data shown in Figure A5-1 and the related evaluations are especially informative (NTP 2020a, p. 235)."
- Finally, consideration of the use and value of data from observational studies relative to randomized clinical trials (RCTs) in meta-analyses has been empirically studied, and Cochrane analyses have repeatedly shown that there is little evidence for significant effect estimate differences between observational studies and RCTs regardless of specific observational study design, heterogeneity, or inclusion of studies of pharmacological interventions (Anglemyer et al. 2014; Benson and Hartz 2000; Schwingshackl et al. 2021).

6b.B: In the present meta-analysis, the majority of studies are from one country: China. There are plausible mechanisms that might create the appearance of an association between fluoridation and IQ scores. For example, people in rural communities may exhibit lower IQ test scores than those in urban areas; they may also be more likely to drink tap water or well water as opposed to bottled water or other beverages. Thus, they might consume more fluoride. This would induce a non-causal correlation between fluoridation and IQ scores. [redacted] are presenting this scenario not as a fact, but to suggest that there are plausible explanations for a spurious correlation between fluoride and IQs in observational studies.

Response: Disagree (no change)

- The point of a risk-of-bias assessment is to evaluate whether the design or conduct of a study compromised the credibility of the link between exposure and outcome (Higgins and Green 2011, IOM 2011, Viswanathan et al. 2012). The concern in this comment appears to be related to potential bias due to confounding in individual studies. This issue is addressed in the meta-analysis through a rigorous assessment of risk of bias, which included an extensive evaluation of potential bias due to confounding in each individual study, addressing situations exactly like the example presented in the comment. (See eFigure 2a and 2b for risk-of-bias summaries, links to assessments of individual studies, and Appendix E of prepublication 2022 NTP Monograph for more detail.)
- We would also like to note that Chinese studies provide the opportunity to compare the cognitive abilities of children in villages of similar size, SES, and other relevant characteristics where drinking water sources differ widely in their level of naturally occurring fluoride. These variations in fluoride levels can be larger than those found in the other areas, including most of the United States, and therefore provide greater power to detect an effect.

6b.C: Given that most of the studies in this meta-analysis are in China, whose environmental policies could explain a spurious association, [redacted] might expect to see different results in countries with policies more aligned with those of western nations. In fact, that is exactly what [redacted] see in this meta-analysis. Broadbent (2015) and Green (2019) are studies in New Zealand and Canada, arguably the two countries most comparable to the United States. Figure 2 of the manuscript shows narrow confidence intervals centered on no effect in these two studies. This is consistent with the idea that the apparent association between fluoride and IQs may not be causal.

Response: Disagree (no change)

- It's not clear what the comments refer to when citing environmental policies that would explain a spurious effect. However, a spurious association is unlikely given the included studies span broad geographical regions and time periods (1989–2021) and cover a range of methods for outcome and exposure assessment (including different exposure metrics). In addition, potential confounders and co-exposures to other possible neurotoxicants were extensively considered in the risk-of-bias assessment and evaluation of each study.
- [redacted] is correct that, in the *mean-effects meta-analysis*, the SMDs for children's urinary fluoride (CUF) and children's IQ in Broadbent et al. (2015) and Green et al. (2019) were non-

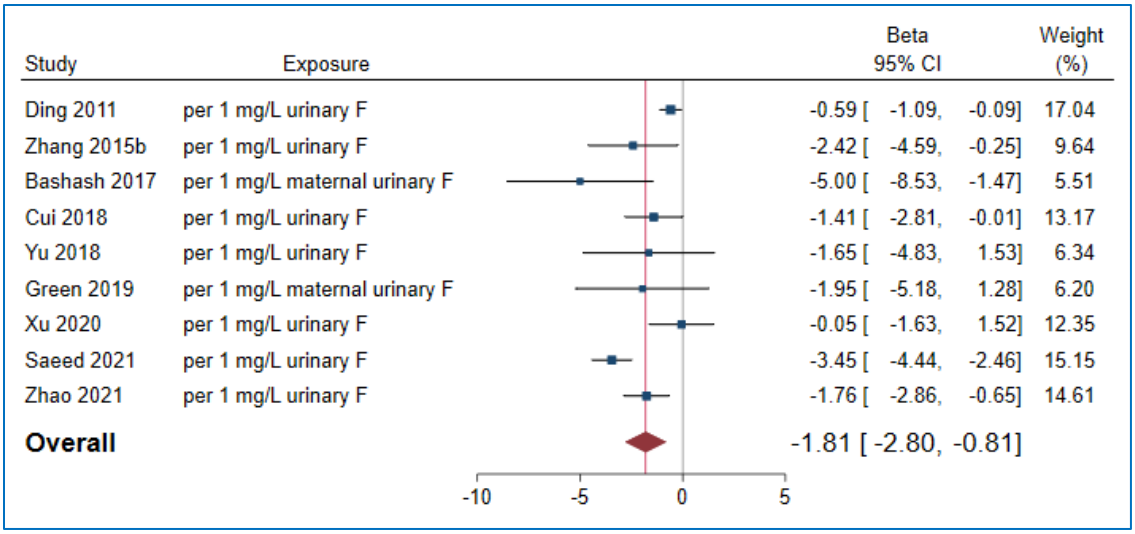
negative; the SMD for Bashash et al. was also non-negative. We clearly describe these non-negative effect estimates in the *Results* section of the manuscript:

"The three studies with a non-negative association reported SMD estimates of 0.01 (95% CI: −0.19, 0.21),[6] 0.01 (95% CI: −0.19, 0.22),[38] and 0.13 (95% CI: −0.16, 0.42).[5]"

- In both Broadbent et al. (2015) and Green et al. (2019), levels of fluoride in water were low, even in communities with fluoridated drinking water. So, when using group-level exposure data (as opposed to individual-level exposure data), as was done in the *mean-effects meta-analysis*, the power to detect an effect may be limited. We note that [redacted] comment ignores the results of the *regression slopes meta-analysis*, which used individual-level maternal urinary fluoride (MUF) for the Canadian (Green et al. 2019) and Mexican (Bashash et al. 2017) studies (MUF levels were comparable in these two studies [Till et al. 2018]) and found an inverse association between MUF and children's IQ as shown in eFigure 19 (provide below; see Bashash et al. 2017 and Green et al. 2019). Green et al. (2019) also reported a statistically significant inverse association between maternal water fluoride levels and children's IQ as shown in eFigure 23 (provided below).
- In response to a comment from the NASEM Committee, we added text to the supplemental materials to identify likely reasons why results from the three studies differed from results of the other studies, as follows:

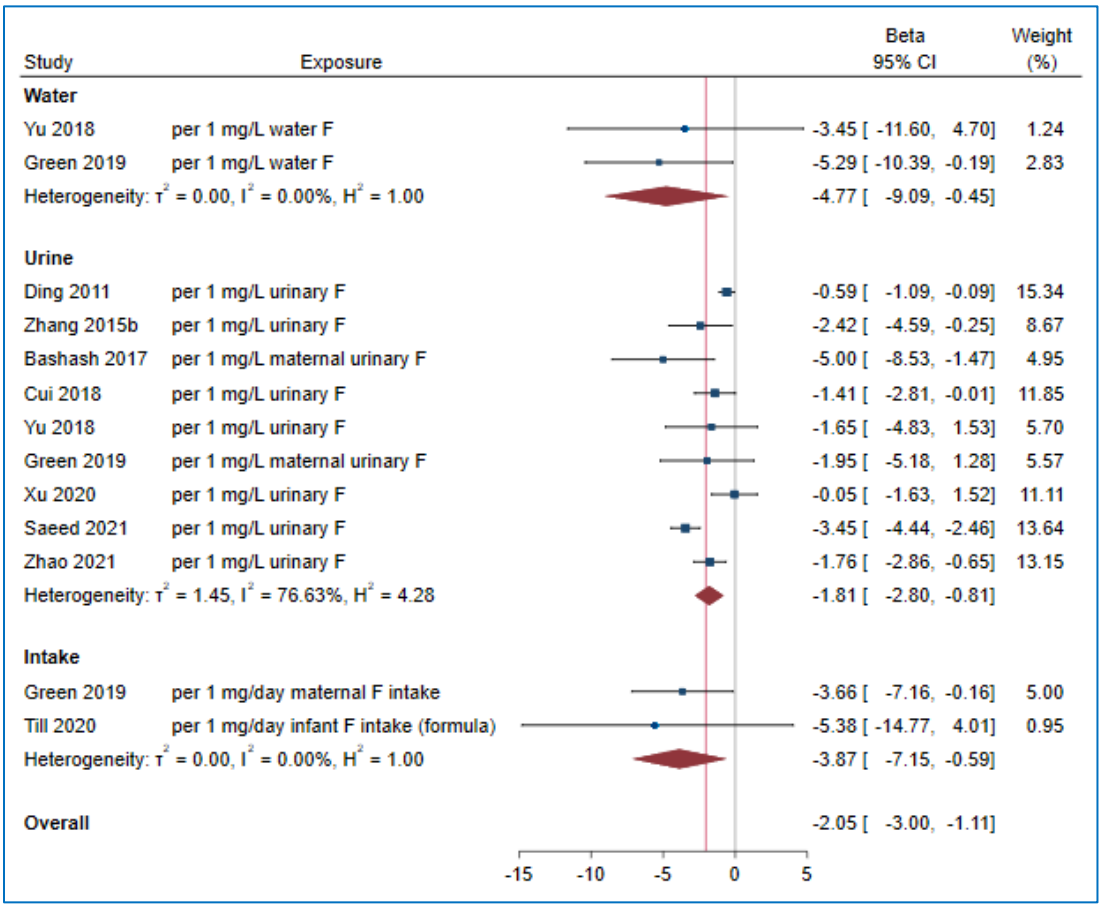
  "The three studies with non-negative associations reported SMD estimates of 0.01 (95% CI: −0.19, 0.21),[113] 0.01 (95% CI: −0.19, 0.22),[25] and 0.13 (95% CI: −0.16, 0.42).[112] Two of the three studies with non-negative SMDs compare mean IQs in children living in fluoridated vs. non-fluoridated areas in Canada,[113] or in New Zealand.[25] No other studies included in the main mean-effects meta-analysis made comparisons between fluoridated vs. non-fluoridated areas. In both studies, levels of fluoride in water were low, even in communities with fluoridated drinking water, likely limiting the power to detect an effect.

  In Bashash et al.,[112] the SMD compares mean IQ scores in children with urinary fluoride levels below vs. above 0.80 mg/L in Mexico.[112] Unlike other studies in the mean-effects meta-analysis which compared mean IQ scores between fluoridated vs. non-fluoridated areas, or areas with high vs. low fluoride exposures (see eTable 2), the Bashash et al.[112] study was not designed to measure fluoride exposure by geographical area. However, since the mean IQ scores were provided in the manuscript for children with urinary fluoride levels below vs. above 0.80 mg/L, we included them in this analysis. It's worth noting that there was no significant difference when comparing MUF levels between the groups of children with urinary fluoride levels above or below 0.80 mg/L, however when children's IQs were regressed against MUF, a statistically significant inverse association was found."



eFigure 1. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Overall Analysis.

eFigure 19 note: Estimates (betas) for individual studies are shown with solid boxes representing the weight, and the pooled estimate is shown as a solid diamond. Horizontal lines represent 95% Cis for the study-specific betas.



eFigure 23. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Exposure Type

6b.D: A meta-analysis of randomized controlled trials with similar results bolsters the evidence of an intervention effect, but a meta-analysis of observational studies, all subject to the same biases, increases the probability of a misleading result. The p-value will become smaller by virtue of increased sample size, but not because of any true cause and effect relationship. The bottom line is that potentially confounding effects on IQ are not randomly assigned, and that makes tenuous any conclusion of a causal effect of fluoride on IQs.

Response: Disagree (no change)

- Unfortunately, as is the case with most studies of potentially harmful exposures, there are no randomized controlled trials assessing the association between exposure to fluoride and children's intelligence (likely due to ethical concerns about randomizing pregnant women and/or children to fluoride). Therefore, observational studies are the best source of available information.
- This comment also repeats the RCT argument relative to observational studies. See prior responses regarding support of the value of observational studies in the public health, systematic review, and environmental epidemiological communities. Other public health conclusions and practices have long been supported by observational studies. For example, the evidence showing that community water fluoridation protects against tooth decay was largely based on observational or "association" studies, most of which were conducted prior to the introduction of fluoridated toothpaste in the early 1970s (Iheozor-Ejiofor et al. 2015).
- The assumption that all observational studies in a meta-analysis suffer from the same biases is unfounded. As mentioned in an earlier response to comment, risk of bias was systematically assessed for each individual study. Multiple potential sources of bias (including confounding bias, selection bias, exposure characterization, and outcome assessment) were extensively evaluated for each individual study, and results of those assessments are presented in Appendix E of the prepublication 2022 NTP Monograph.

6b.E: Next, [redacted] consider the weaknesses identified by the authors themselves. To their credit, the authors attempted to assess the impact and ameliorate the consequences of these weaknesses through analytical approaches such as funnel plots and Egger's test to detect publication bias, trim and fill methods to correct for publication bias, the I^2 and Q statistic for detection of heterogeneity, and subgroup and sensitivity analyses to try to explain the heterogeneity. Unfortunately, these approaches cannot correct for the use of inappropriate study design in a meta-analysis.

Response: Disagree (no change)

- We disagree that meta-analyses of observational studies are not appropriate. See prior responses regarding support of this in the public health, systematic review, and environmental epidemiological communities.
- Furthermore, we conducted the meta-analysis in response to the NASEM Committee's peer review of the 2019 draft NTP Monograph, which stated that the "committee strongly recommends that NTP reconsider its decision not to perform a meta-analysis."

  Additionally, as mentioned in a previous response to comment, the NASEM Committee supported our approach and described the information presented in the meta-analysis as valuable and informative.

- Note: A quick search[1] for meta-analyses of observational studies in PubMed alone identifies over 20,000 studies, which indicates how prevalent meta-analyses of observational studies are in the scientific literature.

6b.F: As recommended in the Cochran Handbook of Systematic Reviews, concerns such as high heterogeneity should preclude the use of the meta-analysis in the first place. A meta-analysis can be performed with data from as few as two studies and therefore it is best to consider only clinically and methodologically similar and sound studies. When results from studies with substantial differences in design, exposure, outcome measures, and risk of bias are combined, the effects of any exposure are more likely to be overestimated. This result was evident in the manuscript when comparing adjusted and non-adjusted findings, and findings from high vs. low risk-of-bias (RoB) studies. For example, the primary effect estimate of differences in children's IQ (Standardized Mean Difference, SMD) shifted from a medium effect size for all studies combined (-0.49) to a small effect size among the low risk of bias studies (-0.24). At the very least, adjusted results and findings only from those studies with a low risk of bias should be emphasized, and the reader should be given a clear interpretation of what the SMD values reflect.

Response: Disagree (no change)

- The Cochrane Handbook (Higgins et al. 2021) does not say that high heterogeneity should preclude the use of meta-analysis, as is suggested in the comment. In fact, Cochrane Section 10.10.3 (Deeks et al. 2021) says that sources of heterogeneity should be explored using prespecified subgroup analyses. Therefore, the meta-analysis followed the Cochrane Handbook recommendations as reflected in the protocol (which underwent peer review) that identified prespecified potential sources of heterogeneity for later analyses. These prespecified potential sources of heterogeneity were then appropriately explored in the subgroup analyses.
- Note that, in addition to recommending NTP conduct a meta-analysis (see response to previous comment), the NASEM Committee, in their 2020 peer review report on the 2019 draft NTP Monograph, stated that a properly conducted meta-analysis can account for heterogeneity in exposure measurements and other aspects of study design.
- Furthermore, in its peer review of the 2020 draft NTP Monograph, the NASEM Committee supported the subgroup analyses NTP used in this evaluation, finding them informative and directly responsive to some of the Committee's previous concerns. They also recommended additional subgroup and sensitivity analyses that were subsequently added to the manuscript:

  "As part of its meta-analysis, NTP presents several subgroup and sensitivity analyses. The committee finds them very informative; several are directly responsive to some of the committee's previous concerns. However, **NTP should also include subgroup or sensitivity**

[1]The following search string was used for the "quick" search because it identifies a high percentage of appropriate studies: (("meta-analysis"[Publication Type] AND ("meta-analysis of observational studies"[tiab] OR "meta-analyses of observational studies"[tiab] OR "observational studies as topic"[MeSH Terms] OR "Observational Studies"[title] OR "Observational Study"[title] OR "Cohort Studies"[Mesh] OR "Cohort Study"[Title] OR "Cohort Studies"[Title] OR "Case-Control Studies"[Mesh] OR "Case-Control Study"[Title] OR "Case-Control Studies"[Title] OR "Cross-Sectional Studies"[Mesh] OR "Cross-Sectional Study"[Title] OR "Cross-Sectional Studies"[Title] OR "Ecological Study"[Title] OR "Ecological Studies"[Title] OR "Interrupted Time Series Analysis"[Mesh] OR "Time Series Analysis"[Title] OR "Time Series Analyses"[Title] OR "Time Series Study"[Title] OR "Time Series Studies"[Title])) NOT "Randomized Controlled Trial"[publication type]).

analyses that respond to the committee's concerns about blinding, complex sampling designs, and statistical analyses that account for clustered study designs.… The additional subgroup or sensitivity analyses noted could help to alleviate some of the committee's current concerns."

- As stated in the protocol, when available, we used the adjusted effect estimates in the meta-analyses. Also, in contrast to what the comment implies, the adjusted versus non-adjusted sensitivity analysis found no difference in results [Adjusted β (95% CI)= −1.81 (−2.80, −0.81); unadjusted β (95% CI) = −1.81 (−2.81, −0.83)]. As the comment points out, the direction of the association was consistent across study quality from the high risk-of-bias to the low risk-of-bias studies, and the effect estimate was smaller among the low risk-of-bias studies. This may be due to lower levels of exposure and/or smaller differences in exposure between "high" and "low" exposure groups among the low risk-of-bias studies. The comment fails to note that there are other stratified estimates that would be considered to underestimate the effect estimate in both the *mean-effects* and *regression slopes meta-analyses*.

6b.G: Additional interpretation and explanation of the subgroup analyses are also needed. Using the results of sub-group analyses to investigate and explain heterogeneity does not accomplish that goal. In most subgroups, there seem to be subgroup effects implying interactions between SMD and investigated factors such as gender, country, and risk of bias (although between sub-groups p-values are not supplied to determine whether these interactions were significant or not). There was also significant unexplained heterogeneity among studies that needs to be investigated further.

Response: Agree (change made)

- We agree that the manuscript benefited from additional discussion of the results of the subgroup analyses. To be responsive to [redacted] comment, we added the following new text in the *Results* section for the *mean-effects meta-analysis*:

  "The subgroup and meta-regression analyses did not explain a large amount of the overall heterogeneity; however, the degree of heterogeneity was lower for studies restricted to Iran (I^2=56%), children ages 10 and older (I^2=68%), and girls (I^2=76%)".

- In the *Results* section for the *regression slopes meta-analysis*, we added the following new text:

  "The observed heterogeneity in the overall effect estimate was explained by the subgroup analyses, with no significant heterogeneity remaining in analyses of low-risk-of bias studies, by sex, by country, by assessment type, and by exposure timing (Table 3)."

- In the *Discussion* section, we added the following new text with further interpretations of the subgroup analyses:

  "With a couple exceptions, the subgroup analyses in the mean-effects meta-analysis did not explain a large amount of the overall heterogeneity. However, the heterogeneity in the regression slopes meta-analysis was explained by subgroup analyses. This suggests that the aggregate nature of the mean-effects meta-analysis might not be sufficiently sensitive to capture potential sources of heterogeneity, as seen possible when using studies with individual-level data in the regression slopes meta-analysis. However, the large number of studies included in the mean-effects meta-analysis and the consistency in the direction of the association across the analyses make this is less of a concern."

- As recommended in the comment, we also further investigated potential sources of heterogeneity by conducting a meta-regression analysis using mean age in years and year of publication in each study. In the supplemental materials we added:

  "The results of the meta-regression models indicate that year of publication and mean age of study children did not explain a large degree of heterogeneity as neither were significant predictors of the relationship between fluoride and children's intelligence, and the residual I^2 remained high (85% and 87%, respectively). Year of publication (SMD = 0.01, 95% CI: -0.01, 0.02) and mean age (SMD = -0.04, 95% CI: -0.13, 0.04) explained relatively little between-study variance (adjusted R^2 of 12% and 5%, respectively). When both year of publication and mean age were included in the model, there were no notable improvements to the amount of between-study variance explained (adjusted R^2 = 13%) or percent residual variation due to heterogeneity (residual I^2 = 85%).

  Excluding the outlier study34 resulted in a slightly lower heterogeneity for the overall effect estimate (I^2=84%) and for the India-specific effect estimate (I^2=69%). The meta-regression indicates that mean age is a significant predictor of the effect (SMD = -0.06, 95% CI: −0.12, −0.01, p-value =0.025), explaining 9% of the between-study variance. Year of publication (SMD = 0.01, 95% CI: 0.001, 0.02, p-value=0.028) explained a larger degree of between-study variance (R^2 = 19 %)."

6b.H: These inconsistencies create uncertainty regarding the validity and significance of the exposure effect estimate for each subgroup. There were fewer than ten studies in many subgroups, thereby reducing their ability to identify statistically significant differences.

Response: Disagree (no change)

- We disagree with [redacted] that there are inconsistencies that would create uncertainty regarding the validity and significance of the exposure effect estimate for each subgroup.
- Also, as previously mentioned, the purpose of the subgroup analyses was to explore sources of potential heterogeneity, not to detect differences between the groups or "interactions between SMD and investigated factors." However, except for certain countries and for studies with other sources of fluoride exposure, all the subgroup analyses of the *mean-effects meta-analysis* included at least 10 studies. In addition, all the subgroup analyses with fewer than 10 studies but more than 1 study (subgroups: India, Iran, and dental fluorosis) reported statistically significant estimates, as shown in the excerpt of Table 2 below.
- As mentioned previously, the NASEM Committee agreed with our use of the prespecified subgroup analyses to investigate sources of heterogeneity, finding them informative and directly responsive to some of the Committee's previous concerns. They also recommended additional subgroup and sensitivity analyses that were subsequently added to the manuscript.

Excerpt of Table 2. Pooled SMDs and 95% CIs for Children's IQ Scores and Exposures to Fluoride.

Analysis	Number of Studies	SMD (95% CI)	Heterogeneity	
			p-value	I^2
Subgroup Analyses				
Country				
China	39	−0.43 (−0.52, −0.34)	<0.001	85%
India	8	−0.99 (−1.55, −0.43)	<0.001	93%
Iran	4	−0.68 (−0.99, −0.38)	0.077	56%
Canada	1	0.01 (−0.19, 0.21)	NA	NA
Mexico	1	0.13 (−0.16, 0.42)	NA	NA
New Zealand	1	0.01 (−0.19, 0.22)	NA	NA
Pakistan	1	−0.25 (−0.65, 0.16)	NA	NA
Assessment Type				
CRT-RC tests	29	−0.36 (−0.46, −0.27)	<0.001	82%
Non-CRT-RC tests	26	−0.60 (−0.78, −0.42)	<0.001	89%
Raven's tests	10	−0.76 (−1.10, −0.43)	<0.001	91%
Other tests	16	−0.52 (−0.74, −0.29)	<0.001	89%
Exposure Type				
Water fluoride	32	−0.37 (−0.46, −0.27)	<0.001	82%
Dental fluorosis	7	−0.99 (−1.57, −0.41)	<0.001	96%
Other exposures[b]	16	−0.54 (−0.71, −0.37)	<0.001	81%

6b.I: Furthermore, the "dose-response" relationship assessments yielded conflicting conclusions, ranging from "non-linear" for fluoride water study to "linear" for urine studies, to "no effect" for other exposure groups, lacking biologic plausibility and casting additional doubt on the overall assessment.

Response: Disagree (no change)

- We disagree that there are conflicting conclusions. The direction of the observed association was consistent across both the water and urine *dose-response meta-analyses*. There are, however, differences in which model was the best fit for the data. Given the heterogeneity and the fact that the individual studies contributing to the water and urine *dose-response meta-analyses* were different, differences in model fit are expected.

6b.J: In summary, while the results of this meta-analysis imply a statistical link between fluoride exposure and IQ, they should be interpreted and communicated with great caution due to the potential for bias from observational studies, the lack of an underlying biologic or scientific plausibility, numerous methodological and statistical issues, and the potential for detriment to the public's health caused by the effect on public perception and policy caused by improperly attributing a putative adverse health effect to an intervention with significant known benefits.

Response: Agree (no change)

- We agree that the results of this analysis require careful and clear communication, which is why we are working closely with the NIEHS Office of Communications to draft relevant communications. We agree that public perceptions around exposures to fluoride are very important and think that this meta-analysis and the prepublication 2022 NTP Monograph should be used to inform a careful analysis of data concerning the potential risks as well as benefits of fluoride. We have provided detailed responses to [redacted] critique concerning risk of bias from observational studies elsewhere in a previous response. We discuss biological plausibility of the studies included in this meta-analysis in the prepublication 2022 NTP Monograph.

6b.K: The authors made laudable attempts to mitigate the impact of these problems, but no statistical approach can solve all of the problems caused by the inappropriate choice of meta-analysis. As indicated in the Cochrane Handbook, results from the investigation of high heterogeneity studies that is designed after heterogeneity is identified can at best lead to hypotheses generation and to support proposals for additional studies. They should be interpreted with caution and should generally not be listed among the conclusions of a review (Cochran Handbook, Section 10.10.3). They should certainly not be used as the rationale for changing public policy.

Response: Disagree (no change requested)

- Section 10.10.3 of the Cochrane Handbook has been misrepresented in the above comment. The sentence from the Cochrane Handbook immediately before the one referenced in the comment states: "Reliable conclusions can only be drawn from analyses that are truly pre-specified before inspecting the studies' results, and even these conclusions should be interpreted with caution." We again point out that all the analyses investigating potential sources of heterogeneity were planned a priori as reflected in the protocol or were added at the recommendation of the NASEM Committee or other peer reviewers.

6b.L: **A more scientifically justifiable conclusion for this review is that extensive, rigorous, and reproducible research in both animals and humans is needed to address the important question of causal influences of fluoride on human cognition.**

Response: Disagree (no change)

- It's always easy to call for more research and we agree that targeted research can certainly add clarity to the existing data—particularly at lower exposure levels. However, hundreds of human and animal studies have been published on this topic. Although these comments are on a previous draft of the meta-analysis, we would like to point out that a recent update of the literature identified 10 new studies that were subsequently added to the database (and are included in the current draft). These new studies were published in the past 2 years and their addition left the findings of the analysis essentially unchanged. Our meta-analysis now includes 60 studies of children's cognition and fluoride exposure, 13 of which are high quality. Many high-quality meta-analyses have been based on fewer studies and the current meta-analysis includes more than double the number of studies of any previous meta-analysis of fluoride.

6b.M: Introduction:

The manuscript described three different types of meta-analyses, using Standardized Mean Difference (SMD) as the effect estimate for each study's outcome (IQ), and assessed and addressed issues related to heterogeneity and publication bias.

First, *mean effect meta-analysis* of group-level fluoride measurement studies (n=46) was conducted to investigate putative associations between fluoride exposure and a child's IQ, with the conclusion that there was "an inverse association between fluoride exposure and children's IQ" (pooled SMD for all studies: −0.49; 95% CI: −0.60, −0.38; p-value < 0.001). However, there was evidence of high heterogeneity (I^2 = 89%, p-value < 0.001) and publication bias (funnel plot and Egger's p-value < 0.001, Begg's p = 0.04), both of which militate against the use of meta-analysis.

Second, *dose-response meta-analysis* of group-level fluoride measurement studies (n=46) was conducted to assess dose-response relationships between fluoride and IQ, with the conclusion that "associations for drinking water appeared to be non-linear and associations for urine appeared to be linear." Heterogeneity and publication bias issues were not reported in this section of the manuscript.

Third, *meta-analysis of regression slopes* for the individual-level urine studies (n=6) was conducted to assess study outcomes with respect to a 1-mg/L unit increase in urinary fluoride. There was moderate heterogeneity (I^2=48%, p=0.09) and indication of publication bias. The manuscript concluded that, after adjustment for publication bias using a trim and fill approach, "a 1-mg/L increase in urinary fluoride was associated with lower IQ, with an adjusted pooled effect estimate of −0.87 (95% CI: −1.93, 0.19; p-value = 0.302)". A p-value of 0.302 indicates that chance may be a reasonable explanation for this finding. The critiques of meta-analyses that follow are categorized by the major issues mentioned above.

Response: Disagree (no change)

- As described in responses to earlier comments, we disagree that evidence of heterogeneity and publication bias militate against the use of meta-analysis. We conducted the meta-analysis in response to the NASEM Committee's peer review of the 2019 draft NTP Monograph. The NASEM Committee urged us not to avoid conducting a meta-analysis because of heterogeneity: "The committee strongly recommends that NTP reconsider its decision not to perform a meta-analysis and, if it still decides not to do a meta-analysis, that it provide a more thorough and convincing justification for its decision…A properly conducted meta-analysis can account for heterogeneity in exposure measurements and other aspects of study design, so it is not clear why heterogeneity was listed as a reason for not performing one."
- The *dose-response meta-analysis* used the same studies that were used in the *mean-effects meta-analysis*. The *mean-effects meta-analysis* already describes heterogeneity and publication bias issues. Therefore, it would be redundant to describe them again.
- After updating the *regression slopes meta-analysis* with new studies from the updated literature search, there was no longer evidence of publication bias, so the quoted text has been removed from the manuscript.

6b.N: Issues related to using SMD as an effect estimate

From the manuscript:

> "The effect estimates in the primary mean-effects meta-analysis were the standardized mean differences (SMDs) for heteroscedastic population variances."

Comment:

The overall treatment effect [in terms of SMD] can be difficult to interpret as it is reported in units of standard deviation rather than in units of any of the measurement scales used in review (Egger et al., 2008). Why would the true effect of fluoride (assuming there is one) depend on the standard deviation? If the reason for using the SMD instead of the more interpretable difference in IQs is that different tests were used to assess intelligence, it is an indication that combining such disparate studies may be inappropriate. *"There is a price for standardization—the SMD does not have any meaningful units. Instead, it can only indicate whether there is any statistical significance of pooled results." (Mickael and Merja, 2021).* Nonetheless, interpreting SMD values of 0.2, 0.5, and 0.8 as small, medium, and large effect sizes, respectively, is a widely accepted role of thumb (Cohen, 1988). Accordingly, using these terms throughout will help clarify the meaning of this estimate. There is also concern that the inclusion of studies with both high RoB and large sample size leads to overstated estimates of the effect sizes.

Response: Disagree (no change)

- The difference in tests is an appropriate reason to use the SMD as the unit of measure in our meta-analysis. The SMD is commonly used in meta-analysis when the studies all assess the same outcome (e.g., intelligence) but measure it in a variety of ways (e.g., WISC-R, Combined Raven's Test for Rural China, etc.). It is necessary to standardize the results of the studies to a uniform scale before they can be combined (Higgins et al. 2021). To address the concern of combining studies that used different tests to assess intelligence, we conducted subgroup analyses that stratified by type of IQ assessment. We also acknowledge limitations of the *mean-effects meta-analysis* in our discussion.
- In addition, in its peer review of the 2019 draft NTP Monograph, the NASEM Committee supported the use of SMDs in this meta-analysis as the Committee recommended that NTP update the Choi et al. (2012) meta-analysis (which used SMDs) with more recent papers. The protocol, which clearly describes these methods, was also peer reviewed.
- Also, as previously mentioned, in the peer review of the 2020 draft NTP Monograph, the NASEM Committee agreed with the methods used in the meta-analysis: "The overall approach appears to be sound in comparing mean IQ scores for the most and least highly exposed to fluoride even though the absolute fluoride concentrations are not comparable among studies", and "The meta-analysis applied standard, broadly accepted methods, and the data shown in Figure A5-1 and the related evaluations are especially informative (NTP 2020a, p 235)."
- We appreciate the suggestion regarding interpretation of the SMDs; however, because the standard deviations of measured IQs are specific to the study population from which they are measured, and the meta-analyses pools the results of many different study populations, we did not translate the pooled SMD into IQ points nor did we characterize them as small, medium, or large. In addition, the Cochrane guidance (Cochrane Section 12.6.2) states that

"…some methodologists believe that such interpretations are problematic because patient importance of a finding is context-dependent and not amenable to generic statements." Also, the SMD interpretations based on cutoffs mentioned by [redacted] are values used in social sciences research (as cited in the 1988 Cohen book "Statistical Power for the Behavioral Sciences") and the utility of those values in analyzing observational environmental health studies has not been demonstrated.

- The concern about combining results from high and low risk-of-bias studies was addressed by the subgroup analyses stratified by risk of bias. As the comment points out, the effect estimate was smaller among the low risk-of-bias studies. This may be due to lower levels of exposure and/or smaller differences in exposure between "high" and "low" exposure groups among the low risk-of-bias studies. The comment fails to note that there are other stratified estimates that would be considered to underestimate, rather than overestimate, the pooled effect estimates in both the *mean-effects* and *regression slopes meta-analyses*. As for studies with large sample sizes, we performed the meta-analyses using random effects models which account for study-specific sample sizes.
- Finally, we would like to note that this comment completely ignores that this manuscript was not restricted to an SMD meta-analysis. The manuscript also includes a *regression slopes meta-analysis* (which has not been previously done in the fluoride and IQ literature). The *regression slopes meta-analysis* does not have the same limitations as the SMD analysis. It uses individual-level exposure data, and the regression coefficient can be directly interpreted as the expected change in IQ points in the study population per 1-mg/L increase in urinary fluoride.

High heterogeneity among studies

6b.O:

From the manuscript:

> The heterogeneity among the group-level fluoride measurement studies was high for all studies (n=46) as well as for the low risk of bias studies (n=9).

Comments:

- This can be seen not only in statistics such as I^2 and the p-value for heterogeneity, but in the figures as well. For example, if there were homogeneity of effects, then approximately 5% of SMDs should be outside of the dotted lines in supplemental eFigure 3. Instead, more than a third of them are outside the dotted lines. A similar phenomenon can be seen in supplemental eFigure 8, even for the studies with low risk of bias. This indicates an unacceptably large level of heterogeneity, such that mean-effect and dose-response meta-analyses should not be conducted in first place. *"High heterogeneity can potentially lead to misleading and non-generalizable results and may indicate that meta-analysis is contra-indicated. A group of studies needs to be similar enough clinically and methodologically to be pooled in a meta-analysis before considering their statistical heterogeneity." (Cochrane handbook, Section 9.5).*

Response: Disagree (no change)

- [redacted] refers to eFigure 3, which is a funnel plot of the included studies in the *mean-effects meta-analysis*. The funnel plots are not used to illustrate or evaluate

homogeneity as implied by [redacted], but to evaluate the potential for publication bias. To evaluate heterogeneity, we performed and reported results of statistical tests for heterogeneity, while also transparently discussing limitations of such tests in the *Discussion* section.

- We disagree with [redacted] that the *mean-effects* and *dose-response meta-analyses* should not have been conducted. As previously explained, high heterogeneity is not a valid rationale for not conducting a meta-analysis. Again, the NASEM Committee agreed with the methods used in the meta-analysis in its peer review. Additionally, in meta-analyses of observational studies, especially those using SMDs as effect measures, high levels of heterogeneity are to be expected. Our protocol outlined the study inclusion criteria which were carefully evaluated to ensure that the appropriate studies were included in the meta-analyses. The protocol also outlined the subgroup analyses that were to be performed to investigate potential sources of heterogeneity.
- The select quote from Section 9.5 misrepresents the totality of the Cochrane guidance, particularly on heterogeneity. Section 10 of Cochrane is "Analyzing data and undertaking a meta-analysis" and section 10.10.3 is on heterogeneity, where Cochrane recommends exploring heterogeneity by conducting subgroup analyses. We transparently presented the heterogeneity results and investigated potential sources of heterogeneity.
- In the *Discussion* section, we clearly outline the limitations of the *mean-effects meta-analysis* and the unexplained heterogeneity. We also added new text to point out that:

  "…the aggregate nature of the mean-effects meta-analysis might not be sufficiently sensitive to capture potential sources of heterogeneity, as seen possible when using studies with individual-level data in the regression slopes meta-analysis."

6b.P:

- Miranda et al. (2021) performed a similar meta-analysis with their results, pointing to an association between fluoride and IQ. However, due to the high heterogeneity among the existing studies they concluded that current evidence is inadequate to support such a conclusion, even at high fluoride levels.

Response: Disagree (no change)

- The Miranda et al. (2021) meta-analysis was not similar to our meta-analysis, which was different in both scope and methodological approach. For example, the systematic review by Miranda et al. (2021) had very limited inclusion criteria, which did not allow for studies using individual-level fluoride exposure measurements to be included. Their analysis only included cross-sectional studies, while our meta-analysis also included prospective cohort studies. In addition, their analysis was much smaller (n = 10 studies) than our meta-analysis (n = 60 studies) and had a very different methodological approach, as it was limited to studies from which crude (unadjusted) odds ratios could be calculated. Finally, Miranda et al. (2021) was limited to one analysis that found a strong association between high fluoride exposure and decreased IQ (unadjusted OR = 3.88, 95% CI 2.41−6.23; $p < 0.00001$). Our analysis found consistent results across different analysis types (*mean-effects, dose-response, and regression slopes meta-analyses)* and across multiple prespecified subgroup analyses.

- The observed level of heterogeneity (77%) in Miranda et al. (2021) is not unusual in small meta-analyses such as theirs; however, it is also worth noting that the authors did not attempt to investigate any sources of heterogeneity in their analysis.

6b.Q:

- Since meta-analyses can be performed with data from as few as two studies, it is more appropriate to include only studies that are clinically and methodologically similar and sound. Pooling results from all studies with significant differences and biased results is not appropriate (ref. Cochrane Handbook) and is likely to overestimate the effects of exposure. This seems to be the case for the meta-analysis of regression slopes (urine level studies, n=6) which reported medium heterogeneity among low-risk, individual-level fluoride measurement.

  Response: Disagree (no change)

  - The purpose of the meta-analysis is to combine results from multiple studies with a variety of features to examine data collectively and more precisely quantify the overall (pooled) association. The Cochrane Handbook does not say it is "not appropriate" to pool results from studies with differences in design and potential sources of bias. Rather, they recommend that differences in study design, study biases, variation in exposure characterization and outcome assessment across studies, and reporting biases be carefully considered. Moreover, excluding studies from systematic reviews or excluding studies from systematic reviews with meta-analyses is not considered a best practice in the systematic review community (Higgins et al. 2021). As a well-documented systematic review and meta-analysis, this evaluation followed a protocol where inclusion and exclusion criteria were defined a priori. Among included studies, our risk-of-bias assessment carefully considered study-specific potential for bias. Our analyses stratified results by risk-of-bias status to evaluate the potential impact on the overall effect estimates from studies that have high potential for bias versus studies that have low potential for bias. We carefully considered other differences between studies by conducting additional prespecified subgroup analyses by factors such as exposure type, outcome assessment, and country.
  - In addition, as previously mentioned, the NASEM Committee agreed with the methods used in the meta-analysis, which includes pooling results from included studies:

    "The critical information regarding comparison of study results comes from the new meta-analysis, which seeks to extract and integrate comparable findings from selected studies as discussed further below. The overall approach appears to be sound in comparing mean IQ scores for the most and least highly exposed to fluoride even though the absolute fluoride concentrations are not comparable among studies."

    "The meta-analysis applied standard, broadly accepted methods, and the data shown in Figure A5-1 and the related evaluations are especially informative."

6b.R: Three strategies were used to assess/address the heterogeneity:

Random-effects models to address heterogeneity:

<u>From the manuscript:</u>

"Data from individual studies were pooled using a random-effects model."

Comments:

- Random-effects models, as opposed to fixed-effect models, are typically used in meta-analyses when there is unexplained heterogeneity. Such models assume that the effects estimated within each study are not identical, but do follow a specific distribution (Cochrane handbook, Section 9.5). However, according to the Cochrane Handbook, random-effect models can only be used "if the heterogeneity cannot be explained clinically or methodologically. It does not remove heterogeneity, so results need to be carefully interpreted." (Cochrane handbook, Section 9.5). This is relevant because the decision to use random effect models seems to be based on statistical findings of heterogeneity, with no attempt to ensure the clinical and methodologic similarity among the studies before conducting the meta-analyses. The State of the Science document mentioned that "heterogeneity within the available evidence was evaluated to determine if a quantitative synthesis (i.e., meta-analysis) is appropriate." (p.19) but no presentation or discussion of the outcomes of this process can be found.

  Response: Disagree (no change)
  - We are unable to find the quote that [redacted] cites or any text in the Cochrane Handbook that states that random-effects models can only be used if the heterogeneity cannot be explained clinically or methodologically.
  - We followed Cochrane guidance (Section 10.10) that recommends using a random-effects model instead of a fixed-effects model when the assumption of a common (fixed) effect size is not appropriate. Figure 2 shows heterogeneity in study-specific effect estimates, clearly indicating that a fixed effect is not appropriate in our meta-analysis (see Figure 2 below). In addition, as recommended by the Cochrane Handbook, even though we used random-effects models, we still investigated potential sources of heterogeneity (Cochrane, Section 10.10; Deeks et al. 2021).

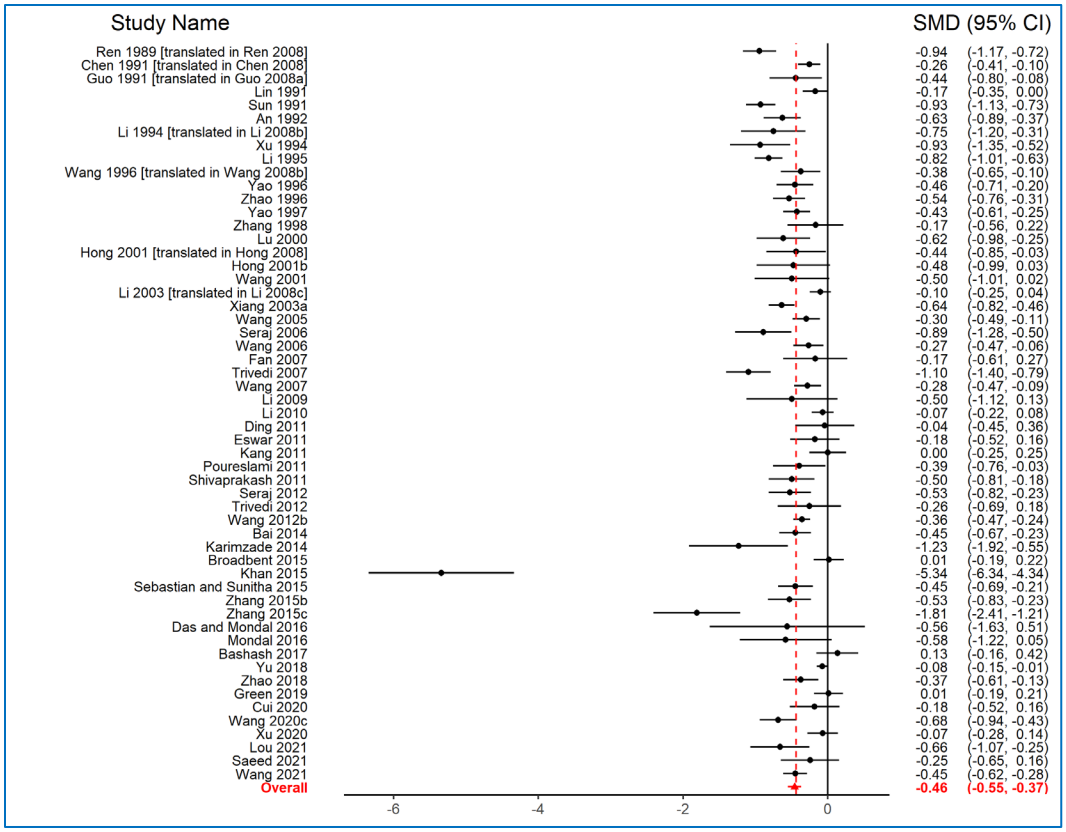


Figure 2. Association Between Fluoride Exposure and IQ Scores in Children

6b.S:

- According to the Cochrane Handbook, "*a pragmatic approach is to plan to undertake both a fixed-effect and a random-effects meta-analysis, with an intention to present the random-effects result if there is no indication of funnel plot asymmetry. If there is an indication of funnel plot asymmetry, then both methods are problematic*." (Cochrane Handbook, Section 10.10.4.1).

Response: Disagree (no change)

- This quote comes from a list of considerations for authors to contemplate when making a choice about whether to use a fixed-effects model or random-effects model. The Handbook is clear that there are a variety of factors to consider and that there is no universal recommendation on which model to use. As explained in the previous response, we have demonstrated that the random-effects model was the appropriate choice for these data.

6b.T:

A fundamental assumption of the random effects model is that the true effects in different studies represent a random sample from some population. The majority of studies are from one country, with unique environmental, economic, and sociopolitical conditions that can hardly be regarded as a random sample that allows generalization to other countries.

Response: Agree (no change requested)

- We agree, which is why we investigated potential sources of heterogeneity, including country.

6b.U: Using sensitivity analyses to address heterogeneity

From the manuscript:

"Multiple sensitivity analyses were conducted as part of the mean-effect meta-analysis (e-table 3) and meta-analysis of regression slopes (e-table 6). Four additional analyses were conducted as per NASEM's recommendation (not shown). The authors concluded that no substantial changes in the pooled SMD estimate were revealed when studies were excluded."

Comment:

Sensitivity analyses do appear necessary, as there is at least one very clear outlier (Khan, 2015) in Figure 2 and in several figures in the supplementary materials. Nonetheless, removal of one or two studies does not eliminate the heterogeneity of results; the magnitude and direction of the effect remain unknown because of lack of adequate testing for heterogeneity.

Response: Disagree (no change)

- We disagree that there was a lack of adequate testing for heterogeneity. As previously mentioned in an earlier response, the NASEM Committee agreed with our use of the prespecified subgroup analyses to investigate sources of heterogeneity, finding them informative and directly responsive to some of the Committee's previous concerns. They also recommended additional subgroup and sensitivity analyses that were subsequently added to the manuscript. Also, as previously explained, the goal of these subgroup analyses was not to eliminate heterogeneity.
- We also disagree that the direction of effect remains unknown. The NASEM Committee agreed with our conclusion that the results (i.e., the direction of the association) were consistent: "As noted in the revised monograph, 44 of the 46 studies represented in that figure had effect estimates to the left of zero-results that indicate an association between higher fluoride exposures and lower IQ. Those results highlight the marked consistency in the current epidemiological literature on fluoride and childhood IQ."
- The NASEM Committee also commented that "NTP notes that 44 of the 46 studies (96%) in its meta-analysis of childhood IQ have effect estimates to the left of zero. That finding should be emphasized more, and its meaning with respect to evaluating and quantifying heterogeneity should be mentioned. To assess heterogeneity, NTP primarily used the Cochran's Q test. However, heterogeneity can also be assessed by providing a count or percentage of the number of studies to the right or left of the null value. Some would consider that a much simpler, more intuitive, and perhaps more useful way of assessing heterogeneity, especially in light of the marked differences between the studies in design, study populations, exposure and outcome assessment methods, and statistical analyses."

6b.V: Using subgroup analyses to address heterogeneity

From the manuscript:

"Subgroup analyses were performed to investigate sources of heterogeneity"

Comment:

Sub-group analyses described in the manuscript do not appear to be pre-specified or justified at the protocol stage based on a clear theoretical, biological, or clinical basis. According to the Cochrane Handbook, "subgroup analysis should be kept to a minimum, and pre-specified and justified at the protocol stage of the review. The planned analyses should be followed at review stage (if sufficient data are available) to minimize selective reporting or over-interpretation of the results based on findings." (Cochrane Handbook, Section 9.6). Furthermore, "reliable conclusions can only be drawn from analyses that are truly pre-specified before inspecting the studies' results, and even these conclusions should be interpreted with caution."

Response: Disagree (no change)

- We disagree with the implication that the subgroup analyses were not based on a biological or other scientific basis. These analyses were based on established scientific evidence which includes the National Research Council's 2006 report (NRC 2006), two previous meta-analyses by Choi et al. (2012) and Duan et al. (2018), and the prepublication 2022 NTP Monograph.
- As previously mentioned in an earlier response, the NASEM Committee agreed with our use of the prespecified subgroup analyses, finding them informative and directly responsive to some of their previous concerns. They also recommended additional subgroup and sensitivity analyses that were subsequently added to the manuscript.
- We agree that subgroup and sensitivity analyses should be prespecified and followed to minimize selective reporting or over-interpretation of the results based on findings, and that results should be interpreted carefully. All our subgroup and sensitivity analyses were prespecified in the protocol or included at the recommendation of peer review.

6b.W:

From the manuscript:

"Sub-group analyses suggested that our conclusions were consistent across high and low risk-of-bias studies."

Comments:

- Results are clearly not numerically consistent within high- and low-risk of bias studies (effect size -0.55 versus -0.24). "Such a difference [in effect estimates between high and low risk of bias studies] is a common finding because biased studies are more likely to overestimate the effects of treatment." (Harrer et al (2021). Furthermore, there is substantial heterogeneity even in the low risk of bias studies. Five of the eight points lie outside the 95% pseudo confidence interval on the right side of eFigure 8. Moreover, several pairs of confidence intervals among the 9 low

risk of bias studies in eFigure 7 have completely non-overlapping confidence intervals, a strong indication that the true effects are different in different studies.

Response: Disagree (no change)

- Variation in the exact numerical estimate is expected in subgroup analyses, as different individual studies contribute to the different pooled effect estimates. However, the direction of the association, which we consider a more important indication of consistency in the literature, was consistent. The NASEM Committee, in its peer review report of the 2020 draft NTP Monograph, agreed with our statements on consistency:

  "As noted in the revised monograph, 44 of the 46 studies represented in that figure had effect estimates to the left of zero—results that indicate an association between higher fluoride exposures and lower IQ. Those results highlight the marked consistency in the current epidemiologic literature on fluoride and childhood IQ."

  "NTP notes that 44 of the 46 studies (96%) in its meta-analysis of childhood IQ have effect estimates to the left of zero. That finding should be emphasized more".

- We are not clear on the point of the Harrer et al. (2021) quote, as we were able to demonstrate differences in effect estimates by using subgroup and sensitivity analyses. As described above in response to another comment, studies that reported unadjusted and adjusted effect estimates did not provide evidence that higher potential for bias (as would be reflected in unadjusted effect estimates) resulted in an overestimation of the effect. This is captured in our sensitivity analysis for the *regression slopes meta-analysis* that used unadjusted effect estimates from Bashash et al. (2017), Cui et al. (2018), Green et al. (2019), and Yu et al. (2018) (see excerpt of eTable 6 below). Also, the comment fails to note that there are other stratified estimates that would be considered to underestimate, rather than overestimate, the pooled effect estimates in both the *mean-effects* and *regression slopes meta-analyses*.

- Note: We acknowledge that there is high heterogeneity in both the high and low risk-of-bias subgroups in the subgroup analysis by risk-of-bias status in the *mean-effects meta-analysis*. In a previous response above, we describe the new text added to the *Discussion* section with further interpretations of these subgroup analyses.

- [redacted] refers to eFigure 8, which is a funnel plot by risk of bias in the *mean-effects meta-analysis*. The funnel plots are not used to illustrate or evaluate heterogeneity, as implied by [redacted], but to evaluate potential for publication bias. We performed and reported on statistical tests for heterogeneity, while also transparently discussing limitations of such tests in the *Discussion* section.

- As [redacted] noticed, eFigure 7 illustrates that there is heterogeneity in the low risk-of-bias studies as well, as illustrated by the I^2 of 83% reported in Table 2.

Excerpt of eTable 1. Regression Slopes Meta-analysis

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimate				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Sensitivity Analyses				
Using unadjusted estimates from Bashash et al. (2017),[112] Cui et al. (2018),[76] Green et al. (2019)[113], Yu et al. (2018)[3]				
Full-scale IQ	9	−1.82 (−2.81, −0.83)	<0.001	76%

6b.X:

- In addition, sub-group analyses are "purely observational, so we should always keep in the mind that effect differences may also be caused by confounding variables." (Harrer et al., 2021). Of course, the studies in this meta-analysis are all observational, so confounding is a major concern even if there were no subgroup analyses. It is possible that fluoride or a combination of other factors is to blame for these differences. There is no sub-group analysis by exposure to fluoride or known neurotoxic chemicals such as lead and arsenic.

  Response: Disagree (no change)
  - As previously mentioned, potential confounding, including concurrent exposure to other neurotoxic chemicals (e.g., lead and arsenic), was assessed extensively as a key component of the risk-of-bias assessment. In addition, our analysis includes a subgroup analysis by exposure to other chemicals (such as arsenic, iodine, coal). Also, concerns about confounding among individual studies may be minimized or ruled out if consistent results are seen across different study populations, study designs, exposure settings, and studies that adjust for different sets of confounders (Arroyave et al. 2020; Steenland et al. 2020).

6b.Y:

From the manuscript:

"Heterogeneity remained low or moderate (I^2 < 48%) for all subgroup analyses except gender (I^2 > 52%)."

Comments:

- Most sub-group analyses, as shown in Table 2, had between-studies heterogeneity of 79% or higher, suggesting significant unexplained heterogeneity among sub-groups that needs to be further investigated. Subgroup analyses are intended to explain some of the heterogeneity, not to introduce more heterogeneity.

  Response: Disagree (no change)
  - In response to a previous [redacted] comment, we further investigated potential sources of heterogeneity. The results are presented in the supplemental materials.
  - We disagree that the heterogeneity within the subgroup analyses "needs to be further investigated." As stated before, the subgroup analyses were planned a priori to

investigate potential sources of heterogeneity in the overall effect estimate. It would not be informative nor is it common practice to then start investigating sources of heterogeneity within subgroup analyses.

6b.Z:

- Several of the included studies have overlapping high and low fluoride groups (i.e., what is labeled as low exposure groups in one study is considered high in another), which likely contributed to the study's high heterogeneity. Sub-group analysis based on precise cut-off points for exposure levels may help explain the considerable variability in the studies.

  Response: Disagree (no change)
  - This comment is referring to the *mean-effects meta-analysis* which was not designed to evaluate dose-response. However, we did address fluoride exposure levels separately in the *dose-response meta-analysis* using studies included in the *mean-effects meta-analysis*. The *dose-response meta-analysis* includes subgroup analyses based on precise cut-offs points for exposure levels (0 to <4 mg/L, 0 to <2 mg/L, and 0 to <1.5 mg/L). Within each of these subgroup analyses (i.e., <4 mg/L, <2 mg/L, and <1.5 mg/L), the data do not overlap.

6b.AA:

- According to Harrer et al. (2021), sub-group meta-analyses may lack power to detect small differences between groups. One solution is performing a subgroup statistical power analysis beforehand to determine the minimum detectable effect size difference with a subgroup analysis. An alternative approach would be to include a minimum of 10 studies for each level/unit (e.g. , country, gender, exposure type, etc.) analyzed in a sub-group analysis. However, many of the level/unit subgroup analyses in the manuscript included fewer than 10 studies.

  Response: Disagree (no change)
  - As we mentioned in a previous response and explained in our protocol, the purpose of the subgroup analyses was to explore potential sources of heterogeneity, not to detect small differences between subgroups. Even so, we would like to note that, except for certain countries and for studies with other sources of fluoride exposure, all the subgroup analyses of the *mean-effects meta-analysis* included at least 10 studies. We would also like to note that all the subgroup analyses across both the *mean-effects meta-analysis* and the *regression slopes meta-analysis* with 2–9 studies per group detected a statistically significant association.

6b.BB: Addressing publication bias

From the manuscript:

- Using funnel plot and Egger regression, evidence of publication bias was observed in all analyses including studies with a high risk of bias, but not those using studies with a low risk of bias.

- Adjusting for possible publication bias through trim-and-fill analysis resulted in an adjusted pooled SMD of – 0.36 (95% CI: −0.48, −0.24) under the mean effect meta-analysis, and an adjusted pooled effect estimate of −0.87 (95% CI: −1.93, 0.19; p value = 0.302) in the metanalysis of regression slopes.

Comment:

Since all but three studies show negative association to begin with, adjusting for publication bias would only center the studies around a distinctly negative pooled effect estimate, with incomplete representation of potential studies showing an association between fluoride and lower IQ. The effect size from all published studies (SMD= -0.49) is 26% larger than the adjusted effect size that imputes unpublished or excluded studies (SMD= -0.36). This imputation produces an adjusted estimate that is closer to a small effect than a medium effect based on Cohen's d. These variable results should be emphasized in the discussion/abstract, given the possibility that there is a bias against the publication of studies that have neutral outcomes.

Response: Disagree (no change)

- Accounting for publication bias is meant to account for potentially missing studies with neutral or positive effects, likely shifting the effect towards 0, as illustrated in the pooled SMD. As [redacted] points out, this was the case. In addition, the pooled SMD remained statistically significant even after the trim-and-fill analysis, which highlights the consistency of the overall association between fluoride exposure and lower IQ in children. We have reported these results in the *Results* section and provided more information in the supplemental materials; we disagree that more emphasis is needed.
- Note that, because there was no longer evidence of publication bias once the literature was updated in November 2021, the trim-and-fill analysis for the *regression slopes meta-analysis* was removed from the current draft.

6b.CC: Issues with the dose-response analyses

From the manuscript:

There is "a dose-response relationship between mean children's IQ and group-level fluoride exposure measures."……and that… “associations for drinking water appeared to be non-linear and associations for urine appeared to be linear.”

Comments:

- Associations for studies using urine fluoride exposure levels appeared to be linear only for all studies combined, and not for the low risk-of-bias studies. The relationship is non-linear for low RoB studies and in fact, the non-linear relationship appears to include a supra-linearity component when looking at low RoB studies of fluoride from both water and urine studies (See e-Table 4), indicating a dose-response curve that corresponds to greater effects at low doses than implied by linearity. For example, drinking water studies with a fluoride level of <2 mg/L (beta=-0.34) have a greater change in SMD than studies with a fluoride level of 4 mg/L. (beta=-0.22). There was no physiologic explanation provided.

Response: Disagree (no change)

- Because of small difference in AICs between the different models, and for ease of interpretability, only the linear model results were reported for the low risk-of-bias studies, so it is unclear why [redacted] says that the low risk-of-bias studies have non-linear relationships or what they mean by evidence of "supra-linearity." However, we do not find it surprising, considering the heterogeneity discussed earlier, that doubling the number of studies included in the <4-mg/L model would result in a different beta coefficient. We do not consider the cited example as convincing evidence of supra-linearity.

6b.DD:

- Regarding the mechanistic or physiologic explanation of supra-linear relationships between environmental measures and IQ, Bowers and Beck (2006) asserted that, "one must take care when interpreting statistical relationships with unexpected results that have no apparent underlying biological or other scientific basis… [and that] consistency of findings in numerous epidemiological studies is an insufficient basis for concluding that the finding is of biological significance, *as all studies share a common alternative explanation*."

Response: Disagree (no change)

- We disagree with the implication that the results were unexpected or have no apparent underlying biological or other scientific basis. Our hypothesis, that higher fluoride exposure would be associated with lower IQ in children, was based on established scientific evidence. This evidence includes the National Research Council's 2006 report (NRC 2006), two previous meta-analyses by Choi et al. (2012) and Duan et al. (2018), and the prepublication 2022 NTP Monograph.

6b.EE:

From the manuscript:

"We also examined whether there was a dose-response relationship at lower exposure levels that corresponded with the U.S. Environmental Protection Agency drinking water standards and World Health Organization drinking water guidelines."

Comment(s):

- Fluoride levels and their studies included in the dose-response assessment appear to be overlapping (eTable 4), i.e., studies with <1.5 mg/L are also included in the < 2mg/L group, and studies with <2mg/L and <1.5mg/L are included in the <4mg/L category. When exposure categories overlap, interpretation of results comparing fluoride exposures and IQ outcomes difficult or impossible.

Response: Disagree (edited for clarity)

- [redacted] may be misinterpreting what the different columns in eTable 4 (and eTable 5) represent. Each column presents results from a separate dose-response analysis restricted to a specified fluoride exposure range [i.e., all data, <4 mg/L (i.e., 0 to <4 mg/L), <2 mg/L (i.e., 0 to <2 mg/L), and <1.5 mg/L (i.e., 0 to <1.5 mg/L). Each row of the

two tables report dose-response results by statistical model used (i.e., linear, quadratic, or restricted cubic spline) for each of these different fluoride exposure ranges. It would be incorrect to interpret eTables 4 and 5 as each row showing results from one dose-response analysis where a trend across different exposure groups (i.e., different columns) could be evaluated. The following new text in the supplemental materials describes eTable 4 as providing results of different dose-response analyses based on restrictions to various fluoride exposure ranges:

"When analyses were restricted to exposed groups with <4 mg/L (i.e., 0 to <4 mg/L) fluoride in drinking water (n = 21 publications [6 low and 15 high risk-of-bias studies]), there was a statistically significant inverse association between fluoride exposure and children's IQ (SMD: −0.22; 95% CI: −0.27, −0.17; p-value < 0.001) (eTable 4). When restricted to <2 mg/L (i.e., 0 to <2 mg/L) in drinking water (n = 7 publications [3 low and 4 high risk-of-bias studies]), the magnitude of the effect estimate did not substantially change (SMD: −0.15; 95% CI: −0.41, 0.12; p-value = 0.274). However, when restricted to exposed groups with <1.5 mg/L (i.e., 0 to <1.5 mg/L) in drinking water (n = 7 publications [3 low and 4 high risk-of-bias studies]), there was no longer an association between fluoride in drinking water and children's IQ (SMD: 0.05; 95% CI: −0.36, 0.45; p-value = 0.816). When analyses were further restricted to low risk-of-bias publications at <4 mg/L, <2 mg/L, and <1.5 mg/L, the associations remained in the same direction and were larger in magnitude compared to when data from both low and high risk-of-bias studies were combined (eTable 4 and eTable 5)."

6b.FF:

- It is also valuable to establish the lowest fluoride dose that can trigger a response. According to the supplement (p.33), "when assessment was restricted to exposed groups with <1.5 mg/L in drinking water for all studies and again for low risk of bias studies there was no longer an association between fluoride in drinking water and children's IQ." However, at this low dose, there seems to be F-IQ association in urine studies. This appears contradictory since urine studies represent measurement of total fluoride intake. i.e., for individuals with fluoride level of 1.5 mg/L in urine, fluoride level in drinking water must be less than 1.5 mg/L, since intake from water is only a portion of the total intake.

  Response: Disagree (no change)

  - Establishing the "lowest fluoride dose that can trigger a response" is beyond the purpose or scope of this analysis. Also, the data used in the *dose-response meta-analyses* assessing <1.5 mg/L fluoride in urine and <1.5 mg/L fluoride in drinking water come from completely different sets of studies and have different total sample sizes (2,935 and 4,317, respectively), which could easily account for differences in the results. As [redacted] previously pointed out, the planned analyses should be followed and post-hoc explorations should be limited, which is what we did.

References [see missing and additional references below]

1. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Doing meta-analysis with R: a hands-on guide.

2. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane handbook for systematic reviews of interventions. John Wiley & Sons; 2019 Sep 23.
3. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiol Drug Saf. 2006 May;15(5):291-303. doi: 10.1002/pds.1200. PMID: 16447304.
4. Shi, L., & Lin, L. (2019). The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses. *Medicine*, *98*(23), e15987. https://doi.org/10.1097/MD.0000000000015987
5. https://opal.latrobe.edu.au/articles/journal_contribution/Heterogeneity_and_subgroup_analysis/6818882
6. Miranda GH, Alvarenga MO, Ferreira MK, Puty B, Bittencourt LO, Fagundes NC, Pessan JP, Buzalaf MA, Lima RR. A systematic review and meta-analysis of the association between fluoride exposure and neurological disorders. Scientific reports. 2021 Nov 22;11(1):1-5.
7. Atal I, Porcher R, Boutron I, Ravaud P. The statistical significance of meta-analyses is frequently fragile: definition of a fragility index for meta-analyses. J Clin Epi 111 (2019) 32-40.

Missing References

The following references were cited in the comments without a citation.

- Cohen, 1988

  Possible citation: Cohen, 1988: Statistical Power for the Behavioral Sciences. Second Edition. Lawrence Erlbaum Associates, Publishers. 567 pages.

- Bowers and Beck, 2006

  Possible citation: Bowers TS and Beck BD 2007. What is the meaning of non-linear dose-response relationships between blood lead concentrations and IQ? Neurotoxicology 27 (4) 520-524.

- Mickael and Merja, 2021

  Unclear which reference the comments were referring to.

- Egger et al., 2008

  Unclear which reference the comments were referring to.

Additional References

Anglemyer A, Horvath HT, Bero L. 2014. Health care outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database of Systematic Reviews, Issue 4; Art No.: MR000034. DOI: 10.1002/14651858.MR000034.pub2.

Arroyave WD, Mehta SS, Guha N, Schwingl P, Taylor KW, Glenn B, Radke EG, Vilahur N, Carreón T, Nachman RM, Lunn RM. 2021. Challenges and recommendations on the conduct of systematic reviews of observational epidemiologic studies in environmental and occupational health. J Expo Sci Environ Epidemiol. 31(1):21-30.

Benson K, Hartz A. 2000. A comparison of observational studies and randomized, controlled trials. New England Journal of Medicine. 342(25):1878-86.

Choi AL, Sun G, Zhang Y, Grandjean P. 2012. Developmental fluoride neurotoxicity: A systematic review and meta-analysis. Environ Health Perspect. 120:1362-1368.

Deeks, JJ, Higgins JPT, Altman, DG on behalf of the Cochrane Statistical Methods Group (2021) Cochrane Handbook Chapter 10: Analysing data and undertaking meta-analyses https://training.cochrane.org/handbook/current/chapter-10

Duan Q, Jiao J, Chen X, Wang X. 2018. Association between water fluoride and the level of children's intelligence: A dose-response meta-analysis. Public Health. 154:87-97. https://doi.org/10.1016/j.puhe.2017.08.013

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). 2021. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane. Available from www.training.cochrane.org/handbook

Iheozor-Ejiofor Z, Worthington HV, Walsh T, O'Malley L, Clarkson JE, Macey R, Alam R, Tugwell P, Welch V, Glenny A-M. 2015. Water fluoridation for the prevention of dental caries. Cochrane Database Syst Rev (6):CD010856.

National Research Council (NRC). 2006. Committee on fluoride in drinking water, board on environmental studies and toxicology. Fluoride in drinking water: A scientific review of EPA's standards.: National Research Council. http://www.nap.edu/catalog/11571/fluoride-in-drinking-water-a-scientific-review-of-epas-standards.

Schwingshackl, L., Balduzzi, S., Beyerbach, J., Bröckelmann, N., Werner, S. S., Zähringer, J., Nagavci, B., & Meerpohl, J. J. 2021. Evaluating agreement between bodies of evidence from randomised controlled trials and cohort studies in nutrition research: meta-epidemiological study. BMJ (Clinical research ed.), 374, n1864. https://doi.org/10.1136/bmj.n1864

Steenland K, Schubauer-Berigan MK, Vermeulen R, Lunn RM, Straif K, Zahm S, Stewart P, Arroyave WD, Mehta SS, Pearce N. 2020. Risk of Bias Assessments and Evidence Syntheses for Observational Epidemiologic Studies of Environmental and Occupational Exposures: Strengths and Limitations. Environ Health Perspect. 128(9): 095002. https://doi.org/10.1289/EHP6980

Till C, Green R, Grundy JG, Hornung R, Neufeld R, Martinez-Mier EA, Ayotte P, Muckle G, Lanphear B. 2018. Community water fluoridation and urinary fluoride concentrations in a national sample of pregnant women in Canada. Environ Health Perspect. 126(10):107001. https://doi.org/10.1289/ehp3546

In June 2022, the [redacted] provided comments to NIEHS/DNTP on the prepublication 2022 *NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* and a draft manuscript on a meta-analysis of fluoride exposure and IQ in children. This document contains a subset of the overall [redacted] comments related to the meta-analysis manuscript along with the NIEHS/DNTP responses. The meta-analysis-related comments from the [redacted] are reproduced here in black text, and the NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font. Formatting has been applied to aid in reading.

[redacted]

June 2022

Association between fluoride exposure and children's intelligence: A systematic review and meta-analysis

7.A:

1) (Questions control) In science, one of the hardest things to do is to frame and ask the question to have not necessarily the intended impact, but the optimal impact. In this case, the study examines a previously reported association between fluoride exposure and children's intelligence. The next step particular to federal research is not affirming prior association but building a model to examine the overall cost benefit of fluoride exposure and oral health, given that a decrement in intelligence might be a factor (among others). The proposed analysis seems like the next academic step rather than a federal, public health grounded one. Ideally, the next step could be tackled in a larger context of overall well-being and include dimensions of behavioral health (mental health and substance use challenges) and address the challenges and indeed meaning of measuring intelligence. Updating the evidence base without shifting to the relevant more public health question presents communications and policy risks that might actually decrease overall health.

Response: Agree (no change)

- We agree that the question of whether exposure to fluoride at any level can influence cognitive and neurobehavioral development is not new. The prepublication 2022 NTP Monograph and meta-analysis manuscript point out the evolving concern over this issue by first describing a prior 2006 review of this question by a committee convened by the National Academy of Sciences (NRC 2006). The monograph and meta-analysis manuscript go on to describe and further evaluate the rapidly expanding database of human epidemiological studies with improved quality and precision. The prepublication 2022 *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* provides the most comprehensive assessment of this literature to date and explains the reasoning behind our determination of moderate confidence in the evidence base for an association between higher fluoride exposures and lower IQ in children.
- However, we do not agree that, prior to this assessment, federal research had affirmed, or in fact even formally examined, the question of whether fluoride exposure could lead to decrements in cognition or neurodevelopment.

- We do agree that a federal effort to examine the overall cost-benefit (or risk-benefit) of current fluoride exposure and oral health is an appropriate next step, and there is a precedent for this. In 2010, accumulating evidence of increasing prevalence of dental fluorosis from the CDC National Survey of Oral Health in U.S. School Children, and an earlier CDC Division of Oral Health publication estimating the attributable risk of water fluoridation to dental fluorosis among children (Griffin et al. 2002), prompted the Public Health Service to convene a panel to suggest a strategy to reverse this trend. The committee met multiple times over several months and ultimately proposed to decrease the recommended level of fluoride added to community water systems from a range of 0.7 to 1.2 mg/L (depending on ambient temperature) to a single consistent level of 0.7 mg/L (HHS 2015). Note, that the 2002 CDC publication (Griffin et al. 2002) that was used as the genesis for the Public Health Service panel, did not attempt to examine the cost-benefit relationship between reducing fluorosis and concomitantly diminishing oral health. Rather, the 2002 CDC publication pointed out and quantified the problem, as do our prepublication 2022 NTP Monograph and meta-analysis manuscript, respectively.
- The meta-analysis was a recommendation of the NASEM Committee that evaluated an earlier (2019) draft of the NTP Monograph. In a subsequent review, the Committee provided constructive criticisms of the meta-analysis that we performed. The current version of the meta-analysis manuscript has been revised in response to NASEM Committee suggestions and provides a quantitative estimate of risk. In addition to the prepublication 2022 NTP Monograph, the results of our meta-analyses would be a necessary component of a comprehensive effort to quantify risks in any larger public health risk-benefit evaluation of fluoride. Furthermore, NIEHS/DNTP has provided comprehensive responses to the Committee's comments (see Sup01_Meta-analysis for responses to the NASEM Committee's comments on the meta-analysis) and considers the meta-analysis manuscript is ready for, and appropriate for, submission for further peer review by a journal.

7.B:

2) (Methods) Others have commented on the quality of the research included in the meta-analysis, generalizability to the U.S. populous, and statistical methodology. [redacted] recognizes the value in these constructive criticisms.

Response: Agree (no change requested)

- Like [redacted], we recognize the value of constructive comments. Our responses to those comments have addressed the concerns that were raised, and the meta-analysis manuscript has been improved through responding to peer-review comments.

7.C:

3) (Casual inference) The strength of the work rests in strong grounding in toxicology science but yet the focus on the independent variable is out of balance with a parallel focus on the dependent one, intelligence. The science of measuring intelligence is vast and complex, and its meaning in living a healthier life, unclear. That perspective is absent from the question structure and thin in the concluding interpretation. Although the authors make a case with references 46-51, the larger net effect given issues of oral health are not incorporated, nor seem to be a

principal motivation of prospective question formation, critical in population-based studies. This approach risks being a one-sided toxicology story without the balance of the harm from decreased oral health.

Response: Disagree (no change)

- While we recognize that the science of measuring intelligence is complex, the field has evolved to become more standardized in many respects (see NIEHS 2022), and psychometric test results have played increasingly important roles in the regulation of environmental neurotoxins such as methylmercury (EPA 2001). We also recognize that further examination of the relationship between cognitive and oral health effects related to fluoride would be valuable, as we are unaware of any population-based study that has attempted to assess both the benefits of decreasing fluoride exposures on improved cognition and the concomitant potential risks to oral health. However, as indicated above, the results of our analyses would be necessary components of a comprehensive effort to quantify risks in any larger public health risk-benefit evaluation of fluoride. It is our view that the topic is of such high public health importance that the integration of our confidence assessment of the complete evidence base on increased fluoride exposure and neurodevelopmental and cognitive health effects with an assessment of the potential risk to oral health from decreased fluoride exposure would require a collective effort by the larger public health community that also considers the appropriate method and timing of population exposures to fluoride to benefit oral health.

7.D:

4) (Communication science) It is hard to definitively discern the applicable operative range of exposure levels and how they correspond to exposure in children in the United States. While there is mention of the issue in stating there are levels over 2 mg/L occurring through natural exposure, this not referenced, mapped, along with more local quantitative assessment of the burden in the United States in a fashion that readily enables an interpretive impact. The monograph makes clear this is a low percentage of people.

Response: Agree (change made)

- We agree that the lack of both U.S. studies of fluoride exposures in relation to children's cognition and the absence of publicly available U.S. data on total fluoride exposures for children make it difficult to directly apply our findings to fluoride exposures in the United States. The absence of U.S. studies is currently identified as a limitation in the meta-analysis manuscript. In response to this comment, we have updated the reference to CDC data and added a citation to the manuscript for a publication that maps fluoride concentrations in untreated groundwater from public supply and domestic wells (McMahon et al. 2020).

7.E:

5) (Limitations, bias) Methods do not clearly delineate the question history and approach to multiple comparisons (other measures of cognition, other age groups) which ultimately might undermine the integrity of the measured inference. This is an important limitation not described.

Response: Disagree (no change)

- We are unclear about what [redacted] is identifying as a limitation. Adjustments for multiple tests are not commonly used in meta-analyses and the Cochrane Handbook advises against their use (Section 16.7.2). We did not rely solely on p-values when describing results, and the prespecified subgroup analyses included stratification by intelligence assessment type. In addition, many of the studies in the meta-analysis specifically excluded children with obvious cognitive disabilities, and the findings with respect to IQ deficits were reported over an age range from as young as 3–4 years old to as old as 16–18 years old, suggesting that deficits persist. Measures of neurodevelopmental and cognitive effects other than IQ, such as ADHD behaviors, were evaluated in relation to fluoride exposure in the prepublication 2022 NTP Monograph. However, there was low confidence in the evidence of an association between these other effects and fluoride, suggesting that other measures of neurodevelopment and cognition were not responsible for the IQ findings.

7.F:

6) (Statistics, math) Is there an assessment and explanation of the variation in outcomes as it might affect the interpretation of the statistical measurement (drawing a line with a positive slope through a scattergram).

Response: Agree (no change)

- The authors of studies with individual-level measures of exposure and outcome frequently attempted to apply linear and non-linear regression models to determine the best fit to the scatterplots. In a meta-analysis, a funnel plot illustrating the effect estimate against the inverse of the standard error is equivalent to the scattergram suggested by [redacted]. Our *regression slopes meta-analysis* found that non-linear models did not provide a significantly better fit than linear models and, therefore, we elected to accept an assumption of linearity for the purposes of discussion. The overall pooled effect estimate was determined from studies that reported individual urine levels, which is considered a reasonable estimate of total exposure to fluoride from all sources. Presumably, this effect estimate would be more precise than the effect estimates derived from the individual studies.

7.G:

7) Each point begins with the type of concern raised listed in parentheses. The details of each point is one example of each type, where upon are at times others.

Response: Agree (no change requested)

- This comment describes the structure of the [redacted] comments in this file, which we have found helpful. We have responded to each comment above.

Note: The [redacted] comments on the prepublication 2022 NTP Monograph are not reproduced here as they are not relevant to the meta-analysis. See DocD_Monograph for the monograph-relevant comments and responses.

References

Griffin SO, Beltrán ED, Lockwood SA, Barker LK. 2002. Esthetically objectionable fluorosis attributable to water fluoridation. Community Dent Oral Epidemiol. 30(3):199-209. doi: 10.1034/j.1600-0528.2002.300306.x. PMID: 12000343.

McMahon PB, Brown CJ, Johnson TD, Belitz K, Lindsey BD. 2020. Fluoride occurrence in United States groundwater. Sci Total Environ. 732:139217. doi:https://doi.org/10.1016/j.scitotenv.2020.139217

National Research Council (NRC). 2006. Committee on fluoride in drinking water, board on environmental studies and toxicology. Fluoride in drinking water: A scientific review of EPA's standards.: National Research Council. http://www.nap.edu/catalog/11571/fluoride-in-drinking-water-a-scientific-review-of-epas-standards.

U.S. Department of Health and Human Services (HHS) Federal Panel on Community Water Fluoridation. 2015. U.S. Public Health Service Recommendation for Fluoride Concentration in Drinking Water for the Prevention of Dental Caries. Public Health Reports. July-August;130: 318-331.

U.S. Environmental Protection Agency (EPA). 2001. Methylmercury (MeHg); CASRN 22967-92-6: Chemical Assessment Summary. Retrieved from: https://iris.epa.gov/ChemicalLanding/&substance_nmbr=73

In November 2021, the [redacted] provided comments to NIEHS/DNTP on the 2021 *Draft NTP Monograph on the State of the Science concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* (the NTP Monograph) and a draft manuscript on a meta-analysis of fluoride exposure and IQ in children (the meta-analysis manuscript). NIEHS/DNTP prepared responses and shared those responses back to [redacted] in April 2022.

In July 2022, the [redacted] provided two sets of comments to NIEHS/DNTP, again on the NTP Monograph and the meta-analysis manuscript.

- The first set of [redacted] comments were provided as a new layer of input on top of the original [redacted] comments (from November 2021) and NIEHS/DNTP responses. This document contains a subset of the overall [redacted] comments (from November 2021 and July 2022) related to the meta-analysis manuscript along with the NIEHS/DNTP responses. The meta-analysis-related comments from the [redacted] are reproduced below in black text and the NIEHS/DNTP responses have been inserted in blue text following each of the comments beginning with the word "**Response**" in bold font.

- The second set of the [redacted] comments were provided in track changes embedded in the draft meta-analysis manuscript in a Microsoft Word document. The full text of [redacted] comments has been reproduced below verbatim in black text along with the specific sentence referred to by [redacted] as quotes under a heading for the specific section of the document (e.g., "Abstract section"). When the [redacted] comments were inserted on a particular word or phrase, that word or phrase is highlighted in grey within the quoted sentence. Again, the NIEHS/DNTP responses have been added in blue text following each of the comments beginning with the word "**Response**" in bold font.

Formatting has been applied to aid in reading

[redacted] comments from November 2021 and July 2022

Summary of [redacted] comments on the *"Draft NTP monograph on the state of the science concerning fluoride exposure and neurodevelopmental and cognitive health effects: a systematic review"* ("SoS document") and draft Taylor *et al. Association between fluoride exposure and children's intelligence: A systematic review and meta- analysis* manuscript ("meta-analysis document")

Note: [redacted] comments on the monograph are not reproduced here as they are not relevant to the meta-analysis. See DocB1_Monograph for the monograph-relevant comments and responses.

8.A:

- [redacted] **comment on SoS document (November 2021):** The revised NTP monograph seems to address concerns from prior comments as NTP removed the hazard assessment and is now calling this a "state of the science" document. However, the meta-analysis that NTP removed from the original monograph is now being published independently. Although it will be in a scientific review publication (JAMA pediatrics), [redacted] think that this may raise questions regarding exposure levels and neurodevelopmental effects, as the publication does not seem to put the exposure levels into context.

  Response: Disagree (no change)
  - We appreciate the need to provide context concerning fluoride exposure levels and neurodevelopmental effects and presume that this comment concerns fluoride

exposures in the United States. As the comment points out, this topic is more fully addressed in the NTP state of the science monograph, which is referenced in the meta-analysis, and we have added reference to the U.S. Public Health Service recommendations for optimal water fluoridation in the meta-analysis manuscript; however, we also stress that the subject of our fluoride monograph and meta-analysis is total fluoride exposures from all sources. The November 2022 literature search update of the meta-analysis includes a number of new non-U.S. studies that further inform the relationship between IQ deficits in children and exposures to fluoride that were not available for inclusion in the 2020 draft NTP monograph. These studies have provided additional information to sharpen the dose-response mean-effects estimates and improve the *regression slopes meta-analysis*. Although the clarity of effects at lower fluoride exposures, which are presumed to be applicable to exposures in the United States, is improving, providing further context is speculative because there are no studies of the potential association between fluoride exposures and IQ in children in the United States, and nationally representative urinary fluoride levels are not available. These facts make it difficult to make more specific statements about the relevance of our meta-analysis findings to the U.S. population.

8.B:

- [redacted] **comment on SoS document (July 2022):** The systematic review finds, with moderate confidence, that higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] is consistently associated with lower IQ in children, and that more studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ. In this regard:

  a) What is the overall confidence in the conclusions of the meta-analysis? No study was excluded from the meta-analysis based on concerns for risk of bias – how does this affect the overall confidence in the conclusions?

  Response: No change requested

  - The meta-analysis itself does not have confidence conclusions, but the finding of moderate confidence in the body of evidence reviewed in the prepublication 2022 NTP Monograph has been added to the *Discussion* section of the revised meta-analysis manuscript as follows:

    "This meta-analysis complements a larger systematic review[8] that concluded moderate confidence in the body of evidence that fluoride exposure is associated with lower IQ in children."

  - While the meta-analysis includes more studies than the prepublication 2022 NTP Monograph, which resulted from a literature search update for the meta-analysis in November 2021, our review of these additional studies has given us no reason to believe that they would either increase or decrease our confidence in that body of evidence.

8.C:

b) For the dose-response mean-effects meta-analysis and regression slopes meta-analysis, were subgroup analyses stratified by risk of bias (high or low), study location (e.g., country), outcome assessment, exposure matrix (e.g., urine or water), pre- or post-natal exposure, sex, and age group conducted? If not, is there a reason why?

Response: No change requested

- The *dose-response meta-analysis* was stratified by risk of bias and exposure level as was pre-specified in the protocol. Because the purpose of the subgroup analyses was to explore sources of heterogeneity, and the *dose-response meta-analysis* included many of the same studies included in the *mean-effects meta-analysis*, there was no reason to add further subgroup analyses post-hoc. The *regression slopes meta-analysis* was stratified by risk of bias, exposure type, country, outcome assessment type, sex, and pre- and post-natal exposure.

8.D:

- [redacted] **comment on SoS and meta-analysis documents (July 2022):** [redacted] raised concerns regarding exposure measurement in previous comments. The current Discussion sections in each document cover some exposure measurement limitations but may not sufficiently address [redacted] previous comments or other important issues potentially impacting individual and group urinary fluoride measurement, such as variation in period of urine collection, variations/transient increases in excretion, variations in clearance times, as well as total fluoride exposure by age, sex, developmental stage, and over time.

  Note: The above comment refers to previous [redacted] concerns regarding exposure measurement that were focused on the monograph and are, therefore, not reproduced here. See DocB1_Monograph for the monograph-relevant comments and responses. The meta-analysis-relevant response is provided immediately below.

  Response: Disagree (edited for clarity)

  - In responses to earlier comments from [redacted] and others, we have pointed out reasons we consider these concerns are overstated and speculative. We have addressed exposure measurements as part of the evaluations in both documents. These include our requirement for creatinine or specific gravity adjustments for measurements of urinary fluoride to be considered lower risk of bias for exposure. We also cite studies reporting reasonable agreements between 24-hour and repeated volume corrected spot urine fluorides in the monograph. We also would point out that to account for the consistent direction of effect of an inverse relationship between fluoride in urine and children's IQ would require that one or all of the cited factors would need to affect children's IQ, as well as produce the speculated spurious correlated fluoride measurements. We are happy to entertain such evidence if [redacted] wish to provide.
  - However, we acknowledge that the type and timing of urinary sample collection is important to consider and have extended the *Discussion* section of the meta-analysis to acknowledge concerns related to the issues associated with individual and group urinary fluoride measurement:

"Another limitation of the mean-effects meta-analyses is that exposure values are assumed to be the same for each child in an exposure group, either because the study used a community-level water fluoride measure or a median, mean, or midpoint in water or urine as the exposure value. Fluoride exposure may vary considerably depending on individual behaviors and is best captured by individual-level measures of total exposure, such as urinary fluoride measures. Because drinking water measures capture only some of a person's total exposure to fluoride, it is reasonable to assume that some children in the meta-analysis had higher exposure to fluoride and those children may have skewed the mean IQ deficits of the entire group. Urinary fluoride levels include all ingested fluoride and are considered a valid measure to estimate total fluoride exposure.[61, 62] When compared with 24-hour urine samples, spot urine samples are more prone to the influence of timing of exposure (e.g., when water was last consumed, when teeth were last brushed) and can also be affected by differences in dilution. However, correlations between urinary fluoride concentrations from 24-hour samples and spot samples adjusted for urinary dilution have been described,[63] and with one exception[35] all studies in the regression slopes meta-analysis, accounted for dilution."

8.E:

- [redacted] **comment on meta-analysis document (July 2022):** Given that in the Results section, heterogeneity was evaluated and found to be high, suggest that the Discussion section should address those findings with some coverage of potential sources of high heterogeneity. This would be consistent with the objectives outlined in the cited protocol [*National Toxicology Program (NTP). Protocol for systematic review of effects of fluoride exposure on neurodevelopment. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health; 2020*], where it is indicated that one of the specific aims of the meta-analysis is to "[s]ynthesize the evidence across studies that assessed learning and memory using a narrative approach or meta-analysis (if appropriate) and evaluate sources of heterogeneity."

  Response: Agree (change made)

  - In the *Discussion* section of the manuscript, we added the following new text with further interpretations of the subgroup analyses as they relate to potential sources of heterogeneity:

    "With a couple exceptions, the subgroup analyses in the mean-effects meta-analysis did not explain a large amount of the overall heterogeneity. However, the heterogeneity in the regression slopes meta-analysis was explained by subgroup analyses. This suggests that the aggregate nature of the mean-effects meta-analysis might not be sufficiently sensitive to capture potential sources of heterogeneity, as seen possible when using studies with individual-level data in the regression slopes meta-analysis. However, the large number of studies included in the mean-effects meta-analysis and the consistency in the direction of the association across the analyses make this is less of a concern."

  - We also further investigated potential sources of heterogeneity by conducting a meta-regression analysis using mean age in years and year of publication in each study. In the supplemental material we added:

"The results of the meta-regression models indicate that year of publication and mean age of study children did not explain a large degree of heterogeneity as neither were significant predictors of the relationship between fluoride and children's intelligence, and the residual I^2 remained high (85% and 87%, respectively). Year of publication (SMD = 0.01, 95% CI: -0.01, 0.02) and mean age (SMD = -0.04, 95% CI: -0.13, 0.04) explained relatively little between-study variance (adjusted R^2 of 12% and 5%, respectively). When both year of publication and mean age were included in the model, there were no notable improvements to the amount of between-study variance explained (adjusted R^2 = 13%) or percent residual variation due to heterogeneity (residual I^2 = 85%).

Excluding the outlier study34 resulted in a slightly lower heterogeneity for the overall effect estimate (I^2=84%) and for the India-specific effect estimate (I^2=69%). The meta-regression indicates that mean age is a significant predictor of the effect (SMD = -0.06, 95% CI: −0.12, −0.01, p-value =0.025), explaining 9% of the between-study variance. Year of publication (SMD = 0.01, 95% CI: 0.001, 0.02, p-value=0.028) explained a larger degree of between-study variance (R^2 = 19 %)."

[redacted] comments embedded in the Microsoft Word meta-analysis document July 2022

[redacted] *general comments*

8.F: Abstract Section: "Water and water-based beverages are the main source of systemic fluoride intake; however, an individual's total exposure to fluoride also reflects contributions from other sources such as food, dental products, industrial emissions, and some pharmaceuticals and pesticides."

[redacted] **comment:** Although this statement is true, it may relay a misleading impression that the authors also measured these other sources of exposure in the study. The authors may want to check in the pertinent EPA documentation whether this [pesticides] is a significant source of exposure worth mentioning here since sulfuryl fluoride is a fumigant and dissipates rapidly. [**Note**: the text in brackets has been added by NIEHS/DNTP to clarify [redacted] comment.]

Response: Agree (change made)

- We agree that sulfuryl fluoride is a minor contributor to a person's total fluoride exposure and have removed "pesticides" from the list of contributors to fluoride exposure in the manuscript.

8.G: *Abstract section:* (the underlined text was inserted by [redacted]) **"**To perform a systematic review and meta-analysis to investigate associations between fluoride exposure, based primarily on urinary and water fluoride levels, and children's intelligence."

[redacted] **comment:** [redacted] suggested [redacted] add "based primarily on urinary and water fluoride levels" to the sentence.

Response: Disagree (no change)

- The objective describes the intent of the systematic review, which was to include studies that captured any source of fluoride exposure, not just from urinary and water

levels. However, our results make it clear that urinary and water fluoride levels were the main measures of fluoride exposure in the relevant studies.

8.H: ***Results section:*** *"*For studies that had more than one exposed group (n = 17), a sensitivity analysis was performed to evaluate the impact of combining all exposed groups and comparing them to the reference group."

[redacted] **comment:** A decision was made to use the highest exposed groups in comparison to the reference group. Was a sensitivity analysis performed on this decision?

Response: Agree (edited for clarity)

- We have revised this sentence to clarify that the sensitivity analysis to evaluate the impact of combining all exposed groups to compare them to the reference group was applied to all studies in the *mean-effects meta-analysis* and not limited to these 17 studies. We clarified this aspect in the Results section as follows:

  "The sensitivity analysis to evaluate the impact of combining all exposed groups and comparing them to the reference group did not appreciably change the effect estimates".

- However, we are unclear if [redacted] is suggesting an additional sensitivity analysis using the second highest exposed group compared to the reference. If so, because the number of studies that have a second highest exposed group (n=17) is the same as the sensitivity analysis combining all exposed groups, it is unlikely such an analysis would provide further value.

8.I: ***Results section:*** *"*The dose-response mean-effects meta-analysis combining data from 29 studies with group-level fluoride measurements in drinking water (23 high risk-of-bias and 6 low risk-of-bias studies) and 17 studies with children's group-level mean urinary fluoride levels (9 high risk-of-bias and 8 low risk-of-bias studies) show statistically significantly lower children's IQ scores with increasing fluoride exposures."

[redacted] **comment:** Is the small number of low risk of bias studies [in the drinking water dose-response meta-analysis] of concern? [**Note**: the text in brackets has been added by NIEHS/DNTP to clarify [redacted] comment.]

Response: Disagree (no change)

- We are unclear on the exact concern to which the comment is referring. The results of the *dose-response meta-analysis* are presented in the supplemental material and are not emphasized in the main manuscript. However, the six low risk-of-bias studies in the drinking water fluoride analysis includes 4,355 children and the nine low risk of bias studies in the urinary fluoride analysis includes 5,713 children. For perspective one might consider the NHANES assessments. The annual sample size for NHANES is 5,000 people (all ages) and is considered a representative sample of the United States population. The number of participants who provide biomonitoring samples is about 1/3 of that total, so it is recommended at least 4 years of data (2 NHANES cycles) be

combined to obtain a sample size with an acceptable level of reliability for most of the sampling domains. This usually works out to ~3-5k participants.

8.J: ***Results section:*** "Adjusting for possible publication bias through trim-and-fill analysis supports the conclusion that a 1-mg/L increase in urinary fluoride was associated with lower IQ, with an adjusted pooled effect estimate of −0.87 (95% CI: −1.93, 0.19; p- value = 0.302) (**eFigure 22**). The results for fluoride intake and water fluoride levels are available in **Supplemental Materials**."

[redacted] **comment:** How significant is this change [1-mg/L increase in urinary fluoride] in relation to "normal" urinary fluoride levels? Indicating what these levels are would help contextualizing this conclusion. [**Note**: the text in brackets has been added by NIEHS/DNTP to clarify [redacted] comment.]

Response: Disagree (no change)

- What constitutes "normal" urinary fluoride levels depends entirely on the population being examined. However, using data from Green et al. (2019), the difference between water fluoride concentration in a fluoridated area v. non-fluoridated area is roughly half of 1mg/L.
- Note: After updating the *regression slopes meta-analysis* with new studies from the updated literature search, there was no longer evidence of publication bias, so the specific quoted text related to the adjusted pooled effect estimate has been removed from the manuscript.

[redacted] *comments on defining "higher":*

8.K: ***Abstract section:*** "The meta-analysis of 46 studies (N = 15,538 children) with group-level exposures found that children exposed to higher fluoride levels had lower mean IQ scores (pooled SMD: −0.49; 95% CI:−0.60, −0.38; p-value < 0.001)."

[redacted] **comment:** Please specify the meaning of "higher". For example, "greater than XX mg/mL"

Response: Agree (change made)

- The *mean-effects meta-analysis* pooled results from individual studies that compared mean IQ in "higher" vs. "lower" fluoride areas. Each individual study had its own definition of what exposure level constituted "higher" and "lower", so the data do not support defining a threshold for "higher" in the pooled SMD of the *mean-effects meta-analysis*. To clarify that "higher" exposure is simply being used relative to "lower", we have revised the quoted sentence as follows (please note the numbers have changed due to a literature search update):

  "The meta-analysis of 55 studies (N = 18,845 children) with group-level exposures found that, when compared to children exposed to lower fluoride levels, children exposed to higher fluoride levels had lower mean IQ scores (pooled SMD: −0.46; 95% CI: −0.55, −0.37; p-value < 0.001)."

- For transparency, Table 1 (excerpt below) includes the exposure levels that were compared in the *mean-effects meta-analysis*. Please note that we did explore lower levels of fluoride exposure in the *dose-response meta-analysis* (see supplemental materials).

Excerpt of Table 1. Characteristics of Studies Included in the Meta-analysis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure: Assessment (Metric, Exposure Groups)	Fluoride Exposure: Levels	Intelligence Assessment	Overall RoB Rating	Confounders Considered
Ren et al. (1989)[66] [translated in Ren et al. 2008][me, o] *Cross-sectional*	China	8–14	No fluoride measurement Low iodine village/high fluoride and low iodine village	Not specified	Wechsler Intelligence Scale for Children	High	Sex; iodine
Chen et al. (1991)[68] [translated in Chen et al. 2008][me, w] *Cross-sectional*	China	7–14	Drinking water Nonendemic/endemic fluorosis village	0.89 mg/L (nonendemic) 4.55 mg/L (endemic)	Chinese Standardized Raven Test	High	Age; sex
Guo et al. (1991)[70] [translated in Guo et al. 2008a][me, o] *Cross-sectional*	China	7–13	Serum Reference area using wood/coal burning-related fluoride endemic area	0.1044 ± 0.0652 mg/L (reference) 0.1483 ± 0.0473 mg/L (endemic)	Chinese Binet Intelligence Test	High	Age; sex; SES

8.L: ***Abstract section:*** (the underlined text was inserted by [redacted] and the strikethrough text was deleted by [redacted]) "Our meta-analysis confirms results of previous meta-analyses and extends them by including newer, more precise studies with individual-level exposure measures to assess associations between high fluoride exposure (e.g., >1.5 mg/L, the World Health Organization Guideline for Drinking-water Quality) and lower IQ levels of children. Associations between lower fluoride exposure (e.g., < 1.5 mg/L) and children's IQ remain uncertain. ~~The data support a consistent inverse association between fluoride exposure and children's IQ~~. Additional prospective cohort studies with individual urinary fluoride measures, along with studies conducted in the United States, would increase the confidence in this body of evidence."

[redacted] **comment:** The term "high" is used throughout the manuscript to characterize exposure and should be defined.

Response: Disagree (no change)

- [redacted] has suggested we define "high" as greater than 1.5mg/L and "low" as less than 1.5 mg/L and references 1.5 mg/L, the WHO Guideline for fluoride in drinking water. In the prepublication 2022 NTP Monograph we used this description of "higher" because, in that qualitative assessment of the epidemiology literature, the WHO Guideline represented a useful total fluoride exposure equivalent metric. However, in the meta-analysis, we were able to explore lower exposure levels by limiting the dose-response analyses to include study groups where exposure levels were equal to or lower than the U.S. Environmental Protection Agency drinking water standards (i.e., <4mg/L and <2mg/L)[20] and World Health Organization drinking water guidelines (<1.5mg/L).

8.M: ***Results section:*** *"The meta-analysis of 46 studies (37 high risk-of-bias studies and 9 low risk-of-bias studies) that provided mean IQ scores shows that children exposed to higher fluoride levels had statistically significantly lower IQ scores (random-effects pooled SMD, −0.49; 95% CI: −0.60, −0.38; p-value < 0.001) (**Figure 2**)."*

[redacted] **comment:** Please define ["higher"] [**Note**: the text in brackets has been added by NIEHS/DNTP to clarify [redacted] comment.]

Response: Agree (change made)

- To clarify that "higher" exposure is simply being used relative to "lower", we have revised the quoted sentence as follows (please note the numbers have changed due to a literature search update):

  "The meta-analysis of 55 studies (45 high risk-of-bias studies and 10 low risk-of-bias studies) that provided mean IQ scores shows that, when compared to children exposed to lower levels of fluoride, children exposed to higher fluoride levels had statistically significantly lower IQ scores (random-effects pooled SMD,¬ −0.46; 95% CI: −0.55, −0.37; p value < 0.001) (Table 2, Figure 2)."

8.N: ***Discussion section:*** *"The results of our mean-effects meta-analysis are consistent with two previous meta-analyses that reported statistically significantly lower IQ scores in children exposed to higher fluoride levels (p < 0.001) (**Table 2**)."*

[redacted] **comment:** Please define ["higher"] [**Note**: the text in brackets has been added by NIEHS/DNTP to clarify [redacted] comment.]

Response: Agree (change made)

- To clarify that "higher" exposure is simply being used relative to "lower", we have revised the quoted sentence as follows (please note the numbers have changed due to a literature search update):

  "The results of the mean-effects meta-analysis are consistent with two previous meta-analyses that, when comparing children exposed to lower fluoride levels, reported statistically significantly lower IQ scores in children exposed to higher fluoride levels (p < 0.001) (Table 2)."

8.O: ***Discussion section:*** *"Therefore*, the data support a consistent inverse association between fluoride exposure and children's IQ."

[redacted] **comment:** At all levels or only documented at certain levels? It is important to contextualize this statement.

Response: Disagree (no change)

- Based on previous comments, our interpretation of this comment is that [redacted] is recommending we contextualize the sentence with a threshold (e.g., >1.5mg/L) which is why we disagree with the comment. To answer [redacted] question, the meta-analysis includes fluoride exposures at all levels, some of which were below 1.5mg/L. Therefore, the evidence does not support excluding lower levels from this statement.

8.P: ***Discussion section:*** *"Although the estimated decreases in IQ may seem small, research on other neurotoxicants has shown that subtle shifts in IQ at the population level can have a profound*

impact on the number of people who fall within the high and low ranges of the population's IQ distribution."

[redacted] **comment:** Does this imply that fluoride causes a shift in intelligence at all levels of exposure (e.g., including at 0.7 mg/L)? If that is not the intent, this passage could be misleading.

Response: Disagree (no change)

- We do not consider this statement to be misleading. Using [redacted] example, total fluoride exposure among individuals living in optimally fluoridated areas (0.7mg/L in drinking water) may be higher than 0.7mg/L, dependent on personal behaviors and habits. We discuss the potential for this type of variation in the manuscript.

[redacted] *comments or questions on defining a threshold:*

8.Q: ***Abstract section:*** "*The meta-analysis of the association between individual-level measures of fluoride and children's IQ found a decrease of 1.58 IQ points (95% CI: −2.63, −0.53; p-value = 0.003) per 1-mg/L increase in urinary fluoride*."

[redacted] **comment:** Was there a threshold for this effect?

Response: No change requested

- Because we used a linear model, there is no threshold for this effect.
- Note: After adding new studies the sentence has been updated:

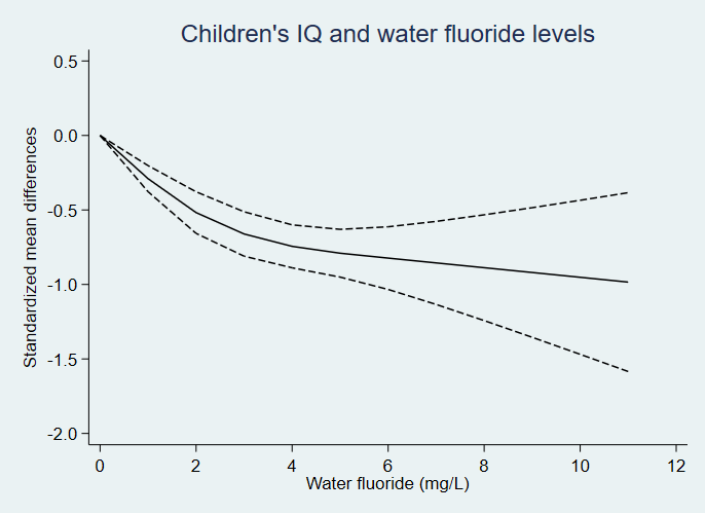
  "The meta-analysis of studies that reported individual-level measures of fluoride and children's IQ scores found a decrease of 1.81 points (95% CI: −2.80, −0.81; p-value < 0.001) per 1-mg/L increase in urinary fluoride."

8.R: ***Discussion section:*** "There is also evidence of a dose- response relationship between lower children's IQ and higher fluoride exposures."

[redacted] **comment:** Was a threshold for such relationship considered?

Response: No change requested

- As previously mentioned, results for the dose-response relationship restricted to lower fluoride exposure levels (i.e., <4mg/L and <2mg/L, <1.5mg/L) in both drinking water and urine are reported in the supplemental materials.
- The restricted cubic splines model for water fit slightly better than the linear model, however there was no obvious threshold as illustrated by the figure at either of the modelled knots.



eFigure 17. Pooled Dose-Response Association Between Fluoride in Water and Standardized Mean Differences in Children's IQ

eFigure 17 note: Water fluoride levels were modeled with quadratic restricted cubic splines terms in a random-effects model (solid line). Dashed lines represent the 95 % confidence intervals for the quadratic spline model.

8.S: ***Discussion section:*** *"Associations for drinking water appeared to be non-linear and associations for urine appeared to be linear. The Duan et al.[4] meta-analysis reported a significant non-linear dose-response relationship above 3 ppm [3 mg/L] in water."*

[redacted] **comment:** Was there a threshold for water?

Response: No change requested

- Results for the dose-response relationship restricted to lower fluoride levels (i.e., <4mg/L and <2mg/L, <1.5mg/L fluoride in drinking water) are reported in the supplemental materials.
- As described in the previous response, there was no obvious threshold for water.

8.T: ***Discussion section:*** *(underlined text inserted by* [redacted]*)* "However, among the low risk-of-bias cross-sectional studies, most provided information to indicate that exposure preceded the outcome (e.g., only including children who had lived in the area since birth, children had dental fluorosis, linked to fluoride levels greater than XX)."

[redacted] **comment:** [redacted] recommended adding "linked to fluoride levels greater than XX".

Response: Disagree (no change)

- The "e.g.," of this sentence is meant to provide examples for how cross-sectional studies provided information that establishes temporality and is not linked to any fluoride level.

8.U: ***Discussion section:*** *"*In addition, there is inconsistency in which model is the best fit at lower exposure levels (**eTable 4** and **eTable 5**) leading to uncertainty in the shape of the dose-response curve at these levels."

[redacted] **comment:** Was a threshold considered?

Response: No change requested

- See above responses concerning the various models explored for best model fit.

Supplemental Materials

NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review

NTP Monograph 08

May 2022

National Toxicology Program
Public Health Service
U.S. Department of Health and Human Services
ISSN: 2378-5144

Research Triangle Park, North Carolina, USA

Foreword

The National Toxicology Program (NTP), established in 1978, is an interagency collaboration within the Public Health Service of the U.S. Department of Health and Human Services. Its activities are executed through a partnership of the National Institute for Occupational Safety and Health (part of the Centers for Disease Control and Prevention), the Food and Drug Administration (primarily at the National Center for Toxicological Research), and the National Institute of Environmental Health Sciences (part of the National Institutes of Health), where this virtual program is administratively located. NTP's work focuses on the testing, research, and analysis of agents of concern to identify toxic and biological effects, provide information that strengthens the science base, and inform decisions by health regulatory and research agencies to safeguard public health. NTP also works to develop and apply new and improved methods and approaches that advance toxicology and better assess health effects from environmental exposures.

Literature-based evaluations are one means by which NTP assesses whether exposure to environmental substances (e.g., chemicals, physical agents, and mixtures) may be associated with adverse health effects. These evaluations result in hazard conclusions or characterize the extent of the evidence and are published in the NTP Monograph series, which began in 2011. NTP monographs serve as an environmental health resource to provide information that can be used to make informed decisions about whether exposure to a substance may be of concern for human health.

These health effects evaluations follow prespecified protocols that apply the general methods outlined in the "Handbook for Conducting a Literature-Based Health Assessment Using the OHAT Approach for Systematic Review and Evidence Integration."[†] The protocol describes project-specific procedures tailored to each systematic review in a process that facilitates evaluation and integration of scientific evidence from published human, experimental animal, and mechanistic studies.

Systematic review procedures are not algorithms, and the methods require scientific judgments. The key feature of the systematic review approach is the application of a transparent framework to document the evaluation methods and the basis for scientific judgments. This process includes steps to comprehensively search for studies, select relevant evidence, assess individual study quality, rate confidence in bodies of evidence across studies, and then integrate evidence to develop conclusions for the specific research question. Draft monographs undergo external peer review prior to being finalized and published.

NTP monographs are available free of charge on the NTP website and cataloged in PubMed, a free resource developed and maintained by the National Library of Medicine (part of the National Institutes of Health). Data for these evaluations are included in the Health Assessment and Workspace Collaborative.

For questions about the monographs, please email NTP or call 984-287-3211.

[†]OHAT is the abbreviation for Office of Health Assessment and Translation, which has become the Health Assessment and Translation group in the Integrative Health Assessment Branch of the Division of the National Toxicology Program at the National Institute of Environmental Health Sciences.

Table of Contents

Tables

Figures

About This Review

National Toxicology Program[1]

[1]Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Collaborators

Kyla W. Taylor, John R. Bucher, Robyn B. Blain, Christopher A. Sibrizzi, Pamela A. Hartman, Kristen Magnuson, Sorina E. Eftim, and Andrew A. Rooney

Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
Developed protocol and conducted systematic review through oversight and review of literature search, screening, data extraction, risk-of-bias assessment, and meta-analysis; wrote and edited monograph
Kyla W. Taylor, Ph.D., Project Lead
John R. Bucher, Ph.D.
Andrew A. Rooney, Ph.D.

ICF, Fairfax, Virginia, USA
Conducted literature screening, data extraction, risk-of-bias assessment, and meta-analysis; reviewed data, results, and analyses; wrote, edited, and formatted monograph; provided database and HAWC support
Robyn B. Blain, Ph.D.
Sorina E. Eftim, Ph.D.
Pamela A. Hartman, M.E.M.
Kristen Magnuson, M.E.S.M., Deputy Work Assignment Manager
Christopher A. Sibrizzi, M.P.H., Lead Work Assignment Manager

Contributors

Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
Provided oversight for external peer review
Mary S. Wolfe, Ph.D.

Provided literature search, screening, Distiller and HAWC support
Vickie R. Walker, B.S.

Reviewed meta-analysis protocol
Suril Mehta, Dr.P.H.

Conducted a technical review of the draft monograph
Freya Kamel, Ph.D. (Retired)

Conducted literature screening
Kris Thayer, Ph.D. (currently Chemical and Pollution Assessment Division, United States Environmental Protection Agency, Washington, D.C., USA)

Division of Extramural Research and Training, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
Conducted literature screening
Heather King, Ph.D.

Kelly Government Services, Research Triangle Park, North Carolina, USA
Supported external peer review
Elizabeth A. Maull, Ph.D. (retired from NIEHS, Research Triangle Park, North Carolina, USA)

ICF, Fairfax, Virginia, USA
Provided contract oversight
David F. Burch, M.E.M., Principal Investigator
Cara Henning, Ph.D.
Jessica A. Wignall, M.S.P.H.

Reviewed data, results, and analyses
William Mendez, Ph.D.

Reviewed data, results, and analyses, conducted literature screening and risk-of-bias assessment, and provided database and HAWC support
Anna Engstrom, Ph.D.

Provided database and HAWC support
Steven Black, M.P.H.
Natalie Blanton, M.P.H.
Kaitlin Geary, B.S.
Courtney Lemeris, B.A.
Rachel McGill, B.S.
Daniel Peters, B.A.
Alessandria Schumacher, B.A.
Anna Stamatogiannakis, B.S.
River Williams, B.S.

Conducted literature screening
Carlye A. Austin, Ph.D.
Ryan Cronk, Ph.D.
Lauren Fitzharris, M.P.H.
Samantha Goodman, M.S.
Angela Hensel, B.A.
Melinda Hoang, M.P.H.
Allison Killius, B.S.
Alex Kliminsky, M.S.
Shannon McGinnis, Ph.D.

Devon Morgan, M.S.
Joei Robertson, M.Sc.
Courtney Rosenthal, M.S.
Amanda Ross, M.Ed.
Karen E. Setty, Ph.D.
Kelly Shipkowski, Ph.D.
Raquel A. Silva, Ph.D.
Parnian Soleymani, M.S.

Conducted literature screening and data extraction
Katie Duke, Ph.D.
Susan Goldhaber, M.P.H.
Revathi Muralidharan, B.S.
Jennifer Seed, Ph.D.
Codi Sharp, M.P.H.

Conducted literature screening, data extraction, and risk-of-bias assessment
Johanna Rochester, Ph.D.

Conducted literature screening and risk-of-bias assessment
Pam Ross, M.S.P.H.

Conducted data extraction
Robert Shin, M.H.S.

Conducted data extraction and risk-of-bias assessment
Tao Hong, Ph.D.
Alex Lindahl, M.P.H.
Samantha J. Snow, Ph.D., DABT

Conducted risk-of-bias assessment
Louise Assem, Ph.D.
Kristin Bornstein, Ph.D.
Jennifer Hsieh, M.S.P.H.
Yi Lu, Ph.D.

Conducted risk-of-bias assessment and reviewed data to support meta-analysis
Alexandra Goldstone, M.P.H.

Reviewed data to support meta-analysis
Jonathan Cohen, Ph.D.
Hannah Eglinton, B.A.
Cynthia J. Lin, Ph.D.

Supported meta-analysis visualizations
Rebecca Gray, M.P.H.

Coordinated peer review
Canden N. Byrd, B.S.
Lindsey Green, M.P.H

Designed and executed literature searches, and managed references
Michelle Cawley, M.A., M.L.S.
Jeremy S. Frye, M.S.L.S.
Nicole Vetter, M.S.L.S.

Edited and formatted monograph
Penelope Kellar, M.S.
Whitney Mitchell, B.S.
Kevin O'Donovan, B.A.

Integrated Laboratory Systems, LLC, Research Triangle Park, North Carolina, USA
Conducted a technical review of the draft monograph
Cynthia J. Willson, D.V.M., Ph.D., DACVP

Boston University, Boston, Massachusetts, USA
Reviewed the systematic review protocol
Thomas Webster, Ph.D.

Brown University, Providence, Rhode Island, USA
Reviewed the systematic review protocol
Joseph Braun, Ph.D.

Columbia University, New York, New York, USA
Reviewed the systematic review protocol
Gail Wasserman, Ph.D.

Health Research Board, Dublin, Ireland
Reviewed the systematic review protocol
Marie Sutton, Ph.D.

University of Massachusetts, Amherst, Massachusetts, USA
Reviewed the systematic review protocol
Thomas Zoeller, Ph.D.

University of Colorado, Denver, Colorado, USA
Reviewed meta-analysis protocol
Tianjing Li, Ph.D.

Peer Review

The National Toxicology Program (NTP) conducted a peer review of the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review* by letter in December 2021. Reviewer selection and document review followed established NTP practices. The reviewers were charged to:

(1) Comment on the technical accuracy and whether the draft *NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* is clearly stated and objectively presented.

(2) Determine whether the scientific evidence supports the NTP's confidence ratings for the bodies of evidence regarding neurodevelopmental and cognitive health effects associated with exposure to fluoride.

NTP carefully considered reviewer comments in finalizing this monograph.

Peer Reviewers

Ethan Balk, M.D., M.P.H.
Associate Professor (Research)
Brown University School of Public Health
Providence, Rhode Island, USA

Pam Factor-Litvak, Ph.D.
Professor of Epidemiology
Columbia University Medical Center
New York, New York, USA

Erin Haynes, Dr.P.H.
Kurt W. Deuschle Professor in Preventive Medicine and Environmental Health
University of Kentucky College of Public Health
Lexington, Kentucky, USA

Julie Obbagy, Ph.D., RD
Nutritionist, Center for Nutrition Policy and Promotion
Office of Nutrition Guidance and Analysis
U.S. Department of Agriculture
Medfield, Massachusetts, USA

Heather Volk, Ph.D.
Associate Professor
Johns Hopkins Bloomberg School of Public Health
Baltimore, Maryland, USA

Publication Details

Publisher: National Toxicology Program

Publishing Location: Research Triangle Park, NC

ISSN: 2378-5144

DOI: https://doi.org/10.22427/NTP-MGRAPH-8

Report Series: NTP Monograph Series

Report Series Number: 08

Official citation: National Toxicology Program (NTP). 2022. NTP monograph on the state of the science concerning fluoride exposure and neurodevelopmental and cognitive health effects: a systematic review. Research Triangle Park, NC: National Toxicology Program. NTP Monograph 08.

Acknowledgements

This work was supported by the Intramural Research Program (ES103316, ES103317) at the National Institute of Environmental Health Sciences, National Institutes of Health and performed for the National Toxicology Program, Public Health Service, U.S. Department of Health and Human Services under contract GS00Q14OADU417 (Order No. HHSN273201600015U).

Conflict of Interest

Individuals who reviewed the systematic review protocol or meta-analysis protocol, conducted a technical review of the draft monograph, or served on the peer review panel have certified that they have no known real or apparent conflict of interest related to fluoride exposure or neurodevelopmental and cognitive health effects.

Abstract

Background: Fluoride is a common exposure in our environment that comes from a variety of sources and is widely promoted for its dental and overall oral health benefits. A 2006 evaluation by the National Research Council (NRC) found support for an association between consumption of high levels of naturally occurring fluoride in drinking water and adverse neurological effects in humans and recommended further investigation. The evidence reviewed at that time was from dental and skeletal fluorosis-endemic regions of China. Since the NRC evaluation, the number and location of studies examining cognitive and neurobehavioral effects of fluoride in humans have grown considerably, including several recent North American prospective cohort studies evaluating prenatal fluoride exposure.

In 2016, the National Toxicology Program (NTP) published a systematic review of the evidence from experimental animal studies on the effects of fluoride on learning and memory. That systematic review found a low-to-moderate level of evidence that deficits in learning and memory occur in non-human mammals exposed to fluoride.

Objective: To conduct a systematic review of the human, experimental animal, and mechanistic literature to evaluate the extent and quality of the evidence linking fluoride exposure to neurodevelopmental and cognitive effects in humans.

Method: A systematic review protocol was developed and utilized following the standardized OHAT systematic review approach for conducting literature-based health assessments. This monograph presents the current state of evidence associating fluoride exposure with neurocognitive or neurodevelopmental health effects and incorporated predefined assessments of study quality and confidence levels. Benefits of fluoride with respect to oral health are not addressed in this monograph.

Results: The current bodies of experimental animal studies and human mechanistic evidence do not provide clarity on the association between fluoride exposure and neurocognitive or neurodevelopmental human health effects.

This systematic review identified studies that assessed the association between fluoride exposure and cognitive or neurodevelopmental effects in both adults and children, which were evaluated separately. In adults, only two high-quality cross-sectional studies examining cognitive effects were available. The literature in children was more extensive and was separated into studies assessing intelligence quotient (IQ) and studies assessing other cognitive or neurodevelopmental outcomes. Eight of nine high-quality studies examining other cognitive or neurodevelopmental outcomes reported associations with fluoride exposure. Seventy-two studies assessed the association between fluoride exposure and IQ in children. Nineteen of those studies were considered to be high quality; of these, 18 reported an association between higher fluoride exposure and lower IQ in children. The 18 studies, which include 3 prospective cohort studies and 15 cross-sectional studies, were conducted in 5 different countries. Forty-six of the 53 low-quality studies in children also found evidence of an association between higher fluoride exposure and lower IQ in children.

Discussion: Existing animal studies provide little insight into the question of whether fluoride exposure affects IQ. In addition, studies that evaluated fluoride exposure and mechanistic data in humans were too heterogenous and limited in number to make any determination on biological plausibility. The body of evidence from studies in adults is also limited and provides low

confidence that fluoride exposure is associated with adverse effects on adult cognition. There is, however, a large body of evidence on IQ effects in children. There is also some evidence that fluoride exposure is associated with other neurodevelopmental and cognitive effects in children; although, because of the heterogeneity of the outcomes, there is low confidence in the literature for these other effects. This review finds, with moderate confidence, that higher fluoride exposure (e.g., represented by populations whose total fluoride exposure approximates or exceeds the World Health Organization Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride) is consistently associated with lower IQ in children. More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ.

Preface

The National Toxicology Program (NTP) conducted a systematic review of the published scientific literature because of public concern regarding the potential association between fluoride exposure and adverse neurodevelopmental and cognitive health effects.

NTP initially published a systematic review of the experimental animal literature in 2016 that was subsequently expanded to include human epidemiological studies, mechanistic studies, and newer experimental animal literature. Because of the high public interest in fluoride's benefits and potential risks, NTP asked the National Academies of Sciences, Engineering, and Medicine (NASEM) to conduct an independent evaluation of the draft *NTP Monograph on Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* (2019 draft monograph dated September 6, 2019) and the revised draft (2020 draft monograph dated September 16, 2020), which addressed the NASEM committee's recommendations for improvement. The NASEM committee determined that, "Overall the revised monograph seems to include a wealth of evidence and a number of evaluations that support its main conclusion, but the monograph falls short of providing a clear and convincing argument that supports its assessments…." Thus, NTP has removed the hazard assessment step and retitled this systematic review of fluoride exposure and neurodevelopmental and cognitive health effects as a "state-of-the-science" document to indicate the change. This state-of-the-science document does not include the meta-analysis of epidemiological studies or hazard conclusions found in previous draft monographs; however, it provides a comprehensive and current assessment of the scientific literature on fluoride as an important resource to inform safe and appropriate use.

NTP has responded to the NASEM committee's comments on the revised draft (September 16, 2020) in a separate document (placeholder for URL) and revised relevant sections of this monograph.

Introduction

Fluoride is a common exposure in our environment from a variety of sources and is widely promoted for its dental and overall oral health benefits. Approximately 67% of the U.S. population receives fluoridated water through a community water system (CDC 2013). In other countries, fluoride supplementation has been achieved by fluoridating food products such as salt or milk. Fluoride supplementation has been recommended to prevent bone fractures (Jones et al. 2005). Fluoride also can occur naturally in drinking water. Other sources of human exposure include other foods and beverages, industrial emissions, pharmaceuticals, and pesticides (e.g., cryolite, sulfuryl fluoride). Soil ingestion is another source of fluoride exposure in young children (US EPA 2010).

The U.S. Public Health Service (PHS) first recommended that communities add fluoride to drinking water in 1962. PHS guidance is advisory, not regulatory, which means that while PHS recommends community water fluoridation as a public health intervention, the decision to fluoridate water systems is made by state and local governments. For many years, most fluoridated community water systems used fluoride concentrations ranging from 0.8 to 1.2 milligrams/liter (mg/L) (US DHHS 2015). For community water systems that add fluoride, PHS now recommends a fluoride concentration of 0.7 mg/L (equal to 0.7 parts per million [ppm]). Under the Safe Drinking Water Act, the U.S. Environmental Protection Agency (EPA) sets maximum exposure level standards for drinking water quality. The current enforceable drinking water standard for fluoride, or the maximum contaminant level (MCL), is 4.0 mg/L. This level is the maximum amount of fluoride contamination (naturally occurring, not from water fluoridation) that is allowed in water from public water systems and is set to protect against increased risk of skeletal fluorosis, a condition characterized by pain and tenderness of the major joints. EPA also has a non-enforceable secondary drinking water standard of 2.0 mg/L of fluoride, which is recommended to protect children against the tooth discoloration and/or pitting that can be caused by severe dental fluorosis during the formative period prior to eruption of teeth. Although the secondary standard is not enforceable, EPA requires that public water systems notify the public if and when average fluoride levels exceed 2.0 mg/L (NRC 2006). The World Health Organization (WHO) set a safe water guideline of 1.5 mg/L of fluoride in drinking water (first established in 1984 and reaffirmed in 1993 and 2011), which is recommended to protect against increasing risk of dental and skeletal fluorosis (WHO 2017).

As of April 2020, 1.08% of persons living in the United States (~3.5 million people) were served by community water systems (CWS) containing ≥1.1 mg/L naturally occurring fluoride. CWS supplying water with ≥1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population (~1.9 million people), and systems supplying water with ≥2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people) (CDC Division of Oral Health 2020).

Commonly cited health concerns related to fluoride are bone fractures and skeletal fluorosis, lower intelligence quotient (IQ) and other neurological effects, cancer, and endocrine disruption.

Effects on neurological function, endocrine function (e.g., thyroid,[1] parathyroid, pineal), metabolic function (e.g., glucose metabolism), and carcinogenicity were assessed in the 2006 NRC report, *Fluoride in Drinking Water: A Scientific Review of EPA's Standards* (NRC 2006). The NRC review considered adverse effects of water fluoride, focusing on a range of concentrations (2–4 mg/L) above the current 0.7-mg/L recommendation for community water fluoridation. The NRC report concluded that the Maximum Contaminant Level Goal (MCLG), 4 mg/L, should be lowered to protect against severe enamel fluorosis and reduce the risk of bone fractures associated with skeletal fluorosis (NRC 2006). Other than severe fluorosis, NRC did not find sufficient evidence of negative health effects at fluoride levels below 4 mg/L; however, it concluded that the consistency of the results of IQ deficits in children exposed to fluoride at 2.5 to 4 mg/L in drinking water from a few epidemiological studies of Chinese populations appeared significant enough to warrant additional research on the effects of fluoride on intelligence. The NRC report noted several challenges to evaluating the literature, including deficiencies in reporting quality, lack of consideration of all sources of fluoride exposure, incomplete consideration of potential confounding, selection of inappropriate control subject populations in epidemiological studies, absence of demonstrated clinical significance of reported endocrine effects, and incomplete understanding of the biological relationship between histological, biochemical, and molecular alterations with behavioral effects.

In 2016, the National Toxicology Program (NTP) 2016, NTP published a systematic review of the evidence from experimental animal studies on the potential effects of fluoride exposure on learning and memory (NTP 2016). That systematic review found a low-to-moderate level of evidence that deficits in learning and memory occur in experimental animals exposed to fluoride. Given these findings, NTP decided to conduct additional animal studies before carrying out this full systematic review and integrate human, animal, and potentially relevant mechanistic evidence in order to reach human health hazard identification conclusions for fluoride and learning and memory effects. As the NTP (2016) report on the experimental animal evidence focused on learning and memory and developed confidence ratings for bodies of evidence by life stage of exposure (i.e., exposure during development or adulthood), this monograph also evaluates two different age groups in humans (i.e., children and adults) with a focus on cognitive neurodevelopmental effects in children and cognitive effects in adults in order to address potential differences in health impacts based on time frame of exposure (i.e., during development or during adulthood). The evaluation of experimental animal studies in this monograph has been conducted separately from the 2016 experimental animal assessment; however, like the 2016 assessment, it assessed mainly learning and memory effects in experimental animal studies to determine whether the findings inform the assessment of cognitive neurodevelopmental effects in children and cognitive effects in adults.

A committee convened by the National Academies of Sciences, Engineering, and Medicine (NASEM) reviewed earlier drafts of this monograph (September 6, 2019, and September 16, 2020) (NASEM 2020; 2021). The current document incorporates changes stemming from those reviews, and responses to the 2020 review are available at (placeholder to cite NTP 2021

[1]The current review has evaluated the fluoride literature with an eye toward potential thyroid effects because a large literature base has accumulated examining the interaction of fluoride with iodine uptake by the thyroid gland and consequential effects on synthesis of thyroid hormones, which are recognized to play significant roles in neurodevelopment in utero and during early childhood. This literature, along with a detailed proposed mechanism of action, was recently reviewed by Waugh (2019).

Response to NASEM comments). See Appendix B, Table B-1 for a timeline of key activities contributing to this 2022 NTP monograph, including document review activities that have occurred since 2016.

Objective and Specific Aims

Objective

The overall objective of this evaluation was to undertake a systematic review to develop NTP human health hazard identification conclusions on the association between exposure to fluoride and neurodevelopmental and cognitive effects based on assessing levels of evidence from human and non-human animal studies with consideration of the degree of support from mechanistic data. However, the NASEM Committee's reviews (NASEM 2020; 2021) of the 2019 and 2020 drafts of the monograph indicated that, "Overall the revised monograph seems to include a wealth of evidence and a number of evaluations that support its main conclusion, but the monograph falls short of providing a clear and convincing argument that supports its assessments…." For this reason, our methods were revised to remove the hazard assessment step (i.e., the section "Integrate Evidence to Develop Hazard Identification Conclusions" and the associated section "Translate Confidence Ratings into Level of Evidence for Health Effect"). In addition, a meta-analysis of the epidemiological studies examining children's IQ in relation to fluoride exposure added to the 2020 draft in response to NASEM comments (NASEM 2020) will be published separately and is not part of this document.

Therefore, the objective of this monograph is to undertake a systematic review of the literature concerning the association between fluoride exposure and neurodevelopmental and cognitive effects and to determine the level of confidence in that evidence. The assessment was based on evidence from human and non-human animal studies with consideration of mechanistic information.

Specific Aims

- Identify literature that assessed neurodevelopmental and cognitive health effects, especially outcomes related to learning, memory, and intelligence, following exposure to fluoride in human, animal, and relevant in vitro/mechanistic studies.
- Extract data on potential neurodevelopmental and cognitive health effects from relevant studies.
- Assess the internal validity (risk of bias) of individual studies using pre-defined criteria.
- Assess effects on thyroid function to help evaluate potential mechanisms of impaired neurobehavioral[2] function.
- Summarize the extent and types of health effects evidence available.

[2]The specific aim in the protocol refers to "impaired neurological function"; however, it was changed to "impaired neurobehavior function" in this document to use more precise terminology. The overall aim from the protocol remained the same for this evaluation.

- Describe limitations of the systematic review, strengths and limitations of the evidence base, identify areas of uncertainty, as well as data gaps and research needs for neurodevelopmental and cognitive health effects of fluoride.

Depending on the extent and nature of the available evidence:

- Synthesize the evidence using a narrative approach.
- Rate confidence in the body of evidence for human and animal studies separately according to one of four statements: High, Moderate, Low, or Very Low/No Evidence Available.

Methods

Problem Formulation and Protocol Development

The research question and specific aims stated above were developed and refined through a series of problem formulation steps, including:

(1) receipt of a nomination from the public in June 2015 to conduct analyses of fluoride and developmental neurobehavioral toxicity;

(2) analysis of the extent of evidence available and the merit of pursuing systematic reviews, given factors such as the extent of new research published since previous evaluations and whether these new reports address or correct the deficiencies noted in the literature (OEHHA 2011; NRC 2006; SCHER 2011);

(3) request for information in a Federal Register notice (dated October 7, 2015);

(4) consideration of comments providing a list of studies to review through Federal Register notice and public comment period from October 7, 2015, to November 6, 2015;

(5) release of draft concept titled *Proposed NTP Evaluation on Fluoride Exposure and Potential for Developmental Neurobehavioral Effects* in November 2015;

(6) presentation of draft concept at the NTP Board of Scientific Counselors (BSC) meeting on December 1–2, 2015;

(7) consideration of comments on NTP's draft concept from the NTP BSC meeting in December 2015; and

(8) consideration of input on the draft protocol from review by technical advisors.

The protocol used to conduct this systematic review was posted in June 2017 with updates posted in May 2019 and September 2020 (https://ntp.niehs.nih.gov/go/785076).[3] The protocol served as the complete set of methods followed for the conduct of this systematic review. The OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673) is a source of general systematic review methods that were selected and tailored in developing this protocol. Options in the OHAT handbook that were not specifically referred to in the protocol were not part of the methods for the systematic review.

A brief summary of the methods is presented below. Although the methods were revised to remove the hazard assessment step and meta-analysis from this document, the protocol was not further revised.

PECO Statements

PECO (**P**opulation, **E**xposure, **C**omparators and **O**utcomes) statements were developed as an aid to identify search terms and appropriate inclusion/exclusion criteria for addressing the overall research question (effects on neurodevelopmental or cognitive function and thyroid associated

[3]NTP conducts systematic reviews following prespecified protocols that describe the review procedures selected and applied from the general methods outlined in the OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673). The protocol describes project-specific procedures tailored to each systematic review that supersede the methods in the OHAT Handbook.

with fluoride exposure) for the systematic review (Higgins and Green 2011). The PECO statements are listed below for human, animal, and in vitro/mechanistic studies (see Table 1, Table 2, and Table 3).

Using the PECO statements, the evaluation searched human studies, controlled exposure animal studies, and mechanistic/in vitro studies for evidence of neurodevelopmental or cognitive function and thyroid effects associated with fluoride exposure. Mechanistic data can come from a wide variety of studies that are not intended to identify a disease phenotype. This source of experimental data includes in vitro and in vivo laboratory studies directed at cellular, biochemical, and molecular mechanisms and attempt to explain how a substance produces particular adverse health effects. The mechanistic data were first organized by general categories (e.g., biochemical effects in the brain and neurons, neurotransmitters, oxidative stress) to evaluate the available information. Categories focused on were those with more robust data at levels of fluoride more relevant to human exposure. The intent was not to develop a mechanism for fluoride induction of effects on learning and memory but to evaluate whether a plausible series of mechanistic events exists to support effects observed in the low-dose region (below approximate drinking-water-equivalent concentrations of 20 ppm for animal studies) that may strengthen a hazard conclusion if one is derived.

Table 1. Human PECO (Population, Exposure, Comparator and Outcome) Statement

PECO Element	Evidence
Population	Humans without restriction as to age or sex, geographic location, or life stage at exposure or outcome assessment
Exposure	Exposure to fluoride based on administered dose or concentration, biomonitoring data (e.g., urine, blood, other specimens), environmental measures (e.g., air, water levels), or job title or residence. Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; Chemical Abstracts Service Registry Number [CASRN] 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable populations not exposed to fluoride or exposed to lower levels of fluoride (e.g., exposure below detection levels)
Outcomes	Neurodevelopmental outcomes, including learning, memory, intelligence, other forms of cognitive behavior, other neurological/neurobehavioral[4] outcomes (e.g., anxiety, aggression, motor activity), and biochemical changes in the brain or nervous system tissue; measures of thyroid function, biochemical changes, or thyroid tissue pathology

Table 2. Animal PECO Statement

[4]The human PECO statement in the protocol refers to "neurological outcomes"; however, it was changed to "neurological/neurobehavioral outcomes" in this document to use more precise terminology for the outcomes included.

PECO Element	Evidence
Population	Non-human mammalian animal species (whole organism)
Exposure	Exposure to fluoride based on administered dose or concentration and biomonitoring data (e.g., urine, blood, other specimens). Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; CASRN 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable animals that were untreated or exposed to vehicle-only treatment
Outcomes	Neurodevelopmental outcomes, including learning, memory, intelligence, other forms of cognitive behavior, other neurological/neurobehavioral[5] outcomes (e.g., anxiety, aggression, motor activity), and biochemical changes in the brain or nervous system tissue; measures of thyroid function, biochemical changes, or thyroid tissue pathology

Table 3. In Vitro/Mechanistic PECO Statement

PECO Element	Evidence
Population	Human or animal cells, tissues, or biochemical reactions (e.g., ligand binding assays)
Exposure	Exposure to fluoride based on administered dose or concentration. Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; CASRN 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable cells or tissues that were untreated or exposed to vehicle-only treatment
Outcomes	Endpoints related to neurological and thyroid function, including neuronal electrophysiology; mRNA, gene, or protein expression; cell proliferation or death in brain or thyroid tissue/cells; neuronal signaling; synaptogenesis, etc.

Literature Search

Main Literature Search

Search terms were developed to identify all relevant published evidence on developmental neurobehavioral toxicity or thyroid-related health effects potentially associated with exposure to fluoride by reviewing Medical Subject Headings for relevant and appropriate neurobehavioral

[5]The animal PECO statement in the protocol refers to "neurological outcomes"; however, it was changed to "neurological/neurobehavioral outcomes" in this document to use more precise terminology for the outcomes included.

and thyroid-related terms and by extracting key neurobehavioral and thyroid-related health effects and developmental neurobehavioral terminology from reviews and a sample of relevant studies.[6] Combinations of relevant subject headings and keywords were subsequently identified. A test set of relevant studies was used to ensure the search terms retrieved 100% of the test set. Six electronic databases were searched (see Main Literature Database Search) using a search strategy tailored for each database (specific search terms used for the PubMed search are presented in Appendix B; the search strategy for other databases are available in the protocol (https://ntp.niehs.nih.gov/go/785076). A search of PubChem indicated that sodium fluoride was not found in either the Tox21 or ToxCast databases; therefore, these databases were not included in the search. No language restrictions or publication-year limits were imposed. These six databases were searched in December 2016, and the search was regularly updated during the review process through April 1, 2019.

An additional search was conducted on May 1, 2020, where human epidemiological studies with primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence) were prioritized during screening. The review of the 2020 search results focused only on the human studies because they formed the basis of the confidence ratings (see Figure 1 for framework to assess confidence) and conclusions in the September 6, 2019, draft. A supplemental literature search of Chinese-language databases (described below) was also conducted. See Appendix B, Table B-1 for a timeline of key activities contributing to this 2022 NTP monograph, including information relevant to the timing of multiple literature searches.

Publications identified in these searches are categorized as "references identified through database searches" in Figure 2. Studies identified from other sources or manual review that might impact conclusions are considered under "references identified through other sources" in Figure 2. Literature searches for this systematic review were conducted independently from the literature search conducted for NTP (2016). The current literature search strategy was based on the search terms used for NTP (2016) and refined for the current evaluation, including the addition of search terms to identify human studies. Although the review process identified experimental animal studies prior to 2015, the current assessment did not evaluate these studies and relied on the NTP (2016) assessment. The focus of the literature searches for this systematic review was to identify and evaluate relevant animal studies that were published since completion of the literature searches for the NTP (2016) assessment in addition to the human and mechanistic data that were not previously evaluated.

Supplemental Chinese Database Literature Search

In order to identify non-English-language studies that might not appear in databases for the main literature search, additional searches were developed for non-English-language databases. No definitive guidance was found on the most comprehensive, highest quality, or otherwise most appropriate non-English-language databases for health studies of fluoride. Therefore, databases were chosen that identified non-English-language studies that were not captured in searches of databases from the main literature search—those previously identified from other resources (see the Searching Other Resources section below). Multiple non-English-language databases were explored before two were identified, CNKI and Wanfang, that covered studies previously

[6]The terms "study" and "publication" are used interchangeably in this document to refer to a published work drawn from an original body of research conducted on a defined population.

identified from other sources. These two Chinese electronic databases were searched in May 2020 with no language restrictions or publication year limits. Search terms from the main literature search were refined to focus on human epidemiological studies. The CNKI and Wanfang databases have character limits in the search strings; therefore, key terms were prioritized using text analytics to identify the most prevalent terms from neurodevelopmental or cognitive human epidemiological studies previously identified as relevant. Search strings were designed to capture known relevant studies that were previously identified from searching other resources without identifying large numbers of non-relevant studies (the search strategy for both databases is available in the protocol [https://ntp.niehs.nih.gov/go/785076]). Publications retrieved were compared with publications retrieved from the main literature search, and duplicates were removed. The remaining relevant publications are categorized as "references identified through database searches" in Figure 2.

New animal and mechanistic references retrieved were scanned for evidence that might extend the information currently in the September 6, 2019, draft. Although additional studies were identified, data that would materially advance the animal and mechanistic findings were not identified; therefore, these studies were not extracted nor were they added to the draft. A primary goal of the screening of the newly retrieved human references in the supplemental search of Chinese databases was to identify studies that evaluated primary neurodevelopmental or cognitive outcomes (i.e., learning, memory, and intelligence) that may have been missed in previous searches that did not include the Chinese databases. A secondary goal was to examine whether the non-English-language studies on the Fluoride Action Network website (http://fluoridealert.org/)—a site used as another resource to identify potentially relevant studies because it is known to index fluoride publications—had been selectively presented to list only studies reporting effects of fluoride. Newly retrieved human references were reviewed to identify studies that may have been missed using previous approaches. Studies identified that evaluated primary neurodevelopmental or cognitive outcomes were included and either translated or reviewed by an epidemiologist fluent in Chinese.

Databases Searched

Main Literature Database Search

- BIOSIS (Thomson Reuters)
- EMBASE
- PsycINFO (APA PsycNet)
- PubMed (NLM)
- Scopus (Elsevier)
- Web of Science (Thomson Reuters, Web of Science indexes the journal Fluoride)

Supplemental Chinese Database Literature Search

- CNKI
- Wanfang

Searching Other Resources

The reference lists of all included studies; relevant reviews, editorials, and commentaries; and the Fluoride Action Network website (http://fluoridealert.org/) were manually searched for additional relevant publications.

Unpublished Data

Although no unpublished data were included in the review, unpublished data were eligible for inclusion, provided the owner of the data was willing to have the data made public and peer reviewed (see protocol for more details: https://ntp.niehs.nih.gov/go/785076).

Study Selection

Evidence Selection Criteria

In order to be eligible for inclusion, studies had to satisfy eligibility criteria that reflect the PECO statements in Table 1, Table 2, and Table 3.

The following additional exclusion criteria were applied (see protocol for additional details: https://ntp.niehs.nih.gov/go/785076):

(1) Case studies and case reports. Although there are various definitions of 'case study' and 'case report,' the terms are used here to refer to publications designed to share health-related events on a single subject or patient with a disease, diagnosis, or specific outcome in the presence of a specific exposure.

(2) Articles without original data (e.g., reviews, editorials, or commentaries). Reference lists from these materials, however, were reviewed to identify potentially relevant studies not identified from the database searches. New studies identified were assessed for eligibility for inclusion.

(3) Conference abstracts, theses, dissertations, and other non-peer-reviewed reports.

Screening Process

References retrieved from the literature search were independently screened by two trained screeners at the title and abstract level to determine whether a reference met the evidence selection criteria. Screening procedures following the evidence-selection criteria in the protocol were pilot tested with experienced contract staff overseen by NTP. For citations with no abstract or non-English abstracts, articles were screened based on title relevance (the title would need to indicate clear relevance); number of pages (articles ≤2 pages were assumed to be conference reports, editorials, or letters unlikely to contain original data); and/or PubMed Medical Subject Headings (MeSH). Using this approach, literature was manually screened for relevance and eligibility against the evidence selection criteria using a structured form in SWIFT-Active Screener (Sciome) (Howard et al. 2020). While the human screeners review studies, SWIFT-Active Screener aids in this process by employing a machine-learning software program to priority-rank studies for screening (Howard et al. 2020). SWIFT-Active Screener also refines a statistical model that continually ranks the remaining studies according to their likelihood for inclusion. In addition, SWIFT-Active Screener employs active learning to continually incorporate user feedback during title and abstract screening to predict the total number of

included studies, thus providing a statistical basis for a decision about when to stop screening (Miller et al. 2016). Title and abstract screening was stopped once the statistical algorithm in SWIFT-Active Screener estimated that 98% of the predicted number of relevant studies were identified.

Studies that were not excluded during the title and abstract screening were further screened for inclusion with a full-text review by two independent reviewers using DistillerSR® (Evidence Partners), a web-based, systematic-review software program with structured forms and procedures to ensure standardization of the process. Screening conflicts were resolved through discussion and consultation with technical advisor(s), if necessary. During full-text review, studies that were considered relevant were tagged to the appropriate evidence streams (i.e., human, animal, and/or in vitro). Studies tagged to human or animal evidence streams were also categorized by outcome as primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence); secondary neurobehavioral outcomes (anxiety, aggression, motor activity, or biochemical); or related to thyroid effects. In vitro data were tagged as being related to neurological effects or thyroid effects. Translation assistance was sought to assess the relevance of non-English studies. Following full-text review, the remaining studies were "included" and used for the evaluation.

Evaluation of SWIFT-Active Screener Results

During the initial title and abstract screening of 20,883 references using SWIFT-Active Screener, approximately 38%[7] of the studies were manually screened in duplicate to identify an estimated 98.6% of the predicted number of relevant studies using the software's statistical algorithm (13,023 references were not screened). SWIFT-Active Screener predicted that there were 739 relevant studies during the initial title and abstract screening, of which 729 were identified and moved to full-text review. The SWIFT-Active Screener statistical algorithm predicted that 10 relevant studies at the title and abstract level (10 represents 1.4% × 739 predicted relevant studies; or 739 predicted relevant studies minus 729 identified relevant studies during screening) were not identified by not screening the remaining 13,023 studies.

To further consider the impact of using SWIFT-Active Screener for this systematic review, the evaluation team assessed the SWIFT-Active screening results to gain a better understanding of the relevance of the last group of studies that was screened before 98% predicted recall (i.e., 98% of the predicted number of relevant studies were identified). The goal was to determine the likelihood of having missed important studies by not screening all of the literature. To do this, the evaluation team examined subsets of studies screened in SWIFT-Active Screener for trends and followed those studies through to full-text review for a final determination of relevance and potential impact (i.e., whether the studies had data on primary outcomes). Based on this evaluation, it was estimated that the use of SWIFT-Active Screener may have resulted in missing one to two relevant human studies and one to two relevant animal studies with primary neurodevelopmental or cognitive outcomes. Therefore, the use of SWIFT-Active Screener saved

[7]Howard et al. (2020) evaluated the performance of the SWIFT-Active Screener methods for estimating total number of relevant studies using 26 diverse systematic review datasets that were previously screened manually by reviewers. The authors found that on average, 95% of the relevant articles were identified after screening 40% of the total reference list when using SWIFT-Active Screener. In the document sets with 5,000 or more references, 95% of the relevant articles were identified after screening 34% of the available references, on average, using SWIFT-Active Screener.

considerable time and resources and is expected to miss very few potentially relevant publications.

Screening of the May 2020 Literature Search Update

For the May 1, 2020, literature search, only primary human epidemiological studies were identified for data extraction. The study screening and selection process was focused on the human studies with primary outcomes for the evaluation because they form the basis of the confidence ratings and conclusions. Animal in vivo, human secondary outcome-only, and human and animal mechanistic references were identified as part of the screening process. These studies were then scanned for evidence that might extend the information in the September 6, 2019, draft. All included studies from the May 2020 literature search update appear in Appendix C; however, other than the primary human epidemiological studies, data from the new studies were not extracted unless they would materially advance the findings.

Note that NTP is aware of a conference abstract by Santa-Marina et al. on a Spanish cohort study that looked at fluoride exposure and neuropsychological development in children (Santa-Marina et al. 2019). The evaluation team conducted a targeted literature search in April 2021 to see whether the data from this study had been published. When no publication was found, the evaluation team contacted the study authors to inquire about the publication of their data. The response from the study authors indicated that the study report was being finalized but had not yet been sent to a journal for review; therefore, it was not considered here.[8]

Supplemental Chinese Database Searches and Human Epidemiological Studies

Supplemental searches were conducted in non-English-language databases (CNKI and Wanfang). Of the 910 references that were identified in the supplemental Chinese database searches, 13 relevant studies published in Chinese with primary neurobehavioral or cognitive outcomes were identified during title and abstract screening (which were not identified through the main literature searches). Full texts were not found for four studies after an extensive search. The remaining nine studies for which full texts were retrieved were included and were either professionally translated or evaluated by an epidemiologist fluent in Chinese for the data extraction and quality assessment steps described below. If necessary, author inquiries were conducted in Chinese to obtain missing information relevant to the assessment of the key risk-of-bias questions described below.

[8]NTP is aware that this study was published after April 2021 (Ibarluzea et al. 2021) and, therefore, is not included in this monograph because it is beyond the dates of the literature search. Even if it had been published earlier, the study would not have contributed to the body of evidence on children's IQ because the authors assessed other neurodevelopmental or cognitive effects, specifically the association between fluoride exposure and neuropsychological development in children aged 1 year using the Mental Development Index (MDI) of the Bayley Scales of Infant Development and in children aged 4 years using the General Cognitive Index (GCI) of the McCarthy Scales of Children’s Abilities (MSCA). The study will be examined as part of the NTP meta-analysis, which is being prepared as a separate report for publication.

Data Extraction

Extraction Process

Data were collected (i.e., extracted) from included studies by one member of the evaluation team and checked by a second member for completeness and accuracy. Any discrepancies in data extraction were resolved by discussion or consultation with a third member of the evaluation team.

Data Availability

Data extraction was completed using the Health Assessment Workspace Collaborative (HAWC), an open-source and freely available web-based application.[9] Data extraction elements are listed separately for human, animal, and in vitro studies in the protocol (https://ntp.niehs.nih.gov/go/785076). Data for primary and secondary outcomes, as well as thyroid hormone level data, were extracted from human studies. Studies evaluating only goiters or thyroid size were not extracted because they do not provide specific information on thyroid hormone levels that would inform whether a thyroid-mediated mechanism was involved in fluoride-associated changes in neurodevelopment. All primary outcomes and functional neurological secondary outcomes (e.g., motor activity) were extracted from animal studies identified since the NTP (2016) report. For animal mechanistic data, studies were tiered based on exposure dose (with preference given to fluoride drinking-water-equivalent exposures, which were calculated using the method described in the NTP (2016) report, of 20 ppm or less as deemed most relevant to exposures in humans), exposure duration or relevant time window (i.e., developmental), exposure route (with preference given to oral exposures over injection exposures), and commonality of mechanism (e.g., inflammation, oxidative stress, changes in neurotransmitters, and histopathological changes) were considered pockets of mechanistic data. Thyroid data were not extracted for animal studies due to inconsistency in the available data in humans. In vitro studies were evaluated, although data were not extracted from these studies as none of the findings were considered informative with respect to biological plausibility. The data extraction results for included studies are publicly available and can be downloaded in Excel format through HAWC (https://hawcproject.org/assessment/405/). Methods for transforming and standardizing dose levels and results from behavioral tests in experimental animals are detailed in the protocol (https://ntp.niehs.nih.gov/go/785076).

In 2016, NTP published a systematic review of the evidence from experimental animal studies on the potential effects of fluoride exposure on learning and memory (NTP 2016). The literature searches for the current assessment identified and evaluated relevant animal studies published since the 2016 assessment and also included human and mechanistic data that were not previously evaluated. Although literature search activities for the current assessment identified experimental animal studies prior to 2015, the current assessment did not re-evaluate animal studies published prior to 2015 because these were reviewed in the NTP (2016) assessment.

[9]HAWC (Health Assessment Workspace Collaborative): A Modular Web-based Interface to Facilitate Development of Human Health Assessments of Chemicals (https://hawcproject.org/portal/).

Quality Assessment of Individual Studies

Risk of bias was assessed for individual studies using the OHAT risk-of-bias tool (https://ntp.niehs.nih.gov/go/riskbias) that outlines a parallel approach to evaluating risk of bias from human, animal, and mechanistic studies to facilitate consideration of risk of bias across evidence streams with common terms and categories. The risk-of-bias tool is comprised of a common set of 11 questions that are answered based on the specific details of individual studies to develop risk-of-bias ratings for each question. Study design determines the subset of questions used to assess risk of bias for an individual study (see Table 4). When evaluating the risk of bias for an individual study, the direction and magnitude of association for any specific bias is considered.

Assessors were trained with an initial pilot phase undertaken to improve clarity of rating criteria and to improve consistency among assessors. Studies were independently evaluated by two trained assessors who answered all applicable risk-of-bias questions with one of four options in Table 5 following prespecified criteria detailed in the protocol (https://ntp.niehs.nih.gov/go/785076). The criteria describe aspects of study design, conduct, and reporting required to reach risk-of-bias ratings for each question and specify factors that can distinguish among ratings (e.g., what separates “definitely low” from “probably low” risk of bias).

Key Risk-of-bias Questions

In the OHAT approach, some risk-of-bias questions or elements are considered potentially more important when assessing studies because these issues are generally considered to have a greater impact on estimates of the effect size or on the credibility of study results in environmental health studies. There are three Key Questions for observational human studies: confounding, exposure characterization, and outcome assessment. Based on the complexity of the possible responses to these questions in epidemiological studies, considerations made and methods used for evaluating the Key Questions are provided below. There are also three Key Questions for experimental animal studies: randomization, exposure characterization, and outcome assessment. In addition, for animal developmental studies, failure to consider the litter as the unit of analysis was also a key risk-of-bias concern. When there was not enough information to assess the potential bias for a risk-of-bias question and authors did not respond to an inquiry for further information, a conservative approach was followed, and the studies were rated probably high risk of bias for that question.

Risk-of-bias Considerations for Human Studies

The risk of bias of individual studies in the body of evidence was considered in developing confidence ratings. The key risk-of-bias questions (i.e., confounding, exposure characterization, and outcome assessment for human studies) are discussed in the consideration of the body of evidence. For this assessment, the key risk-of-bias questions, if not addressed appropriately, are considered to have the greatest potential impact on the results. The other risk-of-bias questions, including selection of study participants, were also considered and were used to identify any other risk-of-bias concerns that may indicate serious issues with a study that could cause it to be considered high risk of bias. No study was excluded based on concerns for risk of bias; however, the low risk-of-bias studies generally drive the ratings on confidence in the results across the

body of evidence. Human evidence was evaluated with and without high risk-of-bias studies to assess the impact of these studies on confidence in the association.

High risk-of-bias studies: Studies rated probably high risk of bias for at least two key risk-of-bias questions or definitely high for any single question are considered studies with higher potential for bias (i.e., high risk-of-bias studies) and to be of low quality. Studies could also be considered high risk of bias if rated probably high risk of bias for one key risk-of-bias question along with other concerns, including potential for selection bias and concerns with statistical methods.

Low risk-of-bias studies: The remaining studies (i.e., other than the high risk-of-bias studies) were considered to have lower potential for bias (i.e., low risk of bias) and to be of high quality. Appendix E describes strengths and limitations of the low risk-of-bias/high-quality studies identified during the assessment and clarifies why they are considered to pose low risk of bias. Details on the statistical analyses are provided in the "Other potential threats" domain in order to evaluate the adequacy of the statistical approach for individual studies.

Given the number of non-English-language studies in this assessment, the potential for the translation to introduce bias was examined as described below, and it was determined that translation of non-English-language studies did not impact evaluation of risk of bias. Thirty-two of 100 studies included in the entire human body of evidence on neurodevelopmental and cognitive effects were initially published in a foreign language (Chinese) and were either translated and published in volume 41 of the journal *Fluoride* (n = 19) or were translated by the Fluoride Action Network (n = 13) (http://fluoridealert.org/researchers/translations/complete_archive/). Most of these studies were considered to have high potential for bias due to lack of information across the key risk-of-bias questions. Therefore, in order to assess whether the lack of information relevant to key risk-of-bias concerns was the result of a loss in translation, the original Chinese publications and the translated versions of the five studies that had the most potential for being included in the low risk-of-bias group of studies were reviewed by a team member fluent in Chinese to determine whether any of the risk-of-bias concerns could be addressed (An et al. 1992; Chen et al. 1991 [translated in Chen et al. 2008]; Du et al. 1992 [translated in Du et al. 2008]; Guo et al. 1991 [translated in Guo et al. 2008a]; Li et al. 2009). For all five studies, the translations were determined to be accurate, and there was no impact of the translations on the key risk-of-bias concerns.

Confounding

Covariates were determined a priori based on factors that are associated with neurodevelopment or cognition and could be related to fluoride exposure. Covariates that were considered key for all studies, populations, and outcomes included age, sex, and socioeconomic status (e.g., maternal education, household income, marital status, crowding). Additional covariates considered important for this evaluation, depending on the study population and outcome, included race/ethnicity; maternal demographics (e.g., maternal age, body mass index [BMI]); parental behavioral and mental health disorders (e.g., attention deficit hyperactivity disorder [ADHD], depression); smoking (e.g., maternal smoking status, secondhand tobacco smoke exposure); reproductive factors (e.g., parity); nutrition (e.g., BMI, growth, anemia); iodine deficiency/excess; minerals and other chemicals in water associated with neurotoxicity (e.g., arsenic, lead); maternal and paternal IQ; and quantity and quality of caregiving environment

(e.g., Home Observation Measurement of the Environment [HOME] score). To be assigned a rating of probably low risk of bias for the key risk-of-bias question regarding the confounding domain, studies were not required to address every important covariate listed; however, studies were required to address the three key covariates for all studies, the potential for co-exposures, if applicable (e.g., arsenic and lead, both of which could affect cognitive function), and any other potential covariates considered important for the specific study population and outcome. For example, studies of populations in China, India, and Mexico, where there is concern about co-exposures to high fluoride and high arsenic, were required to address arsenic. If the authors did not directly specify that arsenic exposures were evaluated, groundwater quality maps were evaluated (https://www.gapmaps.org/Home/Public) in order to identify areas of China, India, and Mexico where arsenic is a concern (Podgorski and Berg 2020). If no arsenic measurements were available for the area, the arsenic groundwater quality predictions from the global arsenic 2020 map were used (Podgorski and Berg 2020). If an area had less than 50% probability of having arsenic levels greater than 10 µg/L (the WHO guideline concentration), the area was considered not to have an issue with arsenic that needed to be addressed by the study authors; however, it should be noted that arsenic may be associated with neurodevelopmental effects at concentrations below 10 µg/L.

Exposure

Fluoride ion is rapidly absorbed from the gastrointestinal tract and is rapidly cleared from serum by distribution into calcified tissues and urinary excretion (IPCS 2002). There is general consensus that the best measures of long-term fluoride exposure are bone and/or tooth measurements, and other than measures of dental fluorosis, these were not performed in any of the studies reviewed in this document. Prolonged residence in an area with a given fluoride content in drinking water has been considered in many studies as a proxy for long-term exposure.

Exposure was assessed using a variety of methods in the human body of evidence. Studies provided varying levels of details on the methods used and employed different exposure characterization methods to group study subjects into exposed and reference groups. Exposure metrics included spot urine (from children or mothers during at least one trimester of gestation), serum, individual drinking water, intake from infant formula, estimated total exposure dose, municipal drinking water (with residence information), evidence of dental or skeletal fluorosis, area of residence (endemic versus a non-endemic fluorosis area, with or without individual validation of exposure), burning coal (with or without fluoride), and occupation type.

Urinary fluoride levels measured during pregnancy and in children include all ingested fluoride and are considered a valid measure to estimate total fluoride exposure (Villa et al. 2010; Watanabe et al. 1995); however, the type and timing of urinary sample collection are important to consider. Urinary fluoride is thought to reflect recent exposure but can be influenced by the timing of exposure (e.g., when water was last consumed, when teeth were last brushed). When compared with 24-hour urine samples, spot urine samples are more prone to the influence of timing of exposure and can also be affected by differences in dilution; however, many studies attempted to account for dilution either by using urinary creatinine or specific gravity. Good correlations between 24-hour samples and urinary fluoride concentrations from spot samples adjusted for urinary dilution have been described (Zohouri et al. 2006). Despite potential issues with spot urine samples, if authors made appropriate efforts to reduce the concern for bias (e.g.,

accounting for dilution), studies that used this metric were generally considered to have probably low risk of bias for exposure.

Analytical methods to measure fluoride in biological or water samples also varied, some of which included atomic absorption, ion-selective electrode methods, colorimetric methods, or the hexamethyldisiloxane microdiffusion method. Individual-level measures of exposure were generally considered more accurate than group-level measures; however, using group-level measures (e.g., endemic versus non-endemic area) in an analysis was less of a concern if the study provided water or urinary fluoride levels from some individuals to verify that there were differences in the fluoride exposure between groups. Studies that provided results by area and also reported individual urinary or serum fluoride concentrations or other biochemical measures, including dental fluorosis in the children or urinary levels in mothers during pregnancy, were considered to have probably low risk of bias. Ideally, these studies would still need to consider and adjust for area-level clustering; however, these concerns are captured in evaluations of other potential threats to internal validity.

Outcome

Studies included in this evaluation used a wide variety of methods to measure IQ and other cognitive effects. Measures of IQ were generally standardized tests of IQ; however, for these standardized methods to be considered low potential for bias, they needed to be conducted in the appropriate population or modified for the study population. Because results of many of the tests to measure neurodevelopment and cognitive function can be subjective, it was important that the outcome assessors were blind to the fluoride exposure when evaluating the results of the tests. If the study reported that the assessor was blind to the exposure, this was assumed to mean that the outcome assessor did not have any knowledge of the exposure, including whether the study subjects were from high-fluoride communities. If cross-sectional studies collected biomarker measurements at the time of an IQ assessment, this was considered indirect evidence that the outcome assessor would not have knowledge of the fluoride exposure unless there was also potential for the outcome assessor to have knowledge of varying levels of fluoride by study area. In cases wherein the study did not specify that the outcome assessors were blind, the study authors were contacted and asked whether the outcome assessors were, in fact, blind to exposure. When authors responded and indicated that outcome assessors were blind to exposure or that it was not likely that they would have had knowledge of exposure, this was considered direct or indirect evidence, respectively, that blinding was not a concern for those studies.

Any discrepancies in ratings between assessors were resolved by a senior technical specialist and through discussion when necessary to reach the final recorded risk-of-bias rating for each question along with a statement of the basis for that rating. Members of the evaluation team were consulted for assistance if additional expertise was necessary to reach final risk-of-bias ratings based on specific aspects of study design or performance reported for individual studies. Study procedures that were not reported were assumed not to have been conducted, resulting in an assessment of “probably high” risk of bias. Authors were queried by email to obtain missing information, and responses received were used to update risk-of-bias ratings.

Table 4. OHAT Risk-of-bias Questions and Applicability by Study Design

Risk-of-bias Questions	Experimental Animal[a]	Human Controlled Trials[b]	Cohort	Case-control	Cross-sectional[c]	Case Series
1. Was administered dose or exposure level adequately randomized?	X	X				
2. Was allocation to study groups adequately concealed?	X	X				
3. Did selection of study participants result in the appropriate comparison groups?			X	X	X	
4. Did study design or analysis account for important confounding and modifying variables?			X	X	X	X
5. Were experimental conditions identical across study groups?	X					
6. Were research personnel blinded to the study group during the study?	X	X				
7. Were outcome data complete without attrition or exclusion from analysis?	X	X	X	X	X	
8. Can we be confident in the exposure characterization?	X	X	X	X	X	X
9. Can we be confident in the outcome assessment (including blinding of outcome assessors)?	X	X	X	X	X	X
10. Were all measured outcomes reported?	X	X	X	X	X	X
11. Were there no other potential threats to internal validity?	X	X	X	X	X	X

[a]Experimental animal studies are controlled exposure studies. Non-human animal observational studies can be evaluated using the design features of observational human studies such as cross-sectional study design.
[b]Human Controlled Trials are studies in humans with controlled exposure (e.g., randomized controlled trials, non-randomized experimental studies).
[c]Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).

Answers to the risk-of-bias questions result in one of the following four risk-of-bias ratings:

Table 5. The Four Risk-of-bias Rating Options

Symbol	Description
++	***Definitely Low*** *risk of bias*: There is direct evidence of low risk-of-bias practices.
+	***Probably Low*** *risk of bias*: There is indirect evidence of low risk-of-bias practices, **OR** it is deemed that deviations from low risk-of-bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias.
– NR	***Probably High*** *risk of bias*: There is indirect evidence of high risk-of-bias practices (indicated with "–"), **OR** there is insufficient information provided about relevant risk-of-bias practices (indicated with "NR" for not reported). Both symbols indicate probably high risk of bias.
—	***Definitely High*** *risk of bias*: There is direct evidence of high risk-of-bias practices.

Organizing and Rating Confidence in Bodies of Evidence

Health Outcome Categories for Neurodevelopmental and Cognitive Effects

After data were extracted from all studies, the health effects results within the category of neurodevelopmental or cognitive effects were grouped across studies to develop bodies of evidence or collections of studies with data on the same or related outcomes. The grouping of health effect results was not planned a priori. The vast majority of the human studies evaluated IQ in children as the single outcome; therefore, the discussion of cognitive neurodevelopmental effects in children focuses on IQ studies with supporting information from data on other endpoints. Cognitive function in adults was evaluated separately. Consistent with the NTP (2016) assessment, the primary focus within the animal study body of evidence was on animal studies with endpoints related to learning and memory.

Considerations for Pursuing a Narrative or Quantitative Evidence Synthesis

This evaluation provides only a narrative review of the data; however, heterogeneity within the available evidence was evaluated to determine whether a quantitative synthesis (i.e., meta-analysis) would be appropriate. Choi et al. (2012) and Duan et al. (2018) conducted meta-analyses and found that high fluoride exposure was associated with lower IQ scores. Choi et al. (2012) was able to determine a risk ratio for living in an endemic fluorosis area but was unable to develop a dose-response relationship. Duan et al. (2018) reported a significant non-linear dose-response relationship between fluoride dose and intelligence with the relationship stated as most evident with exposures from drinking water above 4 mg/L (or 4 ppm) fluoride. Duan et al. (2018) found similar results as Choi et al. (2012) for the standardized mean difference; however, the majority of the available studies in both analyses compare populations with high fluoride exposure to those with lower fluoride exposure (with the lower exposure levels frequently in the range of drinking water fluoridation in the United States). The meta-analysis conducted in

association with this systematic review further informs this issue and will be published separately.

Confidence Rating: Assessment of Body of Evidence

The quality of evidence for neurodevelopmental and cognitive function outcomes was evaluated using the GRADE system for rating the confidence in the body of evidence (Guyatt et al. 2011; Rooney et al. 2014). More detailed guidance on reaching confidence ratings in the body of evidence as "high," "moderate," "low," or "very low" is provided in the protocol (https://ntp.niehs.nih.gov/go/785076). In brief, available human and animal studies on a particular health outcome were initially grouped by key study design features, and each grouping of studies was given an initial confidence rating by those features. Starting at this initial rating (see column 1 of Figure 1), potential downgrading of the confidence rating was considered for factors that decrease confidence in the results (see column 2 of Figure 1). Potential upgrading of the confidence rating was considered for factors that increase confidence in the results (see column 3 of Figure 1). Short descriptions of the factors that can decrease or increase confidence in the body of evidence for human studies are provided below (see protocol [https://ntp.niehs.nih.gov/go/785076] for additional details related to the human body of evidence, as well as considerations for experimental animal studies).

Factors to Consider for Potential Downgrading

- Risk of bias: Addresses whether the body of evidence did not account for critical factors in study quality or design, including confounding bias, selection bias, exposure assessment, and outcome assessment. Consideration for downgrading the confidence rating is based on the entire body of evidence, and the evidence is downgraded when there is substantial bias across most studies that could lead to decreased confidence in the results and when the studies without substantial bias could not support the confidence rating. Individual studies are evaluated for risk of bias based on a set of criteria (as discussed above); magnitude and direction of the bias are also considered.
- Unexplained inconsistency: Addresses inconsistencies in results across studies of similar populations and design that can be determined by assessing similarity of point estimates and extent of overlap between confidence intervals or more formally through statistical tests of heterogeneity. Sensitivity analysis can be used to assess the impact of specific variables on the outcome. Inconsistencies that can be plausibly explained by characteristics of the studies (e.g., sex-associated differences) are typically not used to support a downgrade. A downgrade would only be applied when there is an inconsistency that cannot be explained and results in reduced confidence in the body of evidence.
- Indirectness: Addresses generalizability and relevance to the objective of the assessment. As outlined in the Objective and consistent with the population specified in the PECO statement, this systematic review evaluated the extent and quality of the evidence linking fluoride exposure to neurodevelopmental and cognitive effects in humans without restriction as to age, sex, geographic location, or life stage at exposure or outcome assessment. Furthermore, the review did not exclude subjects exposed in occupational settings. All exposure levels and scenarios encountered in

human studies are considered direct (i.e., applicable, generalizable, and relevant to address the objective of the assessment); therefore, a downgrade for indirectness would not be applied to bodies of evidence from human studies.

- Imprecision: Addresses confidence associated with variability in quantitative measures such as effect sizes. Typically, 95% confidence intervals are used as the primary method to assess imprecision, but considerations can also be made on whether studies were adequately powered. Meta-analyses can also be used to determine whether the data are imprecise. When a meta-analysis is not appropriate or feasible, imprecision can be based on variability around the effect estimate. A downgrade would occur if the body of evidence was considered to be imprecise based on a meta-analysis, or if serious or very serious imprecision was consistently present in the body of evidence. A downgrade is especially likely if imprecision raised questions as to whether an overall effect was significant.
- Publication bias: Addresses evidence of biased publication practices. Downgrade if one strongly detects publication bias. Publication bias is difficult to detect but may be evident if major sections of the research community are not publishing (e.g., absence of industry, academic, or government studies) on a topic or if there are multiple instances wherein data from conference abstracts are never published in peer-reviewed journals. In addition, there are methods included in conducting a meta-analysis to detect whether there is potential for publication bias, including the use of fit-and-trim models, which help identify how publication bias may affect the results of the meta-analysis. Although a meta-analysis is not included in this systematic review, there are two published meta-analyses (Choi et al. 2012; Duan et al. 2018) in addition to the one associated with this systematic review (manuscript in progress) that can be used to address publication bias.

Factors to Consider for Potential Upgrading

- Large magnitude of effect: Factors to consider include the outcome being measured and the dose or exposure range assessed. The confidence can be upgraded if the body of evidence is suggestive of a large magnitude of effect. GRADE provides guidance on what can be considered a large magnitude of effect based on relative risk (i.e., suggests one upgrade in confidence if relative risk is greater than 2 and two upgrades in confidence if greater than 5). However, not all studies provide data as a risk estimate, and smaller changes, such as increases in blood pressure, may have greater impact on health at the population level. Consideration for an upgrade is not based on a single study, and what constitutes a large magnitude of effect will depend on the outcome and the potential public health impact.
- Dose response: Patterns of dose response are evaluated within and across studies. Confidence in the body of evidence can be increased when there is sufficient evidence of a dose-response pattern across multiple studies.
- Consistency: Does not apply in this evaluation. The consideration of a potential upgrade for consistency is primarily for non-human animal evidence in which it would be applied to address increased confidence based on an observation of consistent effects across multiple non-human animal species. For human evidence, this factor would generally not be applied. Human studies are instead evaluated for

issues of consistency that could result in downgrading confidence for unexplained inconsistency (see "Factors to Consider for Potential Downgrading" above).

- Consideration of residual confounding: Applies to observational studies and refers to consideration of unmeasured determinants that are likely to be distributed unevenly across groups. Residual confounding can push results in either direction, but confidence in the results is increased when the body of evidence is biased by factors that counter the observed effect and would cause an underestimation of the effect. Confounding that would cause an overestimation of the effect is considered under the risk-of-bias considerations for decreasing confidence.

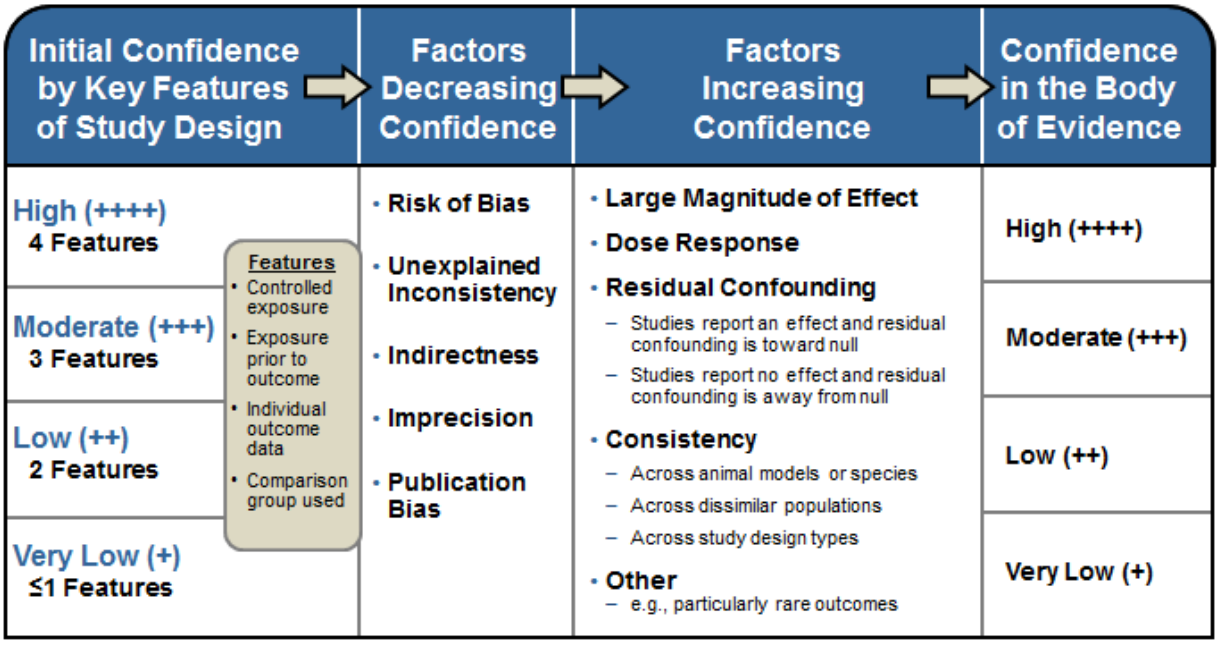


Figure 1. Assessing Confidence in the Body of Evidence

Confidence ratings were assessed by the evaluation team for accuracy and consistency, and discrepancies were resolved by consensus and consultation with technical advisors as needed. Confidence ratings for the primary outcomes are summarized in evidence profile tables for each outcome.

Results

Literature Search Results

The electronic database searches retrieved 25,450 unique references with 11 additional references[10] identified by technical advisors or obtained by manually searching the Fluoride Action Network website or reviewing reference lists of published reviews and other included studies. During title and abstract screening, 1,036 references were moved to full-text review and 24,425 were excluded (11,402 by manual screening for not satisfying the PECO criteria and 13,023 based on the SWIFT-Active Screener algorithm). Among the 1,036 references that underwent full-text review, 547 studies were considered PECO-relevant (see Appendix C for list of included studies). A few studies assessed data for more than one evidence stream (human, non-human mammal, and/or in vitro), and several studies assessed more than one type of outcome (e.g., primary and secondary outcomes). Included studies break down as follows:

- 167 human studies (84 primary only; 13 secondary only; 5 primary and secondary; 8 primary and thyroid; 2 secondary and thyroid; and 55 thyroid only);
- 339 non-human mammal studies (7 primary only; 186 secondary only; 67 primary and secondary; 6 primary, secondary, and thyroid; 4 secondary and thyroid; and 69 thyroid only); and,
- 60 in vitro/mechanistic studies (48 neurological and 12 thyroid).

Additional details on the screening results are provided in Appendix C. These screening results are outlined in a study selection diagram that reports numbers of studies excluded at each stage and documents the reason for exclusion at the full-text review stage (see Figure 2) [using reporting practices outlined in Moher et al. (2009)].

[10]These 11 studies (9 human and 2 animal studies) were not identified through the electronic database searches, as they were not indexed in any of the electronic databases searched. Note that the supplemental search of non-English-language databases was designed in part to identify non-English-language studies that are not indexed in traditional bibliographic databases such as PubMed. It was successful in this goal, as multiple studies that were initially only identified through “other sources” were subsequently captured in the supplemental Chinese database search, leaving only 11 as identified through other sources.

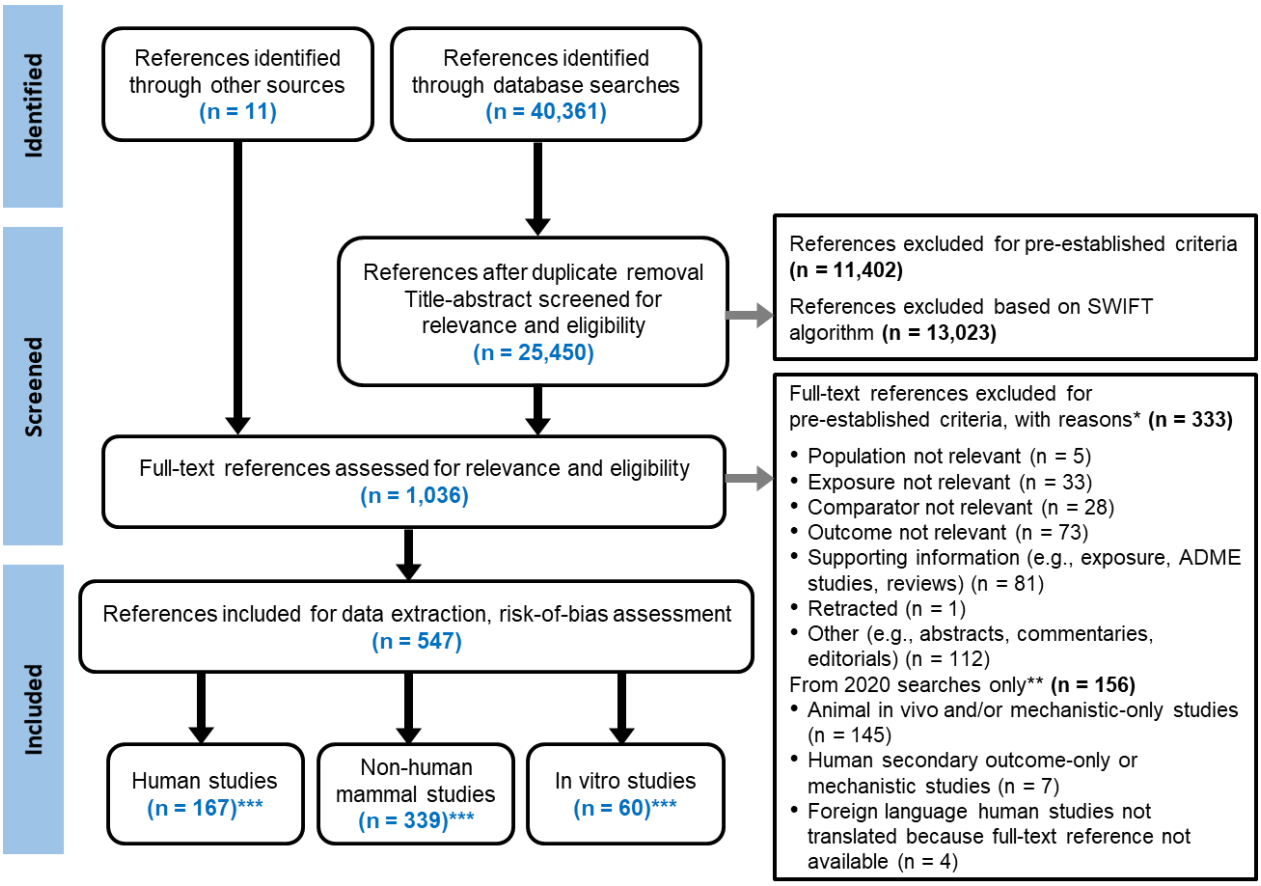


Figure 2. Study Selection Diagram[a]

[a]An interactive reference flow diagram is available here: https://hawcproject.org/summary/visual/assessment/405/Figure-2/.
*Includes studies from all literature searches conducted during the review; see the Methods section for extraction and search update information. Studies may have been excluded for more than one reason; the first reason identified was recorded.
**Includes all studies from all 2020 literature searches not otherwise excluded for pre-established criteria; see the Methods section for extraction and search update information.
***Publications may contain more than one evidence stream, so the numbers will not total the 547 included studies.

Human Neurodevelopmental and Cognitive Data

The body of literature that evaluates the association between fluoride exposure and neurodevelopmental and cognitive effects in humans is relatively robust with a large number of studies (n = 100) that cover a wide array of endpoints (see Figure 3). Seventy-two human studies investigated IQ in children. Additional studies evaluated learning and memory (n = 9 studies) or other cognitive developmental effects (e.g., total neurobehavioral scores and total mental capacity index in children, cognitive impairment in adults; n = 15 studies).[11] For this review, the evidence in children and adults was evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood.

[11]Some studies are included in more than one endpoint category (e.g., IQ and other cognitive developmental effects); therefore, these counts are not mutually exclusive.

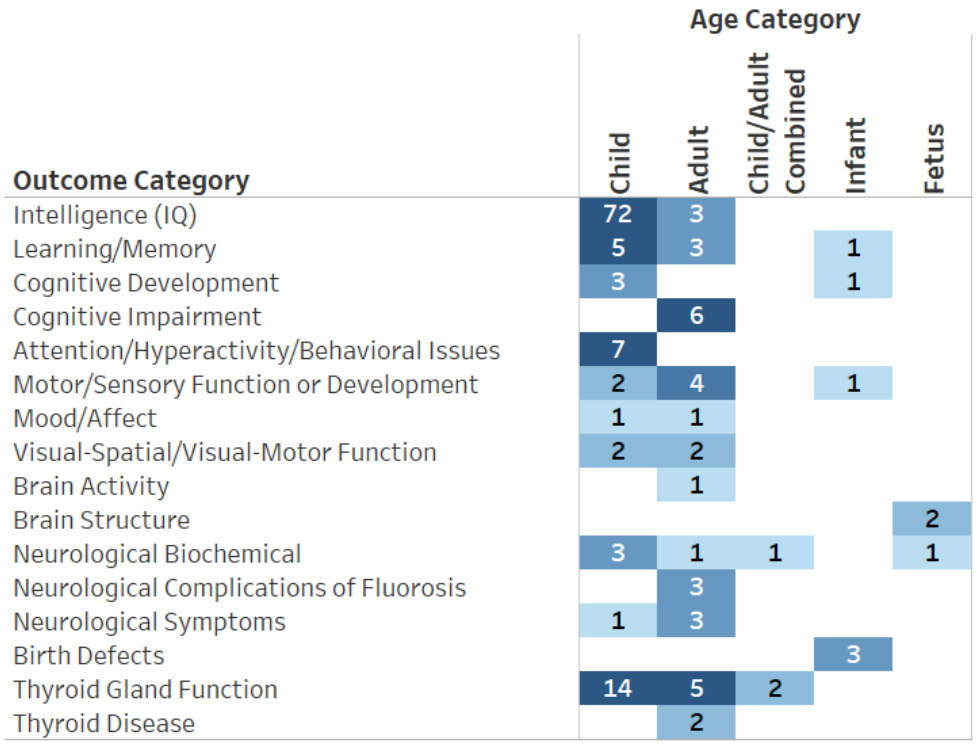


Figure 3. Number of Epidemiological Studies by Outcome and Age Categories[a]

[a]Interactive figure and additional study details in Tableau®. (https://public.tableau.com/app/profile/ntp.visuals/viz/Fluoride_Epi_2022Update/Figure3?publish=yes)
Choi et al. (2015) used subtests of the omnibus IQ test reported by the authors as Wechsler Intelligence Scale for Children-Revised (WISC-IV) to evaluate visuospatial abilities (using block design) and executive function (using digit span). These endpoints are included in the intelligence (IQ) outcome category as they are subsets of the IQ tests.
Three additional publications based on subsamples (i.e., 50–60 children) of the larger Yu et al. (2018) cohort were identified (Zhao et al. 2019; Zhao et al. 2020; Zhou et al. 2019) and are not included in the counts of this figure.

Because the majority of studies evaluated intelligence, the following section focuses on IQ effects in children followed by separate discussions on other measures of cognitive function and neurobehavioral effects in children and cognitive effects in adults. Studies that evaluated mechanistic data in humans, including effects on the thyroid, are discussed in the Mechanistic Data in Humans section. Note that a few studies were identified on congenital neurological malformations and neurological complications of fluorosis; however, they are not considered further due to the limited number of studies and the heterogeneity of outcomes evaluated in those studies.

IQ in Children

Seventy-two epidemiological studies were identified that evaluated the association between fluoride exposure and children's IQ. Nineteen of the 72 IQ studies were determined to have low potential for bias (i.e., were of high quality). Looking across the literature, there has been a progression over the years in the quality of studies conducted to assess the association between fluoride exposure and IQ in children, with more recent studies including better study designs, larger sample sizes, and more sophisticated statistical analysis. Older studies often had limitations related to study design or methods, and most of the high risk-of-bias studies (i.e.,

studies of low quality) were published prior to the 2006 NRC evaluation of fluoride in drinking water. In contrast, 18 of the low risk-of-bias studies were published after the 2006 NRC evaluation of fluoride in drinking water, and over half of those were published between 2015 and 2020 (Figure 4).

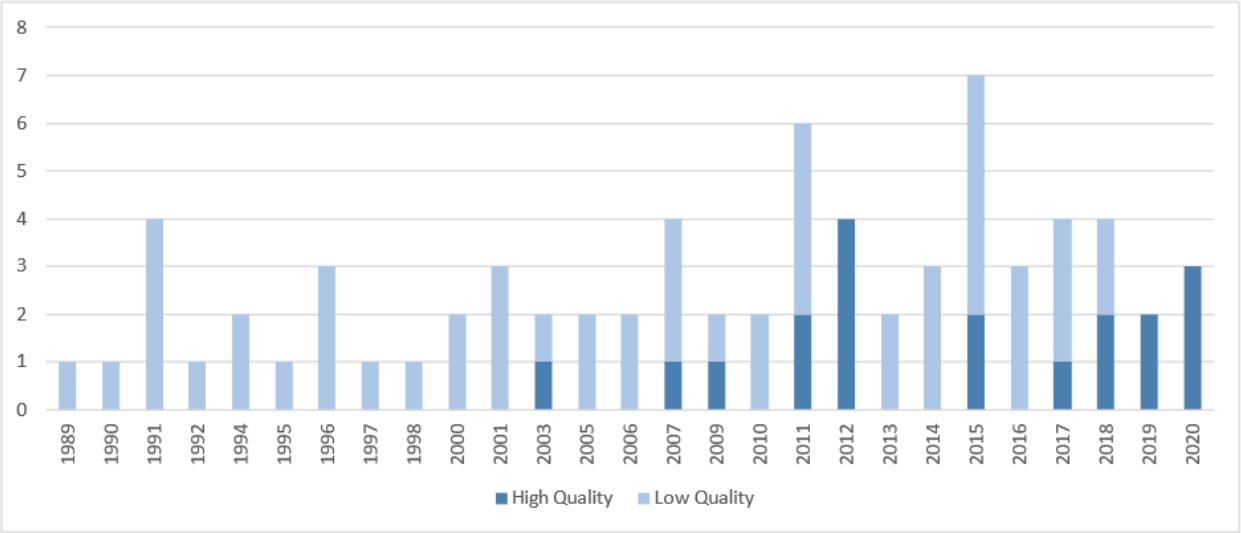


Figure 4. Number of High- and Low-quality Studies of Fluoride Exposure and IQ in Children by Year of Publication

Several characteristics of recent studies contribute to higher study quality in the overall body of literature on children's IQ and fluoride, including:

- Demonstration that exposure occurred prior to outcome assessment (an important factor when considering confidence in study results; see Figure 1) either by study design (e.g., for prospective cohort studies) or analysis (e.g., prevalence of dental fluorosis in children, limiting study populations to children who lived in the same area for long periods of time).
- Improved reporting of key study details that are necessary to evaluate study quality and allow for a more precise analysis of risk of bias.
- Increased consideration of key covariates (e.g., socioeconomic status) including potential co-exposures (e.g., arsenic or lead intake).
- Increased use of individual-level exposure measures (urine or water) as well as prenatal fluoride exposure to assess either individual-level fluoride exposure or—if still using group-level data—to confirm that regions being compared had differences in fluoride exposure.
- Utilization of more sophisticated sampling techniques for the study populations (e.g., stratified multistage random sampling).
- Application of more sophisticated regression approaches (e.g., piecewise linear regression models, multi-level regression with random effects, or generalized additive models for longitudinal measurements of fluoride).

- For studies using individual-level exposure measures, application of more sophisticated regression techniques to account for clustering at the cohort level by using cohort as a fixed or random effect and by accounting for numerous covariates that capture the cohort effect.

In addition, newer studies represent more diverse study populations across several countries (Figure 5), whereas all identified peer-reviewed studies that were published prior to 2006 took place in a single country (China). The majority of high-quality, low risk-of-bias studies exhibit these important study design and analysis characteristics, as discussed further in subsequent sections.

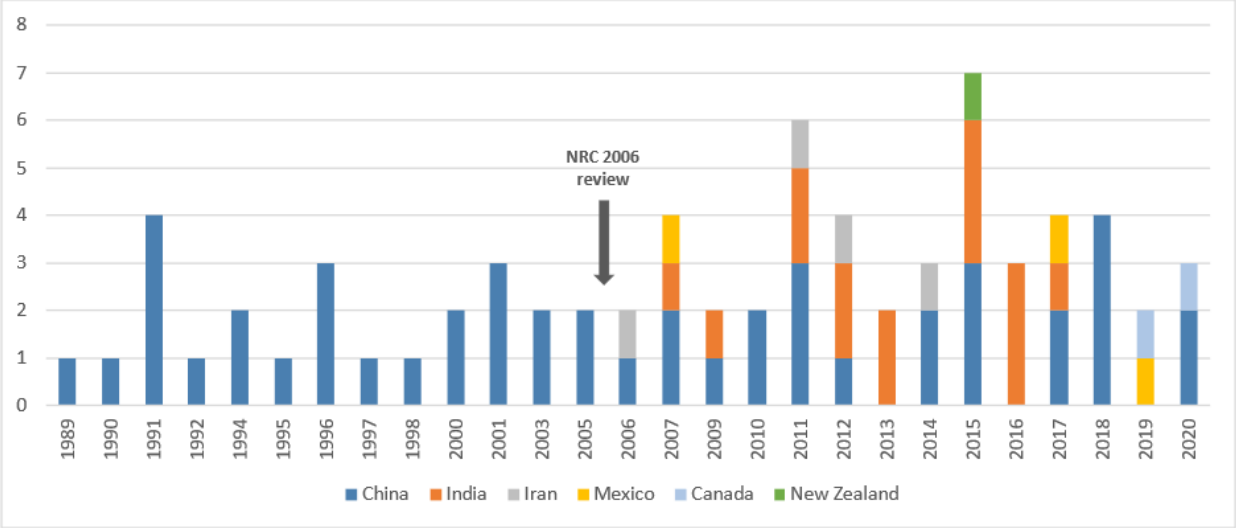


Figure 5. Number of Studies of Fluoride Exposure and IQ in Children by Country and Year of Publication

All available studies were considered in this evaluation; however, review of the body of evidence focused on the high-quality, low risk-of-bias studies for two main reasons. First, there are fewer limitations and greater confidence in the results of the high-quality studies. Second, there are a relatively large number of high-quality studies (n = 19), such that the body of evidence from these studies could be used to evaluate confidence in the association between fluoride exposure and changes in children's IQ. Therefore, the remainder of the discussion on IQ in children focuses on the 19 studies with low risk of bias. The high risk-of-bias studies are discussed briefly relative to their overall support of findings from the low risk-of-bias studies.

Low Risk-of-bias IQ Studies

Overview of Studies

Nineteen studies (3 longitudinal prospective cohort and 16 cross-sectional studies) with low potential for bias evaluated the association between fluoride exposure and IQ in children (see Quality Assessment of Individual Studies section for methods on determining which studies pose low risk of bias). These IQ studies were conducted in 15 study populations across 5 countries

and included more than 7,000 children. Specifically, of the 19 low risk-of-bias studies of IQ in children:

- ten were conducted in four areas of China on seven study populations,[12]
- three were conducted in three areas of Mexico on three study populations,
- two were conducted in Canada using the same study population,
- three were conducted in three areas of India on three study populations, and
- one was conducted in Iran.

Most studies measured fluoride in drinking water (n = 15) and/or urine (child or maternal) (n = 15). Two studies measured fluoride in serum. The IQ studies used a variety of tests to measure IQ. Because IQ tests should be culturally relevant, the tests used often differed between studies, reflecting adjustments for the range in populations studied (e.g., western vs. Asian populations). In some cases, different IQ tests were used to study similar populations. Overall, these studies used IQ tests that were population- and age-appropriate.

Table 6 provides a summary of study characteristics and key IQ and fluoride findings for the 19 low risk-of-bias studies. Several of these studies conducted multiple analyses and reported results on multiple endpoints. The purpose of the table is to summarize key findings (independent of whether an association is indicated) from each study and is not meant to be a comprehensive summary of all results from each study. For each study, results are summarized for each exposure measure assessed, but results from multiple analyses using the same exposure measure may not be presented for all studies unless multiple analyses yielded conflicting results. See Appendix E for additional information on each study in Table 6, including strengths and limitations, clarifications for why studies are considered to pose low risk of bias, and information regarding statistical analyses, important covariates, exposure assessment, and outcome assessment.

[12]In this document, "study population" refers to a defined population on which an original body of research was conducted. The published work drawn from that original body of research is often referred to as a "study." IQ studies that report on the same study populations are identified in Table 6.

Table 6. Studies on IQ in Children[a]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
China					
Xiang et al. (2003a)[d]	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children [512]	Drinking water Mean (SD): 0.36 (0.15) (control), 2.47 (0.79) (high fluoride) mg/L Children's urine Mean (SD): 1.11 (0.39) (control), 3.47 (1.95) (high fluoride) mg/L Village of residence (non-endemic vs. endemic fluorosis)	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significant dose-related association of fluoride on IQ score based on drinking water quintile levels with significantly lower IQ scores observed at water fluoride levels of 1.53 mg/L or higher; % of subjects with IQ <80 was significantly increased at water levels 2.46 mg/L or higher; significant inverse correlation between IQ and urinary fluoride (Pearson correlation coefficient of −0.164); mean IQ scores for children in non-endemic region (100.41 ± 13.21) significantly higher than endemic region (92.02 ± 13.00) No statistical adjustment for covariates
Ding et al. (2011)	Cross-sectional Inner Mongolia (Hulunbuir City)/elementary school children [331]	Children's urine Range: 0.1–3.55 mg/L Drinking water (reported but not used in analyses) Mean (SD): 1.31 (1.05) mg/L	Children (ages 7–14 years)	IQ: Combined Raven's Test for Rural China	Significant association between urinary fluoride and IQ score (each 1-mg/L increase was associated with a decrease in IQ score of 0.59 points; 95% CI: −1.09, −0.08) Adjusted for age
Xiang et al. (2011)[d]	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children [512]	Children's serum Mean (SD): 0.041 (0.009) (control), 0.081 (0.019) (high fluoride) mg/L	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significant linear trend across quartiles of serum fluoride and children's IQ score <80 (adjusted ORs for Q1 and Q2; Q1 and Q3; and Q1 and Q4, respectively: 1; 2.22 [95% CI: 1.42, 3.47]; and 2.48 [95% CI: 1.85, 3.32]); significant associations at ≥0.05 mg/L serum fluoride Adjusted for age and sex

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Wang et al. (2012)[d]	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children [526]	Children's total fluoride intake Mean (SD): 0.78 (0.13) (control), 3.05 (0.99) (high fluoride) mg/day Village of residence (non-endemic vs. endemic fluorosis) Drinking water (reported for villages but not used in analyses) Mean (SD): 0.36 (0.11) (control), 2.45 (0.80) (high fluoride) mg/L	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significantly lower mean IQ in the endemic versus non-endemic regions, as reported in Xiang et al. (2003a); when high-exposure group was broken into four exposure groups based on fluoride intake, a dose-dependent decrease in IQ and increase in % with low IQ observed; significant correlation between total fluoride intake and IQ ($r = -0.332$); for IQ <80, adjusted OR of total fluoride intake per 1-mg/(person/day) was 1.106 (95% CI: 1.052, 1.163) Adjusted for age and sex
Choi et al. (2015)	Cross-sectional Mianning County/1st grade children [51]	Drinking water GM: 2.20 mg/L Children's urine GM: 1.64 mg/L Severity of fluorosis (Dean Index)	Children (ages 6–8 years)	IQ: WISC-IV (block design and digit span)	Compared to normal/questionable fluorosis, presence of moderate/severe fluorosis significantly associated with lower total (adjusted $\beta = -4.28$; 95% CI: −8.22, −0.33) and backward (adjusted $\beta = -2.13$; 95% CI: −4.24, −0.02) digit span scores; linear associations between total digit span and log-transformed urinary fluoride (adjusted $\beta = -1.67$; 95% CI: −5.46, 2.12) and log-transformed drinking water fluoride (adjusted $\beta = -1.39$; 95% CI: −6.76, 3.98) observed but not significant; forward digit span had similar results as backward and total but was not statistically significant; block design (square root transformed) not significantly associated with any measure of fluoride exposure Adjusted for age and sex, parity, illness before 3 years old, household income last year, and caretaker's age and education

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Zhang et al. (2015b)	Cross-sectional Tianjin City (Jinnan District)/school children [180]	Drinking water Mean: 0.63 (control), 1.40 (endemic fluorosis) mg/L (SD not reported) Children's urine Mean (SD): 1.1 (0.67) (control), 2.4 (1.01) (endemic fluorosis) mg/L Children's serum Mean (SD): 0.06 (0.03) (control), 0.18 (0.11) (endemic fluorosis) mg/L	Children (ages 10–12 years)	IQ: Combined Raven's Test for Rural China	Significant correlation between IQ score and children's serum fluoride ($r = -0.47$) and urinary fluoride ($r = -0.45$); significant difference in mean IQ score for high-fluoride area (defined as >1 mg/L in drinking water; 102.33 ± 13.46) compared with control area (109.42 ± 13.30); % of subjects with IQ <90 significantly increased in high-fluoride area (28.7%) vs. low-fluoride area (8.33%); not significantly correlated with water fluoride Adjusted for age and sex, if applicable
Cui et al. (2018)	Cross-sectional Tianjin City (districts Jinghai and Dagang)/school children [323]	Children's urine Median (Q1–Q3): 1.3 (0.9–1.7) mg/L (boys), 1.2 (0.9–1.6) mg/L (girls)	Children (ages 7–12 years)	IQ: Combined Raven's Test for Rural China	Significant association between IQ score and log-transformed urinary fluoride (adjusted $\beta = -2.47$; 95% CI: −4.93, −0.01) Adjusted for age, mother's education, family member smoking, stress, and anger

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Yu et al. (2018)[e,f]	Cross-sectional Tianjin City (7 towns)/children [2,886]	Drinking water Mean (SD): 0.50 (0.27) (normal), 2.00 (0.75) (high) mg/L Children's urine Mean (SD): 0.41 (0.49) (normal), 1.37 (1.08) (high) mg/L	Children (ages 7–13 years)	IQ: Combined Raven's Test for Rural China	Significant difference in mean IQ scores in high water fluoride areas (>1.0 mg/L; 106.4 ± 12.3 IQ) compared to the normal water fluoride areas (≤1.0 mg/L; 107.4 ± 13.0); distribution of the IQ scores also significantly different (p = 0.003); every 0.5-mg/L increase in water fluoride was associated with a decrease of 4.29 in IQ score (95% CI: −8.09, −0.48) when exposure was between 3.40 and 3.90 mg/L; no significant association between 0.2 and 3.40 mg/L; every 0.5-mg/L increase in urinary fluoride was associated with a decrease of 2.67 in IQ score (95% CI: −4.67, −0.68) between 1.60 and 2.50 mg/L but not at levels of 0.01–1.60 mg/L or 2.50–5.54 mg/L. Adjusted for age and sex, maternal education, paternal education, and low birth weight
Cui et al. (2020)	Cross-sectional Tianjin City (all districts)/school children (potentially some overlap with Cui et al. (2018)) [498]	Children's urine <1.6–≥2.5 mg/L	Children (ages 7–12 years)	IQ: Combined Raven's Test	Decreasing mean (± SD) IQ score with increasing urinary fluoride levels (statistical significance not reached based on a one-way ANOVA) <1.6 mg/L: 112.16 ± 11.50 1.6–2.5 mg/L: 112.05 ± 12.01 ≥2.5 mg/L: 110 ± 14.92 No statistical adjustment for covariates

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Wang et al. (2020b)[e]	Cross-sectional Tianjin City (villages not specified)/school children [571]	Drinking water Mean (SD): 1.39 (1.01) mg/L Children's urine Mean (SD): 1.28 (1.30) mg/L	Children (ages 7–13 years)	IQ: Combined Raven's Test for Rural China	Significant associations between IQ and water and urinary fluoride concentrations in boys and girls combined based on both quartiles and continuous measures (water: 1.587 decrease in IQ score per 1-mg/L increase; urine: 1.214 decrease in IQ score per 1-mg/L increase); no significant effect modification of sex Adjusted for age and sex, BMI, maternal education, paternal education, household income, and low birth weight
Mexico					
Rocha-Amador et al. (2007)	Cross-sectional Moctezuma and Salitral in San Luis Potosi State and 5 de Febrero of Durango State /elementary school children [132]	Drinking water Mean (SD): 0.8 (1.4), 5.3 (0.9), 9.4 (0.9) mg/L (3 rural areas) Children's urine Mean (SD): 1.8 (1.5), 6.0 (1.6), 5.5 (3.3) mg/L (3 rural areas)	Children (ages 6–10 years)	IQ: WISC-Revised Mexican Version	Significant associations between log-transformed fluoride and IQ scores (full IQ adjusted βs of −10.2 [water] and −16.9 [urine]; CIs not reported); arsenic also present, but the association with arsenic was smaller (full-scale IQ adjusted βs of −6.15 [water] and −5.72 [urine]; CIs not reported) Adjusted for blood lead, mother's education, SES, height-for-age z-scores, and transferrin saturation

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Bashash et al. (2017)	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [299] IQ analysis [211]	Maternal urine during pregnancy Mean (SD): 0.90 (0.35) mg/L Children's urine Mean (SD): 0.82 (0.38) mg/L	Children (ages 6–12 years)	IQ: WASI-Spanish Version	Significantly lower child IQ score per 0.5-mg/L increase in maternal urinary fluoride (adjusted β = −2.50; 95% CI: −4.12, −0.59); no significant association with children's urine Adjusted for sex, gestational age; weight at birth; parity (being the first child); age at outcome measurement; and maternal characteristics, including smoking history (ever smoked during the pregnancy vs. nonsmoker), marital status (married vs. not married), age at delivery, education, IQ, and cohort
Soto-Barreras et al. (2019)	Cross-sectional Chihuahua/school children [161]	Children's urine Range: 0.11–2.10 mg/L Drinking water Range: 0.05–2.93 mg/L Fluoride exposure dose (summary statistics not reported) Fluorosis index (summary statistics not reported)	Children (ages 9–10 years)	IQ: Raven's Colored Progressive Matrices	No significant difference in urinary fluoride, drinking water fluoride, fluoride exposure dose, or fluorosis index in subjects across different IQ grades No statistical adjustment for covariates

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Canada					
Green et al. (2019)[g]	Cohort (prospective) 10 cities/Maternal-Infant Research on Environmental Chemicals (MIREC) [512] Non-fluoridated [238] Fluoridated [162] Boys [248] Girls [264]	Maternal urine during pregnancy Mean (SD): 0.51 (0.36) mg/L (0.40 [0.27] mg/L in non-fluoridated areas and 0.69 [0.42] mg/L in fluoridated areas) Maternal fluoride intake during pregnancy Mean (SD): 0.54 (0.44) mg/day (0.30 [0.26] and 0.93 [0.43] mg/day, respectively) Drinking water Mean (SD): 0.31 (0.23) mg/L (0.13 [0.06] and 0.59 [0.08] mg/L, respectively)	Children (ages 3–4 years)	IQ: full-scale, performance, and verbal using Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	Significantly lower full-scale IQ (adjusted β = −4.49; 95% CI: −8.38, −0.60) and performance IQ (adjusted β = −4.63; 95% CI: −9.01, −0.25) per 1-mg/L increase in maternal urinary fluoride in boys but not girls (adjusted β = 2.40; 95% CI: −2.53, 7.33 and adjusted β = 4.51; 95% CI: −1.02, 10.05, respectively) or boys and girls combined (adjusted β = −1.95; 95% CI: −5.19, 1.28 and adjusted β = −1.24; 95% CI: −4.88, 2.40, respectively); significantly lower full-scale IQ (adjusted β = −3.66; 95% CI: −7.16, −0.15) per 1-mg increase in maternal fluoride intake (no sex interaction); significantly lower full-scale IQ (adjusted β = −5.29; 95% CI: −10.39, −0.19) per 1-mg/L increase in water fluoride concentration (no sex interaction); no significant associations observed between measures of fluoride and verbal IQ Adjusted for sex, city, HOME score, maternal education, race, and prenatal secondhand smoke exposure

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Till et al. (2020)[g]	Cohort (prospective) 10 cities/ MIREC [398] Non-fluoridated [247] Fluoridated [151] Breastfed as infants [200] Formula-fed as infants [198]	Drinking water Mean (SD) For breastfed infants: 0.13 (0.06) mg/L in non-fluoridated areas and 0.58 (0.08) mg/L in fluoridated areas For formula-fed infants: 0.13 (0.05) mg/day in non-fluoridated areas and 0.59 (0.07) mg/L in fluoridated areas Infant fluoride intake Mean (SD) For breastfed infants: 0.02 (0.02) mg/day in non-fluoridated areas and 0.12 (0.07) mg/day in fluoridated areas For formula-fed infants: 0.08 (0.04) mg/day in non-fluoridated areas and 0.34 (0.12) mg/day in fluoridated areas Maternal urine during pregnancy	Children (ages 3–4 years)	IQ: full-scale, performance, and verbal using Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	Drinking water Breastfed infants: Lower (not significant) full-scale IQ (adjusted β = −1.34, 95% CI: −5.04, 2.38) per 0.5-mg/L increase in water fluoride concentration; significantly lower performance IQ (adjusted β = −6.19, 95% CI: −10.45, −1.94) Formula-fed infants: Significantly lower full-scale IQ (adjusted β = −4.40, 95% CI: −8.34, −0.46) per 0.5-mg/L increase in water fluoride concentration; significantly lower performance IQ (adjusted β = −9.26, 95% CI: −13.77, −4.76) Infant fluoride intake Breastfed: No results reported Formula-fed: Lower (not significant) full-scale IQ (adjusted β = −2.69, 95% CI: −709, 3.21) per 0.5-mg/L increase in fluoride intake from formula; significantly lower performance IQ (adjusted β = −8.76, 95% CI: −14.18, −3.34) Maternal urine during pregnancy+

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
		Mean (SD) Breastfed: 0.42 (0.28) mg/L in non-fluoridated areas and 0.70 (0.39) mg/L in fluoridated areas Formula-fed: 0.38 (0.27) mg/L in non-fluoridated areas and 0.64 (0.37) mg/L in fluoridated areas			Lower (not significant) full-scale IQ (adjusted β = −1.08, 95% CI: −1.54, 0.47) per 0.5-mg/L increase in maternal urinary fluoride++; lower (not significant) performance IQ (adjusted β = −1.31, 95% CI: −3.63, 1.03)++ Lower (not significant) performance IQ (adjusted β = −1.50, 95% CI: −3.41, 0.43) per 0.5-mg/L increase in maternal urinary fluoride+++; significantly lower full-scale IQ (adjusted β = −2.38, 95% CI: −4.62, −0.27)+++ No association between verbal IQ scores and any measure of fluoride exposure +Maternal urinary fluoride analyzed as covariate in the drinking water and infant fluoride intake from formula models and not in an individual model ++After additional adjustment for drinking water and breastfeeding status +++After additional adjustment for infant fluoride intake from formula All models adjusted for maternal education, maternal race, age at IQ testing, sex, HOME total score, and secondhand smoke status in the child's home (separate analysis also adjusted for mother's urinary fluoride)

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
India					
Sudhir et al. (2009)	Cross-sectional Nalgonda District (Andhra Pradesh)/school children [1,000]	Drinking water Level 1: <0.7 mg/L Level 2: 0.7–1.2 mg/L Level 3: 1.3–4.0 mg/L Level 4: >4.0 mg/L	Children (ages 13–15 years)	IQ: Raven's Standard Progressive Matrices	Significant increase in mean and distributions of IQ grades (i.e., increase in proportion of children with intellectual impairment) with increasing drinking water fluoride levels No statistical adjustment for covariates
Saxena et al. (2012)	Cross-sectional Madhya Pradesh/school children [170]	Drinking water ≥1.5 mg/L (high fluoride group) Children's urine Range: 1.7–8.4 mg/L	Children (age 12 years)	IQ: Raven's Standard Progressive Matrices	Significant correlations between IQ grade and water (r = 0.534) and urinary (r = 0.542) fluoride levels; in adjusted analyses, significant increase in mean IQ grade (i.e., increase in proportion of children with intellectual impairment) with increasing urinary fluoride; no significant differences in the levels of urinary lead or arsenic in children with the different water fluoride exposure levels Covariates included in the analysis were not reported
Trivedi et al. (2012)	Cross-sectional Kachchh, Gujarat/school children (6th and 7th grades) [84]	Mean (SE) Low-fluoride villages: drinking water: 0.84 (0.38) mg/L Children's urine: 0.42 (0.23) mg/L High fluoride villages: drinking water: 2.3 (0.87) mg/L Children's urine: 2.69 (0.92) mg/L	Children (ages 12–13 years)	IQ: questionnaire prepared by Professor JH Shah (97% reliability rating)	Significantly lower mean IQ score in high fluoride villages (92.53 ± 3.13) compared to the low-fluoride villages (97.17 ± 2.54); differences significant for boys and girls combined, as well as separately No statistical adjustment for covariates

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[b,c]
Iran					
Seraj et al. (2012)	Cross-sectional Makoo/school children [293]	Drinking water Mean (SD): 0.8 (0.3) (normal), 3.1 (0.9) (medium), 5.2 (1.1) (high) mg/L	Children (ages 6–11 years)	IQ: Raven's Colored Progressive Matrices	Significant association between water fluoride and IQ score (adjusted β = −3.865 per 1-mg/L increase in water fluoride); CIs not reported); significantly higher mean IQ score in normal area (97.77 ± 18.91) compared with medium (89.03 ± 12.99) and high (88.58 ± 16.01) areas Adjusted for age, sex, child's education level, mother's education level, father's education level, and fluorosis intensity

ANOVA = analysis of variance; GM = geometric mean; HOME = Home Observation Measurement of the Environment; IQ = intelligence quotient; Q1, Q3 = first and third quartiles; SD = standard deviations; WASI = Wechsler Abbreviated Scale of Intelligence (Spanish version); WISC-IV = Wechsler Intelligence Scale for Children-Revised (as reported by Choi et al. 2015).

[a]Includes low risk-of-bias studies.

[b]Associations between IQ and fluoride levels were reported quantitatively, when possible. For studies with multiple analyses and results, the table summarizes key findings and is not a comprehensive summary of all findings. Results also indicate when a study reported no association between IQ and fluoride, provided as a qualitative statement of no association.

[c]See Figure A-1 through Figure A-8 for additional study results.

[d]Xiang et al. (2003a), Xiang et al. (2011), and Wang et al. (2012) are based on the same study population.

[e]Yu et al. (2018) and Wang et al. (2020b) are based on the same study population.

[f]Three additional publications based on a subsample (i.e., 50–60 children) of the larger Yu et al. (2018) cohort were identified (Zhao et al. 2019; Zhao et al. 2020; Zhou et al. 2019); however, these publications focused on mechanistic considerations and are not included in the study totals for IQ because the main study by Yu et al. (2018) is considered a better representation of the IQ results.

[g]Green et al. (2019) and Till et al. (2020) are based on the same study population.

Summary of Results

Overall Findings

The results from 18 of the 19 high-quality (low risk-of-bias) studies (3 longitudinal prospective cohort studies from 2 different study populations and 15 cross-sectional studies from 13 different study populations) that evaluated IQ in children provide consistent evidence that higher fluoride exposure is associated with lower IQ scores (see "Summary of IQ Results" in Table 6) (Bashash et al. 2017; Choi et al. 2015; Cui et al. 2018; Ding et al. 2011; Green et al. 2019; Rocha-Amador et al. 2007; Saxena et al. 2012; Seraj et al. 2012; Sudhir et al. 2009; Till et al. 2020; Trivedi et al. 2012; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Xiang et al. 2011; Yu et al. 2018; Zhang et al. 2015b). Only one study (Soto-Barreras et al. 2019) did not observe an association between fluoride exposure and IQ; however, results were not provided in a manner that allowed for a direct comparison with other low risk-of-bias studies (see Appendix E for details). A strength of the findings across 18 of 19 low risk-of-bias studies was the consistent association between higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximated or exceeded the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] and lower IQ scores among studies of varying study designs, exposure measures, and study populations. In studies that analyzed the sexes separately (n = 5 studies with 2 studies reporting on the same study population), consistent findings of lower IQ associated with fluoride exposure were generally reported for both sexes. There is some indication of differential susceptibility between sexes, but ultimately, due to too few high-quality studies that analyzed exposure and outcome by sex separately and a lack of consistent findings that one sex is more susceptible, it is unclear whether one sex is more susceptible to the effects of fluoride exposure than the other. The body of evidence from the 19 low risk-of-bias studies is described in further detail below. Prospective cohort studies are discussed first, as this study design can establish a temporal relationship between exposure and outcome, which would contribute to demonstrating causality and, therefore, providing the strongest evidence for an association between fluoride exposure during development and IQ in children.

Results by Study Design – Prospective Cohort Studies

As noted above, three longitudinal prospective cohort studies, conducted in Mexico and Canada, were identified and considered to reflect a low risk for bias. All three prospective cohort studies found an association between increasing maternal or child fluoride exposure and lower IQ in children (Bashash et al. 2017; Green et al. 2019; Till et al. 2020). Two of the studies (Green et al. 2019; Till et al. 2020) were based on the same Canadian study population, but one evaluated prenatal fluoride exposure and the other evaluated postnatal fluoride exposure. Green et al. (2019) included maternal urinary fluoride, maternal fluoride intake, and water fluoride concentrations, while Till et al. (2020) used fluoride intake from formula or water concentrations in formula-fed versus breastfed infants. Multiple analyses were conducted in each prospective study, and results by analysis for the three prospective studies are discussed below. In summary, although not every analysis found a statistically significant association, together the three studies provided consistent evidence that increasing maternal fluoride levels were associated with lower IQ scores in the children.

In the Early Life Exposures in Mexico to Environmental Toxicants cohort, Bashash et al. (2017) observed a statistically significant association (p-value = 0.01) between lower IQ scores in children and prenatal fluoride exposure measured by maternal urinary fluoride (measured during

all three trimesters and included if at least one measurement was available). An increase of 0.5 mg/L of maternal urinary fluoride was associated with a 2.5-point decrease in IQ score [95% CI: −4.12, −0.59] in boys and girls combined (see Figure A-8). This study also reported an inverse association between IQ level and children's urinary fluoride levels (single spot urine sample); however, this specific result did not achieve statistical significance (a 0.5-mg/L increase of child urinary fluoride was associated with a 0.89-point decrease in IQ score [95% CI: −2.63, 0.85]) (Bashash et al. 2017).

In the Maternal-Infant Research on Environmental Chemicals cohort, consisting of 10 cities in Canada, Green et al. (2019) also reported inverse associations between IQ scores in children and multiple measures of prenatal fluoride exposure, including maternal urinary fluoride, maternal fluoride intake, and water fluoride concentrations. Green et al. (2019) observed a statistically significantly lower IQ for boys associated with maternal urinary fluoride averaged across trimesters (4.49-point decrease in IQ score [95% CI: −8.38, −0.60; p-value = 0.02] per 1-mg/L increase in maternal urinary fluoride); however, results were not significant in boys and girls combined (1.95-point decrease in IQ [95% CI: −5.19, 1.28]) and were positive but not significant in girls (2.40-point increase in IQ [95% CI: −2.53, 7.33]). Other measures of prenatal exposure (maternal fluoride intake or water fluoride concentrations) were associated with lower IQ scores in boys and girls combined; the authors found no significant effect measure modification between child sex and fluoride exposure in these analyses so they did not report boys and girls separately (Green et al. 2019). Specifically, when evaluating the association between estimated maternal fluoride intake based on maternal water and beverage consumption during pregnancy and IQ in children, a 1-mg increase in daily maternal consumption of fluoride during pregnancy was associated with a significantly decrease in IQ score of 3.66 points in boys and girls combined (95% CI: −7.16, −0.15; p-value = 0.04). Similarly, water fluoride concentrations for pregnant women from fluoridated areas (mean water fluoride levels of 0.59 ± 0.08 mg/L) versus pregnant women from non-fluoridated areas (mean water fluoride levels of 0.13 ± 0.06 mg/L) were associated with a significant 5.29-point decrease in IQ score per 1-mg/L increase in fluoride in both boys and girls combined (95% CI: −10.39, −0.19; p-value <0.05) (Green et al. 2019).

In a study of the same study population as Green et al. (2019) that used fluoride intake from formula or water concentrations in formula-fed versus breastfed infants, Till et al. (2020) observed significantly lower performance IQ scores with higher fluoride regardless of the comparison used (p-values ≤0.004). They did not observe any association with verbal IQ, and full-scale IQ was only significantly lower in formula-fed infants using water fluoride concentrations as the exposure measure (p-value = 0.03). Breastfed infants and fluoride intake from formula also showed inverse associations but were not significant.

Taken together, the three prospective cohort studies (based on two North American study populations) indicate consistency in results across different types of analysis and across two study populations that higher fluoride exposure during development is associated with lower IQ scores.

Results by Study Design – Cross-sectional Studies

As with the prospective cohort studies, the cross-sectional studies reported a consistent association between fluoride exposure and lower IQ scores in children. Fifteen of the 16 low risk-of-bias cross-sectional studies [i.e., all with the exception of Soto-Barreras et al. (2019)]

consistently demonstrate that exposure to fluoride is associated with lower IQ scores. Fourteen of these 15 studies [with the exception of Cui et al. (2020)] reported significant associations.

Cross-sectional studies can have limitations, as the study design often cannot ensure that exposure preceded outcome. This uncertainty reduces confidence in study findings compared with prospective cohort studies—which, by design, establish that exposure occurred prior to outcome—and is captured in the outcome assessment. In some cases, cross-sectional studies do provide indicators of prior exposure (e.g., prevalence of dental fluorosis, limiting study populations to subjects who lived in the same area for long periods of time). Evidence that exposure occurred prior to the outcome of interest increases the confidence in results and any potential association reported in these studies. Of the 16 low risk-of-bias cross-sectional studies, 12 established that exposure preceded the outcome assessment (Choi et al. 2015; Ding et al. 2011; Rocha-Amador et al. 2007; Saxena et al. 2012; Seraj et al. 2012; Soto-Barreras et al. 2019; Sudhir et al. 2009; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Xiang et al. 2011; Yu et al. 2018). Five studies from different study populations indicated that a large portion of the exposed children had dental fluorosis (ranging from 43% to 100%) at the time of assessment (Choi et al. 2015; Ding et al. 2011; Seraj et al. 2012; Sudhir et al. 2009; Yu et al. 2018). Because dental fluorosis occurs when fluoride is consumed during enamel formation (usually during the first 6–8 years of life), the presence of dental fluorosis suggests that exposures to fluoride occurred prior to the outcome assessment. Nine studies from six study populations (including Yu et al. (2018) and Sudhir et al. (2009) listed above) excluded subjects who had not lived in the study area for a specified period of time, sometimes since birth (Rocha-Amador et al. 2007; Saxena et al. 2012; Soto-Barreras et al. 2019; Sudhir et al. 2009; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Xiang et al. 2011; Yu et al. 2018). Because these areas were generally known to be fluoride-endemic for long periods of time, it can generally be assumed that in these nine studies, exposure occurred prior to the outcome. Taken together, 12 cross-sectional studies from 9 study populations provide indicators of prior exposure.

Results by Study Design – Cross-sectional Study Variations

Overall, the cross-sectional studies consistently provide evidence that fluoride exposure is associated with lower IQ scores in children. Several cross-sectional studies conducted multiple analyses (e.g., reported results for multiple exposure metrics, endpoints, subpopulations). Although some of these variations are heterogeneous and are not comparable across studies, the consistency of the results across multiple metrics contributes to the confidence in the data. Table 6 summarizes key results for each of the low risk-of-bias cross-sectional studies, and a few examples of the within-study variations in results are provided below.

Nine cross-sectional studies (from six study populations) assessed the association between IQ and multiple exposure measures (Choi et al. 2015; Rocha-Amador et al. 2007; Saxena et al. 2012; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Xiang et al. 2011; Yu et al. 2018; Zhang et al. 2015b). Lower IQ was consistently observed across exposure measures in these studies; however, Choi et al. (2015), a small pilot study (n = 51), did not achieve statistical significance in all results by exposure measure. Specifically, the authors reported a consistent association between all fluoride exposure measures assessed (drinking water, children's urine, and severity of fluorosis) and digit span measures (subtest of the WISC-IV omnibus IQ test); however, results were only statistically significant when fluoride exposure was based on moderate or severe dental fluorosis in children (see Figure A-7). Choi et al. (2015) also observed

some variation in results by outcome assessed (i.e., square root transformed block design and digit span [forward, backward, and total]). It was the only cross-sectional study that did not provide a full IQ score but instead provided results by specific subtests. The study authors consistently observed an inverse association between fluoride exposure and results from the digit span subtest (which specifically assesses executive function); however, results from the block design (square root transformed), a subtest of the WISC-IV omnibus IQ test that specifically assesses visuospatial function, was not associated with fluoride exposure. Note that Rocha-Amador et al. (2009) also assessed visuospatial function, and the authors reported a significant association (p-value <0.001) between fluoride exposure and decreased visuospatial constructional ability using the Rey-Osterrieth Complex Figure (ROCF) Test. Ultimately, too few studies were identified that reported results by subtest of omnibus IQ tests or assessed domains other than IQ (e.g., visuospatial function) to examine or explain the variation by outcome observed in Choi et al. (2015). The only other studies that provided a breakdown of the full IQ score were the prospective cohort studies by Green et al. (2019) and Till et al. (2020), which provided results for full-scale IQ as well as results for performance and verbal IQ. In both of these studies, lower verbal IQ was not associated with fluoride exposure, but lower performance and full-scale IQ were associated with fluoride exposure. There are too few studies to evaluate whether there is a specific aspect of IQ testing that is affected by exposure to fluoride, but the studies nonetheless consistently provide evidence that fluoride exposure is associated with lower IQ.

Yu et al. (2018) reported an overall association between lower IQ and higher fluoride exposure across multiple analyses but observed some variation in IQ results by urinary exposure level. The authors reported inverse associations between IQ and children's medium- and high-range urinary fluoride levels (1.60–2.50 mg/L and 2.50–5.54 mg/L, respectively), although change in IQ score was greater in the medium-range group (2.67 points decrease [95% CI: −4.67, −0.68]) for every 0.5-mg/L increase of urinary fluoride than in the high-range group (0.84 points decrease [95% CI: −2.18, 0.50]) (see Figure A-7). No association was reported at low-range urinary fluoride levels (0.01–1.60 mg/L). Note that Yu et al. (2018) also reported an inverse association between IQ and drinking water fluoride levels at 3.40–3.90 mg/L (4.29-point decrease in IQ score [95% CI: −8.09, −0.48]) for every 0.5-mg/L increase in water fluoride; a 0.04-point decrease in IQ score [95% CI: −0.33, 0.24] was observed for 0.5-mg/L increase in water fluoride at levels of 0.20–3.40 mg/L). The variation by exposure level in urine could not be verified in the analysis of drinking water exposures because there were only two water exposure groups (low and high). In a second study (Wang et al. 2020b), authors conducted a categorical analysis using urinary fluoride quartiles with reported betas per quartile. As observed in Yu et al. (2018), there were decreasing trends in IQ within each quartile; however, unlike Yu et al. (2018), Wang et al. (2020b) observed a larger decrease in IQ with each increasing urinary quartile and observed similar results using water fluoride quartiles (Wang et al. 2020b). Note that Wang et al. (2020b) cannot be compared directly to Yu et al. (2018) for evaluation at the higher exposure levels because the two studies do not use the same categorical exposure ranges. Although additional studies may have looked at different exposure levels, none of these studies provided results in the same manner as Yu et al. (2018) and Wang et al. (2020b) (i.e., betas by exposure category). Instead, these other studies provided an overall beta or mean IQ scores by exposure level. Despite the noted variations among these studies, the overall results still consistently support an association between fluoride exposure and lower IQ.

Two studies (Cui et al. 2018; Zhang et al. 2015b) observed associations between lower IQ in children and exposure to fluoride, with variations in results in subpopulations of children with different polymorphisms (see Figure A-7). These were the only two studies that considered polymorphism as a sub-analysis. Cui et al. (2018) observed a significant association between log-transformed children's single spot urinary fluoride and lower IQ scores (2.47-point decrease in IQ scores [95% CI: −4.93, −0.01; p-value = 0.049] per ln-mg/L increase in urinary fluoride), and the association was strongest in subjects with a TT polymorphism (compared with children with a CC or CT polymorphism) in the dopamine receptor D2 (DRD2) gene (12.31-point decrease in IQ score [95% CI: −18.69, −5.94; p-value <0.001] per ln-mg/L increase in urinary fluoride), which, according to the authors, probably resulted in a reduced D2 receptor density (Cui et al. 2018). Similarly, Zhang et al. (2015b) observed a significant association between lower IQ scores and children's single spot urinary fluoride (2.42-point decrease in IQ scores [95% CI: −4.59, −0.24; p-value = 0.030] per 1-mg/L increase in urinary fluoride), and the association was strongest in subjects with a val/val polymorphism (compared with children who carried the heterozygous or homozygous variant genotypes [met/val or met/met]) in the catechol-O-methyltransferase (COMT) gene (9.67-point decrease in IQ score [95% CI: −16.80, −2.55; p-value = 0.003] per 1-mg/L increase in urinary fluoride).

Overall, the cross-sectional studies consistently support a pattern of findings that higher fluoride exposure is associated with lower IQ scores in children. Slight within-study variations occur that may be associated with study variables such as IQ domains or subsets of IQ tests in a few studies that conducted multiple analyses, but these variations are heterogenous and cannot be further explored with the available studies. Despite these few variations, the overall evidence of an association with lower IQ is apparent.

Exposure Measure and Study Population Factors

Low risk-of-bias studies provide consistent evidence that higher fluoride exposure is associated with lower IQ scores across studies using different exposure measures. In addition to water fluoride levels, studies measured fluoride exposure using single serum samples in children (Xiang et al. 2011; Zhang et al. 2015b), single spot urine samples in children (Cui et al. 2018; Ding et al. 2011; Rocha-Amador et al. 2007; Saxena et al. 2012; Wang et al. 2020b; Xiang et al. 2003a; Yu et al. 2018; Zhang et al. 2015b), and prenatal maternal urinary measures (Bashash et al. 2017; Green et al. 2019), all of which were demonstrated to be consistently associated with lower IQ scores (see Figure A-6, Figure A-7, and Figure A-8). Urine levels encompass all sources of fluoride exposure and provide a better measure of the totality of exposure. As noted previously, even though some studies measured single spot samples, which may not be representative of peak exposure, these studies generally provided evidence that fluoride exposure had been occurring for some time. The consistency in the results across studies that used different measures of fluoride exposure and different life stages at which fluoride was measured strengthens the body of evidence.

The low risk-of-bias studies consistently provide evidence that higher fluoride exposure is associated with lower IQ scores across studies of different study populations. These 19 high-quality studies represent diverse populations (n = 15 study populations) across 5 countries. Eighteen of the 19 studies conducted in Canada (n = 2), China (n = 10), India (n = 3), Iran (n = 1), and Mexico (n = 2) provide evidence that exposure to fluoride is associated with lower IQ scores; 1 study conducted in Mexico did not observe an association but reported results in a

manner that did not allow for a direct comparison with the other studies (see Appendix E for details). The overall consistency in the study results across study populations adds strength to the body of evidence.

Exposure Levels

As described in this section, the body of evidence for studies assessing the association between fluoride exposure and IQ in children consistently provides evidence of an association between higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] and lower IQ in children; however, there is less certainty in the evidence of an association in populations with lower fluoride exposures. In the September 6, 2019, draft of this monograph, NTP conducted a qualitative analysis of children's IQ studies that 1) evaluated lower fluoride exposures (<1.5 mg/L) in drinking water and/or urine and 2) provided information to evaluate dose response (i.e., provided three or more fluoride exposure groups or a dose-response curve in their publication) in the lower fluoride exposure range. Nine low risk-of-bias studies met these criteria, which includes the three prospective cohort studies discussed in this section. Based on the qualitative review of these studies, the evidence of an association between fluoride exposure below 1.5 mg/L and lower IQ in children appeared less consistent than results of studies at higher exposure levels.

A draft quantitative dose-response meta-analysis was prepared and included in the September 16, 2020, draft monograph (NTP 2020). This meta-analysis is undergoing further refinement in preparation for separate publication and may further inform a discussion on the association between fluoride exposure levels and IQ in children.

Sex Considerations

Recent literature suggests that adverse neurodevelopmental effects of early-life exposure to fluoride may differ depending on timing of exposure and sex of the exposed subject. In a review of the human and animal literature, Green et al. (2020) concluded that, compared with females, male offspring appear to be more sensitive to prenatal but not postnatal exposure to fluoride, with several potential sex-specific mechanisms.

Sex differences were examined in five of the low risk-of-bias studies (in four study populations) (Green et al. 2019; Trivedi et al. 2012; Wang et al. 2012; Wang et al. 2020b; Xiang et al. 2003a). In general, sex differences were difficult to assess for trends within different study populations because few studies in the body of evidence analyzed exposure and stratified results by sex. Although these five studies reported IQ scores separately for boys and girls, only two of these studies analyzed fluoride exposure for boys and girls separately (Green et al. 2019; Wang et al. 2020b), which is essential for evaluating whether a differential change in IQ by sex may be related to higher susceptibility in one sex or higher exposure in that sex. The remaining three studies stratified results by sex (Trivedi et al. 2012; Wang et al. 2012; Xiang et al. 2003a), but the analyses were based on area-level exposure data (e.g., low-fluoride village compared with high fluoride village) and not drinking water or urinary fluoride concentrations. In the five studies that reported results by sex separately, consistent findings of lower IQ associated with fluoride exposure were generally reported for both sexes. There was some variation in the results between sexes across study populations and exposure measures, but there is insufficient evidence

to determine whether one sex is more susceptible to the effects of fluoride exposure than the other.

Green et al. (2019) observed a significant inverse association between maternal urinary fluoride levels and IQ scores in boys (p-values ≤0.04) but not girls in a Canadian population. Green et al. (2019) did not find any sex differences in the association between IQ and water fluoride concentrations. Wang et al. (2020b) evaluated Chinese boys and girls separately and combined and observed statistically significant decreasing trends in IQ in all groups by urinary fluoride quartiles (p-values for trend ≤0.035) (see Figure A-7). Similarly, when evaluated as a continuous variable, spot urinary fluoride levels (per 1-mg/L increase) were significantly associated with lower IQ scores in girls (−1.379 [95% CI: −2.628, −0.129; p-value = 0.031]), boys (−1.037 [95% CI: −2.040, −0.035; p-value = 0.043]), and in the sexes combined (−1.214 [95% CI: −1.987, −0.442; p-value = 0.002]). According to water fluoride quartiles, Wang et al. (2020b) found that there was a significant trend in the sexes combined, although the decreasing trend in boys and girls separately did not achieve statistical significance (p-values = 0.077 and 0.055, respectively). When water fluoride levels were evaluated as a continuous variable (per 1-mg/L increase), there were significant associations with lower IQ scores in girls (−1.649 [95% CI: −3.201, −0.097]; p-value = 0.037), boys (−1.422 [95% CI: −2.792, −0.053; p-value = 0.042]), and the sexes combined (−1.587 [95% CI: −2.607, −0.568]; p-value = 0.002).

The remaining three studies that reported results by sex-based comparisons of areas of high and low urinary or water fluoride did not report exposure levels separately for boys and girls, which decreases the utility of the data to evaluate differential susceptibility by sex. Trivedi et al. (2012) observed significantly lower IQ in children in high fluoride Indian villages compared with low-fluoride villages with decreases observed in boys and girls separately or combined (p-values ≤0.05) (see Figure A-2). Xiang et al. (2003a) and Wang et al. (2012) provide data on the same study population in China. There was a significantly lower IQ in the high fluoride area compared with the low-fluoride area in boys and girls separately and in the sexes combined (p-values <0.01), although the difference was greater in girls. Because fluoride exposure was not analyzed for boys and girls separately, it is unclear whether the greater change in IQ scores in girls could be attributed to higher susceptibility to fluoride exposure or differences in fluoride exposure by sex.

In summary, it is unclear whether one sex is more susceptible to the effects of fluoride exposure than the other due to the limited number of studies that analyzed exposure and outcome by sex and the lack of a consistent pattern of findings that one sex is more susceptible. Green et al. (2019) did not observe an association between maternal urinary fluoride levels and IQ scores in girls but did observe a significant association in boys. Although this is an indication of higher sensitivity in boys in this analysis, the authors did not detect this sex difference using other measures of prenatal exposure (maternal fluoride intake or water fluoride concentrations). Wang et al. (2020b) and Trivedi et al. (2012) reported statistically significant associations in both boys and girls without indication that one sex may be more susceptible. Although Xiang et al. (2003a) and Wang et al. (2012) reported a greater change in IQ in girls than boys, the studies used area-level exposure data, and the authors did not determine whether fluoride exposure differed in boys versus girls. Therefore, it is unclear whether this differential result by sex is an indication of higher susceptibility in girls or whether it could be explained by a difference in exposure by sex. Overall, there are too few studies that analyzed exposure and outcome by sex separately to properly evaluate whether there is differential susceptibility to fluoride exposure by sex, and

results from the five low risk-of-bias studies that do evaluate sex differences indicate that there is no consistent difference by sex across the different study populations.

Summary of Key Findings for Low Risk-of-bias Children's IQ Studies

In summary, the high-quality studies (i.e., studies with low potential for bias) consistently demonstrate lower IQ scores with higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)]. The consistency in association is observed among studies of varying study designs, exposure measures, and study populations. Although some studies that conducted multiple analyses observed within-study variations in results (e.g., differences between subsets of IQ tests), these variations were unique to individual studies and did not detract from the overall consistency in the findings that higher fluoride is associated with lower IQ scores.

High Risk-of-bias IQ Studies

The results from 53 studies with high potential for bias that evaluated IQ in children also consistently provide supporting evidence of decrements in IQ associated with exposures to fluoride. Forty-six of the 53 studies reported an association between high fluoride exposure and lower IQ scores in children.

Risk of Bias for IQ Studies in Children

The confidence in the human body of evidence was based on studies with the lowest potential for bias. A total of 19 studies on IQ in children had little or no risk-of-bias concerns, representing a relatively large body of evidence for low risk-of-bias studies (i.e., 15 study populations across 5 countries evaluating more than 7,000 children). These 19 studies are considered low risk of bias because they were rated probably low or definitely low risk of bias for at least two of the three key risk-of-bias questions and did not have any other risk-of-bias concerns that would indicate serious issues with the studies. Thirteen of the 19 studies were rated definitely low or probably low risk of bias for all risk-of-bias questions, and the remaining 6 studies were rated probably high risk of bias for a single question that was judged to have minimal impact on overall potential for bias. None of the 19 studies had a rating of definitely high risk of bias for any question. Risk-of-bias ratings for individual studies for all questions are available in Figure D-1 through Figure D-4, with risk-of-bias ratings for IQ studies in children available in Figure D-5 through Figure D-8 and Appendix E. Although the low risk-of-bias studies had minimal or no concerns, the studies with high overall potential for bias had a number of risk-of-bias concerns, including potential confounding, poor exposure characterization, poor outcome assessment, and, in many cases, potential concern with participant selection. The key risk-of-bias questions are discussed below.

Confounding for IQ Studies in Children

Low Risk-of-bias Studies

As discussed above, there are 19 studies considered to have low risk of bias when assessed across all risk-of-bias domains. Sixteen of the 19 low risk-of-bias studies [i.e., all with the exception of Cui et al. (2020), Ding et al. (2011), and Soto-Barreras et al. (2019)] were considered to have low potential for bias due to confounding because the authors addressed the three key covariates for all studies (i.e., age, sex, and socioeconomic status) through study design

or analysis. Other important covariates, including health factors, smoking, and parental characteristics, were also addressed in many of the low risk-of-bias studies (see Figure 6).

Co-exposures to arsenic and lead were not considered a concern in 18 of 19 low risk-of-bias studies [i.e., all except for Soto-Barreras et al. (2019)] because the studies addressed the potential co-exposures, the co-exposures were not considered an issue in the study population, or the impact of the potential bias on the results was not a concern. Fifteen of 19 low risk-of-bias studies either addressed potential bias related to co-exposure to arsenic through study design or analysis or co-exposure to arsenic was unlikely in the study area. All 15 studies observed an association between lower IQ and fluoride exposure. Co-exposure to arsenic was not accounted for in the remaining four low risk-of-bias studies and was the main potential concern in these studies; however, three of these studies (Wang et al. 2012; Xiang et al. 2003a; Xiang et al. 2011) were still considered low risk of bias for confounding because although arsenic was observed in the water in the low-fluoride (and not the high-fluoride) comparison areas, which would bias the association toward the null, an association was still observed. In this case, the lack of adjustment for arsenic strengthens the evidence for an association and does not represent a potential concern. The other study did not address arsenic co-exposure and, as noted above, was conducted in an area that had potential for arsenic exposure to occur (Soto-Barreras et al. 2019); it is also the only low risk-of-bias study that did not observe an association between lower IQ and fluoride exposure (see Appendix E for further discussion of the risk-of-bias concern regarding arsenic for this study). Although Soto-Barreras et al. (2019) did not discuss arsenic, there is no direct evidence that arsenic was present in the study area. Fourteen studies accounted for co-exposure to lead through study design or analysis, and all observed an association between lower IQ and fluoride exposure. Five studies did not consider co-exposure to lead; however, for all of these studies, co-exposure to lead was considered unlikely to have an impact in these study populations as there was no evidence that lead was prevalent or occurring in relation to fluoride (Cui et al. 2018; Cui et al. 2020; Soto-Barreras et al. 2019; Till et al. 2020; Trivedi et al. 2012).

There is considerable variation in the specific covariates considered across the 19 low risk-of-bias studies. The consistency of results across these studies suggests that confounding is not a concern in this body of evidence. Each of the 18 low risk-of-bias studies that observed an association between fluoride and IQ (see Summary of Results section above) considered a unique combination of covariates. The findings of these studies consistently provide evidence of an association between lower IQ in children and exposure to fluoride regardless of the inclusion or absence of consideration of any one or combination of covariates of interest. For example, maternal or family member smoking was addressed in 7 of the 19 low risk-of-bias studies, and this did not appear to affect the conclusions. All 7 studies that accounted for smoking found evidence of an association between fluoride exposure and lower IQ scores as did 11 of the 12 studies that did not account for smoking. Similarly, all 16 studies that addressed the three key covariates (age, sex, SES) (16 of 16 studies) and two of the three studies that did not fully account for them also found evidence of an association between fluoride exposure and lower IQ scores. In summary, when considering the impact of each covariate (or combinations of covariates) on the consistency of results, no trends are discernable that would suggest that bias due to confounding has impacted or would explain the consistency in findings across the body of evidence that fluoride exposure is associated with lower IQ in children.

Five of the low risk-of-bias studies confirmed the robustness of the results by conducting sensitivity analyses (Bashash et al. 2017; Green et al. 2019; Till et al. 2020; Wang et al. 2020b;

Yu et al. 2018), and none of the sensitivity analyses adjusting for additional covariates found meaningful shifts in the association between fluoride exposure and IQ or other measures of cognitive function. Bashash et al. (2017) found that adjusting for HOME score increased the association between maternal urinary fluoride and children's IQ. Green et al. (2019) reported that adjusting for lead, mercury, manganese, perfluorooctanoic acid, and arsenic concentrations did not substantially alter the associations with IQ. Sensitivity analyses by Yu et al. (2018) that adjusted for covariates (including age, sex, and socioeconomic status) did not find differences in the results compared with the primary analyses. Wang et al. (2020b) found the results of the sensitivity analysis to be the same as the results from the primary analysis. Till et al. (2020) observed that adjusting for maternal urinary fluoride levels, as a way to consider postnatal exposure, had little impact on the results.

Among the 19 low risk-of-bias studies, three were identified that have potential for bias due to confounding (Cui et al. 2020; Ding et al. 2011; Soto-Barreras et al. 2019). This was mainly due to a lack of details on covariates considered key for all studies (i.e., age, sex, and SES). See Appendix E for further discussion of the risk-of-bias concerns regarding confounding for individual studies. Although these three studies have some potential for bias due to confounding, they are considered to be low risk of bias overall because they have low potential for bias for the other two key risk-of-bias questions (exposure characterization and outcome assessment), and no other major concerns for bias were identified. Consistent with the 16 studies that adequately addressed confounding, two of these three studies also provide evidence of an association between fluoride exposure and lower IQ scores in children.

Taken together and considering the consistency in the results despite the variability across studies in which covariates were accounted for, bias due to confounding is not considered to be a concern in the body of evidence. The potential for the consistency in results to be attributable to bias due to confounding in the 19 low risk-of-bias studies is considered low.

Study (Location)[a]	Potential Covariates Considered[b]																Notes	Reported Association with Fluoride[e]
	Subject Characteristics				Other Exposures					Socioeconomic Factors		Parental Characteristics				Other[c]		
	Age	Sex	Race/Ethnicity	Health Factors[c]	Arsenic	Smoking	Iodine	Lead	Other[c]	SES[d]	Caregiving Environment (e.g., HOME score)	Demographics[c]	Reproductive Factors[c]	Health Factors[c]	IQ			
Overall RoB Rating for Confounding: Probably Low																		
Bashash 2017 (Mexico)	√	√	–	–	√	√	–	√	√	√	√	√	√	–	√	√	Other exposures: Hg, Ca Demographics: maternal age Reproductive: parity, birth order, birth weight, gestational age at delivery Other: cohort	Yes
Choi 2015 (China)	√	√	–	√	√	–	–	√	–	√	–	√	√	√	–	√	Health: subject Fe deficiency, illnesses before age 3, medical history of subject and caretakers Demographics: parental age Reproductive: parity Other: residential history	Yes
Cui 2018 (China)	√	√	√	√	√	√	√	–	–	√	–	√	√	√	–	√	Health: subject BMI, stress/anger/anxiety/depression, psychological trauma, having a cold, in relatives: thyroid diseases, cancer, mental retardation Demographics: maternal age Reproductive: abnormal birth Other: alcohol consumption, proximity to factory, physical activity, various dietary factors, environmental noise	Yes
Green 2019 (Canada)	√	√	√	–	√	√	–	√	√	√	√	√	√	–	–	√	Other exposures: Hg, Mn, PFOA Demographics: parental age, pre-pregnancy BMI Reproductive: parity, weeks of gestation, birth weight, maternal chronic condition during pregnancy Other: alcohol consumption, birth country, voiding interval in urine sampling, breastfeeding duration	Yes
Rocha-Amador 2007 (Mexico)	√	√	–	√	√	–	–	√	–	√	–	–	–	–	–	–	Health: subject height and weight by age, transferrin saturation	Yes
Saxena 2012 (India)	√	√	–	√	√	–	√	√	–	√	–	–	–	–	–	√	Health: subject height for age ratio, weight for height ratio, medical history Other: residential history	Yes
Seraj 2012 (Iran)	√	√	–	–	√	–	√	√	–	√	–	–	–	–	–	√	Other: fluorosis intensity	Yes
Sudhir 2009 (India)	√	√	–	–	√	–	–	√	–	√	–	–	–	–	–	√	Other: staple food consumed, liquids routinely consumed, aids used for oral hygiene maintenance (fluoridated or nonfluoridated)	Yes
Till 2020 (Canada)	√	√	√	–	√	√	–	–	–	√	√	–	–	–	–	√	Other: city	Yes
Trivedi 2012 (India)	√	√	–	–	√	–	√	–	–	√	–	–	–	–	–	–		Yes
Wang 2012 (China)	√	√	–	√	–	–	√	√	–	√	–	–	–	√	–	√	Health: medical conditions Other: transportation, natural environment, and lifestyle	Yes
Wang 2020b (China)	√	√	–	√	√	√	√	√	–	√	–	–	√	–	–	√	Health: subject BMI, exclusions based on diseases affecting intelligence, history of trauma or neurological disorders, positive screening test history Reproductive: low birth weight Other: alcohol consumption	Yes
Xiang 2003 (China)	√	√	–	–	–	–	√	√	–	√	–	–	–	–	–	–		Yes
Xiang 2011 (China)	√	√	–	–	–	–	√	√	–	√	–	–	–	–	–	–		Yes
Yu 2018 (China)	√	√	–	√	√	√	√	√	√	√	–	–	√	–	–	√	Health: subject BMI Reproductive: disease history during pregnancy, delivery conditions Other: dental fluorosis prevalence, consanguineous marriage	Yes
Zhang 2015b (China)	√	√	–	√	√	–	√	√	√	√	–	–	–	–	–	√	Health: subject physical and mental health status Other: thyroid hormone levels, residential history, having knowledge of fluorosis, COMT genotype	Yes
Overall RoB Rating for Confounding: Probably High																		
Cui 2020 (China)	–	√	–	√	√	√	√	–	–	√	–	√	√	√	–	√	Health: stress/anger/anxiety, having a cold, in relatives: mental retardation Demographics: maternal age Reproductive: abnormal birth, smoking and drinking during pregnancy Other: thyroid hormone levels	Yes[f]
Ding 2011 (China)	√	–	–	–	√	–	√	√	–	–	–	–	–	–	–	–		Yes
Soto-Barreras 2019 (Mexico)	√	√	–	–	–	–	–	–	–	√	–	–	–	–	–	–		No

Figure 6. Important Covariates Considered in Low Risk-of-bias IQ Studies Conducted in Children

[a]Includes all low risk-of-bias IQ studies in children. Studies are organized as those with an overall risk-of-bias rating for confounding as probably low (green) followed by those with an overall risk-of-bias rating for confounding as probably high (yellow).
[b]Covariates represented here are those considered important for this evaluation. Depending on the specific study population, individual covariates may be considered a potential confounder, effect measure modifier, and/or co-exposure. See study details provided in HAWC for information on additional covariates.
Factors outlined in blue are key covariates for all studies (subject age, subject sex, SES) and arsenic (which is of particular importance to some study populations).

A √ indicates that a covariate was considered. Examples of what it means for a covariate to be "considered": it was adjusted for in the final model, it was considered in the model but not included in the final model because it did not change the effect estimate, it was reported to have the same distribution in both the exposed and unexposed groups, it was reported to not be associated with the exposure or outcome in that specific study population. For arsenic, a √ might also be used when arsenic was not expected to be an issue because there is no evidence to indicate that the co-exposure was prevalent or occurring in relation to fluoride. See risk-of-bias explanations in Appendix E (or HAWC) for details. A hyphen (-) indicates that the factor was not considered.
[c]See the "Notes" column for additional details.
[d]Covariates considered measures of SES include SES scaled scores, household/family income, child education, caretaker/parental education, and occupation/employment.
[e]Extent of reported associations varies by study. "Yes" indicates that study authors provided evidence of an association between lower IQ scores and fluoride exposure.
[f]Study reported lower IQ scores with increasing fluoride exposure, but the results did not achieve statistical significance.

High Risk-of-bias Studies

Most high risk-of-bias studies (n = 53) considered important covariates to some degree through study design or analysis; however, when considering the full scale of potential concerns of bias due to confounding, all but three of these studies were rated probably or definitely high risk of bias. The majority of high risk-of-bias studies accounted for one or two of the three covariates considered key for all studies (age, sex, SES) but did not address all three and did not address other covariates considered important for the specific study population and outcome. Potential confounding related to important co-exposures (e.g., arsenic) was often not addressed in high risk-of-bias studies. In studies in which there was high exposure to fluoride via drinking water with high naturally occurring fluoride or from the use of coal-containing fluoride, most researchers did not account for potential exposures to arsenic, which is commonly found in coal and drinking water in fluoride-endemic areas of China and Mexico.

Despite the lack of adequate consideration of key covariates in the vast majority of high risk-of-bias studies, the results across most of these studies (46 of 53) consistently provide evidence of an association between fluoride exposure and IQ, supporting the results observed in the low risk-of-bias studies. This finding suggests that confounding is likely less of a concern for the body of evidence as a whole than for any individual study. Although the high risk-of-bias studies may have more potential for bias due to confounding compared with the low risk-of-bias studies, the consistent IQ findings across high and low risk-of-bias studies indicate that the results cannot be explained solely by potential bias due to confounding.

Exposure Characterization in IQ Studies

Low Risk-of-bias Studies

In general, there were few, if any, risk-of-bias concerns regarding exposure characterization in the low risk-of-bias studies. These studies mainly had individual exposure data based on urine or water measures with appropriate analyses. Although there are concerns related to using urine samples (see the Risk-of-bias Considerations for Human Studies section for details), the evidence suggests that urinary fluoride is a reasonable measure of exposure (Villa et al. 2010; Watanabe et al. 1995). Using three methods to account for urine dilution, Till et al. (2018) reported that adjusted risk estimates did not differ from unadjusted estimates. Analyzing the same study population as Till et al. (2018), Green et al. (2019) found that adjusting for time of urine collection or time of collection since last void during pregnancy did not substantially affect associations with IQ results in either boys or girls. In addition, adjusting maternal urinary fluoride for creatinine did not substantially alter the observed association (Green et al. 2019). To provide a more accurate and sensitive measurement of maternal urinary fluoride than a single measurement provides, Green et al. (2019) included only participants with valid fluoride

measurements at all trimesters in their analysis. Other studies also measured urinary fluoride multiple times throughout pregnancy (Bashash et al. 2017). Some studies demonstrated correlations between urinary fluoride and fluoride in drinking water, fluorosis, or estimated dose based on drinking water concentrations and consumption (Choi et al. 2015; Ding et al. 2011; Green et al. 2019; Saxena et al. 2012; Yu et al. 2018; Zhang et al. 2015b). Till et al. (2018) demonstrated that there was a linear association between urinary fluoride concentrations in pregnant women and drinking water fluoride concentrations regardless of method used to correct for urine dilution or whether adjustments were made for dilution. Bashash et al. (2017) excluded exposure outliers and found that doing so did not substantively change the results. Taken together, these studies suggest that urinary fluoride is a reasonable measure of exposure despite some potential issues.

All but one low risk-of-bias study was rated probably or definitely low risk of bias for exposure assessment. Seraj et al. (2012) had potential exposure misclassification and was rated probably high risk of bias for exposure assessment. Villages were categorized as normal (0.5–1 ppm), medium (3.1 ± 0.9 ppm), or high (5.2 ± 1.1 ppm) based on average fluoride content in drinking water in varying seasons over a 12-year period. Mild fluorosis observed in children in the normal fluoride level group indicates that there may have been higher exposure in this group at some point in the past; however, this would bias the results toward the null, and the children in the normal fluoride group had a significantly higher IQ score compared with the medium and high fluoride groups (p-value = 0.001). There were also significant associations between lower IQ scores and fluorosis intensity (p-value = 0.014) and water fluoride concentration when evaluated as a continuous variable (p-values <0.001). Although there is potential for exposure bias, the apparent exposure misclassification and inclusion of children with higher fluoride exposure in the normal group indicate that the association may be greater than what was observed in this study.

High Risk-of-bias Studies

A frequent, critical limitation among the high risk-of-bias studies was lack of information regarding exposure or poor exposure characterization. Many of the high risk-of-bias studies compared only subjects living in two regions with differing levels of fluoride exposure, and although most of them did provide some differentiation in levels of fluoride between the areas, limited or no individual exposure information was reported. Among studies that provided drinking water levels of fluoride in two areas being compared, sufficient information to determine whether the individual study subjects were exposed to these levels was often not reported. Some studies also lacked information on fluoride analysis methods and timing of the exposure measurements. In some cases (n = 3), study areas that were considered endemic for dental and/or skeletal fluorosis were compared with non-endemic areas, or high-fluoride areas were compared with low-fluoride areas, with no other information provided on fluoride levels in the areas (Li et al. 2003 [translated in Li et al. 2008c]; Ren et al. 1989 [translated in Ren et al. 2008]; Sun et al. 1991). Although living in an area endemic for fluorosis could be an indicator of exposure, these studies did not specify whether the study subjects themselves had fluorosis. Another study used only dental fluorosis as a measure of fluoride exposure in subjects who were all from an endemic area with similar drinking water fluoride levels (Li et al. 2010). In one case, multiple sources of fluoride exposure were assessed separately without properly controlling for the other sources of exposure, which could bias the results (Broadbent et al. 2015). Broadbent et al. (2015) assessed fluoride exposure in three ways: use of community water in a fluoridated area

versus a non-fluoridated area, use of fluoride toothpaste (never, sometimes, always), or use of fluoride tablets prior to age 5 (ever, never). The same children were used for each analysis without accounting for fluoride exposure through other sources. For example, there were 99 children included in the non-fluoridated area for the community water evaluation, but there is no indication that these 99 children were not some of the 139 children that had ever used supplemental fluoride tablets or the 634 children that had always used fluoride toothpaste. Therefore, comparing fluoridated areas to non-fluoridated areas without accounting for other sources of exposure that might occur in these non-fluoridated areas would bias the results toward the null.

Outcome Assessment for IQ Studies

Low Risk-of-bias Studies

The low risk-of-bias studies have few concerns regarding outcome assessment. All 19 low risk-of-bias studies used appropriate methods for measuring IQ in the study population being assessed, and blinding of outcome assessors was not a concern in 18 of the 19 studies [i.e., all low risk-of-bias studies except Sudhir et al. (2009)]. Fourteen of these 18 studies reported blinding of the outcome assessors, or correspondence with the study authors confirmed that it was not likely an issue. For the remaining 4 of the 18 studies, it was assumed that the outcome assessors were most likely blind because exposure was assessed via urine or drinking water obtained at the same time as the outcome assessment in the general population studies. One IQ study (Sudhir et al. 2009) had concerns for potential bias in the outcome assessment due to lack of information to determine whether blinding at the time of the outcome assessment was a concern (see Appendix E for details).

High Risk-of-bias Studies

Among the studies with high risk of bias, the main limitation in the outcome assessment was the lack of reporting on blinding of the outcome assessor (i.e., whether the outcome was assessed without knowledge of exposure). Although there is little concern that the children's knowledge of their own exposure would bias the way they took the IQ tests, there is potential for bias if the tests were administered by an interviewer, or if the scoring of results could be subjective (e.g., drawing tests), and the interviewer or scorer had knowledge of the children's exposure. Most of the studies did not provide sufficient information on the person scoring or administering the tests or other information on the assessment methods to alleviate concerns for potential interviewer or reviewer bias.

High risk-of-bias studies were mainly carried out in two separate populations without information provided that the tests were conducted in a central location. In many cases, the methods indicated that the tests were conducted at the schools in the study area (indicating that there was likely knowledge of exposure). In some cases, the outcomes were not considered sensitive measures (e.g., Seguin Form Board Test to test for IQ), or the test was not considered appropriate for the study population (e.g., a test validated in a western population was used on a rural Chinese population).

Confidence Assessment of Findings on IQ in Children

We conclude that there is moderate confidence in the body of evidence that higher fluoride exposure is associated with lower IQ in children. This confidence rating was reached by starting

with an initial confidence rating based on key study design features of the body of evidence and then considering factors that may increase or decrease the confidence in that body of evidence. The initial moderate confidence rating is based on 15 of the 19 low risk-of-bias studies that have 3 of the 4 key study design features shown in Figure 1 (i.e., exposure occurred prior to outcome, individual-based outcomes were evaluated, and a comparison group was used). Three of these studies were prospective cohort studies, and 12 were cross-sectional studies that provided evidence of long-term, chronic fluoride exposure prior to outcome measurement.

There are nine factors to consider for increasing or decreasing the confidence in the body of evidence (provided in Figure 1). Discussion of each of these factors in the body of evidence on fluoride exposure and IQ in children is presented below.

- **Risk of bias**: Only studies that were considered to have low risk of bias were included in the moderate confidence rating; therefore, there was no downgrade for risk-of-bias concerns.
- **Unexplained inconsistencies:** The data are consistent, and there was no downgrade for this factor. Eighteen of the 19 low risk-of-bias studies reported associations between higher fluoride levels and lower IQ scores in children. These studies were conducted in 5 different countries on more than 7,000 children from 15 different study populations. There is consistency in results across prospective and cross-sectional study designs. There is also consistency in results across studies using different fluoride exposure measures, including urinary and drinking water fluoride. The one study that did not observe an association did not provide results in a comparable manner and therefore this body of evidence is not considered to have unexplained inconsistencies.
- **Indirectness:** IQ in humans is a direct measure of the association of interest; therefore, no adjustment in confidence is warranted.
- **Imprecision:** There is no evidence of imprecision that would warrant a downgrade. Eighteen studies reported lower IQ with higher fluoride, and no issues with imprecision were identified to challenge the significance of the effect estimate.
- **Publication bias:** There is no strong evidence of publication bias; therefore, no downgrade was applied for publication bias. Two published meta-analyses (Choi et al. 2012; Duan et al. 2018) did not indicate strong evidence of publication bias. The draft meta-analysis conducted by NTP in the September 16, 2020, draft monograph found no publication bias among the low risk-of-bias studies (NTP 2020). Among high risk-of-bias studies, adjusting for publication bias using the trim-and-fill analysis estimated that, in the absence of publication bias, the inverse direction of association and statistical significance remained, thus indicating that there was no need to downgrade for publication bias.
- **Large magnitude of effect size:** Although some individual studies indicated a large magnitude of effect size, the magnitude of effect was not the same across all studies. Therefore, the overall data would not support an upgrade due to a large magnitude of effect size.
- **Dose response:** Evidence of an exposure-response relationship that could justify an upgrade to the confidence in the body of evidence is not presented in this monograph.

While the overall findings qualitatively appear less clear in the lower exposure range, many of the studies that provide data to evaluate exposure response were judged to be high risk of bias. The meta-analysis conducted in association with this systematic review further informs this issue and will be published separately.

- **Residual confounding:** Xiang et al. (2003a), Xiang et al. (2011), and Wang et al. (2012) studied the same population where arsenic occurred in the area with low fluoride but did not occur in the area with high fluoride. This would have biased the results toward the null, but there were significantly lower IQ scores in the area with high fluoride. The remaining studies do not provide enough information to consider whether residual confounding occurred for the body of evidence. Note that parental IQ has the potential to be an important factor when considering residual confounding based on likely correlations between parental IQ and children's IQ; however, there is not sufficient evidence that parental IQ is associated with water fluoride content. Taken together, the overall data would not support an upgrade due to residual confounding.
- **Consistency:** The consideration of a potential upgrade for consistency in the methods is primarily for non-human animal evidence, where it would be applied to address increased confidence for consistent effects across multiple non-human animal species. For human evidence, it is generally not applied, and the data would only be considered in deciding whether to downgrade for unexplained inconsistency. Therefore, no upgrade is applied for consistency.

As described above, there are no changes in confidence rating based on any of the possible upgrade or downgrade factors. The magnitude of effect size and the overall strength and quality of the human literature base provide moderate confidence in the body of evidence that higher exposure to fluoride is associated with lower IQ in children (see the Discussion section for strengths and limitations of the evidence base). Note that additional, well-designed prospective cohort studies with individual-level exposure data and outcome measures could provide increased confidence in the association between fluoride exposure and lower IQ in children.

Other Neurodevelopmental or Cognitive Effects in Children

Low Risk-of-bias Studies

Overview of Studies

Nine low risk-of-bias studies (three prospective cohort and six cross-sectional studies) evaluated the association between fluoride exposure and cognitive neurodevelopmental effects other than IQ in children. These nine studies were conducted in multiple study populations in three countries, specifically:

- three were conducted in three areas of China on three study populations,
- four were conducted in two areas of Mexico on three study populations, and
- two were conducted in Canada using the same study population.

There is considerable heterogeneity across studies, particularly in the different health outcomes evaluated and ages assessed. Most studies measured fluoride in the drinking water or urine (child or maternal) with one study using severity of dental fluorosis as an exposure measure in addition

to drinking water and children's urine. Two of the studies were conducted on infants, with one evaluating effects within 72 hours of birth (Li et al. 2004 [translated in Li et al. 2008a]) and the other evaluating effects at 3 to 15 months of age (Valdez Jimenez et al. 2017). The remaining studies were conducted in children of varying ages, ranging from 4 to 17 years. Other cognitive neurodevelopmental outcomes assessed include neurobehavioral effects in infants, learning and memory impairment, and learning disabilities such as attention deficit hyperactivity disorder (ADHD). Few studies measured the same health outcomes, used the same outcome assessment methods, or evaluated the same age groups.

Table 7 provides a summary of study characteristics and key findings related to other cognitive neurodevelopmental outcomes and fluoride exposure for the nine low risk-of-bias studies. The different tests conducted and the populations on which the tests were conducted are also indicated in Table 7. Several of these studies conducted multiple analyses and reported results on multiple endpoints. The purpose of the table is to summarize key findings (independent of whether an association was found) from each study and is not meant to be a comprehensive summary of all results. For each study, results are summarized for each exposure measure assessed. Results from multiple analyses using the same exposure measure may not all be presented unless conflicting results were reported. See Appendix E for additional information on studies in Table 7, including strengths and limitations, clarifications for why they are considered to pose low risk of bias, and information regarding statistical analyses, covariates, exposure assessment, and outcome assessment.

Table 7. Studies on Other Neurodevelopmental and Cognitive Function in Children[a]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
China					
Li et al. (2004) [translated in Li et al. 2008a]	Cross-sectional Zhaozhou County, Heilongjiang Province/neonates [91]	Drinking water Range: 0.5–1.0 mg/L (control); 1.7–6.0 mg/L (high) Maternal urine during pregnancy Mean (SD): 1.74 (0.96) mg/L (control); 3.58 (1.47) mg/L (high)	Neonates (24–72 hours after delivery)	Neurodevelopmental: Neonatal behavioral neurological assessment (NBNA)	Significant differences in neurobehavioral assessment total scores between high-fluoride (36.48 ± 1.09) and control groups (38.28 ± 1.10) (subjects divided into high fluoride group and control group based on drinking water fluoride levels in place of residence); significant differences in total score of behavioral capability that includes measures of non-biological visual orientation reaction and biological visual and auditory orientation reaction between the two groups (11.34 ± 0.56 in controls compared to 10.05 ± 0.94 in high-fluoride group) No statistical adjustment for covariates
Choi et al. (2015)	Cross-sectional Mianning County/1st grade children [51]	Drinking water GM: 2.20 mg/L Children's urine GM: 1.64 mg/L Severity of fluorosis (Dean Index)	Children (ages 6–8 years)	Learning and memory: Neuropsychological tests including WRAML Visual motor ability: WRAVMA Motor ability: Finger tapping task Manual dexterity: Grooved pegboard test	Outcomes unrelated to the IQ test not significantly associated with any fluoride exposure measure Adjusted for age, sex, parity, illness before 3 years old, household income last year, and caretaker's age and education
Wang et al. (2020a)	Cross-sectional Tongxu County/school children [325]	Children's urine Mean (SD): 1.54 (0.89) mg/L	Children (ages 7–13 years)	ADHD and behavior measures: Conners' Parent Rating Scale-Revised (Chinese version) (CPRS-48)	Significant association between psychosomatic problems and urinary fluoride level (per 1-mg/L increase; β = 4.01; 95% CI: 2.74, 5.28; OR for T-score >70 = 1.97; 95% CI: 1.19, 3.27); no associations between urinary fluoride level and ADHD index or other behavioral measures Adjusted for age, sex, child's BMI, urinary creatinine, mother migrated, and father migrated

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
Mexico					
Rocha-Amador et al. (2009)	Cross-sectional Durango/elementary school children [80]	Children's urine GM (SD): 5.6 (1.7) mg/L	Children (ages 6–11 years)	Visuospatial organization and visual memory: Rey-Osterrieth Complex Figure Test, children's version	Significant correlation between urinary fluoride and visuospatial organization ($r = -0.29$) and visual memory scores ($r = -0.27$); no significant correlation with arsenic Adjusted for age
Valdez Jimenez et al. (2017)	Cohort (Prospective) Durango City and Lagos de Moreno/infants [65]	Maternal urine Range: 0.16–8.2 mg/L (all trimesters) Drinking water Range: 0.5–12.5 mg/L (all trimesters)	Infants (ages 3–15 months)	Mental development index (MDI): Bayley Scales of Infant Development II (BSDI-II) Psychomotor developmental index (PDI): Bayley Scales of Infant Development II (BSDI-II)	Significant association between log10-mg/L maternal urinary fluoride and MDI score during first trimester (adjusted $\beta = -19.05$; SE = 8.9) and second trimester (adjusted $\beta = -19.34$; SE = 7.46); no significant associations between maternal urinary fluoride and PDI score; analyses of outcomes using drinking water fluoride not performed Adjusted for age, gestational age, marginality index, and type of drinking water
Bashash et al. (2017)[c]	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [299] GCI analysis [287]	Maternal urine during pregnancy Mean (SD): 0.90 (0.35) mg/L Children's urine Mean (SD): 0.82 (0.38) mg/L	Children (age 4 years)	General cognitive index (GCI): McCarthy Scales of Children's Abilities (MSCA)	Significant association between maternal urinary fluoride and offspring GCI score (per 0.5-mg/L increase adjusted $\beta = -3.15$; 95% CI: −5.42, −0.87); associations with children's urine not significant Adjusted for gestational age; weight at birth; sex; parity (being the first child); age at outcome measurement; and maternal characteristics, including smoking history (ever smoked during the pregnancy vs. nonsmoker), marital status (married vs. not married), age at delivery, IQ, education, and cohort

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
Bashash et al. (2018)[c]	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [210]	Maternal urine during pregnancy Mean 0.85 (95% CI: 0.81, 0.90) mg/L	Children (ages 6–12 years)	ADHD: Conners' Rating Scales-Revised (CRS-R)	Significant associations between maternal urinary fluoride (per 0.5-mg/L increase) and CRS-R scores, including Cognitive Problems + Inattention Index (adjusted β = 2.54; 95% CI: 0.44, 4.63), DSM-IV Inattention Index (adjusted β = 2.84; 95% CI: 0.84, 4.84), DSM-IV ADHD Total Index (adjusted β = 2.38; 95% CI: 0.42, 4.34), and ADHD Index (adjusted β = 2.47; 95% CI: 0.43, 4.50) Adjusted for gestational age; birth weight; sex; parity; age at outcome measurement; and maternal characteristics, including smoking history (ever smoked vs. nonsmoker), marital status (married vs. not married), education, socioeconomic status, and cohort
Canada					
Barberio et al. (2017b)[d]	Cross-sectional General population/Canadian Health Measures Survey (Cycles 2 and 3) [2,221]	Children's urine Mean Cycle 2: 32.06 (95% CI: 29.65, 34.46) µmol/L Mean Cycle 3: 26.17 (95% CI: 22.57, 29.76) µmol/L	Children (ages 3–12 years)	Learning disability, ADHD (Cycle 2 only): Parent or child self-report	Significant increase in adjusted OR for learning disability (adjusted OR = 1.02; 95% CI: 1.00, 1.03) per 1-µmol/L increase in unadjusted urinary fluoride when Cycle 2 and 3 were combined; no significant associations found between urinary fluoride and ADHD (only evaluated in Cycle 2); no significant associations found when using creatinine- or specific gravity-adjusted urinary fluoride Adjusted for age and sex, household income adequacy, and highest attained education in the household

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
Riddell et al. (2019)[d]	Cross-sectional General population/Canadian Health Measures Survey (Cycles 2 and 3) [3,745]	Drinking water Mean (SD): 0.23 (0.24) mg/L [non-fluoridated water: 0.04 (0.06) mg/L; fluoridated water: 0.49 (0.22)] Community water fluoridation status (yes or no) Children's urine Mean (SD): 0.61 (0.39) mg/L [non-fluoridated water: 0.46 (0.32) mg/L; fluoridated water: 0.82 (0.54)]	Children (ages 6–17 years)	Hyperactivity/inattention: Strengths and Difficulties Questionnaire (SDQ) ADHD: parent or self-reported physician diagnosis	Significantly increased risk of ADHD with fluoride in tap water (adjusted OR = 6.10 per 1-mg/L increase; 95% CI: 1.60, 22.8) or community water fluoridation status (1.21; 95% CI: 1.03, 1.42) but not with urinary fluoride; similar results observed with attention symptoms based on the SDQ scores Adjusted for age and sex, child's BMI, ethnicity, parental education, household income, blood lead, and smoking in the home

ADHD = attention-deficit/hyperactivity disorder; BMI = body mass index; GCI = General Cognitive Index; GM = geometric mean; HOME = Home Observation Measurement of the Environment; IQ = intelligence quotient; MSCA = McCarthy Scales of Children's Abilities; SD = standard deviation; WASI = Wechsler Abbreviated Scale of Intelligence (Spanish version); WISC-IV = Wechsler Intelligence Scale for Children-Revised (as reported by Choi et al. 2015); WRAML = Wide Range Assessment of Memory and Learning; WRAVMA = Wide Range Assessment of Visual Motor Ability.

[a]Includes low risk-of-bias studies.

[b]Associations between other cognitive neurodevelopmental outcomes in children and fluoride levels were reported quantitatively, when possible. For studies with multiple analyses and results, the table summarizes key findings and is not a comprehensive summary of all findings. Results also indicated when a study reported no association, provided as a qualitative statement of no association.

[c]Bashash et al. (2017) and Bashash et al. (2018) are based on the same study population.

[d]Barberio et al. (2017b) and Riddell et al. (2019) are based on the same study population.

Summary of Results

Overall Findings

Although discussed together in this section, various health outcomes were assessed in the nine low risk-of-bias studies of other neurodevelopmental outcomes, including neurobehavioral scores in infants (two studies), cognitive tests in children other than IQ (three studies), and ADHD or learning disabilities (four studies) in children. Altogether, the results from eight of nine low risk-of-bias studies (three prospective cohort studies and five cross-sectional studies from seven different study populations) provide evidence of significant associations between fluoride exposure and cognitive neurodevelopmental outcomes in children other than decrements in IQ (see Figure A-9 through Figure A-11) (Barberio et al. 2017b; Bashash et al. 2017; Bashash et al. 2018; Li et al. 2004 [translated in Li et al. 2008a]; Riddell et al. 2019; Rocha-Amador et al. 2009; Valdez Jimenez et al. 2017; Wang et al. 2020a). Only one cross-sectional study did not find a significant association between fluoride exposure and a measure of cognitive neurodevelopment (Choi et al. 2015).

Although there is heterogeneity in the outcomes assessed and a limited number of directly comparable studies, the data provide additional evidence (beyond the consistent evidence of an association between fluoride exposure and IQ) of an association between higher fluoride exposure and cognitive or neurodevelopmental effects. The body of evidence from the nine low risk-of-bias studies is described in further detail below and is grouped into outcome categories of studies that are most comparable.

Results in Infants

Two studies evaluated neurobehavioral effects in infants either shortly after birth or at 3 to 15 months of age (Li et al. 2004 [translated in Li et al. 2008a]; Valdez Jimenez et al. 2017). Both studies observed a significant association between higher fluoride exposure and lower neurobehavioral scores. In neonates (1–3 days old), the high fluoride group (3.58 ± 1.47 mg/L fluoride based on spot maternal urine collected just prior to birth) had significantly lower total neurobehavioral assessment scores (36.48 ± 1.09 versus 38.28 ± 1.10 in controls; p-value <0.05) and total behavioral capacity scores (10.05 ± 0.94 versus 11.34 ± 0.56 in controls; p-value <0.05) compared to the control group (1.74 ± 0.96 mg/L fluoride) as measured by a standard neonatal behavioral neurological assessment (NBNA) method (Li et al. 2004 [translated in Li et al. 2008a]). In infants 3 to 15 months of age, the Mental Development Index (MDI)—which measures functions including hand-eye coordination, manipulation, understanding of object relations, imitation, and early language development—was significantly inversely associated with maternal urinary fluoride in both the first and second trimesters (adjusted βs per log10-mg/L increase = −19.05 with standard error of 8.9 for first trimester [p-value = 0.04] and −19.34 with standard error of 7.46 for second trimester [p-value = 0.013]) (Valdez Jimenez et al. 2017). Note that this study did not find an association between maternal fluoride during any trimester and the Psychomotor Developmental Index (PDI), which measures gross motor development (adjusted βs = 6.28 and 5.33 for first and second trimesters, respectively; no standard errors provided) (Valdez Jimenez et al. 2017).

Results for Cognitive Tests Other Than IQ in Children

Three studies conducted tests on cognitive function in children that were not part of an IQ test (Bashash et al. 2017; Choi et al. 2015; Rocha-Amador et al. 2009). None of the studies

conducted the same tests, but two of the three studies (Bashash et al. 2017; Rocha-Amador et al. 2009) observed associations between fluoride exposure and lower test scores. The General Cognitive Index (GCI) of the McCarthy Scales of Children's Abilities (MSCA) in 4-year-old children was significantly inversely associated with maternal creatinine-adjusted urinary fluoride levels during pregnancy (collected during each trimester) (adjusted β per 0.5-mg/L increase = −3.15 [95% CI: −5.42, −0.87; p-value = 0.01] in a model adjusting for main covariates including gestational age, weight at birth, sex, maternal smoking, and indicators of socioeconomic status). The association remained even after adjusting for maternal bone lead (adjusted β per 0.5-mg/L increase = −5.63 [95% CI: −8.53, −2.72; p-value <0.01]) (Bashash et al. 2017) (see Figure A-11). Choi et al. (2015), however, evaluated cognitive function endpoints in addition to IQ and found no significant associations between concurrent log-transformed water or urinary fluoride levels and Wide Range Assessment of Visual Motor Ability (WRAVMA) scores, finger tapping test scores, and grooved pegboard test scores, although there were some significant associations based on degree of fluorosis (see Figure A-11). Another study using visuoconstructional and memory scores from the Rey-Osterrieth Complex Figure Test in children 6–11 years old observed significantly lower scores with increasing concurrent child single spot urinary fluoride even after adjusting for age (partial correlation coefficients, per log-mg/L increase = −0.29 and −0.27 for copy [p-value <0.001] and immediate recall [p-value <0.001], respectively [CIs not reported]) (Rocha-Amador et al. 2009). Although these children were also exposed to arsenic, the presence of arsenic could not explain the changes because, in the area with natural contamination by fluoride and arsenic (F–As), the test scores were not significantly associated with urinary arsenic levels (partial correlation coefficients, per log-mg/L increase = −0.05 and 0.02 for copy and immediate recall, respectively [CIs not reported]). The test scores were only marginally increased from fluoride alone when both fluoride and arsenic were included simultaneously in the model (partial correlation coefficients, per log-mg/L increase = −0.32 and −0.34 for copy and immediate recall, respectively [CIs not reported]) (Rocha-Amador et al. 2009) (see Figure A-10).

Attention-related Disorders Including ADHD and Learning Disabilities in Children

Four studies evaluated attention-related disorders or learning disabilities (Barberio et al. 2017b; Bashash et al. 2018; Riddell et al. 2019; Wang et al. 2020a). All four studies found an association between increased fluoride and increased ADHD or learning disability; however, studies varied in the exposure metrics and outcomes measure. Bashash et al. (2018) evaluated behaviors associated with ADHD in children ages 6–12 years using the Conners Rating Scales-Revised (CRS-R) and observed significant associations between maternal urinary fluoride (measured during each trimester) and ADHD-like symptoms, particularly those related to inattention (an increase in 0.5 mg/L of maternal urinary fluoride was significantly associated with a 2.84-point increase [95% CI: 0.84, 4.84; p-value = 0.0054] in the DSM-IV Inattention Index and a 2.54-point increase [95% CI: 0.44, 4.63; p-value = 0.0178] in the Cognitive Problems and Inattention Index). These two scales contributed to the global ADHD Index and the DSM-IV ADHD Total Index, which were also significantly associated with higher levels of prenatal fluoride exposure (an increase of 0.5 mg/L in maternal urinary fluoride was associated with a 2.38-point increase [95% CI: 0.42, 4.34; p-value = 0.0176] in the DSM-IV ADHD Total Index and a 2.47-point increase [95% CI: 0.43, 4.50; p-value = 0.0175] in the ADHD Index) (see Figure A-11). Significant associations were not observed between maternal urinary fluoride concentrations during pregnancy and child performance on measures of hyperactivity, nor were there any significant results in children using Conners' Continuous Performance Test (CPT-II,

2nd Edition), a computerized test of sustained attention and inhibitory control (Bashash et al. 2018). Wang et al. (2020a) also used Conners' Parent Rating Scale (Chinese version) to assess behavioral outcomes in children ages 7–13 years but found only a significant association between spot urinary fluoride concentrations in children (model adjusted for creatinine) and psychosomatic problems (adjusted OR for T-score >70 per 1-mg/L increase = 1.97 [95% CI: 1.19, 3.27; p-value = 0.009] and adjusted β per 1-mg/L increase = 4.01 [95% CI: 2.74, 5.28; p-value <0.001]). No associations were found between spot urinary fluoride and the ADHD index or other behavioral measures.

Barberio et al. (2017b) evaluated learning disabilities in children 3–12 years of age, including ADHD, attention deficit disorder (ADD), and dyslexia, as part of the Canadian Health Measures Survey and found a small but significantly increased risk in self-reported (children 12 years of age) or parent- or guardian-reported (children 3–11 years of age) learning disabilities associated with higher spot urinary fluoride levels in children (adjusted OR per 1-µmol/L increase = 1.02; 95% CI: 1.00, 1.03; p-value <0.05) (see Figure A-12); however, significant associations were not observed in analyses using creatinine- or specific gravity-adjusted urinary fluoride (Barberio et al. 2017b). Barberio et al. (2017b) also reported no associations between single spot urinary fluoride and ADHD in children ages 3 to 12 years. Riddell et al. (2019) used the same Canadian Health Measured Survey but evaluated children 6–17 years old. Riddell et al. (2019) found a significantly increased risk for ADHD diagnosis with both tap water fluoride (adjusted OR per 1-mg/L increase = 6.10; 95% CI: 1.60, 22.8; p-value <0.05) and community water fluoridation status (adjusted OR per 1-mg/L increase = 1.21; 95% CI: 1.03, 1.42; p-value <0.05). A similar increase in the hyperactivity-inattention symptoms score based on the Strengths and Difficulties Questionnaire was observed with both tap water fluoride (adjusted β per 1-mg/L increase = 0.31; 95% CI: 0.04, 0.58; p-value <0.05) and community fluoridation status (adjusted β per 1-mg/L increase = 0.11; 95% CI: 0.02, 0.20; p-value <0.05). As was observed with Barberio et al. (2017b), Riddell et al. (2019) did not observe associations between specific gravity-adjusted spot urinary fluoride concentrations and either ADHD diagnosis (adjusted OR per 1-mg/L increase = 0.96; 95% CI: 0.63, 1.46) or hyperactivity-inattention symptoms (adjusted β per 1-mg/L increase = 0.31; 95% CI: −0.04, 0.66).

Summary of Key Findings for Low Risk-of-bias Studies of Other Neurodevelopmental and Cognitive Effects in Children

In summary, the high-quality studies (i.e., studies with low potential for bias) provide evidence of an association between fluoride exposure and neurodevelopmental and cognitive effects in children other than IQ; however, the body of evidence is limited by heterogeneity in the outcomes evaluated and few directly comparable studies. Across these outcomes, eight of nine studies reported a significant association between fluoride exposure and a measure of neurodevelopment or cognition other than IQ, which provides support for the consistency in evidence based on children's IQ studies of an association between fluoride exposure and adverse effects on cognitive neurodevelopment.

High Risk-of-bias Studies

High risk-of-bias studies (n = 6) also provide some evidence of associations between fluoride exposure and neurodevelopmental or cognitive effects in children other than effects on IQ, but the results are inconsistent and address different outcomes (Jin et al. 2016; Li et al. 1994

[translated in Li et al. 2008b]; Malin and Till 2015; Morgan et al. 1998; Mustafa et al. 2018; Shannon et al. 1986).

Risk of Bias for Neurodevelopmental or Cognitive Effect Studies in Children

The confidence in the human body of evidence was based on studies with the lowest potential for bias (i.e., studies that rated probably low or definitely low risk of bias for at least two of the three key risk-of-bias questions and did not have any other risk-of-bias concerns that would indicate serious issues with the studies). Each of the nine low risk-of-bias studies on other neurodevelopmental effects in children had little or no risk-of-bias concerns. Four of the nine studies were rated definitely low or probably low risk of bias for all risk-of-bias questions, and the remaining five studies were rated probably high risk of bias for a single question that was judged to have minimal impact on overall potential bias. None of the nine studies had a rating of definitely high risk of bias for any question. Although the nine low risk-of-bias studies had minimal or no concerns, the six studies with high overall potential for bias had several risk-of-bias concerns related to one or more of the three key risk-of-bias questions (confounding, exposure characterization, and outcome assessment). The key risk-of-bias questions are discussed below. Risk-of-bias ratings for other neurodevelopmental effect studies in children are available in Figure D-9 through Figure D-12 and Appendix E for the low and high risk-of-bias studies.

Confounding for Other Neurodevelopmental Studies in Children

Low Risk-of-bias Studies

As discussed above, there are nine studies considered to have low risk of bias when assessed across all risk-of-bias domains. Seven of nine low risk-of-bias studies were considered to have low potential for bias due to confounding because the authors addressed the three key covariates for all studies (age, sex, and socioeconomic status) and also addressed arsenic as a potential co-exposure of concern through study design or analysis. Other important covariates, including health factors, smoking, and parental characteristics, were also addressed in many of the low risk-of-bias studies. One of the studies (Bashash et al. 2018) examined several covariates in sensitivity analyses involving subsets of participants, including HOME scores, child contemporaneous fluoride exposure measured by child urinary fluoride adjusted for specific gravity, and maternal lead and mercury exposures. The authors reported that none of the sensitivity analyses indicated appreciable changes in the fluoride-related association with behaviors related to ADHD, nor was there evidence of effect modification between maternal urinary fluoride and sex.

Among the nine low risk-of-bias studies, two studies were identified that have potential for bias due to confounding (Rocha-Amador et al. 2009; Valdez Jimenez et al. 2017). Although both of these studies adjusted for several covariates through analysis or study design, Valdez Jimenez et al. (2017) did not address a potential concern for co-exposure to arsenic, and Rocha-Amador et al. (2009) does not appear to adjust for SES or address why it would not be a concern in the study population (see Appendix E for further details). Although these two studies have some potential for bias due to confounding, they are considered to have low potential for bias overall because they have low potential for bias for the other two key risk-of-bias questions (exposure characterization and outcome assessment), and no other major concerns for bias were identified.

Consistent with the IQ studies, bias due to confounding is not likely a concern for the low risk-of-bias studies.

High Risk-of-bias Studies

The six high risk-of-bias studies in the human body of evidence did not adequately address important covariates through study design or analysis. The same concerns due to potential confounding noted previously for the high risk-of-bias children's IQ studies were also present in the other neurodevelopmental high risk-of-bias studies, including not addressing the three key covariates for all studies (age, sex, SES) and/or not addressing potential co-exposures (e.g., arsenic) in areas of potential concern.

Exposure Characterization in Other Neurodevelopmental Studies in Children

Low Risk-of-bias Studies

There were no risk-of-bias concerns regarding exposure assessment in the low risk-of-bias studies. All of the low risk-of-bias studies had individual exposure data based on urine or water measures with appropriate analyses, and most of the urinary fluoride studies accounted for urinary dilution when appropriate. Although there are concerns related to the timing of urine samples (see the Risk-of-bias Considerations for Human Studies section for details), the studies that used maternal urine measured urinary fluoride multiple times throughout pregnancy (Bashash et al. 2017; Bashash et al. 2018; Valdez Jimenez et al. 2017). Another study demonstrated correlations between urinary fluoride and fluoride in the drinking water, fluorosis, or estimated dose based on water (Choi et al. 2015). Bashash et al. (2017) excluded exposure measurement outliers but found that doing so did not change the results in a meaningful way.

High Risk-of-bias Studies

A frequent critical limitation among the high risk-of-bias studies was lack of information regarding exposure or poor exposure characterization. In the high risk-of-bias studies that assessed the association between fluoride exposure and other neurodevelopmental and cognitive effects in children, fluoride exposure assessment was based on dental fluorosis, municipality-level water fluoridation prevalence data, number of years living in an area with fluorinated water, or group-level water samples. See the Exposure Characterization in IQ Studies section for further discussion on the limitations of exposure assessments in high risk-of-bias studies.

Outcome Assessment in Other Neurodevelopmental Studies in Children

Low Risk-of-bias Studies

The low risk-of-bias studies have few concerns regarding outcome assessment. Seven of the nine studies [i.e., all low risk-of-bias studies except Barberio et al. (2017b) and Riddell et al. (2019)] used appropriate methods for measuring other neurodevelopmental effects in the study population, and blinding of outcome assessors was either reported or not a concern in eight of the nine studies [i.e., all with the exception of Wang et al. (2020a)].

Among the nine low risk-of-bias studies, three were identified that have a potential for bias due to outcome assessment. One of the studies (Wang et al. 2020a) had potential concern for bias due to lack of information regarding the blinding of outcome assessors. Two of the studies (Barberio et al. 2017b; Riddell et al. 2019) were based on the same study population in Canada, where different questions were asked in Cycles 2 (2009–2011) and 3 (2012–2013) of the Canadian

Health Measures Survey (CHMS) to ascertain learning disabilities including ADHD. In Cycle 2, subjects were asked whether they had a learning disability diagnosed by a health professional and, if yes, were asked what kind. In Cycle 3, CHMS did not ask what kind of learning disability was diagnosed nor was a reason for the question omission provided. Because no reason was provided for the removal of the question, and because a question on learning disability without the specific diagnosis may be more prone to bias, this change in questioning from Cycles 2 to 3 is a potential concern. Blinding was not considered an issue in these two studies, but the methods for obtaining the information are considered to be less than ideal for measuring learning disabilities including ADHD. Although the questionnaire asked about a doctor's diagnosis of a learning disability, there was no confirmation with medical records. Moreover, these questionnaires were not validated like Conners' Rating Scales, which would have been a better method for assessing ADHD. Although the outcome assessment methods are less than ideal, there was no direct evidence that they were conducted incorrectly or that the methods would have biased the results in any specific direction. Because this was the only concern in these studies, they were considered to have low risk of bias overall.

High Risk-of-bias Studies

Among the studies on other neurodevelopmental effects with high potential for bias, there were several reasons for studies to be considered probably or definitely high risk of bias for outcome assessment. One study (Shannon et al. 1986) was considered to have probably high risk of bias based on lack of information regarding blinding of outcome assessors. One study was considered definitely high risk of bias because outcome was assessed based on a parent-completed questionnaire, and the study authors noted that the parents were informed of the study's intent and were requested to provide information on fluoride history. Other studies used outcome assessment methods that were not validated or utilized group-level measurements (i.e., school performance, working memory scores).

Confidence Assessment of Findings on Other Neurodevelopmental Effects in Children

The high-quality studies (i.e., studies with low potential for bias) provide evidence of an association between fluoride exposure and other cognitive neurodevelopmental effects, including lower neurobehavioral scores in infants, cognitive effects other than IQ in children, and increased attention-related disorders including ADHD in children. However, due to limitations in the data set, including the heterogeneity in the outcomes assessed, a limited number of directly comparable studies, and differences in outcome assessment methods even when studies evaluated similar outcomes, there is low confidence based on this body of evidence that fluoride exposure is associated with other cognitive neurodevelopmental effects in children. Due to these limitations, the confidence assessment is not described in the same manner as the IQ in Children section or as outlined in Figure 1. Although there are limitations in the body of evidence, the low risk-of-bias studies demonstrate a relationship between higher fluoride exposure and neurodevelopmental effects, even in very young children, which supports the consistency in evidence shown in children's IQ studies of an association between fluoride exposure and adverse effects on cognitive neurodevelopment.

Cognitive Effects in Adults

Low Risk-of-bias Studies

Overview of Studies

Two low risk-of-bias cross-sectional studies evaluated the association between fluoride exposure and cognitive effect in adults (Jacqmin et al. 1994; Li et al. 2016). These two studies used the same test for cognitive function (i.e., Mini-Mental State or MMS Examination) and used drinking water fluoride levels to assess fluoride exposure. Li et al. (2016) also measured urinary fluoride. Both studies were cross-sectional in design. One was conducted in France (Jacqmin et al. 1994) and the other in China (Li et al. 2016). Both studies were conducted in older populations (i.e., over 60 or 65 years of age).

Table 8 provides a summary of study characteristics and key findings related to fluoride exposure and cognitive effects in adults for the two low risk-of-bias studies. The purpose of the table is to summarize key findings (independent of whether an association was found) from each study and is not meant to be a comprehensive summary of all results. For each study, results are summarized for each exposure measure assessed. Results from multiple analyses using the same exposure measure may not all be presented unless conflicting results were reported.

Table 8. Studies on Cognitive Function in Adults[a]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurobehavioral Outcome Summary[b]
Jacqmin et al. (1994)	Cross-sectional France (Gironde and Dordogne)/elderly adults [3,490]	Drinking water Range: 0.03–2.03 mg	Adults (ages ≥65 years)	Cognitive function: MMS Examination	No significant increase in the prevalence of cognitive impairment with increasing fluoride quartiles No statistical adjustment for covariates for prevalence rates
Li et al. (2016)	Cross-sectional China (Inner Mongolia)/adults [511]	Drinking water daily fluoride intake Mean (SD): 2.23 (2.23) (normal group), 3.62 (6.71) (cognitive impairment group) mg Urine Mean (SD): 1.46 (1.04) (normal group), 2.47 (2.88) (cognitive impairment group) mg/L Fluorosis score Mean (SD): 0.74 (0.98) (normal group), 1.29 (1.01) (cognitive impairment group)	Adults (ages ≥60 years)	Cognitive function: MMS Examination	Subjects with cognitive impairment had a significantly higher skeletal fluorosis score and urinary fluoride concentrations; odds of increasing severity of cognitive impairment increased with urinary fluoride concentrations but were not statistically significant; no significant association with total daily water fluoride intake Adjusted for sex, age, education, marital status (married vs. not married), alcohol consumption (non-drinkers, light drinkers, moderate to heavy drinkers), smoking history (never smoker, ex-smoker, light smoker, heavy smoker), and serum homocysteine levels

GM = geometric mean; MMS = Mini-Mental State.
[a]Includes low risk-of-bias studies.
[b]Associations between cognitive effects in adults and fluoride levels were reported quantitatively, when possible. For studies with multiple analyses and results, the table summarizes key findings and is not a comprehensive summary of all findings. Results also indicate when a study reported no association, provided as a qualitative statement of no association.

Summary of Results

Results from two low risk-of-bias studies in adults did not provide enough evidence to evaluate consistency when assessing evidence for a potential association between fluoride exposure and cognitive impairment (based on the MMS Examination) (Jacqmin et al. 1994; Li et al. 2016). Jacqmin et al. (1994) did not find an association between drinking water fluoride and cognitive impairment in populations in France (n = 3,490) and found prevalence rates of cognitive impairment to be the same regardless of fluoride exposure (see Figure A-13). In contrast, Li et al. (2016) did find significantly higher urinary fluoride levels and skeletal fluorosis scores in the cognitively impaired group compared with the control group in an analysis of 38 cognitively impaired cases and 38 controls matched for several covariates, including age, sex, education, alcohol consumption, and smoking (p-value <0.05). However, the authors found no significant association between cognitive impairment and total daily water fluoride intake (adjusted ORs per 1-mg/day increase = 0.94 [95% CI: 0.85, 1.04] and 0.86 [95% CI: 0.69, 1.06] in the moderate and severe cognitive impairment groups, respectively) or urinary fluoride levels (adjusted ORs per 1-mgL increase = 1.12 [95% CI: 0.89, 1.42] and 1.25 [95% CI: 0.87, 1.81] in the moderate and severe cognitive impairment groups, respectively) in subjects from fluorosis-endemic areas of China (n = 511).

High Risk-of-bias Studies

The results from five out of eight high risk-of-bias studies provide evidence of cognitive impairment in adults associated with exposure to fluoride; however, there was heterogeneity in the outcomes assessed, a limited number of directly comparable studies, and some variability in results (e.g., variation in IQ results across studies). Due to the limited number of low risk-of-bias studies identified that assess cognitive impairment in adults, the results from the high risk-of-bias studies are summarized in greater detail below than had been done in this document for bodies of evidence for IQ in children and other neurodevelopmental and cognitive effects in children.

In aluminum factory workers (exposed to gaseous and particular fluoride emissions during the production of aluminum metal), significant decreases in IQ (Duan et al. 1995), diminished performance on several neurobehavioral core battery tests (NCTBs) (Guo et al. 2001 [translated in Guo et al. 2008b]), and impaired psychomotor performance and memory were observed (Yazdi et al. 2011). One study conducted on adult subjects with fluorosis (dental and skeletal) from a fluorosis-endemic area compared with healthy subjects from a non-endemic area observed significant differences for some cognitive function tests (i.e., tests of speech fluency, recognition, and working memory) but not others and generally did not observe a significant change in IQ except in the operation scores (Shao 2003). One prospective cohort study evaluated exposure to fluoride in children at 5 years of age, based on whether the children resided in areas with community water fluoridation or used fluoride toothpaste or fluoride tablets, and found no clear differences in IQ scores of the subjects at 38 years of age (Broadbent et al. 2015). One additional study suggested that populations living in areas with higher drinking water fluoride had lower levels of dementia (Still and Kelley 1980); however, the study was not focused on effects of fluoride but on whether fluoride was able to reduce the risk associated with aluminum by competing with aluminum and reducing its bioavailability. Therefore, the study was considered inadequate to evaluate the association between fluoride and dementia (Still and Kelley 1980). A more recent study in Scotland evaluated dementia rates associated with aluminum and fluoride drinking water concentrations and observed a significant increased risk of dementia per standard deviation increase in fluoride (p-value <0.001) with the risk of dementia

more than double in the highest quartile of fluoride exposure (56.3 µg/L) compared to the lowest quartile (<44.4 µg/L). The authors also found a significantly increased risk of dementia associated with increased aluminum levels at all quartiles compared with the reference group (p-values <0.05) but found no statistical interaction between aluminum and fluoride levels in relation to dementia (Russ et al. 2019). Conversely, a study in China did not find a significant association between fluoride concentrations in the drinking water and risk for dementia (Liang et al. 2003). In addition to studies that reported on cognitive impairment and exposure to fluoride, two high risk-of-bias studies were identified that reported impaired motor and sensory function (Rotton et al. 1982) and a higher prevalence of self-reported headaches, insomnia, and lethargy (Sharma et al. 2009) associated with fluoride exposure.

Risk of Bias for Cognitive Effect Studies in Adults

Due to the small number of studies with a low potential for bias (see Figure D-13 and Figure D-14), the key risk-of-bias domains (confounding, exposure characterization, outcome assessment) are not discussed separately in respective subsections, as was done for the IQ in Children and Other Neurodevelopmental and Cognitive Effects in Children bodies of evidence. The high risk-of-bias studies had concerns across several domains (see Figure D-15 and Figure D-16), but there were still relatively few studies. Therefore, the discussion for high risk-of-bias studies is also not separated into subsections by key domain.

Low Risk-of-bias Studies

Both low risk-of-bias studies on cognitive effects in adults had little or no risk-of-bias concerns. One study was rated definitely low or probably low risk of bias for all risk-of-bias questions (Li et al. 2016), and the other study was rated probably high risk of bias for a single question that was judged to have minimal impact on overall potential bias (Jacqmin et al. 1994). Jacqmin et al. (1994) had potential concern for bias due to confounding because smoking was not addressed, which has the potential to impact risk for Alzheimer's disease and rates could vary by parish (the target population consisted of men and women from 75 civil parishes in southwestern France).

High Risk-of-bias Studies

There were several issues in the eight studies in adults considered to have high potential for bias. Four of the eight studies had potential concern for bias due to lack of information on the comparison groups, or the comparison groups were considered inappropriate. All eight studies had potential concern for bias regarding covariates not being addressed, including possible co-exposures in occupational studies (e.g., aluminum) and smoking. Five of the eight studies had potential concern for bias due to lack of information regarding exposure characterization or poor exposure characterization with the most utilized exposure measure in these studies being a comparison between exposed and unexposed areas. In one case (Broadbent et al. 2015), multiple sources of fluoride exposure were assessed separately without properly controlling for the other sources of exposure, which could bias the results (see Exposure Characterization in IQ Studies for further details). Five studies also had potential for bias based on limitations in the outcome assessment, which was mainly due to lack of blinding of outcome assessors, lack of validation of the methods, or lack of sufficient details on how the outcomes were assessed.

Confidence Assessment of Findings on Cognitive Effects in Adults

The body of evidence available to examine the association between exposure to fluoride and cognitive effects in adults is limited to two low risk-of-bias cross-sectional studies. Due to the

limited number of studies and a lack of evidence of an effect, there is low confidence based on this body of evidence that fluoride exposure is associated with cognitive effects in adults.

Mechanistic Data in Humans

Eight low risk-of-bias studies that evaluated fluoride exposure and mechanistic data in humans were considered potentially relevant to neurological effects. Effects on the thyroid were specifically evaluated because the NRC 2006 report identified this as a possible effect of fluoride (NRC 2006), and changes in thyroid hormones have been identified as a mechanism for neurodevelopmental effects (Haschek and Rousseaux 1991). These included effects on thyroid hormones in children (Kheradpisheh et al. 2018a; Kheradpisheh et al. 2018b; Malin et al. 2018), adults (Kheradpisheh et al. 2018a; Kheradpisheh et al. 2018b; Malin et al. 2018), or children and adults combined (Barberio et al. 2017a). In addition, some studies evaluated self-reported thyroid conditions in children and adults combined (Barberio et al. 2017a) and thyroid diseases in adults (Kheradpisheh et al. 2018b; Peckham et al. 2015) (see Figure D-17 and Figure D-18). Although the low risk-of-bias studies provide some evidence of mechanistic effects (primarily changes in thyroid stimulating hormone [TSH] levels in children), the studies were too heterogeneous or limited in number to make any determination on mechanism (see Figure 7).

Among the seven low risk-of-bias studies that reported on changes in thyroid hormones, three studies were conducted in children (Kumar et al. 2018; Singh et al. 2014; Zhang et al. 2015b) and reported increases in TSH levels. Zhang et al. (2015b) reported significant increases in TSH in children from a fluorosis-endemic area (median fluoride drinking water concentration = 1.40 mg/L; interquartile range = 1.23–1.57 mg/L) compared with a non-fluorosis-endemic area (median fluoride drinking water concentration = 0.63 mg/L; interquartile range = 0.58–0.68 mg/L), whereas 3,5,3'-triiodothyronine (T_3) or thyroxine (T_4) were not significantly different between the two groups. Similarly, Singh et al. (2014) observed significantly higher TSH levels in children without dental fluorosis who lived in a fluorosis-endemic area (fluoride drinking water concentrations of 1.6–5.5 mg/L) compared with children without dental fluorosis who lived in a non-fluorosis-endemic area (fluoride drinking water concentrations of 0.98–1.00 mg/L). When all children (with and without dental fluorosis) in the endemic area were compared with children from the non-endemic area, the TSH levels were higher in children from the fluorosis-endemic area, although results did not reach statistical significance ($p = 0.057$). Significant differences in T_4 or T_3 were not observed between groups (Singh et al. 2014). Kumar et al. (2018) also observed a significant increase in TSH levels in children from a fluorosis-endemic area (1.5–5.8 mg/L fluoride) compared with a control area (0.94–1.08 mg/L fluoride). There were also decreases in T_3 and T_4, but results were not statistically significant.

Barberio et al. (2017a) evaluated associations between fluoride and TSH levels in children and adults combined and found no relationship between fluoride exposure (measures in urine and tap water) and TSH levels. In the one study that evaluated thyroid hormone levels in adults but not children, Kheradpisheh et al. (2018b) found a significant increase in TSH associated with higher fluoride concentrations in drinking water in both adults with and without thyroid diseases such as hypothyroidism, hyperthyroidism, thyroid nodules, or thyroid cancer. Significant increases in T_3 were associated with higher fluoride in drinking water in adults without thyroid diseases, but increases in T_3 were not significant in adults with thyroid diseases. A significant association

between T_4 and higher fluoride in drinking water was not observed in adults with or without thyroid diseases (Kheradpisheh et al. 2018b).

Other than changes in hormone levels, there is limited evidence of fluoride-related mechanistic effects in the three low risk-of-bias studies that evaluated thyroid-related effects. Barberio et al. (2017a) found no relationship between fluoride exposure and self-reported thyroid conditions in children and adults (children were older than 12). Kheradpisheh et al. (2018b) also found no association between fluoride exposure and hypothyroidism in an adult population in Iran. One study found a significantly higher prevalence of hypothyroidism in areas with higher fluoride concentrations in drinking water (>0.7 mg/L) compared with areas with lower fluoride drinking water concentrations (≤0.7 mg/L) (Peckham et al. 2015).

Sixteen high risk-of-bias studies were available that evaluated mechanistic data in humans associated with fluoride exposure, including effects on thyroid hormones in children (n = 9 studies), thyroid hormones in adults (Michael et al. 1996; Yasmin et al. 2013), catecholamines in adults (Michael et al. 1996) or in subjects of unknown ages (Chinoy and Narayana 1992), acetylcholinesterase (AChE) or serotonin levels in children (Lu et al. 2019; Singh et al. 2013), brain histopathology or biochemistry in aborted fetuses (Du et al. 1992 [translated in Du et al. 2008]; Yu et al. 1996 [translated in Yu et al. 2008]), and mitochondrial fission/fusion molecules in children (Zhao et al. 2019). Similar to the low risk-of-bias studies, the high risk-of-bias studies provide some evidence of mechanistic effects (primarily changes in TSH levels in children); however, the data are insufficient to identify a clear mechanism by which fluoride causes neurodevelopmental or cognitive effects in humans.

Among high risk-of-bias studies (see Figure D-19 and Figure D-20), varying results were reported in 11 studies that evaluated associations between fluoride exposure and thyroid hormones, and a few of these studies (Lin et al. 1991; Wang et al. 2001; Yang et al. 1994 [translated in Yang et al. 2008]) were complicated by high or low iodine in the high fluoride area. When considering fluoride effects on each of the hormones individually, similar to results from low risk-of-bias studies, the most consistent evidence of fluoride-associated effects on a thyroid hormone was reported as changes in TSH levels in children, although there was some variation in the direction of association. Six of the nine high risk-of-bias studies that evaluated changes in TSH levels in children reported increases in TSH levels with higher fluoride (Lin et al. 1991; Susheela et al. 2005; Wang et al. 2001; Yang et al. 1994 [translated in Yang et al. 2008]; Yao et al. 1996; Yasmin et al. 2013). Two of the nine high risk-of-bias studies reported decreases in TSH levels in children with higher fluoride (Khandare et al. 2017; Khandare et al. 2018). One of the nine studies found no significant alterations in TSH levels in children from fluorosis-endemic areas (Hosur et al. 2012) (see Figure 8).

When considering associations between fluoride and TSH, T_3, and T_4 levels together, studies that evaluated changes in all three thyroid hormones reported varying combinations of increases, decreases, or no changes in levels across the three hormones, although among the eight low and high risk-of-bias studies that evaluated associations between fluoride exposure and TSH, T_3, and T_4 levels and reported increases in TSH levels in children, seven of the eight studies found no alterations in T_3 levels (one study found an increase in T_3), and six of the eight studies found no alterations in T_4 levels (two studies found an increase in T_4). Studies also displayed variation by age in the associations between fluoride and TSH, T_3, and T_4. Due to the dynamic relationship between the thyroid gland, the pituitary gland, and the production and clearance of TSH, T_3, and

T_4, the variations in results are not unexpected and do not eliminate the possibility of a mechanistic link between thyroid effects and neurodevelopmental or cognitive effects; however, the data do not support a clear indication that thyroid effects are a mechanism by which fluoride causes these effects in humans.

Endpoint	Direction of Effect ↑	Direction of Effect NS	Grand Total
serum T3	1	3	4
serum T4		4	4
serum TSH	5	2	7
self-reported thyroid condition		1	1
thyroid disease		1	1
hypothyroidism prevalence	1		1

Figure 7. Number of Low Risk-of-bias Studies that Evaluated Thyroid Hormones in Children and Adults by Endpoint and Direction of Association

Interactive figure and additional study details in Tableau® (https://public.tableau.com/app/profile/ntp.visuals/viz/Fluoride_EpiThyroid_UPDATE/Figures7and8?publish=yes). This figure displays study counts for low risk-of-bias studies in both children and adults, as these counts are most relevant to the summary of fluoride-related mechanistic effects in low risk-of-bias studies. Counts for high risk-of bias studies and studies by age (i.e., children, adults, or children/adults combined) can also be accessed in the interactive figure in Tableau®. Study counts are tabulated by significance (unless study footnotes in Tableau indicate that statistical significance was not tested)—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with significantly increased results.

Endpoint	Direction of Effect ↑	Direction of Effect ↓	Direction of Effect NS	Grand Total
serum T3	1	1	6	8
serum T4	3		5	8
serum TSH	6	2	1	9

Figure 8. Number of High Risk-of-bias Studies that Evaluated Thyroid Hormones in Children by Endpoint and Direction of Association

Interactive figure and additional study details in Tableau® (https://public.tableau.com/app/profile/ntp.visuals/viz/Fluoride_EpiThyroid_UPDATE/Figures7and8.). This figure displays study counts for high risk-of-bias studies in children, as these counts are most relevant to the summary of associations between fluoride and thyroid hormones in high risk-of-bias studies. Counts for low risk-of bias studies, studies in adults, or all studies combined, can also be accessed in the interactive figure in Tableau®. Study counts are tabulated by significance (unless study footnotes in Tableau indicate that statistical significance was not tested)—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with significantly increased results.

In addition to evaluating thyroid hormone levels, a few high risk-of-bias studies evaluated other mechanistic data associated with fluoride exposure; however, the data are insufficient to identify a clear mechanism by which fluoride might cause neurodevelopmental or cognitive effects in humans. Serum epinephrine and norepinephrine were significantly increased in a fluoride-endemic region (it was not reported whether subjects were children or adults) compared with a non-endemic region (Chinoy and Narayana 1992). Serum adrenaline and noradrenaline were

significantly increased in adults in a fluoride-endemic area (fluoride in the drinking water ranged from 1.0–6.53 ppm) compared with a control area (fluoride in the drinking water ranged from 0.56–0.72 ppm) (Michael et al. 1996). Serum AChE was significantly reduced in children from a high fluoride region compared with a lower fluoride region (Singh et al. 2013). Serum serotonin was significantly increased in children from Turkey who were drinking water containing 2.5 mg/L of fluoride compared with children drinking bottled water or water containing <0.5 mg/L of fluoride (Lu et al. 2019). Aborted fetuses from high fluoride areas in China were found to have histological changes in the brain and significant changes in neurotransmitter levels compared with a control area (Du et al. 1992 [translated in Du et al. 2008]; Yu et al. 1996 [translated in Yu et al. 2008]).

There are also two more recent low risk-of-bias studies that evaluated polymorphisms in dopamine-related genes; however, a determination on mechanism cannot be made at this time due to the limited number of studies. For children (10–12 years old) with a Val158Met polymorphism in the *COMT* gene (i.e., catechol-O-methyltransferase), which results in slower degradation and greater availability of dopamine within the brain, a stronger association between increasing urinary fluoride levels and decreasing IQ was reported (Zhang et al. 2015b). For children (7–12 years old) with a dopamine receptor-2 (DRD2) Taq 1A polymorphism (which is involved in reduced D2 receptor density and availability) and the TT (variant) genotype, a significant inverse association between log urinary fluoride and IQ was observed; however, this significant relationship was not observed in children with the CC (wild-type) or CT (hybrid) genotypes (Cui et al. 2018).

Animal Learning and Memory Data

NTP provided a review of the experimental animal evidence in the earlier draft monographs (NTP 2020) and agrees with the NASEM committee's comments (NASEM 2020; 2021) (placeholder to cite NTP 2021 Response to NASEM comments) that the experimental animal database is of poor quality, with many studies suffering from major reporting deficiencies. NTP acknowledges that further efforts to disentangle the potential for motor activity deficits to influence tests of learning and memory in the fluoride literature are warranted. Overall, these general issues and deficiencies with the experimental animal database led to NTP's conclusion that the animal studies are currently *inadequate* to inform the question of an association between fluoride exposure and neurodevelopmental and cognitive effects in humans. Therefore, this systematic review does not include an experimental animal section.

Mechanistic Data in Animals

There are a wide variety of studies in animals that evaluate mechanistic effects potentially related to neurological changes following oral fluoride exposure (see Appendix F); however, the mechanisms underlying fluoride-associated cognitive neurodevelopmental effects are not well characterized, and review of the data did not identify a mode of action for fluoride effects on IQ in children. Categories of mechanistic endpoints with the largest amount of available data include changes in biochemical components of the brain or neurons, neurotransmitters, oxidative stress, histopathology, and thyroid function. Limiting the data to studies with at least one exposure at or below 20 ppm fluoride drinking water equivalents (gavage and dietary exposures were backcalculated into equivalent drinking water concentrations for comparison) still provided a sufficient number of studies for evaluation of these mechanistic endpoints. This evaluation is

provided in Appendix F. Neurotransmitter and biochemical changes in the brain and neurons were considered the mechanistic areas with the greatest potential to demonstrate effects of fluoride on the brain of animals in the lower dose range and provide evidence of changes in the brain that may relate to lower IQ in children (see Appendix F). Histological data can be useful in determining whether effects are occurring in the brain at lower fluoride concentrations; however, author descriptions of these effects may be limited, thereby making it difficult to directly link histological changes in the brain to learning and memory effects. Oxidative stress is considered a general mechanistic endpoint that cannot be specifically linked to neurodevelopmental or cognitive effects in humans; however, like histopathology, it may help in identifying changes in the brain occurring at lower concentrations of fluoride. Although any effects in the brain or neurological tissue at lower concentrations of fluoride may support reduced IQ in humans, it may be difficult to distinguish the potential effects of fluoride on learning and memory functions from other neurological or general health outcomes.

In Vitro Data on Neurodevelopmental or Cognitive Effects

Although in vitro studies were identified as part of the systematic review process, NTP determined that the information on neurological effects from these studies is too general, and results cannot necessarily be attributed to effects on learning and memory or other cognitive functions at this time. The in vitro data may help support specific mechanisms identified from in vivo mechanistic data; however, as described above, no specific mechanism has been determined for fluoride effects on learning and memory or other neurodevelopmental or cognitive outcomes.

Discussion

This systematic review evaluated the available animal and human literature concerning the association between fluoride exposure and cognitive neurodevelopment. The available data on potential mechanisms to evaluate biological plausibility were also assessed. The potential health benefits of fluoride with respect to oral health are acknowledged but are not the focus of this review.

This review extended NTP's previous evaluation of the experimental animal data (NTP 2016). Although the animal data provide some evidence of effects of fluoride on neurodevelopment, they give little insight into the question of whether fluoride influences IQ. This is due to deficiencies identified in the animal body of evidence. Mechanistic studies in humans provide some evidence of adverse neurological effects of fluoride. However, these studies were too heterogenous and limited in number to make any determination on biological plausibility.

The literature on adults is also limited; therefore, it was determined that there is low confidence in the body of evidence from studies that evaluate fluoride exposure and adult cognition. Compared to the literature in adults, there is a much more extensive literature in children.

The literature in children was separated into studies assessing IQ and studies assessing other cognitive or neurodevelopmental outcomes. There is low confidence in the body of evidence from studies that evaluate fluoride exposure and other cognitive or neurodevelopmental outcomes in children. Altogether, the results from eight of nine high-quality studies (three prospective cohort and five cross-sectional studies from seven different study populations) provide some evidence that fluoride is associated with other cognitive or neurodevelopmental outcomes in children. The data also suggest that neurodevelopmental effects occur in very young children. However, the number of studies is limited, and there is too much heterogeneity in the outcomes measured and methods used to directly compare studies of any one outcome. Additional studies on outcomes such as attention-deficit hyperactivity disorder (ADHD) and other attention-related disorders, where there is some evidence of an effect of fluoride exposure, would be necessary to critically assess the data.

Most of the epidemiological studies (n = 72) assessed the association between fluoride exposure and IQ in children. Although all studies, both high- and low-quality, were considered, this evaluation focuses on the high-quality, low risk-of-bias studies in children for two reasons. First, there are fewer limitations and greater confidence in the results of the high-quality studies. Second, there is a relatively large number of high-quality studies (n = 19), such that the body of evidence from these studies could be used to evaluate confidence in the association between fluoride exposure and changes in children's IQ.

This review finds, with moderate confidence, that fluoride exposure is associated with lower IQ in children. The association between higher fluoride exposure and lower IQ in children was consistent across different study populations, study locations, study quality/risk-of-bias determinations, study designs, exposure measures, and types of exposure data (group-level and individual-level). There were 19 low risk-of-bias studies that were conducted in 15 study populations, across 5 countries, and evaluating more than 7,000 children. Of these 19 studies, 18 reported an association between higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximated or exceeded the WHO Guidelines for Drinking-water

Quality of 1.5 mg/L of fluoride (WHO 2017)] and lower IQ. These include 3 prospective cohort studies and 15 cross-sectional studies (12 of which indicated that exposure likely preceded the outcome). Forty-six of 53 low-quality studies in children also reported an association between higher fluoride exposure and lower IQ.

Many studies in this assessment relied on drinking-water fluoride levels (both group-level measures and individual-level measures), rather than measures of total fluoride exposure, to establish exposed versus "unexposed" or reference groups. Although fluoride in water is a major source of exposure [comprising 40% to 70% of total exposure (US EPA 2010)], other sources of fluoride provide variable amounts that depend on personal preferences and habits. The use of dental products containing fluoride and consuming foods and beverages prepared with fluoridated water can also result in measurable exposures (US EPA 2010). Green et al. (2019) suggested that significant exposures occur from black tea consumption. Thus, drinking water fluoride levels may, but usually do not, reflect total fluoride exposure. This could be a potential limitation in studies that rely on water fluoride data to assess fluoride exposure (in particular, earlier studies). However, because water is only part of a person's total exposure to fluoride, this limitation would likely result in an underestimate of exposure to fluoride. In addition, this limitation is less of a concern in areas where fluoride in the drinking water is high because drinking water likely contributes a large proportion of the total fluoride intake in those areas as compared with areas where fluoride in the drinking water is lower.

This review found that the quality of exposure assessment has improved over the years. More recent studies by Valdez Jimenez et al. (2017), Bashash et al. (2017), and Green et al. (2019) used individual measures of urinary fluoride, either maternal urine collected prenatally or children's urine, which confirmed the association between higher total fluoride exposure and lower children's IQ and other cognitive neurodevelopmental effects. Studies using different types of exposure measures reported similar findings of an association, which strengthens confidence in earlier studies that reported IQ deficits with increasing group-level fluoride exposure. However, there is less certainty in the quantitative estimates of the magnitude of IQ deficits from earlier studies that used group-level exposure measures than the estimates from more recent studies that used individual-level exposure measures.

It is worth noting that there are circumstances wherein typical children's water consumption considered with water fluoride levels may substantially underestimate total fluoride exposure. One example is bottle-fed infants wherein nutrition is provided by powdered formula that is rehydrated with fluoridated water (Till et al. 2020). To decrease an exclusively formula-fed infant's exposure to fluoride, for the purpose of reducing risk of dental fluorosis, the Centers for Disease Control and Prevention recommends using low-fluoride bottled water to mix with infant formula (CDC 2015). A few studies also support the possibility of heightened sensitivities to the detrimental cognitive effects of fluoride exposure in individuals with certain genetic polymorphisms in dopamine receptor D2 or catechol-O-methyltransferase (Cui et al. 2018; Zhang et al. 2015b), potentially impacting dopamine catabolism and receptor sensitivity. Differential exposures to fluoride and genetic susceptibilities of children to fluoride may represent special situations that would appear to warrant further research.

The following section briefly recaps the strength of the epidemiological evidence for an association between fluoride exposure and cognitive neurodevelopmental deficits. This is followed by a more detailed listing of limitations of the evidence base and limitations of the

systematic review, with some suggestions of areas where further research may be most beneficial.

Strengths of the Evidence Base

Strengths in the epidemiological evidence base include:

- There are 72 studies directly addressing the relationship between fluoride exposure and children's IQ.
- There are 12 high-quality cross-sectional studies with low risk of bias providing evidence that exposure occurred prior to outcome assessment in those studies.
- Studies are from diverse geographic locations that included data for more than 7,000 children.
- There are 19 high-quality studies evaluating the same outcome (i.e., IQ) and 9 evaluating other neurodevelopmental outcomes.
- Reported responses to fluoride exposure are consistent in studies of both low and high quality.
- Reported responses to fluoride exposure are consistent across different study populations, study designs, and exposure measures.
- Findings of studies with group- and individual-level information on exposure and outcomes are similar.
- A wide variety of important covariates are either addressed by study design or captured across the evidence base, with no consistent patterns that would suggest an alternative explanation.

Limitations of the Evidence Base

Limitations in the epidemiological studies with low risk of bias include:

- Few studies are available that assessed the association between fluoride exposure and cognitive function (particularly IQ) in adults and attention-related disorders including ADHD in children and adults.
- Heterogeneity in outcomes was assessed for other neurobehavioral outcomes, limiting the assessment of other possible effects in children.
- Studies rarely separated the results by sex or provided information to indicate that sex was not a modifying factor.
- Associations between lower total fluoride exposure [e.g., represented by populations whose total fluoride exposure was lower than the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] and children's IQ remain unclear. More studies at lower exposure levels are needed to fully understand potential associations in ranges typically found in the United States (i.e., <1.5 mg/L in water). However, it should be noted that, as of April 2020, CWS supplying water with ≥1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population (~1.9 million people) (CDC Division of Oral Health 2020).

- No studies investigating the association between fluoride exposure and neurodevelopmental or cognitive effects in adults or children have been conducted in the United States.
- No studies are available to evaluate fluoride exposure over a child's lifetime and neurodevelopmental or cognitive changes over time.
- The database does not allow for comparison of ages and possible changes at different developmental stages in children to assess if there is a delay in development or if associations persist.
- The database does not allow for establishing clear correlations between prenatal and postnatal exposures.

Limitations in the epidemiological studies with high risk of bias include:

- Many of the original publications were in a non-English language and provided limited details on methodology.
- Studies lacked information regarding exposure and/or had serious limitations in the exposure assessment. Exposure assessment concerns include limited individual exposure information, a lack of information on fluoride sampling methods and timing of the exposure measurements, a lack of quantitation of levels of fluoride in drinking water in a few studies, and a lack of individual-level information on fluorosis in areas reported to be endemic for fluorosis.
- The comparison groups in studies conducted in areas endemic for fluorosis still may have been exposed to high levels of fluoride or levels similar to those used in water fluoridation in the United States. This factor may have limited the ability to detect true effects.
- Studies did not provide sufficient direct information (e.g., participation rates or methods for selection) to evaluate selection bias.
- Failure to address important covariates was an issue for many studies. Some studies conducted simple statistical analyses without accounting for any covariates in the analysis, although many noted similarities between the study populations. In cases where adjustments in analyses were made, often these studies did not account for covariates considered critical for that study population and outcome including co-exposures.
- Studies conducted in areas with high, naturally occurring fluoride levels in drinking water often did not account for potential exposures to arsenic or iodine deficiencies in study subjects in areas where these substances were likely to occur.
- Studies lacked information on whether the outcome assessors were blind to the exposure group, including studies that examined children in their schools and subjects from high-fluoride communities.

Limitations in the animal and mechanistic evidence base include:

- The overall quality of the experimental animal studies is poor, and there are relatively few well-designed and well-performed studies at lower fluoride exposure levels (i.e., <20 ppm, which is roughly equivalent to human exposure of <4 ppm).

- The understanding of the specific molecular events responsible for fluoride's adverse effects on neurobehavioral function is poor.

A key data gap in the human and animal bodies of evidence includes the need for mechanistic insight into fluoride-related neurodevelopmental or cognitive changes.

Limitations of the Systematic Review

This systematic review has few limitations. The human body of evidence included a large database of observational studies. Most of the observational studies were cross-sectional; however, 12 of these were considered to provide sufficient evidence that exposure occurred prior to the outcome. In addition, the systematic review covered a wide range of study designs, populations, and measures of fluoride exposure. The systematic review was designed to cover reports on all potential mechanistic data including effects on the thyroid. After review of the studies evaluating thyroid effects, studies that only evaluated goiters and other effects on thyroid size were not considered in this review. This is not considered a limitation because these studies did not include specific information on thyroid hormones that could indicate a mechanism for thyroid involvement in neurodevelopment. In addition, review of the mechanistic data was limited to in vivo studies with at least one concentration below 20 ppm. This is not considered a limitation for the systematic review because the mechanistic body of evidence was used to evaluate biological plausibility for the effects observed in humans; therefore, data were limited to concentrations that would be more reflective of human exposures. The decision to not more closely evaluate the in vitro data is not considered a limitation because there were sufficient in vivo data, and no key events were identified where in vitro data would provide additional insight.

The supplemental literature search for non-English-language studies not indexed in traditional databases supports the comprehensive nature of the literature search strategy for this systematic review. In the absence of guidance on the most complete non-English-language databases that may contain health studies of fluoride, databases were selected that identified non-English-language studies of fluoride that we were aware of and were not captured in searches of databases from the main literature search. This informed approach influenced the selection process; however, this is not considered a limitation because it provided an objective measure by which to compare databases. Following the recommendation of the NASEM committee in its review of the September 16, 2020, draft monograph, the experimental animal section has been removed and is not included in this monograph. Although the deficiencies identified in the animal body of evidence support this removal (see Animal Learning and Memory Data for further explanation), NTP acknowledges that the absence of the experimental animal data is a limitation of this systematic review. For the purpose of this review, NTP considers the experimental animal data to be *inadequate* to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans.

Summary

This systematic review evaluated the available animal and human literature concerning the association between fluoride exposure and cognitive neurodevelopment. The available data on potential mechanisms to evaluate biological plausibility were also assessed. Existing animal studies provide little insight into the question of whether fluoride exposure affects IQ. Human mechanistic studies were too heterogenous and limited in number to make any determination on biological plausibility. The body of evidence from studies on adults is also limited and provides low confidence that fluoride exposure is associated with adverse effects on adult cognition. There is, however, a large body of evidence on IQ effects in children. There is also some evidence that fluoride exposure is associated with other neurodevelopmental and cognitive effects; although, because of the heterogeneity of the outcomes, there is low confidence in the literature for these other effects. This review finds, with moderate confidence, that higher fluoride exposure [e.g., represented by populations whose total fluoride exposure approximates or exceeds the WHO Guidelines for Drinking-water Quality of 1.5 mg/L of fluoride (WHO 2017)] is consistently associated with lower IQ in children. More studies are needed to fully understand the potential for lower fluoride exposure to affect children's IQ.

References

Adedara IA, Abolaji AO, Idris UF, Olabiyi BF, Onibiyo EM, Ojuade TD, Farombi EO. 2017a. Neuroprotective influence of taurine on fluoride-induced biochemical and behavioral deficits in rats. Chem Biol Interact. 261:1-10.

Adedara IA, Olabiyi BF, Ojuade TD, Idris UF, Onibiyo EM, Farombi EO. 2017b. Taurine reverses sodium fluoride-mediated increase in inflammation, caspase-3 activity, and oxidative damage along the brain-pituitary-gonadal axis in male rats. Can J Phys Pharmacol. 95:1019-1029. https://doi.org/10.1139/cjpp-2016-0641

Akinrinade ID, Memudu AE, Ogundele OM. 2015a. Fluoride and aluminium disturb neuronal morphology, transport functions, cholinesterase, lysosomal and cell cycle activities. Pathophysiology. 22:105-115.

Akinrinade ID, Memudu AE, Ogundele OM, Ajetunmobi OI. 2015b. Interplay of glia activation and oxidative stress formation in fluoride and aluminium exposure. Pathophysiology. 22:39-48.

An J, Mei S, Liu A, Fu Y, Wang C. 1992. [Effect of high level of fluoride on children's intelligence]. Chin J Control Endem Dis. 7(2):93-94.

Arbuckle TE, Fraser WD, Fisher M, Davis K, Liang CL, Lupien N, Bastien S, Velez MP, von Dadelszen P, Hemmings DG et al. 2013. Cohort profile: The maternal-infant research on environmental chemicals research platform. Paediatr Perinat Epidemiol. 27(4):415-425. https://doi.org/10.1111/ppe.12061

Baba NA, Raina R, Verma PK, Sultana M. 2014. Alterations in plasma and tissue acetylcholinesterase activity following repeated oral exposure of chlorpyrifos alone and in conjunction with fluoride in Wistar rats. Proc Indian Natl Sci Acad B Biol Sci. 84:969-972.

Banala R, Nagapuri K, Mohd K, Reddy M, Karnati P. 2018. Carica Papaya leaf extract as a neuroprotective agent against behavioral and neurotransmitter changes in brain of the rat treated with sodium fluoride in pre- and post-natal periods. Pharmacogn Mag. 14(55):123-131. https://doi.org/10.4103/pm.pm_378_17

Barberio AM, Hosein FS, Quinonez C, McLaren L. 2017a. Fluoride exposure and indicators of thyroid functioning in the Canadian population: Implications for community water fluoridation. J Epidemiol Community Health. 71:1019-1025. https://doi.org/10.1136/jech-2017-209129

Barberio AM, Quinonez C, Hosein FS, McLaren L. 2017b. Fluoride exposure and reported learning disability diagnosis among Canadian children: Implications for community water fluoridation. Can J Public Health. 108:229-239. https://doi.org/10.17269/cjph.108.5951

Bartos M, Gumilar F, Gallegos CE, Bras C, Dominguez S, Monaco N, Esandi MDC, Bouzat C, Cancela LM, Minetti A. 2018. Alterations in the memory of rat offspring exposed to low levels of fluoride during gestation and lactation: Involvement of the alpha7 nicotinic receptor and oxidative stress. Reprod Toxicol. 81:108-114. https://doi.org/10.1016/j.reprotox.2018.07.078

Bashash M, Thomas D, Hu H, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Ettinger AS, Wright R, Zhang Z et al. 2017. Prenatal fluoride exposure and cognitive outcomes in

children at 4 and 6-12 years of age in Mexico. Environ Health Perspect. 125(9):1-12. https://doi.org/10.1289/ehp655

Bashash M, Marchand M, Hu H, Till C, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Green R, Schnaas L et al. 2018. Prenatal fluoride exposure and attention deficit hyperactivity disorder (ADHD) symptoms in children at 6-12 years of age in Mexico City. Environ Int. 121(Pt 1):658-666. https://doi.org/10.1016/j.envint.2018.09.017

Bhatnagar M, Rao P, Sushma J, Bhatnagar R. 2002. Neurotoxicity of fluoride: Neurodegeneration in hippocampus of female mice. Indian J Exp Biol. 40:546-554.

Bhatnagar M, Sukhwal P, Suhalka P, Jain A, Joshi C, Sharma D. 2011. Effects of fluoride in drinking water on NADPH-diaphorase neurons in the forebrain of mice: A possible mechanism of fluoride neurotoxicity. Fluoride. 44:195-209.

Broadbent JM, Thomson WM, Moffitt TE, Poulton R. 2015. Community water fluoridation and intelligence response. Am J Public Health. 105:3-4.

California Office of Environmental Health Hazard Assessment (OEHHA). 2011. Meeting synopsis and slide presentations: carcinogen identification committee meeting held on October 12, 2011. California Office of Environmental Health Hazard Assessment. http://oehha.ca.gov/prop65/public_meetings/cic101211synop.html. [19 August 2019]

CDC Division of Oral Health. 2020. Personal communication. September 3, 2020.

Centers for Disease Control and Prevention (CDC). 2013. Community water fluoridation: Fluoridation statistics. Atlanta, GA: Centers for Disease Control and Prevention. https://www.cdc.gov/fluoridation/statistics/2012stats.htm. [19 August 2019]

Centers for Disease Control and Prevention (CDC). 2015. Community water fluoridation FAQs: Infant formula Atlanta, GA: Centers for Disease Control and Prevention. https://www.cdc.gov/fluoridation/faqs/infant-formula.html. [22 September 2021]

Chen Y. 2012. Organophosphate-induced brain damage: Mechanisms, neuropsychiatric and neurological consequences, and potential therapeutic strategies. Neurotox. 33:391-400.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 1991. [Research on the intellectual development of children in high fluoride areas]. Chin J Control Endem Dis. 6(Suppl):99-100.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 2008. Research on the intellectual development of children in high fluoride areas. Fluoride. 41:120-124.

Chinoy NJ, Narayana MV. 1992. Studies on fluorosis in Mehsana District of North Gujarat. Proc Zool Soc. 45:157-161.

Choi AL, Sun G, Zhang Y, Grandjean P. 2012. Developmental fluoride neurotoxicity: A systematic review and meta-analysis. Environ Health Perspect. 120:1362-1368.

Choi AL, Zhang Y, Sun G, Bellinger DC, Wang K, Yang XJ, Li JS, Zheng Q, Fu Y, Grandjean P. 2015. Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study. Neurotoxicol Teratol. 47:96-101.

Chouhan S, Flora SJS. 2008. Effects of fluoride on the tissue oxidative stress and apoptosis in rats: Biochemical assays supported by IR spectroscopy data. Toxicology. 254:61-67.

Chouhan S, Lomash V, Flora SJ. 2010. Fluoride-induced changes in haem biosynthesis pathway, neurological variables and tissue histopathology of rats. J Appl Toxicol. 30:63-73.

Cui Y, Zhang B, Ma J, Wang Y, Zhao L, Hou C, Yu J, Zhao Y, Zhang Z, Nie J et al. 2018. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. Ecotoxicol Environ Saf. 165:270-277. https://doi.org/10.1016/j.ecoenv.2018.09.018

Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. 2020. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. Neurosci Lett. 729:134981. https://doi.org/10.1016/j.neulet.2020.134981

Ding Y, Sun H, Han H, Wang W, Ji X, Liu X, Sun D. 2011. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. J Hazard Mater. 186:1942-1946.

Du L, Wan C, Cao X, Liu J. 1992. [The effect of fluorine on the developing human brain]. Chin J Pathol. 21(4):218-220.

Du L, Wan C, Cao X, Liu J. 2008. The effect of fluorine on the developing human brain. Fluoride. 41:327-330.

Duan J, Zhao M, Wang L, Fang D, Wang Y, Wang W. 1995. A comparative analysis of the results of multiple tests in patients with chronic industrial fluorosis. Guizhou Med J. 18(3):179-180.

Duan Q, Jiao J, Chen X, Wang X. 2018. Association between water fluoride and the level of children's intelligence: A dose-response meta-analysis. Public Health. 154:87-97. https://doi.org/10.1016/j.puhe.2017.08.013

Gais S, Schonauer M. 2017. Untangling a cholinergic pathway from wakefulness to memory. Neuron. 94(4):696-698.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008a. [Level of oxidative stress in rat brains and learning and memory function of rats with chronic fluorosis]. Chin J Endemiol. 27:371-373.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008b. [Effects of fluoride on learning and memory and cholinesterase activity in rat brains]. Chin J Endemiol. 27:128-130.

Gao Q, Liu YJ, Guan ZZ. 2009. Decreased learning and memory ability in rats with fluorosis: Increased oxidative stress and reduced cholinesterase activity in the brain. Fluoride. 42:277-285.

Garman RH, Li AA, Kaufmann W, Auer RN, Bolon B. 2016. Recommended methods for brain processing and quantitative analysis in rodent developmental neurotoxicity studies. Toxicol Pathol. 44(1):14-42. https://doi.org/10.1177/0192623315596858

Green R, Lanphear B, Hornung R, Flora D, Martinez-Mier EA, Neufeld R, Ayotte P, Muckle G, Till C. 2019. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. JAMA Pediatr.E1-E9.

Green R, Rubenstein J, Popoli R, Capulong R, Till C. 2020. Sex-specific neurotoxic effects of early-life exposure to fluoride: A review of the epidemiologic and animal literature. Current Epidemiology Reports. 7(4):263-273. https://doi.org/10.1007/s40471-020-00246-1

Guner S, Uyar-Bozkurt S, Haznedaroglu E, Mentes A. 2016. Dental fluorosis and catalase immunoreactivity of the brain tissues in rats exposed to high fluoride pre- and postnatally. Biol Trace Elem Res. 174:150-157.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 1991. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. Chin J Epidemiol. 10(2):98-100.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 2008a. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. Fluoride. 41:125-128.

Guo ZY, He YH, Zhu QX. 2001. [Research on the neurobehavioral function of workers occupationally exposed to fluoride]. Ind Health & Occup Dis. 27(6):346-348.

Guo ZY, He YH, Zhu QX. 2008b. Research on the neurobehavioral function of workers occupationally exposed to fluoride. Fluoride. 41:152-155.

Guyatt GH, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, Debeer H et al. 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 64(4):383-394. https://doi.org/10.1016/j.jclinepi.2010.04.026

Haschek W, Rousseaux C. 1991. Handbook of toxicologic pathology. 1st ed.: Academic Press.

Health Canada. 2015. Third report on human biomonitoring of environmental chemicals in Canada - Results of the Canadian Health Measures Survey Cycle 3 (2012–2013). Ottawa, Ontario. https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/ewh-semt/alt_formats/pdf/pubs/contaminants/chms-ecms-cycle3/chms-ecms-cycle3-eng.pdf.

Higgins JP, Green S. 2011. Cochrane handbook for systematic reviews of interventions. New York, NY: John Wiley & Sons.

Hosur MB, Puranik RS, Vanaki S, Puranik SR. 2012. Study of thyroid hormones free triiodothyronine (FT3), free thyroxine (FT4) and thyroid stimulating hormone (TSH) in subjects with dental fluorosis. Eur J Dent. 6:184-190.

Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, Sedykh A, Thayer K, Merrick BA, Walker V et al. 2020. SWIFT-Active Screener: Accelerated document screening

through active learning and integrated recall estimation. Environ Int. 138:105623. https://doi.org/10.1016/j.envint.2020.105623

Ibarluzea J, Gallastegi M, Santa-Marina L, Jiménez Zabala A, Arranz E, Molinuevo A, Lopez-Espinosa MJ, Ballester F, Villanueva CM, Riano I et al. 2021. Prenatal exposure to fluoride and neuropsychological development in early childhood: 1-to 4 years old children. Environ Res. 207:112181. https://doi.org/10.1016/j.envres.2021.112181

International Programme on Chemical Safety (IPCS). 2002. Fluorides. Geneva: World Health Organization, International Programme on Chemical Safety. Environmental Health Criteria 227. https://inchem.org/documents/ehc/ehc/ehc227.htm.

Jacqmin H, Commenges D, Letenneur L, Barberger-Gateau P, Dartigues JF. 1994. Components of drinking water and risk of cognitive impairment in the elderly. Am J Epidemiol. 139:48-57.

Jia B, Zong L, Lee JY, Lei J, Zhu Y, Xie H, Clemens JL, Feller MC, Na Q, Dong J et al. 2019. Maternal supplementation of low dose fluoride alleviates adverse perinatal outcomes following exposure to intrauterine inflammation. Scientific Reports. 9(1):2575. https://doi.org/10.1038/s41598-018-38241-8

Jiang C, Zhang S, Liu H, Guan Z, Zeng Q, Zhang C, Lei R, Xia T, Wang Z, Yang L et al. 2014. Low glucose utilization and neurodegenerative changes caused by sodium fluoride exposure in rat's developmental brain. Neuromolecular Med. 16:94-105.

Jin T, Han T, Wei Y, Wu Y, Wang Z, Zhang H. 2016. Investigation on working memory level of children aged 8-12 years in coal-burning fluorosis area. Journal of environment and health.409-411.

Jones S, Burt BA, Petersen PE, Lennon MA. 2005. The effective use of fluorides in public health. Bull World Health Organ. 83:670-676.

Khan AM, Raina R, Dubey N, Verma PK. 2017. Effect of deltamethrin and fluoride co-exposure on the brain antioxidant status and cholinesterase activity in Wistar rats. Drug Chem Toxicol. 41(2):1-5. https://doi.org/10.1080/01480545.2017.1321009

Khandare AL, Gourineni SR, Validandi V. 2017. Dental fluorosis, nutritional status, kidney damage, and thyroid function along with bone metabolic indicators in school-going children living in fluoride-affected hilly areas of Doda District, Jammu and Kashmir, India. Environ Monit Assess. 189:579. https://doi.org/10.1007/s10661-017-6288-5

Khandare AL, Validandi V, Gourineni SR, Gopalan V, Nagalla B. 2018. Dose-dependent effect of fluoride on clinical and subclinical indices of fluorosis in school going children and its mitigation by supply of safe drinking water for 5 years: An Indian study. Environ Monit Assess. 190:110. https://doi.org/10.1007/s10661-018-6501-1

Kheradpisheh Z, Mahvi AH, Mirzaei M, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018a. Correlation between drinking water fluoride and TSH hormone by ANNs and ANFIS. J Environ Health Sci Eng. 16(1):11-18. https://doi.org/10.1007/s40201-018-0290-x

Kheradpisheh Z, Mirzaei M, Mahvi AH, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018b. Impact of drinking water fluoride on human thyroid hormones: A case-control study. Sci Rep. 8:2674. https://doi.org/10.1038/s41598-018-20696-4

Kumar V, Chahar P, Kajjari S, Rahman F, Bansal DK, Kapadia JM. 2018. Fluoride, thyroid hormone derangements and its correlation with tooth eruption pattern among the pediatric population from endemic and non-endemic fluorosis areas. J Contemp Dent Pract. 19(12):1512-1516.

Li F, Chen X, Huang R, Xie Y. 2009. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. J Environ Health. 26(4):838-840.

Li J, Yao L, Q.L. S, Wu CY. 2004. [Effects of high fluoride level on neonatal neurobehavioral development]. Chin J Endemiol. 23(5):463-465.

Li J, Yao L, Shao QL, Wu CY. 2008a. Effects of high fluoride level on neonatal neurobehavioral development. Fluoride. 41:165-170.

Li M, Gao Y, Cui J, Li Y, Li B, Liu Y, Sun J, Liu X, Liu H, Zhao L et al. 2016. Cognitive impairment and risk factors in elderly people living in fluorosis areas in China. Biol Trace Elem Res. 172:53-60.

Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. 2010. [Investigation and analysis of children's IQ and dental fluorosis in high fluoride area]. Chin J Pest Control. 26(3):230-231.

Li Y, Li X, Wei S. 1994. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. J West Chin University Med Sci. 25(2):188-191.

Li Y, Li X, Wei S. 2008b. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. Fluoride. 41:331-335.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2003. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. Chin J Public Health Manag. 19(4):337-338.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2008c. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. Fluoride. 41:161-164.

Liang C, Ji R, Cao J, Jiang Y, Yu B, Ma F, Wu Y, Ying B, Zhang Y, Sun S et al. 2003. Study on the relationship between drinking water trace elements and cognitive ability of the elderly. Health Res.436-440.

Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. 1991. [High fluoride and low iodine environment and subclinical cretinism in Xinjiang]. Endem Dis Bull. 6(2):62-67.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. Alterations of nAChRs and ERK1/2 in the brains of rats with chronic fluorosis and their connections with the decreased capacity of learning and memory. Toxicol Lett. 192:324-329.

Lou DD, Guan ZZ, Liu YJ, Liu YF, Zhang KL, Pan JG, Pei JJ. 2013. The influence of chronic fluorosis on mitochondrial dynamics morphology and distribution in cortical neurons of the rat brain. Arch Toxicol. 87:449-457.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S et al. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. Physiol Behav. 206:76-83. https://doi.org/10.1016/j.physbeh.2019.02.017

Ma Q, Huang H, Sun L, Zhou T, Zhu J, Cheng X, Duan L, Li Z, Cui L, Ba Y. 2017. Gene-environment interaction: Does fluoride influence the reproductive hormones in male farmers modified by ERα gene polymorphisms? Chemosphere. 188:525-531. https://doi.org/10.1016/j.chemosphere.2017.08.166

Malin AJ, Till C. 2015. Exposure to fluoridated water and attention deficit hyperactivity disorder prevalence among children and adolescents in the United States: An ecological association. Environ Health. 14:17.

Malin AJ, Riddell J, McCague H, Till C. 2018. Fluoride exposure and thyroid function among adults living in Canada: Effect modification by iodine status. Environ Int. 121(Pt 1):667-674. https://doi.org/10.1016/j.envint.2018.09.026

McPherson CA, Zhang G, Gilliam R, Brar SS, Wilson R, Brix A, Picut C, Harry GJ. 2018. An evaluation of neurotoxicity following fluoride exposure from gestational through adult ages in Long-Evans hooded rats. Neurotoxicol Res.1-18. https://doi.org/10.1007/s12640-018-9870-x

Mesram N, Nagapuri K, Banala RR, Nalagoni CR, Karnati PR. 2016. Quercetin treatment against NaF induced oxidative stress related neuronal and learning changes in developing rats. J King Saud Univ Sci. 29:221-229.

Michael M, Barot VV, Chinoy NJ. 1996. Investigations of soft tissue functions in fluorotic individuals of North Gujarat. Fluoride. 29:63-71.

Miller K, Howard B, Phillips J, Shah M, Mav D, Thayer K, Shah R. 2016. SWIFT-Active screener: Reducing literature screening effort through machine learning for systematic reviews. Cochrane Colloquium Seoul; October 25 2016; Seoul, Korea.

Moher D, Liberati A, Tetzlaff J, Altman D. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med. 6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097

Morgan L, Allred E, Tavares M, Bellinger D, Needleman H. 1998. Investigation of the possible associations between fluorosis, fluoride exposure, and childhood behavior problems. Pediatr Dent. 20:244-252.

Mustafa DE, Younis UM, Elhag SAA. 2018. The relationship between the fluoride levels in drinking water and the schooling performance of children in rural areas of Khartoum State, Sudan. Fluoride. 51(2):102-113.

Nageshwar M, Sudhakar K, Reddy KP. 2018. Quercetin ameliorates oxidative stress, neural damage of brain and behavioral impairment of rat with fluoride exposure. Int J Pharm Sci Res. 9(8):3247-3256. https://doi.org/10.13040/IJPSR.0975-8232.9(8).3247-56

National Academies of Sciences Engineering and Medicine (NASEM). 2020. Review of the draft NTP monograph: Systematic review of fluoride exposure and neurodevelopmental and cognitive health effects. Washington, DC: National Academies of Sciences, Engineering and Medicine. https://doi.org/10.17226/25715.

National Academies of Sciences Engineering and Medicine (NASEM). 2021. Review of the revised NTP Monograph on the systematic review of fluoride exposure and neurodevelopmental and cognitive health effects: A letter report. Washington, DC: National Academies of Sciences, Engineering and Medicine. https://doi.org/10.17226/26030.

National Institute for Occupational Safety and Health (NIOSH). 1984. Fluoride in urine. Washington, DC: National Institute for Occupational Safety and Health. Method 8308.

National Research Council (NRC). 2006. Committee on fluoride in drinking water, board on environmental studies and toxicology. Fluoride in drinking water: A scientific review of EPA's standards.: National Research Council. http://www.nap.edu/catalog/11571/fluoride-in-drinking-water-a-scientific-review-of-epas-standards. [19 August 2019]

National Toxicology Program (NTP). 2016. Systematic literature review on the effects of fluoride on learning and memory in animal studies. Research Triangle Park, NC: National Toxicology Program. NTP Research Report 1. https://ntp.niehs.nih.gov/ntp/results/pubs/rr/reports/01fluoride_508.pdf. [19 August 2019]

National Toxicology Program (NTP). 2020. Revised draft NTP Monograph on the systematic review of fluoride exposure and neurodevelopmental and cognitive health effects. Research Triangle Park, NC: National Toxicology Program. https://www.nationalacademies.org/event/10-19-2020/docs/DDA97C9170D1A255D69C004CEB77B698E8D005011EFB.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. Environ Pollut. 233:889-899. https://doi.org/10.1016/j.envpol.2017.09.015

Nkpaa KW, Onyeso GI. 2018. Rutin attenuates neurobehavioral deficits, oxidative stress, neuro-inflammation and apoptosis in fluoride treated rats. Neurosci Lett. 682:92-99. https://doi.org/10.1016/j.neulet.2018.06.023

Peckham S, Lowery D, Spencer S. 2015. Are fluoride levels in drinking water associated with hypothyroidism prevalence in England? A large observational study of GP practice data and fluoride levels in drinking water. J Epidemiol Community Health. 69:619-624.

Podgorski J, Berg M. 2020. Global threat of arsenic in groundwater. Science. 368(6493):845-850. https://doi.org/10.1126/science.aba1510

Pulungan ZS, Sofro ZM, Partadiredja G. 2016. Sodium fluoride does not affect the working memory and number of pyramidal cells in rat medial prefrontal cortex. Anat Sci Int. 93(1):128-138.

Reddy YP, Tiwari SK, Shaik AP, Alsaeed A, Sultana A, Reddy PK. 2014. Effect of sodium fluoride on neuroimmunological parameters, oxidative stress and antioxidative defenses. Toxicol Mech Methods. 24:31-36.

Ren D, Li K, Liu D. 1989. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. Chin J Control Endem Dis. 4(4):251.

Ren D, Li K, Liu D. 2008. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. Fluoride. 41:319-320.

Riddell JK, Malin AJ, Flora D, McCague H, Till C. 2019. Association of water fluoride and urinary fluoride concentrations with attention deficit hyperactivity disorder in Canadian youth. Environ Int. 133:105190. https://doi.org/10.1016/j.envint.2019.105190

Rocha-Amador D, Navarro ME, Carrizales L, Morales R, Calderon J. 2007. Decreased intelligence in children and exposure to fluoride and arsenic in drinking water. Cad Saude Publica. 23(Suppl 4):S579-587.

Rocha-Amador D, Navarro M, Trejo-Acevedo A, Carrizales L, Perez-Maldonado I, Diaz-Barriga F, Calderon J. 2009. Use of the Rey-Osterrieth Complex Figure Test for neurotoxicity evaluation of mixtures in children. Neurotox. 30:1149-1154.

Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environ Health Perspect. 122(7):711-718.

Rotton J, Tikofsky RS, Feldman HT. 1982. Behavioral effects of chemicals in drinking water. J Appl Psychol. 67:230-238.

Russ TC, Killin LOJ, Hannah J, Batty GD, Deary IJ, Starr JM. 2019. Aluminium and fluoride in drinking water in relation to later dementia risk. Brit J Psychol. 14:1-6. https://doi.org/10.1192/bjp.2018.287

Santa-Marina L, Jimenez-Zabala A, Molinuevo A, Lopez-Espinosa M, Villanueva C, Riano I, Ballester F, Sunyer J, Tardon A, Ibarluzea J. 2019. Fluorinated water consumption in pregnancy and neuropsychological development of children at 14 months and 4 years of age. Environ Epidemiol. 3:386-387. https://dx.doi.org/10.1097/01.EE9.0000610304.33479.18

Saxena S, Sahay A, Goel P. 2012. Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India. J Neurosci Rural Pract. 3:144-149.

Scientific Committee on Health and Environmental Risks (SCHER). 2011. Critical review of any new evidence on the hazard profile, health effects, and human exposure to fluoride and the fluoridating agents of drinking water. Scientific Committee on Health and Environmental Risks. http://ec.europa.eu/health/scientific_committees/environmental_risks/docs/scher_o_139.pdf. [19 August 2019]

Seraj B, Shahrabi M, Shadfar M, Ahmadi R, Fallahzadeh M, Eslamlu HF, Kharazifard MJ. 2012. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. J Dent. 9:221-229.

Shan KR, Qi XL, Long YG, Nordberg A, Guan ZZ. 2004. Decreased nicotinic receptors in PC12 cells and rat brains influenced by fluoride toxicity: A mechanism relating to a damage at the level in post-transcription of the receptor genes. Toxicology. 200:169-177.

Shannon FT, Fergusson DM, Horwood LJ. 1986. Exposure to fluoridated public water supplies and child health and behaviour. New Zeal Med J. 99:416-418.

Shao Q. 2003. [Study of cognitive function impairment caused by chronic fluorosis]. Chin J Endemiol. 22(4):336-338.

Sharma JD, Sohu D, Jain P. 2009. Prevalence of neurological manifestations in a human population exposed to fluoride in drinking water. Fluoride. 42:127-132.

Singh N, Verma KG, Verma P, Sidhu GK, Sachdeva S. 2014. A comparative study of fluoride ingestion levels, serum thyroid hormone & TSH level derangements, dental fluorosis status among school children from endemic and non-endemic fluorosis areas. Springerplus. 3:7.

Singh V, Singh C, Sandeep T, Sandeep K, Vikas G, Mukesh T, Anurag T. 2013. A correlation between serum vitamin, acetylcholinesterase activity and IQ in children with excessive endemic fluoride exposure in Rajasthan, India. Int Res J Medical Sci. 1(3):12-16.

Soto-Barreras U, Escalante-Villalobos KY, Holguin-Loya B, Perez-Aguirre B, Nevarez-Rascon A, Martinez-Martinez RE, Loyola-Rodriguez JP. 2019. Effect of fluoride in drinking water on dental caries and IQ in children. Fluoride. 52:474-482.

Still CN, Kelley P. 1980. The incidence of primary degenerative dementia vs. water fluoride content in South Carolina, USA. Neurotox. 1:125-132.

Sudhakar K, Reddy KP. 2018. Protective effects of Abelmoschus moschatus seed extract on neurotransmitter system of developing brain of Wistar rats with gestational and post-natal exposure of sodium fluoride. Int J Green Pharm. 12:S131-S139.

Sudhir KM, Chandu GN, Prashant GM, Subba Reddy VV. 2009. Effect of fluoride exposure on intelligence quotient (IQ) among 13-15 year old school children of known endemic area of fluorosis, Nalgonda District, Andhra Pradesh. J Indian Assoc Public Health Dent. 2009(13):88-94.

Sun M, Li S, Wang Y, Li F. 1991. [Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis]. J Guiyang Med Coll. 16(3):204-206.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2000. [Effects of high fluoride drinking water on the cerebral function of mice]. Chin J Epidemiol. 19:262-263.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2008. Effects of high fluoride drinking water on the cerebral function of mice. Fluoride. 41:148-151.

Susheela AK, Bhatnagar M, Vig K, Mondal NK. 2005. Excess fluoride ingestion and thyroid hormone derangements in children living in Delhi, India. Fluoride. 38:98-108.

Till C, Green R, Grundy JG, Hornung R, Neufeld R, Martinez-Mier EA, Ayotte P, Muckle G, Lanphear B. 2018. Community water fluoridation and urinary fluoride concentrations in a national sample of pregnant women in Canada. Environ Health Perspect. 126(10):107001. https://doi.org/10.1289/ehp3546

Till C, Green R, Flora D, Hornung R, Martinez-Mier EA, Blazer M, Farmus L, Ayotte P, Muckle G, Lanphear B. 2020. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. Environ Int. 134:105315. https://doi.org/10.1016/j.envint.2019.105315

Trivedi M, Sangai N, Patel R, Payak M, Vyas S. 2012. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. Fluoride. 45(4):377-383.

Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. 2007. Effect of high fluoride water on intelligence of school children in India. Fluoride. 40:178-183.

Tsunoda M, Aizawa Y, Nakano K, Liu Y, Horiuchi T, Itai K, Tsunoda H. 2005. Changes in fluoride levels in the liver, kidney, and brain and in neurotransmitters of mice after subacute administration of fluoride. Fluoride. 38:284-292.

U.S. Department of Health and Human Services (US DHHS). 2015. U.S. Public Health Service recommendation for fluoride concentration in drinking water for the prevention of dental caries. U.S. Department of Health and Human Services. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4547570/. [19 August 2019]

U.S. Environmental Protection Agency (US EPA). 2010. Fluoride: Exposure and relative source contribution analysis. Washington, DC: U.S. Environmental Protection Agency. 820-R-10-015. http://www.epa.gov/dwstandardsregulations/fluoride-risk-assessment-and-relative-source-contribution. [19 August 2019]

Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, Costilla-Salazar R, Calderon Hernandez J, Alcaraz Contreras Y, Rocha-Amador DO. 2017. In utero exposure to fluoride and cognitive development delay in infants. Neurotox. 59:65-70. https://doi.org/10.1016/j.neuro.2016.12.011

Villa A, Anabalon M, Zohouri V, Maguire A, Franco AM, Rugg-Gunn A. 2010. Relationships between fluoride intake, urinary fluoride excretion and fluoride retention in children and adults: An analysis of available data. Caries Res. 44(1):60-68. https://doi.org/10.1159/000279325

Wang A, Duan L, Huang H, Ma J, Zhang Y, Ma Q, Guo Y, Li Z, Cheng X, Zhu J et al. 2020a. Association between fluoride exposure and behavioural outcomes of school-age children: A pilot study in China. Int J Environ Health Res.1-10. https://doi.org/10.1080/09603123.2020.1747601

Wang G, Gao M, Zhang M, Yang M, Xiang Q. 2012. Correlation between total fluoride intake and children's IQ. J Southeast Univ Med Ed.743-746.

Wang M, Liu L, Li H, Li Y, Liu H, Hou C, Zeng Q, Li P, Zhao Q, Dong L et al. 2020b. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. Environ Int. 134:105229. https://doi.org/10.1016/j.envint.2019.105229

Wang X, Wang L, Hu P, Guo X, Luo X. 2001. [Effects of high iodine and high fluorine on children’s intelligence and thyroid function]. Chin J Endemiol. 20(4):288-290.

Watanabe M, Kono K, Orita Y, Dote T, Usuda K, Takahashi Y, Yoshida Y. 1995. Influence of dietary fluoride intake on urinary fluoride concentration and evaluation of corrected levels in spot urine. Fluoride. 28(2):61-70.

Waugh DT. 2019. Fluoride exposure induces inhibition of sodium/iodide symporter (NIS) contributing to impaired iodine absorption and iodine deficiency: Molecular mechanisms of inhibition and implications for public health. Int J Environ Res Public Health. 16(6). https://doi.org/10.3390/ijerph16061086

World Health Organization (WHO). 2008. Guidelines for drinking-water quality [electronic resource]: Incorporating 1st and 2nd addenda. Geneva, Switzerland: World Health Organization. Third Edition. Vol. 1. https://apps.who.int/iris/bitstream/handle/10665/204411/9789241547611_eng.pdf?sequence=1&isAllowed=y.

World Health Organization (WHO). 2017. Guidelines for drinking-water quality. World Health Organization. 4th ed. + 1st add. https://apps.who.int/iris/handle/10665/254637.

Xiang Q, Liang Y, Chen L, Wang C, Chen B, Chen X, Zhou M. 2003a. Effect of fluoride in drinking water on children's intelligence. Fluoride. 36:84-94.

Xiang Q, Liang Y, Chen B, Chen L. 2011. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. Fluoride. 44:191-194.

Xiang Q, Wang Y, Yang M, Zhang M, Xu Y. 2013. Level of fluoride and arsenic in household shallow well water in Wamiao and Xinhuai villages in Jiangsu province, China. Fluoride. 46:192-197.

Xiang QY, Liang YX, Zhou MS, Zang HB. 2003b. Blood lead of children in Wamiao-Xinhuai intelligence study. Fluoride. 36:198-199.

Yang Y, Wang X, X. G, Hu P. 1994. [The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine]. Chin J Epidemiol. 15(4):296-298.

Yang Y, Wang X, Guo X, Hu P. 2008. The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine. Fluoride. 41:336-339.

Yao L, Zhou J, Wang S, Cui K, Lin F. 1996. Analysis on TSH and intelligence level of children with dental Fluorosis in a high fluoride area. Lit Inf Prev Med. 2(1):26-27.

Yasmin S, Ranjan S, D'Souza D. 2013. Effect of excess fluoride ingestion on human thyroid function in Gaya Region, Bihar, India. Toxicol Environ Chem. 95:1235-1243.

Yazdi SM, Sharifian A, Dehghani-Beshne M, Momeni VR, Aminian O. 2011. Effects of fluoride on psychomotor performance and memory of aluminum potroom workers. Fluoride. 44:158-162.

Yu X, Chen J, Li Y, Liu H, Hou C, Zeng Q, Cui Y, Zhao L, Li P, Zhou Z et al. 2018. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. Environ Int. 118:116-124. https://doi.org/10.1016/j.envint.2018.05.042

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 1996. [Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis]. Chin J Endemiol. 15(5):257-259.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 2008. Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis. Fluoride. 41:134-138.

Zhang KL, Lou DD, Guan ZZ. 2015a. Activation of the AGE/RAGE system in the brains of rats and in SH-SY5Y cells exposed to high level of fluoride might connect to oxidative stress. Neurotoxicol Teratol. 48:49-55.

Zhang S, Zhang X, Liu H, Qu W, Guan Z, Zeng Q, Jiang C, Gao H, Zhang C, Lei R et al. 2015b. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. Toxicol Sci. 144:238-245.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C et al. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. Arch Toxicol. 93(3):709-726. https://doi.org/10.1007/s00204-019-02390-0

Zhao Q, Tian Z, Zhou G, Niu Q, Chen J, Li P, Dong L, Xia T, Zhang S, Wang A. 2020. SIRT1-dependent mitochondrial biogenesis supports therapeutic effects of resveratrol against neurodevelopment damage by fluoride. Theranostics. 10:4822-4838. https://doi.org/10.7150/thno.42387

Zhou G, Tang S, Yang L, Niu Q, Chen J, Xia T, Wang S, Wang M, Zhao Q, Liu L et al. 2019. Effects of long-term fluoride exposure on cognitive ability and the underlying mechanisms: Role of autophagy and its association with apoptosis. Toxicol Appl Pharmacol. 378:114608. https://doi.org/10.1016/j.taap.2019.114608

Zhou T, Duan L-J, Ding Z, Yang R-P, Li S-H, Xi Y, Cheng X-M, Hou J-X, Wen S-B, Chen J et al. 2012. Environmental fluoride exposure and reproductive hormones in male living in endemic fluorosis villages in China. Life Sci J. 9(4):1-7.

Zohouri FV, Swinbank CM, Maguire A, Moynihan PJ. 2006. Is the fluoride/creatinine ratio of a spot urine sample indicative of 24-h urinary fluoride? Community Dent Oral Epidemiol. 34(2):130-138. https://doi.org/10.1111/j.1600-0528.2006.00269.x

Appendix A. Data Figures: Neurodevelopmental or Cognitive Effects and Outcomes

Figures

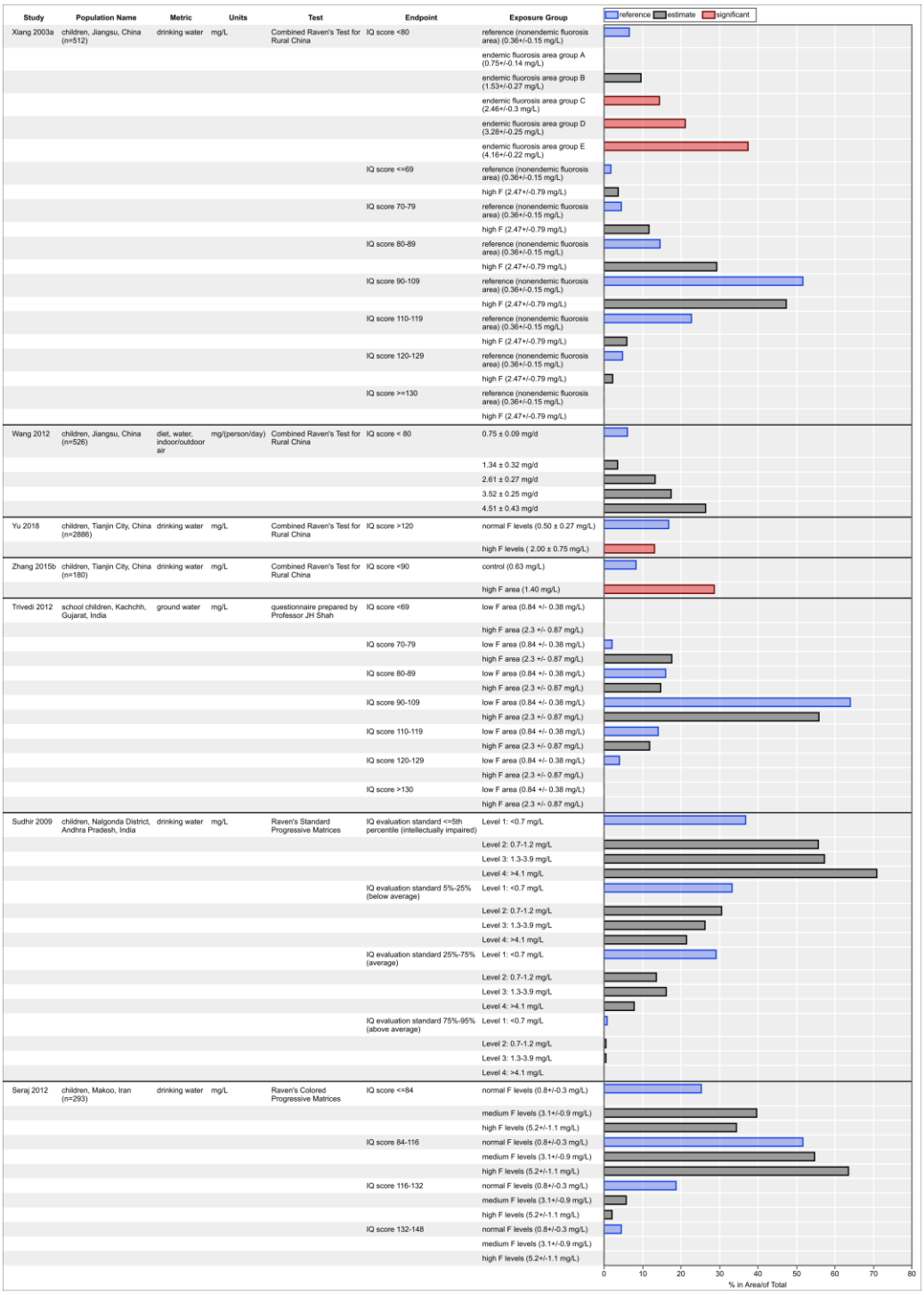


Figure A-1. Distribution of IQ in Children by Fluoride Exposure (Low Risk-of-bias Studies; Presented as % in Area or % of Total Group)

Reference group indicated by blue bars; other bars represent response estimates with red indicating statistical significance compared with the reference group.
An interactive version of Figure A-1 and additional study details in HAWC here. "F" represents fluoride. For IQ distribution results by drinking water fluoride level provided in Xiang et al. (2003a), Trivedi et al. (2012), Sudhir et al. (2009), and Seraj et al. (2012) and rate of low IQ scores by fluoride intake provided in Wang et al. (2012), statistical significance was not evaluated.

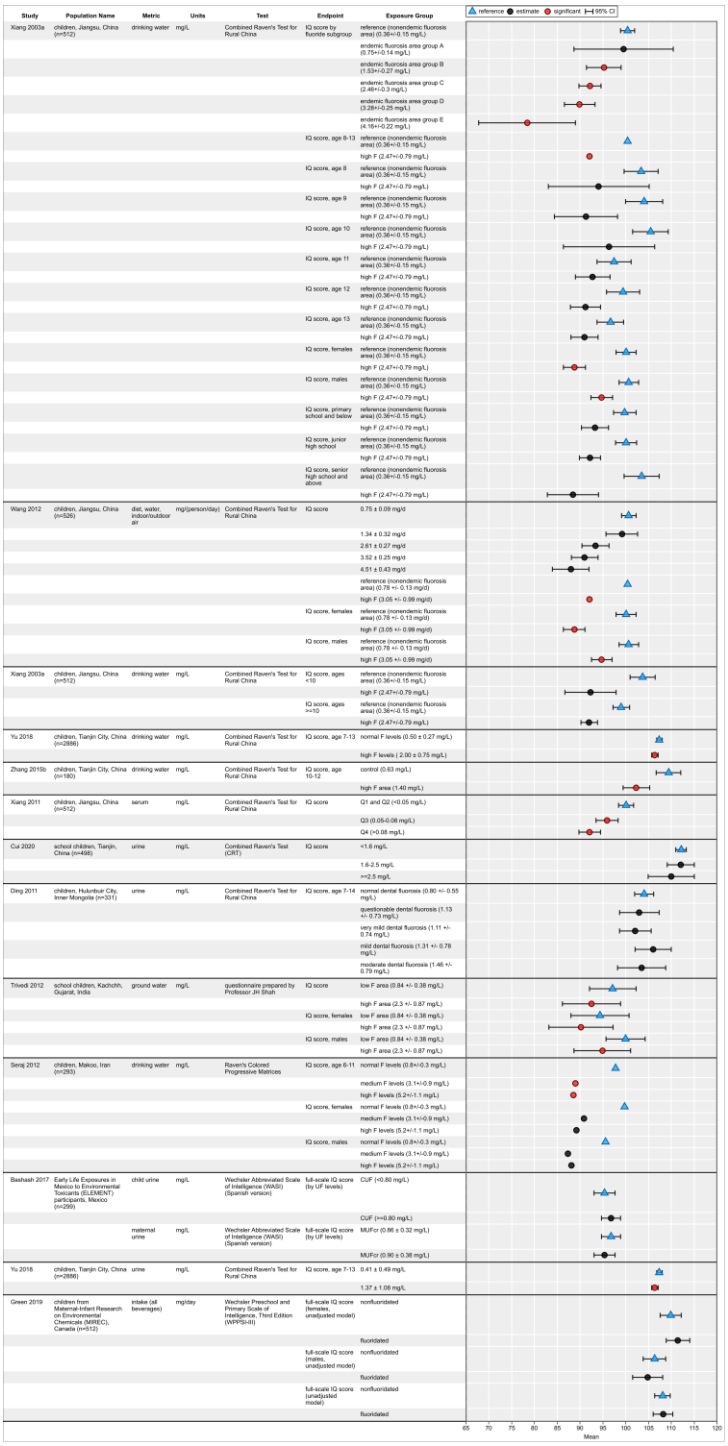


Figure A-2. Mean IQ in Children by Fluoride Exposure (Low Risk-of-bias Studies)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. An interactive version of Figure A-2 and additional study details in HAWC here. "F" represents fluoride. Three additional publications based on subsample of the larger Yu et al. (2018) cohort were identified (Zhao et al. 2019; Zhao et al. 2020; Zhou et al. 2019); however, results from these studies are not presented here. The main study by Yu et al. (2018) is considered a better representation of the IQ results. For all studies, SDs are available and can be viewed in HAWC by clicking the data points within the plot area; however, 95% CIs could not be calculated for Seraj et al. (2012) because Ns are not available for exposure groups.

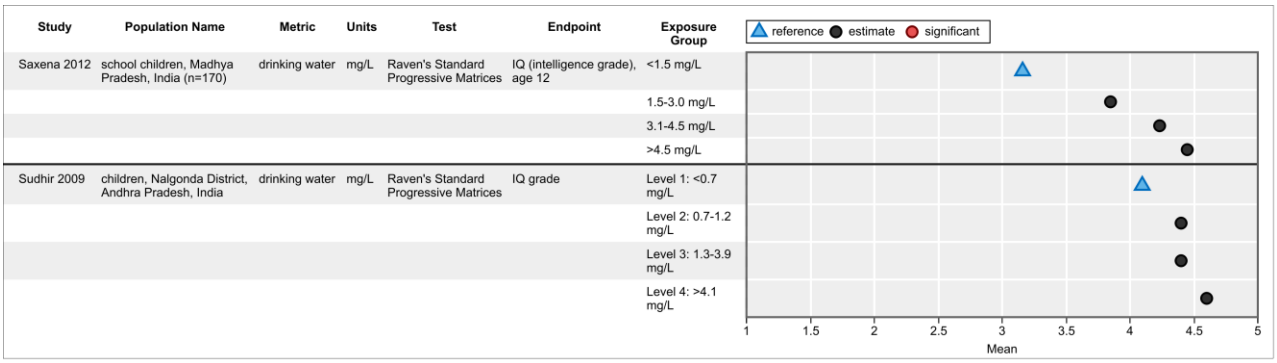


Figure A-3. Intelligence Grade in Children by Fluoride Exposure (Low Risk-of-bias Studies; Presented as Mean)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. An interactive version of Figure A-3 and additional study details in HAWC here. For Saxena et al. (2012), children's intelligence was measured using Raven's Standard Progressive Matrices. Children's scores were converted to percentile, and specific grades were allotted based on the percentiles. Grades ranged from intellectually superior (Grade I) to intellectually impaired (Grade V). Results for Soto-Barreras et al. (2019) are not presented here. Outcomes in the study were presented as levels of fluoride exposure associated with each intelligence grade. Results reported were not significant.

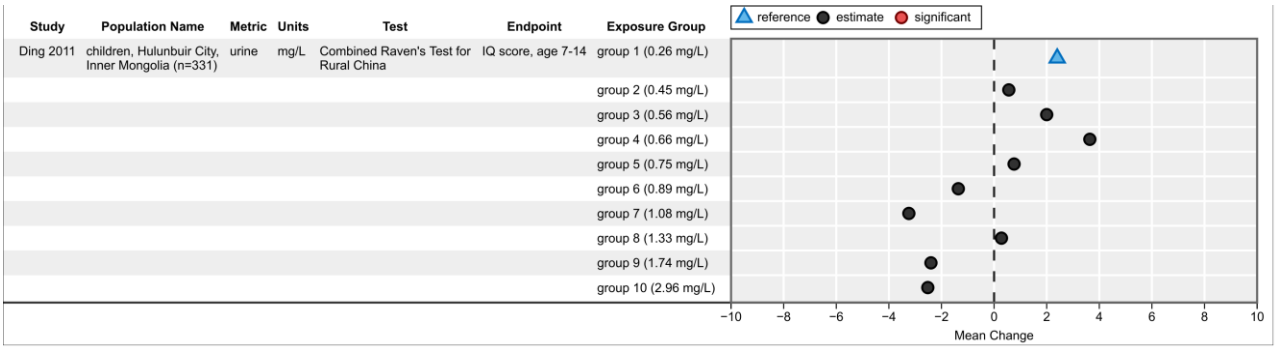
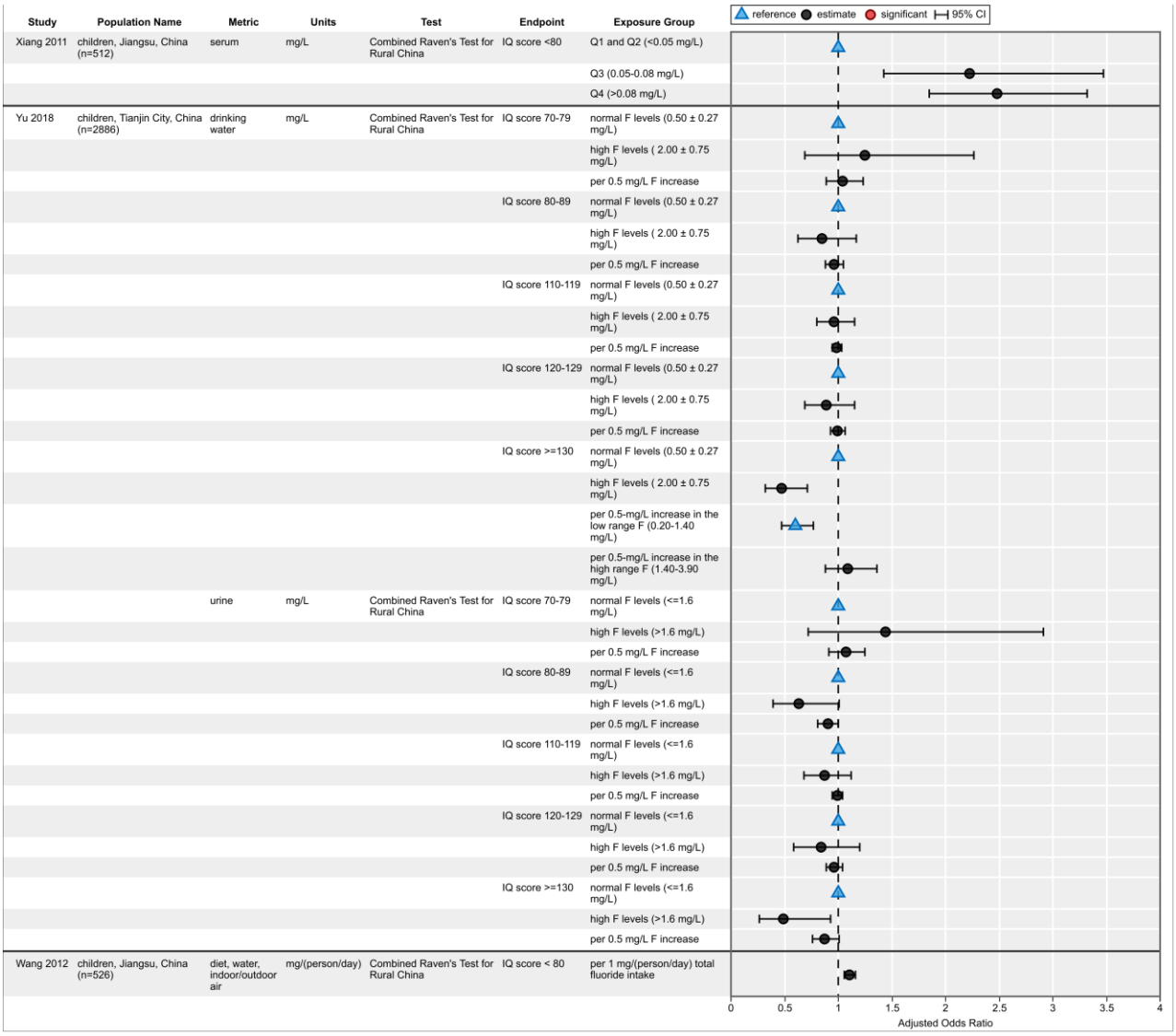


Figure A-4. Mean Change in IQ in Children by Fluoride Exposure (Low Risk-of-bias Studies)

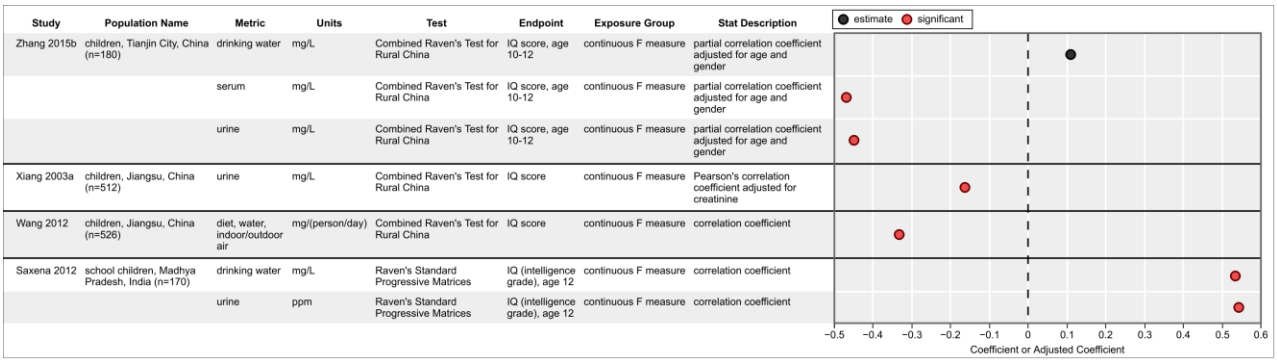
Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. An interactive version of Figure A-4 and additional study details in HAWC here. For Ding et al. (2011), SDs are available and can be viewed in HAWC by clicking the data points within the plot area; however, 95% CIs could not be calculated because Ns for each exposure group are not available.



Study	Population Name	Metric	Units	Test	Endpoint	Exposure Group
Xiang 2011	children, Jiangsu, China (n=512)	serum	mg/L	Combined Raven's Test for Rural China	IQ score <80	Q1 and Q2 (<0.05 mg/L)
						Q3 (0.05-0.08 mg/L)
						Q4 (>0.08 mg/L)
Yu 2018	children, Tianjin City, China (n=2886)	drinking water	mg/L	Combined Raven's Test for Rural China	IQ score 70-79	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score 80-89	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score 110-119	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score 120-129	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score >=130	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5-mg/L increase in the low range F (0.20-1.40 mg/L)
						per 0.5-mg/L increase in the high range F (1.40-3.90 mg/L)
		urine	mg/L	Combined Raven's Test for Rural China	IQ score 70-79	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score 80-89	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score 110-119	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score 120-129	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score >=130	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
Wang 2012	children, Jiangsu, China (n=526)	diet, water, indoor/outdoor air	mg/(person/day)	Combined Raven's Test for Rural China	IQ score < 80	per 1 mg/(person/day) total fluoride intake

Figure A-5. Associations between Fluoride Exposure and IQ Scores in Children (Low Risk-of-bias Studies; Presented as Adjusted OR)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. Cutoffs for the dichotomous outcome are listed in the Endpoint column.
An interactive version of Figure A-5 and additional study details in HAWC here. For Xiang et al. (2011), there was a significant linear trend across different levels of serum fluoride for IQ score <80 ($p < 0.001$). For Yu et al. (2018), significance levels by IQ score were not reported.



Study	Population Name	Metric	Units	Test	Endpoint	Exposure Group	Stat Description
Zhang 2015b	children, Tianjin City, China (n=180)	drinking water	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
		serum	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
		urine	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
Xiang 2003a	children, Jiangsu, China (n=512)	urine	mg/L	Combined Raven's Test for Rural China	IQ score	continuous F measure	Pearson's correlation coefficient adjusted for creatinine
Wang 2012	children, Jiangsu, China (n=526)	diet, water, indoor/outdoor air	mg/(person/day)	Combined Raven's Test for Rural China	IQ score	continuous F measure	correlation coefficient
Saxena 2012	school children, Madhya Pradesh, India (n=170)	drinking water	mg/L	Raven's Standard Progressive Matrices	IQ (intelligence grade), age 12	continuous F measure	correlation coefficient
		urine	ppm	Raven's Standard Progressive Matrices	IQ (intelligence grade), age 12	continuous F measure	correlation coefficient

Figure A-6. Correlations between Fluoride Exposure and IQ Score in Children (Low Risk-of-bias Studies; Presented as Correlation Coefficient)

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-6 and additional study details in HAWC here. "F" represents fluoride. For Saxena et al. (2012), a significant relationship between water fluoride level and intelligence grade was observed. Increasing intelligence grades reflected increasing levels of impairment (reduced intelligence) in children.

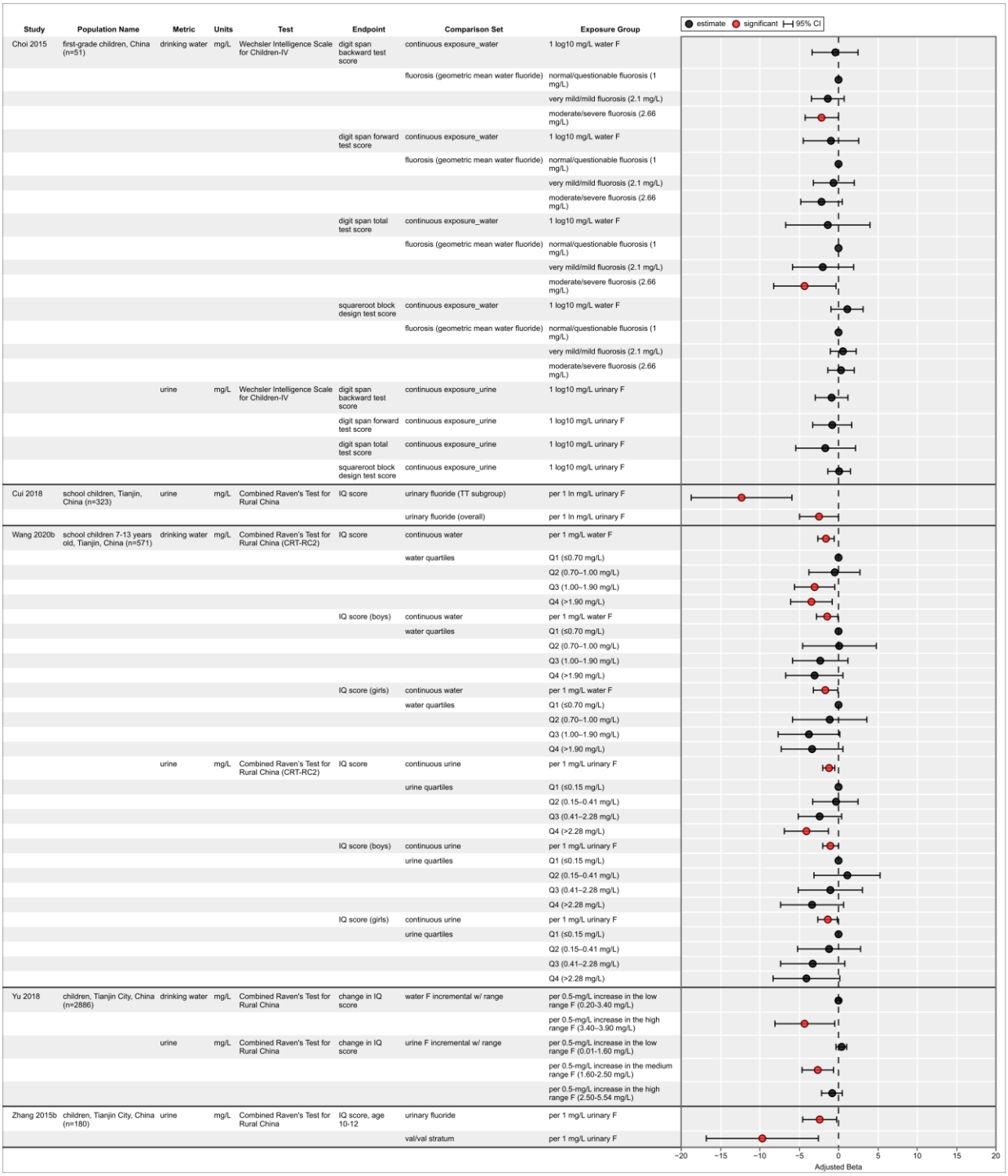


Figure A-7. Associations between Fluoride Exposure and IQ Score in Children (Low Risk-of-bias Studies; Presented as Adjusted Beta)—China

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-7 and additional study details in HAWC here. "F" represents fluoride. For Yu et al. (2018), authors note an obvious decrease in the IQ score at water fluoride exposure levels between 3.40 mg/L and 3.90 mg/L and a similar adverse effect on IQ scores at urinary fluoride exposure levels from 1.60 mg/L to 2.50 mg/L, and so the changes in IQ score are indicated as significant; however, significance levels for change in IQ score were not reported.

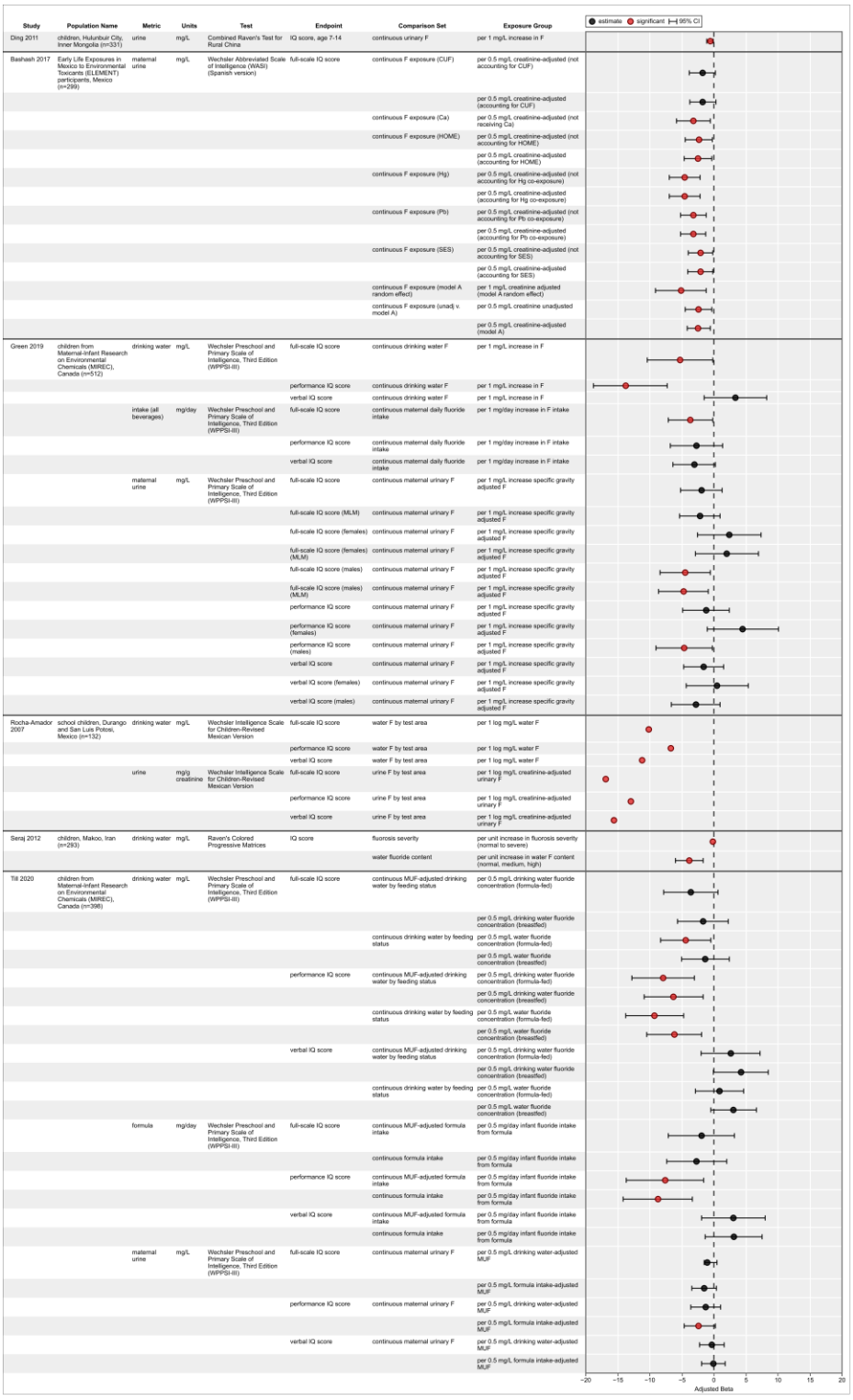


Figure A-8. Associations between Fluoride Exposure and IQ Score in Children (Low Risk-of-bias Studies; Presented as Adjusted Beta)—Areas Other Than China

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-8 and additional study details in HAWC here. "F" represents fluoride.

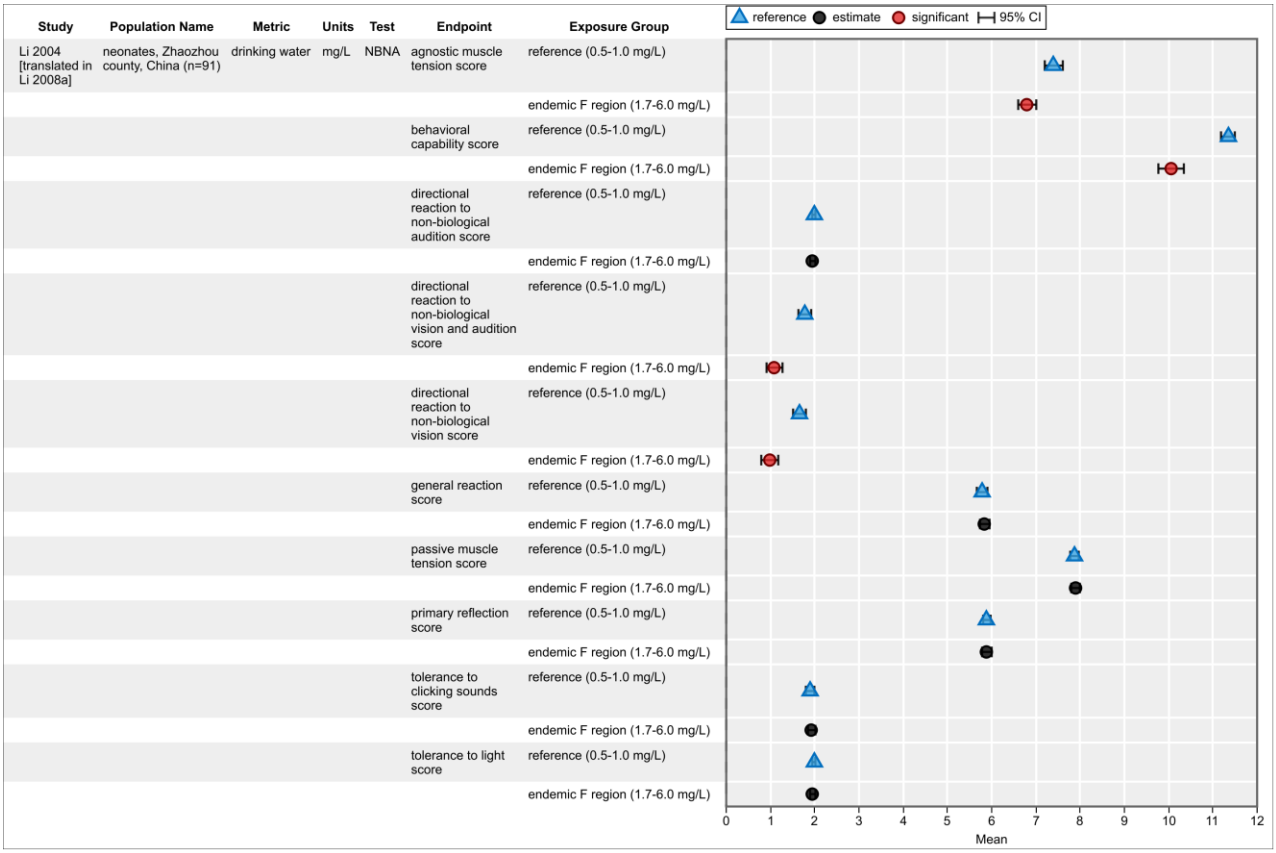


Figure A-9. Mean Motor/Sensory Scores in Children by Fluoride Exposure (Low Risk-of-bias Studies)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-9 and additional study details in HAWC here. "F" represents fluoride. 95% CIs are small and are within figure symbols and may be difficult to see. Values for SDs and 95% CIs can be viewed in HAWC by clicking the data points within the plot area. Total neonatal behavioral neurological assessment (NBNA) score was also significantly reduced in the endemic F region versus reference region (not shown).

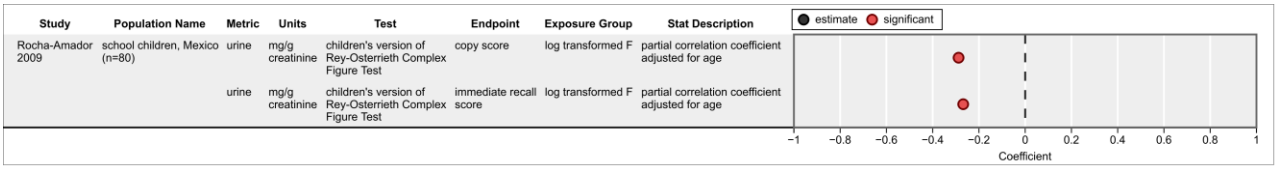


Figure A-10. Correlations between Fluoride Exposure and Other Cognitive Effects in Children (Low Risk-of-bias Studies; Presented as Correlation Coefficient)

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-10 and additional study details in HAWC here. "F" represents fluoride.

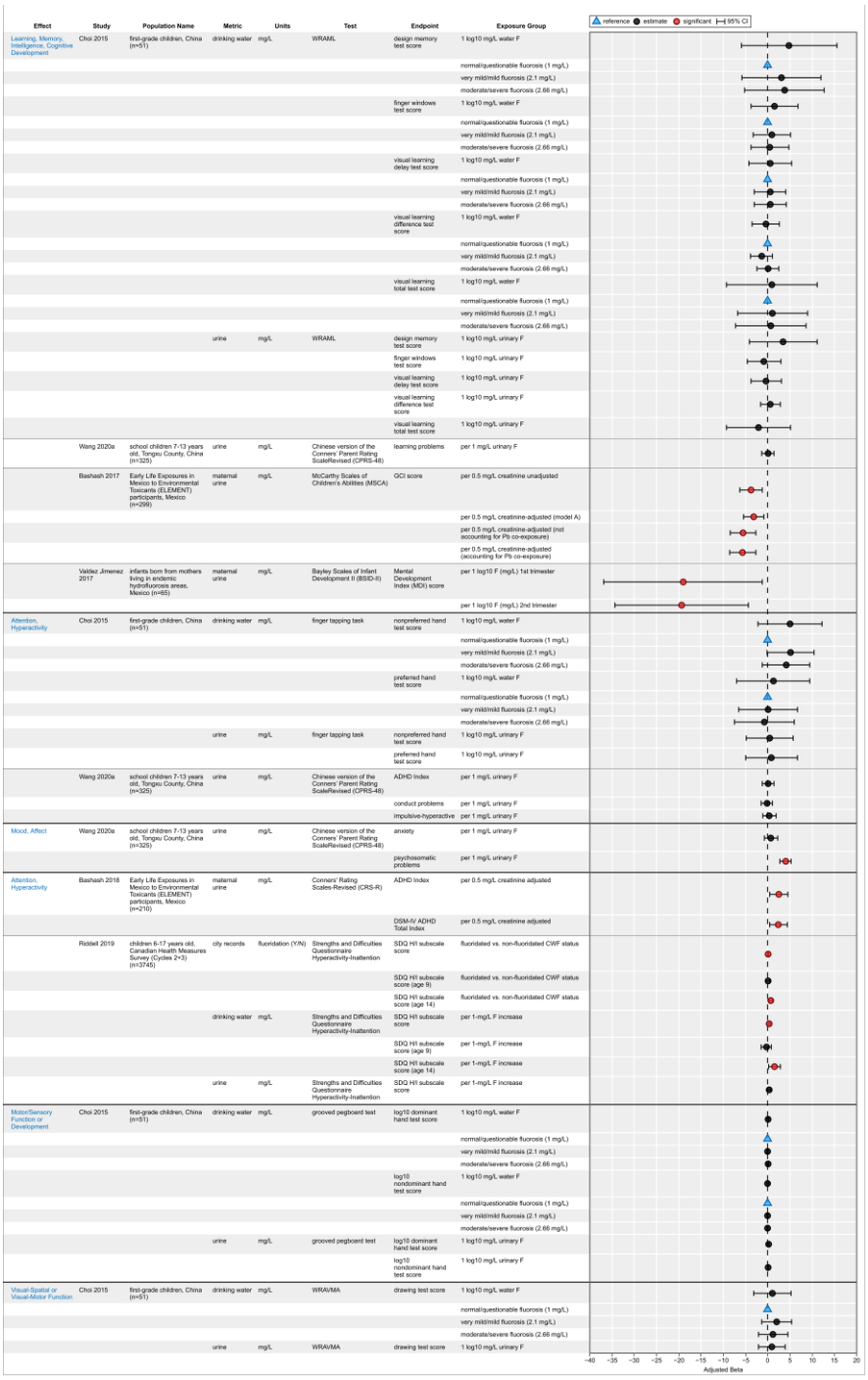


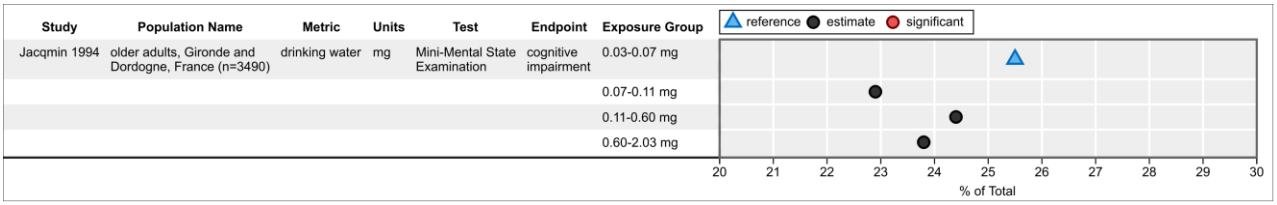
Figure A-11. Associations between Fluoride Exposure and Other Neurodevelopmental Effects in Children (Low Risk-of-bias Studies; Presented as Adjusted Beta)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance. An interactive version of Figure A-11 and additional study details in HAWC here. "F" represents fluoride. Bashash et al. (2018) observed significant associations between maternal urinary fluoride and ADHD-like symptoms related to inattention (an increase in 0.5 mg/L of maternal urinary fluoride was associated with a 2.84-point increase in the DSM-IV Inattention Index and a 2.54-point increase in Cognitive Problems and Inattention Index). These two scales contributed to the global ADHD Index and the DSM-IV ADHD Total Index shown here.

Effect	Study	Population Name	Metric	Units	Test	Endpoint	Comparison Set	Exposure Group
Learning, Memory, Intelligence, Cognitive Development	Wang 2020a	school children 7-13 years old, Tongxu County, China (n=325)	urine	mg/L	Chinese version of the Conners' Parent Rating ScaleRevised (CPRS-48)	learning problems	continuous urinary F	per 1 mg/L urinary F
	Barberio 2017b	children 3-12 years old, Canadian Health Measures Survey (Cycle 2) (n=1120)	urine	umol/L		learning disability	Cycle 2 urine	per 1-umol urinary F
							Cycle 2 SG-adjusted urine	per 1-umol urinary F
				umol/mmol creatinine		learning disability	Cycle 2 creatinine-adjusted urine	per 1-umol urinary F
		children 3-12 years old, Canadian Health Measures Survey (Cycle 3) (n=1101)	drinking water	mg/L		learning disability	Cycle 3 drinking water	per 1 mg/L water F
			urine	umol/L		learning disability	Cycle 3 urine	per 1-umol urinary F
							Cycle 3 SG-adjusted urine	per 1-umol urinary F
				umol/mmol creatinine		learning disability	Cycle 3 creatinine-adjusted urine	per 1-umol urinary F
		children 3-12 years old, Canadian Health Measures Survey (Cycles 2+3) (n=2221)	urine	umol/L		learning disability	pooled urine	per 1-umol urinary F
							pooled SG-adjusted urine	per 1-umol urinary F
				umol/mmol creatinine		learning disability	pooled creatinine-adjusted urine	per 1-umol urinary F
Attention, Hyperactivity	Wang 2020a	school children 7-13 years old, Tongxu County, China (n=325)	urine	mg/L	Chinese version of the Conners' Parent Rating ScaleRevised (CPRS-48)	ADHD Index	continuous urinary F	per 1 mg/L urinary F
						conduct problems	continuous urinary F	per 1 mg/L urinary F
						impulsive-hyperactive	continuous urinary F	per 1 mg/L urinary F
	Barberio 2017b	children 3-12 years old, Canadian Health Measures Survey (Cycle 2) (n=1120)	urine	umol/L		ADHD	Cycle 2 urine	per 1-umol urinary F
							Cycle 2 SG-adjusted urine	per 1-umol urinary F
				umol/mmol creatinine		ADHD	Cycle 2 creatinine-adjusted urine	per 1-umol urinary F
	Riddell 2019	children 6-17 years old, Canadian Health Measures Survey (Cycles 2+3) (n=3745)	city records	fluoridation (Y/N)		ADHD	Cycle 2/3 CWF	fluoridated vs. non-fluoridated CWF status
						ADHD (age 9)	Cycle 2/3 CWF	fluoridated vs. non-fluoridated CWF status
						ADHD (age 14)	Cycle 2/3 CWF	fluoridated vs. non-fluoridated CWF status
			drinking water	mg/L		ADHD	Cycle 3 drinking water	per 1-mg/L F increase
			urine	mg/L		ADHD	Cycle 2/3 SG-adjusted urine	per 1-mg/L F increase
Mood, Affect	Wang 2020a	school children 7-13 years old, Tongxu County, China (n=325)	urine	mg/L	Chinese version of the Conners' Parent Rating ScaleRevised (CPRS-48)	anxiety	continuous urinary F	per 1 mg/L urinary F
						psychosomatic problems	continuous urinary F	per 1 mg/L urinary F

Figure A-12. Associations between Fluoride Exposure and Other Neurodevelopmental Effects in Children (Low Risk-of-bias Studies; Presented as Adjusted OR)

Circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-12 and additional study details in HAWC here. "F" represents fluoride. Drinking water results for Barberio et al. (2017b) have a large confidence interval and are not completely visible in the figure. 95% CIs are 0.068–11.33 and can be viewed in HAWC by clicking the OR within the plot area.



Study	Population Name	Metric	Units	Test	Endpoint	Exposure Group
Jacqmin 1994	older adults, Gironde and Dordogne, France (n=3490)	drinking water	mg	Mini-Mental State Examination	cognitive impairment	0.03-0.07 mg
						0.07-0.11 mg
						0.11-0.60 mg
						0.60-2.03 mg

Figure A-13. Cognitive Impairment in Adults by Fluoride Exposure (Low Risk-of-bias Studies; Presented as % of Total Group)

Reference group indicated by blue triangles; circles represent response estimates with red indicating statistical significance.
An interactive version of Figure A-13 and additional study details in HAWC here. Results from Li et al. (2016) suggested that fluoride exposure may be a risk factor for cognitive impairment in elderly subjects; however, results from the study were not conducive to presentation in this visualization.

Appendix B. Literature Search and Document Review Details

Table of Contents

Tables

B.1. Introduction

NTP initially published a systematic review of the experimental animal literature in 2016 that was subsequently expanded to include human epidemiological studies, mechanistic studies, and newer experimental animal literature. Table B-1 provides a timeline of key activities contributing to the 2022 NTP monograph including the multiple literature searches, draft monographs, and document review activities that have occurred since 2016.

Table B-2 is a summary of the specific search terms used for the PubMed database. In order to ensure inclusion of relevant papers, the strategy for this search was broad for the consideration of neurodevelopmental or cognitive endpoints and comprehensive for fluoride as an exposure or treatment. The specific search strategies for other databases are available in the protocol (https://ntp.niehs.nih.gov/go/785076).

Table B-1. Literature Search and Document Review Timeline

Date	Action
July 2016	Published 2016 NTP monograph of the systematic literature review on the effects of fluoride on learning and memory in animals only
June 2017	Published protocol for a new NTP monograph on systematic review on effects of fluoride on neurodevelopment and cognition from evidence in human, experimental animal, and mechanistic data
April 2019	Completed final literature search for 2019 draft NTP monograph on human, experimental animal, and mechanistic data (i.e., updated through April 2019)
May 2019	Published 2019 revised protocol for 2019 draft NTP monograph
September 2019	Sent 2019 draft NTP monograph for review by NASEM committee
February 2020	Received NASEM committee's review report of 2019 draft NTP monograph; began the following key changes in response to NASEM report: • Expanded literature search to non-English-language databases • Conducted meta-analysis on children's IQ and fluoride exposure • Revised protocol for monograph to include additional information.
May 2020	Completed final literature search for 2020 draft NTP monograph on human experimental animal and mechanistic data (i.e., updated through May 2020 and expanded to include non-English-language databases)
September 2020	Published 2020 revised protocol for 2020 draft NTP monograph
September 2020	Sent 2020 draft NTP monograph for second review by NASEM committee
February 2021	Received NASEM committee's review report of revised 2020 draft NTP monograph; made the following key changes in response to NASEM report: • Removed hazard step and hazard conclusions • Removed meta-analysis to publish separately.
December 2021	Sent 2021 draft NTP monograph on the state of the science for external peer review
April 2022	Published final 2022 NTP monograph on the state of the science

Table B-2. PubMed Search Terms

Database	Search Terms
PUBMED	((Fluorides[mh:noexp] OR fluorides, topical[mh] OR sodium fluoride[mh] OR Fluorosis, Dental[mh] OR fluorosis[tiab] OR fluorid*[tiab] OR flurid*[tiab] OR fluorin*[tiab] OR florin*[tiab]) NOT (18F[tiab] OR f-18[tiab] OR 19F[tiab] OR f-19[tiab] OR f-labeled[tiab] OR "fluorine-18"[tiab] OR "fluorine-19"[tiab] OR pet-scan[tiab] OR radioligand*[tiab])) AND ((Aryl Hydrocarbon Hydroxylases[mh] OR Aryl Hydrocarbon Receptor Nuclear Translocator[mh] OR Behavior and Behavior Mechanisms[mh] OR Gene Expression Regulation[mh] OR Glucuronosyltransferase[mh] OR Intelligence tests[mh] OR Malate Dehydrogenase[mh] OR Mediator Complex Subunit 1[mh] OR Mental disorders[mh] OR Mental processes[mh] OR Monocarboxylic Acid Transporters[mh] OR Myelin Basic Protein[mh] OR nervous system[mh] OR nervous system diseases[mh] OR nervous system physiological phenomena[mh] OR Neurogranin[mh] OR Oligodendroglia[mh] OR Peroxisome Proliferator-Activated Receptors[mh] OR Psychological Phenomena and Processes[mh] OR Receptors, thyroid hormone[mh] OR Receptors, thyrotropin[mh] OR Retinoid X Receptors[mh] OR thyroid diseases[mh] OR thyroid hormones[mh] OR Thyrotropin-releasing hormone[mh] OR Thyroxine-Binding Proteins[mh] OR Pregnane X Receptor[supplementary concept] OR thyroid-hormone-receptor interacting protein[supplementary concept] OR Constitutive androstane receptor[supplementary concept] OR Academic performance[tiab] OR auditory[tiab] OR cortical[tiab] OR delayed development[tiab] OR developmental impairment[tiab] OR developmental-delay*[tiab] OR developmental-disorder*[tiab] OR euthyroid[tiab] OR gait[tiab] OR glia*[tiab] OR gliogenesis[tiab] OR hyperactiv*[tiab] OR impulse-control[tiab] OR iodide peroxidase[tiab] OR IQ[tiab] OR ischemi*[tiab] OR locomotor[tiab] OR mental deficiency[tiab] OR mental development[tiab] OR mental illness[tiab] OR mental-deficit[tiab] OR mobility[tiab] OR mood[tiab] OR morris-maze[tiab] OR morris-water[tiab] OR motor abilit*[tiab] OR Motor activities[tiab] OR motor performance[tiab] OR nerve[tiab] OR neural[tiab] OR neurobehav*[tiab] OR Neurocognitive impairment[tiab] OR neurodegenerat*[tiab] OR Neurodevelopment*[tiab] OR neurodisease*[tiab] OR neurologic*[tiab] OR neuromuscular[tiab] OR neuron*[tiab] OR neuropath*[tiab] OR obsessive compulsive[tiab] OR OCD[tiab] OR olfaction[tiab] OR olfactory[tiab] OR open-field-test[tiab] OR passive avoidance[tiab] OR plasticity[tiab] OR senil*[tiab] OR sociab*[tiab] OR speech*[tiab] OR spelling[tiab] OR stereotypic-movement*[tiab] OR synap*[tiab] OR tauopath*[tiab] OR Thyroglobulin[tiab] OR Thyroid disease*[tiab] OR thyroid gland[tiab] OR thyroid hormone*[tiab] OR thyronine*[tiab] OR visual motor[tiab] OR Visuospatial processing[tiab] OR water maze[tiab]) OR ((active-avoidance[tiab] OR ADHD[tiab] OR alzheimer*[tiab] OR amygdala[tiab] OR antisocial[tiab] OR anxiety[tiab] OR anxious[tiab] OR asperger*[tiab] OR attention deficit[tiab] OR autism[tiab] OR autistic[tiab] OR behavioral[tiab] OR behaviors[tiab] OR behavioural[tiab] OR behaviours[tiab] OR bipolar[tiab] OR cerebellum[tiab] OR cognition[tiab] OR cognitive[tiab] OR communication-disorder*[tiab] OR comprehension[tiab] OR cranial[tiab] OR dementia[tiab] OR dendrit*[tiab] OR dentate-gyrus[tiab] OR depression[tiab] OR dextrothyroxine[tiab] OR diiodothyronine*[tiab] OR diiodotyrosine[tiab] OR down syndrome[tiab] OR dyslexia[tiab] OR entorhinal cortex[tiab] OR epilep*[tiab] OR gangli*[tiab] OR goiter[tiab] OR graves-disease[tiab] OR hearing[tiab] OR hippocamp*[tiab] OR human development[tiab] OR hyperthyroid*[tiab] OR hypothalam*[tiab] OR hypothyroid*[tiab] OR impulsiv*[tiab] OR Intellectual disability[tiab] OR intelligence[tiab] OR language[tiab] OR learning[tiab] OR lewy bod*[tiab] OR long-term potentiation[tiab] OR long-term synaptic depression[tiab] OR memory[tiab] OR mental disorder*[tiab] OR mental recall[tiab] OR monoiodotyrosine[tiab] OR Motor activity[tiab] OR motor skill*[tiab] OR multiple sclerosis[tiab] OR myxedema[tiab] OR Nervous system[tiab] OR nervous-system[tiab] OR neurit*[tiab] OR optic[tiab] OR palsy[tiab] OR panic[tiab] OR parahippocamp*[tiab] OR paranoia[tiab] OR paranoid[tiab] OR parkinson*[tiab] OR perception[tiab] OR perforant*[tiab] OR personality[tiab] OR phobia[tiab] OR problem solving[tiab] OR proprioception[tiab] OR psychomotor[tiab] OR reflex[tiab] OR risk taking[tiab] OR schizophrenia[tiab] OR seizure*[tiab] OR sensation*[tiab] OR sleep[tiab] OR smell[tiab] OR spatial behavior[tiab] OR stroke[tiab] OR substantia-nigra[tiab] OR taste[tiab] OR thyroiditis[tiab] OR thyrotoxicosis[tiab] OR Thyrotropin[tiab] OR thyroxine[tiab] OR triiodothyronine[tiab] OR vision[tiab]) NOT medline[sb]))

Appendix C. Detailed Literature Search Results and List of Included Studies

Table of Contents

C.1. Detailed Literature Search Results

C.1.1. Literature Search Results Counts and Title and Abstract Screening

The electronic database searches retrieved 25,450 unique references in total (20,883 references during the initial search conducted in December 2016, 3,657 references during the literature search updates [including the final updated search conducted for the primary epidemiological studies on May 1, 2020], and 910 references from the supplemental Chinese database searches); 11 additional references were identified by technical advisors or from reviewing reference lists in published reviews and included studies. As a result of title and abstract screening, 1,036 references were moved to full-text review, and 24,425 references were excluded (11,402 by manual screening for not satisfying the PECO criteria and 13,023 based on the SWIFT-Active Screener algorithm).

C.1.2. Full-text Review

Among the 1,036 references that underwent full-text review, 489 were excluded at that stage with reasons for exclusion documented; 333 references were excluded for not satisfying the PECO criteria; and 156 references from the May 2020 searches (main literature search update and supplemental Chinese database searches) were excluded for not including information that would materially advance the human, animal in vivo, or mechanistic findings (see the Main Literature Search section for a description of the methodology). These screening results are outlined in a study selection diagram that reports numbers of studies excluded for each reason at the full-text review stage (see Figure 2) [using reporting practices outlined in Moher et al. (2009)]. After full-text review, 547 studies were considered relevant with primary neurodevelopmental or cognitive outcomes, secondary neurobehavioral outcomes, and/or outcomes related to thyroid function. A few studies assessed data for more than one evidence stream (human, non-human mammal, and/or in vitro), and several human and animal studies assessed more than one type of outcome (e.g., primary and secondary outcomes). The number of included studies is summarized below:

- 167 human studies (84 primary only; 13 secondary only; 5 primary and secondary; 8 primary and thyroid; 2 secondary and thyroid; and 55 thyroid only);
- 339 non-human mammal studies (7 primary only; 186 secondary only; 67 primary and secondary; 6 primary, secondary, and thyroid; 4 secondary and thyroid; and 69 thyroid only); and,
- 60 in vitro/mechanistic studies (48 neurological and 12 thyroid).

One publication contained human, experimental non-human mammal, and in vitro data. Three publications contained both human and experimental non-human mammal data. Fourteen publications contained data relevant to both experimental non-human mammal studies and in vitro studies.

C.2. List of Included Studies

C.2.1. Studies in Humans

As described in Figure 2, 167 human studies were included; however, full data extraction was conducted only on studies with neurological outcomes or thyroid hormone data. Data extraction was completed using HAWC. Data were extracted from a subset of included studies in humans (n = 124) and are available in HAWC based on outcome. The following lists of references are organized as studies that are available in HAWC followed by studies that are not available in HAWC. Specifically, data for primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence) and secondary neurobehavioral outcomes (anxiety, aggression, motor activity, or biochemical changes), as well as thyroid hormone level data, were extracted from included human studies and are available in HAWC. Data for included studies identified through the 2020 literature search update were extracted only for primary neurodevelopmental or cognitive outcomes; a subset of these studies (n = 7) also included secondary neurobehavioral outcomes and/or thyroid hormone level data that were not extracted because those data would not materially advance the human or mechanistic findings. Included human studies that evaluated only other thyroid-related effects such as goiters or thyroid size (n = 43) were not extracted and are not available in HAWC. The list below presents the 167 human studies that were included in the review. An overview of the screening results is outlined in the study selection diagram (Figure 2) that reports numbers of included studies as well as numbers of studies excluded for each reason at the full-text review stage.

C.2.1.1. Studies Available in HAWC

An J, Mei S, Liu A, Fu Y, Wang C. 1992. [Effect of high level of fluoride on children's intelligence]. *Chin J Control Endem Dis* 7(2): 93-94.

Aravind A, Dhanya RS, Narayan A, Sam G, Adarsh VJ, Kiran M. 2016. Effect of fluoridated water on intelligence in 10-12-year-old school children. *J Int Soc Prev Community Dent* 6(Suppl 3): S237-S242.

Bai A, Li Y, Fan Z, Li X, Li P. 2014. [Intelligence and growth development of children in coal-burning-borne arsenism and fluorosis areas: An investigation study]. *Chin J Endemiol* 33(2): 160-163.

Barberio AM, Hosein FS, Quinonez C, McLaren L. 2017. Fluoride exposure and indicators of thyroid functioning in the Canadian population: Implications for community water fluoridation. *J Epidemiol Community Health* 71: 1019-1025.

Barberio AM, Quinonez C, Hosein FS, McLaren L. 2017. Fluoride exposure and reported learning disability diagnosis among Canadian children: Implications for community water fluoridation. *Can J Public Health* 108: 229-239.

Bashash M, Thomas D, Hu H, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Ettinger AS, Wright R, Zhang Z, Liu Y, Schnaas L, Mercado-Garcia A, Tellez-Rojo MM, Hernandez-Avila M. 2017. Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6-12 years of age in Mexico. *Environ Health Perspect* 125(9): 1-12.

Bashash M, Marchand M, Hu H, Till C, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Green R, Schnaas L, Mercado-Garcia A, Hernandez-Avila M, Tellez-Rojo MM. 2018. Prenatal fluoride exposure and attention deficit hyperactivity disorder (ADHD) symptoms in children at 6-12 years of age in Mexico City. *Environ Int* 121(Pt 1): 658-666.

Broadbent JM, Thomson WM, Moffitt TE, Poulton R. 2015. Community water fluoridation and intelligence response. *Am J Public Health* 105: 3-4.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 1991. [Research on the intellectual development of children in high fluoride areas]. *Chin J Control Endem Dis* 6(Suppl): 99-100.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 2008. Research on the intellectual development of children in high fluoride areas. *Fluoride* 41: 120-124.

Chinoy NJ, Narayana MV. 1992. Studies on fluorosis in Mehsana District of North Gujarat. *Proc Zool Soc* 45: 157-161.

Choi AL, Zhang Y, Sun G, Bellinger DC, Wang K, Yang XJ, Li JS, Zheng Q, Fu Y, Grandjean P. 2015. Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study. *Neurotoxicol Teratol* 47: 96-101.

Cui Y, Zhang B, Ma J, Wang Y, Zhao L, Hou C, Yu J, Zhao Y, Zhang Z, Nie J, Gao T, Zhou G, Liu H. 2018. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. *Ecotoxicol Environ Saf* 165: 270-277.

Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. 2020. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. *Neurosci Lett* 729: 134981.

Das K, Mondal NK. 2016. Dental fluorosis and urinary fluoride concentration as a reflection of fluoride exposure and its impact on IQ level and BMI of children of Laxmisagar, Simlapal Block of Bankura District, W.B., India. *Environ Monit Assess* 188: 218.

Ding Y, Sun H, Han H, Wang W, Ji X, Liu X, Sun D. 2011. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. *J Hazard Mater* 186: 1942-1946.

Du L, Wan C, Cao X, Liu J. 1992. [The effect of fluorine on the developing human brain]. *Chin J Pathol* 21(4): 218-220.

Du L, Wan C, Cao X, Liu J. 2008. The effect of fluorine on the developing human brain. *Fluoride* 41: 327-330.

Duan J, Zhao M, Wang L, Fang D, Wang Y, Wang W. 1995. A comparative analysis of the results of multiple tests in patients with chronic industrial fluorosis. Guizhou Med J 18(3): 179-180. Erickson JD, Hay S. 1976. Water fluoridation and congenital malformations: No association. *J Am Dent Assoc* 93: 981-984.

Erickson JD, Hay S. 1976. Water fluoridation and congenital malformations: No association. *J Am Dent Assoc* 93: 981-984.

Erickson JD. 1980. Down syndrome, water fluoridation, and maternal age. *Teratology* 21: 177-180.

Eswar P, Nagesh L, Devaraj CG. 2011. Intelligent quotients of 12-14 year old school children in a high and low fluoride village in India. *Fluoride* 44: 168-172.

Fan Z, Dai H, Bai A, Li P, Li T, Li G. 2007. Effect of high fluoride exposure in children's intelligence. *J Environ Health* 24(10): 802-803.

Green R, Lanphear B, Hornung R, Flora D, Martinez-Mier EA, Neufeld R, Ayotte P, Muckle G, Till C. 2019. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. *JAMA Pediatr*: E1-E9.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 1991. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. *Chin J Epidemiol* 10(2): 98-100.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 2008. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. *Fluoride* 41: 125-128.

Guo ZY, He YH, Zhu QX. 2001. [Research on the neurobehavioral function of workers occupationally exposed to fluoride]. *Ind Hlth & Occup Dis* 27(6): 346-348.

Guo ZY, He YH, Zhu QX. 2008. Research on the neurobehavioral function of workers occupationally exposed to fluoride. *Fluoride* 41: 152-155.

He H, Cheng ZS, Liu WQ. 1989. [Effects of fluorine on the human fetus]. *J Control Endem Dis* 4(3): 136-138.

He H, Cheng ZS, Liu WQ. 2008. Effects of fluorine on the human fetus. *Fluoride* 41: 321-326.

He MX, Zhang CN. 2010. [Investigation of children's intelligence quotient and dental fluorosis in drinking water-type of endemic fluorosis area in Pucheng County, Shaanxi Province before and after drinking water change]. *Chin J Endemiol* 29: 547-548.

Hong F, Wang H, Yang D, Zhang Z. 2001. [Investigation on the intelligence and metabolism of iodine and fluoride in children with high iodine and fluoride]. *Chin J Control Endem Dis* 12-14.

Hong FG, Cao YX, Yang D, Wang H. 2001. [Research on the effects of fluoride on child intellectual development under different environmental conditions]. *Chin Prim Health Care* 15(3): 56-57.

Hong FG, Cao YX, Yang D, Wang H. 2008. Research on the effects of fluoride on child intellectual development under different environmental conditions. *Fluoride* 41: 156-160.

Hosur MB, Puranik RS, Vanaki S, Puranik SR. 2012. Study of thyroid hormones free triiodothyronine (FT3), free thyroxine (FT4) and thyroid stimulating hormone (TSH) in subjects with dental fluorosis. *Eur J Dent* 6: 184-190.

Jacqmin H, Commenges D, Letenneur L, Barberger-Gateau P, Dartigues JF. 1994. Components of drinking water and risk of cognitive impairment in the elderly. *Am J Epidemiol* 139: 48-57.

Jin T, Han T, Wei Y, Wu Y, Wang Z, Zhang, H. 2016. [Investigation on working memory level of children aged 8-12 years in coal-burning fluorosis area]. *J Environ Health* 33(5): 409-411.

Jin T, Wang Z, Wei Y, Wu Y, Han T, Zhang H. 2017. [Investigation on intelligence level of children aged 8-12 years old in coal-burning fluorosis area]. *J Environ Health* 34(3): 229-231.

Kang J, Cheng Y, Wu K, Lin S, He G, Jin Y. 2011. Effect of exposure to fluoride and arsenic in drinking water of Hangjinhouqi on children's intelligence. *Chinese School Health*: 679-681.

Karimzade S, Aghaei M, Mahvi AH. 2014. Investigation of intelligence quotient in 9-12 year-old children exposed to high- and low-drinking water fluoride in West Azerbaijan Province, Iran. *Fluoride* 47: 9-14.

Khan SA, Singh RK, Navit S, Chadha D, Johri N, Navit P, Sharma A, Bahuguna R. 2015. Relationship between dental fluorosis and intelligence quotient of school going children in and around Lucknow District: A cross-sectional study. *J Clin Diagn Res* 9(11): 10-15.

Khandare AL, Gourineni SR, Validandi V. 2017. Dental fluorosis, nutritional status, kidney damage, and thyroid function along with bone metabolic indicators in school-going children living in fluoride-affected hilly areas of Doda District, Jammu and Kashmir, India. *Environ Monit Assess* 189: 579.

Khandare AL, Validandi V, Gourineni SR, Gopalan V, Nagalla B. 2018. Dose-dependent effect of fluoride on clinical and subclinical indices of fluorosis in school going children and its mitigation by supply of safe drinking water for 5 years: An Indian study. *Environ Monit Assess* 190: 110.

Kheradpisheh Z, Mahvi AH, Mirzaei M, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018. Correlation between drinking water fluoride and TSH hormone by ANNs and ANFIS. *J Environ Health Sci Eng* 16(1): 11-18.

Kheradpisheh Z, Mirzaei M, Mahvi AH, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018. Impact of drinking water fluoride on human thyroid hormones: A case-control study. *Sci Rep* 8: 2674.

Kumar V, Chahar P, Kajjari S, Rahman F, Bansal DK, Kapadia JM. 2018. Fluoride, thyroid hormone derangements and its correlation with tooth eruption pattern among the pediatric population from endemic and non-endemic fluorosis areas. *J Contemp Dent Pract* 19(12): 1512-1516.

Kundu H, Basavaraj P, Singla A, Gupta R, Singh K, Jain S. 2015. Effect of fluoride in drinking water on children's intelligence in high and low fluoride areas of Delhi. *J Indian Assoc Public Health Dent* 13(2): 116-121.

Lamberg M, Hausen H, Vartiainen T. 1997. Symptoms experienced during periods of actual and supposed water fluoridation. *Community Dent Oral Epidemiol* 25: 291-295.

Li F, Chen X, Huang R, Xie Y. 2009. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. *J Environ Health* 26(4): 838-840.

Li J, Yao L, Shao QL, Wu CY. 2004. [Effects of high fluoride level on neonatal neurobehavioral development]. *Chin J Endemiol* 23(5): 463-465.

Li J, Yao L, Shao QL, Wu CY. 2008. Effects of high fluoride level on neonatal neurobehavioral development. *Fluoride* 41: 165-170.

Li M, Gao Y, Cui J, Li Y, Li B, Liu Y, Sun J, Liu X, Liu H, Zhao L, Sun D. 2016. Cognitive impairment and risk factors in elderly people living in fluorosis areas in China. *Biol Trace Elem Res* 172: 53-60.

Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. 2010. Investigation and analysis of children's intelligence and dental fluorosis in high fluoride area. *J Med Pest Control* 26(3): 230-231.

Li XS, Zhi JL, Gao RO. 1995. Effect of fluoride exposure on the intelligence of children. *Fluoride* 28: 189-192.

Li Y, Li X, Wei S. 1994. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. *J West Chin University Med Sci* 25(2): 188-191.

Li Y, Li X, Wei S. 2008. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. *Fluoride* 41: 331-335.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2003. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. *Chin J Public Health Manag* 19(4): 337-338.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2008. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. *Fluoride* 41: 161-164.

Liang C, Ji R, Cao J, Jiang Y, Yu B, Ma F, Wu Y, Ying B, Zhang Y, Sun S, Li Y, Emsley CL, Gao S, Hall KS, Hendrie HC. 2003. [Study on the relationship between drinking water trace elements and cognitive ability of the elderly]. *Health Res* 436-440.

Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. 1991. High fluoride and low iodine environment and subclinical cretinism in Xinjiang. *Endem Dis Bull* 6(2): 62-67.

Liu S, Lu Y, Sun Z, Wu L, Wang X, Yan S. 2000. [Report on the intellectual ability of children living in high-fluoride water areas]. *Chin J Control Endem Dis* 15(4): 231-232.

Liu SL, Lu Y, Sun ZR, Wu L, Lu WL, Wang XW, Yan S. 2008. Report on the intellectual ability of children living in high-fluoride water areas. *Fluoride* 41: 144-147.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S, Mentes A, DenBesten PK. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. *Physiol Behav* 206: 76-83.

Lu Y, Sun ZR, Wu LN, Wang X, Lu W, Liu SS. 2000. Effect of high-fluoride water on intelligence in children. *Fluoride* 33: 74-78.

Luo Y, Ma R, Liu Z, Guan Z, Lou D, Zheng D. 2018. [Intelligence investigation and forensic significance of children in coal-burning fluorosis area]. *Chin J Forensic Medicine* 33(6): 590-593.

Malin AJ, Till C. 2015. Exposure to fluoridated water and attention deficit hyperactivity disorder prevalence among children and adolescents in the United States: An ecological association. *Environ Health* 14: 17.

Malin AJ, Riddell J, McCague H, Till C. 2018. Fluoride exposure and thyroid function among adults living in Canada: Effect modification by iodine status. *Environ Int* 121(Pt 1): 667-674.

Michael M, Barot VV, Chinoy NJ. 1996. Investigations of soft tissue functions in fluorotic individuals of North Gujarat. *Fluoride* 29: 63-71.

Mondal D, Dutta G, Gupta S. 2016. Inferring the fluoride hydrogeochemistry and effect of consuming fluoride-contaminated drinking water on human health in some endemic areas of Birbhum District, West Bengal. *Environ Geochem Health* 38: 557-576.

Morgan L, Allred E, Tavares M, Bellinger D, Needleman H. 1998. Investigation of the possible associations between fluorosis, fluoride exposure, and childhood behavior problems. *Pediatr Dent* 20: 244-252.

Mustafa DE, Younis UM, Elhag SAA. 2018. The relationship between the fluoride levels in drinking water and the schooling performance of children in rural areas of Khartoum State, Sudan. *Fluoride* 51(2): 102-113.

Nagarajappa R, Pujara P, Sharda AJ, Asawa K, Tak M, Aapaliya P, Bhanushali N. 2013. Comparative assessment of intelligence quotient among children living in high and low fluoride areas of Kutch, India: A pilot study. *Iran J Public Health* 42: 813-818.

Peckham S, Lowery D, Spencer S. 2015. Are fluoride levels in drinking water associated with hypothyroidism prevalence in England? A large observational study of GP practice data and fluoride levels in drinking water. *J Epidemiol Community Health* 69: 619-624.

Poureslami HR, Horri A, Garrusi B. 2011. A comparative study of the IQ of children age 7-9 in a high and a low fluoride water city in Iran. *Fluoride* 44: 163-167.

Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. 1990. [Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children]. *J Control Endem Dis* 5(4): 203-204.

Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. 2008. Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children. *Fluoride* 41: 115-119.

Ratan SK, Rattan KN, Pandey RM, Singhal S, Kharab S, Bala M, Singh V, Jhanwar A. 2008. Evaluation of the levels of folate, vitamin B12, homocysteine and fluoride in the parents and the affected neonates with neural tube defect and their matched controls. *Pediatr Surg Int* 24: 803-808.

Razdan P, Patthi B, Kumar JK, Agnihotri N, Chaudhari P, Prasad M. 2017. Effect of fluoride concentration in drinking water on intelligence quotient of 12-14-year-old children in Mathura District: A cross-sectional study. *J Int Soc Prev Community Dent* 7: 252-258.

Ren D, Li K, Liu D. 1989. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. *Chin J Control Endem Dis* 4(4): 251.

Ren D, Li K, Liu D. 2008. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. *Fluoride* 41: 319-320.

Riddell JK, Malin AJ, Flora D, McCague H, Till C. 2019. Association of water fluoride and urinary fluoride concentrations with attention deficit hyperactivity disorder in Canadian youth. *Environ Int* 133: 105190.

Rocha-Amador D, Navarro ME, Carrizales L, Morales R, Calderon J. 2007. Decreased intelligence in children and exposure to fluoride and arsenic in drinking water. *Cad Saude Publica* 23 Suppl 4: S579-587.

Rocha-Amador D, Navarro M, Trejo-Acevedo A, Carrizales L, Perez-Maldonado I, Diaz-Barriga F, Calderon J. 2009. Use of the Rey-Osterrieth Complex Figure Test for neurotoxicity evaluation of mixtures in children. *Neurotox* 30: 1149-1154.

Rotton J, Tikofsky RS, Feldman HT. 1982. Behavioral effects of chemicals in drinking water. *J Appl Psychol* 67: 230-238.

Russ TC, Killin LOJ, Hannah J, Batty GD, Deary IJ, Starr JM. 2019. Aluminium and fluoride in drinking water in relation to later dementia risk. *Brit J Psychol* 14: 1-6.

Saxena S, Sahay A, Goel P. 2012. Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India. *J Neurosci Rural Pract* 3: 144-149.

Sebastian ST, Sunitha S. 2015. A cross-sectional study to assess the intelligence quotient (IQ) of school going children aged 10-12 years in villages of Mysore District, India with different fluoride levels. *J Indian Soc Pedod Prev Dent* 33: 307-311.

Seraj B, Shahrabi M, Falahzade M, Falahzade F, Akhondi N. 2006. Effect of high fluoride concentration in drinking water on children's intelligence. *J Dent Med* 19(2): 80-86.

Seraj B, Shahrabi M, Shadfar M, Ahmadi R, Fallahzadeh M, Eslamlu HF, Kharazifard MJ. 2012. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. *J Dent* 9: 221-229.

Shannon FT, Fergusson DM, Horwood LJ. 1986. Exposure to fluoridated public water supplies and child health and behaviour. *New Zeal Med J* 99: 416-418.

Shao Q. 2003. Study of cognitive function impairment caused by chronic fluorosis. *Chin J Endemiol* 22(4): 336-338.

Sharma JD, Sohu D, Jain P. 2009. Prevalence of neurological manifestations in a human population exposed to fluoride in drinking water. *Fluoride* 42: 127-132.

Shivaprakash PK, Ohri K, Noorani H. 2011. Relation between dental fluorosis and intelligence quotient in school children of Bagalkot District. *J Indian Soc Pedod Prev Dent* 29: 117-120.

Singh A, Jolly SS, Devi P, Bansal BC, Singh SS. 1962. Endemic fluorosis: An epidemiological, biochemical and clinical study in the Bhatinda District of Panjab. *Indian J Med Res* 50: 387-398.

Singh N, Verma KG, Verma P, Sidhu GK, Sachdeva S. 2014. A comparative study of fluoride ingestion levels, serum thyroid hormone & TSH level derangements, dental fluorosis status among school children from endemic and non-endemic fluorosis areas. *Springerplus* 3: 7.

Singh V, Singh C, Sandeep T, Sandeep K, Vikas G, Mukesh T, Anurag T. 2013. A correlation between serum vitamin, acetylcholinesterase activity and IQ in children with excessive endemic fluoride exposure in Rajasthan, India. *Int Res J Medical Sci* 1(3): 12-16.

Soto-Barreras U, Escalante-Villalobos KY, Holguin-Loya B, Perez-Aguirre B, Nevarez-Rascon A, Martinez-Martinez RE, Loyola-Rodriguez JP. 2019. Effect of fluoride in drinking water on dental caries and IQ in children. *Fluoride* 52: 474-482.

Still CN, Kelley P. 1980. The incidence of primary degenerative dementia vs. water fluoride content in South Carolina, USA. *Neurotox* 1: 125-132.

Sudhir KM, Chandu GN, Prashant GM, Subba Reddy VV. 2009. Effect of fluoride exposure on intelligence quotient (IQ) among 13-15 year old school children of known endemic area of fluorosis, Nalgonda District, Andhra Pradesh. *J Indian Assoc Public Health Dent* 2009(13): 88-94.

Sun M, Li S, Wang Y, Li F. 1991. Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis. *J Guiyang Med Coll* 16(3): 204-206.

Susheela AK, Bhatnagar M, Vig K, Mondal NK. 2005. Excess fluoride ingestion and thyroid hormone derangements in children living in Delhi, India. *Fluoride* 38: 98-108.

Tamboli BL, Mathur GM, Mathur AP, Lalla SK, Goyal OP. 1980. Prevalence of fluorosis in Pratabpura and Surajpura villages, District Ajmer (Rajasthan). *Indian J Med Res* 71: 57-67.

Till C, Green R, Flora D, Hornung R, Martinez-Mier EA, Blazer M, Farmus L, Ayotte P, Muckle G, Lanphear B. 2020. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. *Environ Int* 134: 105315.

Tripathi P, Sultana N. 2007. Fluoride content of groundwater and prevalence of dental, skeletal and neurological stage of fluorosis in Tehsil Purwa of Unnao. *Indian J Environ Prot* 27: 737-739.

Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. 2007. Effect of high fluoride water on intelligence of school children in India. *Fluoride* 40: 178-183.

Trivedi M, Sangai N, Patel R, Payak M, Vyas S. 2012. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. *Fluoride* 45(4): 377-383.

Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, Costilla-Salazar R, Calderon Hernandez J, Alcaraz Contreras Y, Rocha-Amador DO. 2017. In utero exposure to fluoride and cognitive development delay in infants. *Neurotox* 59: 65-70.

Wang A, Duan L, Huang H, Ma J, Zhang Y, Ma Q, Guo Y, Li Z, Cheng X, Zhu J, Zhou G, Ba Y. 2020. Association between fluoride exposure and behavioural outcomes of school-age children: A pilot study in China. *Int J Environ Health Res*: 1-10.

Wang G, Yang D, Jia F, Wang H. 1996. [A study of the IQ levels of four- to seven-year-old children in high fluoride areas]. *Endem Dis Bull* 11(1): 60-62.

Wang G, Yang D, Jia F, Wang H. 2008. A study of the IQ levels of four- to seven-year-old children in high fluoride areas. *Fluoride* 41: 340-343.

Wang G, Gao M, Zhang M, Yang M, Xiang Q. 2012. Correlation between total fluoride intake and children's IQ. *J Southeast Univ Med Ed*: 743-746.

Wang G, Zhang M, Wang Q, Han A, Gao M, Lin P, Xiang Q. 2017. [Investigation on the relationship between serum fluoride content and IQ of children before and after reducing fluoride to water]. *Capital Public Health* 11(6): 274-277.

Wang M, Liu L, Li H, Li Y, Liu H, Hou C, Zeng Q, Li P, Zhao Q, Dong L, Zhou G, Yu X, Liu L, Guan Q, Zhang S, Wang A. 2020. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. *Environ Int* 134: 105229.

Wang S, Wang L, Hu P, Guo S, Law S. 2001. [Effects of high iodine and high fluorine on children’s intelligence and thyroid function]. *Chin J Endemiol* 20(4): 288-290.

Wang SX, Wang ZH, Cheng XT, Li J, Sang ZP, Zhang XD, Han LL, Qiao XY, Wu ZM, Wang ZQ. 2005. [Investigation and evaluation on intelligence and growth of children in endemic fluorosis and arsenism areas]. *Chin J Endemiol* 24: 179-182.

Wang SX, Wang ZH, Cheng XT, Li J, Sang ZP, Zhang XD, Han LL, Qiao XY, Wu ZM, Wang ZQ. 2007. Arsenic and fluoride exposure in drinking water: Children's IQ and growth in Shanyin County, Shanxi Province, China. *Environ Health Perspect* 115: 643-647.

Wang SY, Zhang HX, Fan W, Fang SJ, Kang PP, Chen XG, Yu MJ. 2005. [The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children]. *J Appl Clin Ped* 20(9): 897-899.

Wang SY, Zhang HX, Fan W, Fang SJ, Kang PP, Chen XG, Yu MJ. 2008. The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children. *Fluoride* 41: 344-348.

Wang Z, Wang S, Zhang X, Li J, Zheng X, Hu C. 2006. Investigation of children’s growth and development under long-term fluoride exposure. *Chin J Control Endem Dis* 21(4): 239-241.

Wei N, Li Y, Deng J, Xu S, Guan Z. 2014. [The effects of comprehensive control measures on intelligence of school-age children in coal-burning-borne endemic fluorosis areas]. *Chin J Endemiol* 33(2): 320-322.

Xiang Q, Liang Y, Chen L, Wang C, Chen B, Chen X, Zhou M. 2003. Effect of fluoride in drinking water on children's intelligence. *Fluoride* 36: 84-94.

Xiang Q, Liang Y, Chen B, Chen L. 2011. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. *Fluoride* 44: 191-194.

Xu Y, Lu C, Zhang X. 1994. Effect of fluoride on children's intelligence. *Endem Dis Bull* 2: 83-84.

Yang Y, Wang X, Guo X, Hu P. 1994. [The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine]. *Chin J Epidemiol* 15(4): 296-298.

Yang Y, Wang X, Guo X, Hu P. 2008. The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine. *Fluoride* 41: 336-339.

Yao L, Zhou J, Wang S, Cui K, Lin F. 1996. Analysis of TSH levels and intelligence of children residing in high fluorosis areas. *Lit Inf Prev Med* 2(1): 26-27.

Yao Y. 1997. Comparative assessment of the physical and mental development of children in endemic fluorosis area with water improvement and without water improvement. *Lit Inf Prev Med* 3(1): 42-43.

Yasmin S, Ranjan S, D'Souza D. 2013. Effect of excess fluoride ingestion on human thyroid function in Gaya Region, Bihar, India. *Toxicol Environ Chem* 95: 1235-1243.

Yazdi SM, Sharifian A, Dehghani-Beshne M, Momeni VR, Aminian O. 2011. Effects of fluoride on psychomotor performance and memory of aluminum potroom workers. *Fluoride* 44: 158-162.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 1996. [Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis]. *Chin J Endemiol* 15(5): 257-259.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 2008. Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis. *Fluoride* 41: 134-138.

Yu X, Chen J, Li Y, Liu H, Hou C, Zeng Q, Cui Y, Zhao L, Li P, Zhou Z, Pang S, Tang S, Tian K, Zhao Q, Dong L, Xu C, Zhang X, Zhang S, Liu L, Wang A. 2018. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. *Environ Int* 118: 116-124.

Zhang J, Yao H, Chen Y. 1998. [Effect of high level of fluoride and arsenic on children's intelligence]. *Chin J Public Health* 17(2): 57.

Zhang P, Cheng L. 2015. [Effect of coal-burning endemic fluorosis on children's physical development and intellectual level]. *Chin J Control Endem Dis* 30(6): 458-459.

Zhang S, Zhang X, Liu H, Qu W, Guan Z, Zeng Q, Jiang C, Gao H, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Cui Y, Yu L, Wang A. 2015. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. *Toxicol Sci* 144: 238-245.

Zhao LB, Liang GH, Zhang DN, Wu XR. 1996. Effect of a high fluoride water supply on children's intelligence. *Fluoride* 29: 190-192.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao Q, Tian Z, Zhou G, Niu Q, Chen J, Li P, Dong L, Xia T, Zhang S, Wang A. 2020. SIRT1-dependent mitochondrial biogenesis supports therapeutic effects of resveratrol against neurodevelopment damage by fluoride. *Theranostics* 10: 4822-4838.

Zhao Y, Cui Y, Yu J, Zhang B, Nie J, Zhao L, Zhang Z, Liu H. 2018. [Study on the relationship between water-borne high iodine and thyroid hormone and children's intelligence level]. *J Environ Health* 35(1): 6-9.

Zhou G, Tang S, Yang L, Niu Q, Chen J, Xia T, Wang S, Wang M, Zhao Q, Liu L, Li P, Dong L, Yang K, Zhang S, Wang A. 2019. Effects of long-term fluoride exposure on cognitive ability and the underlying mechanisms: Role of autophagy and its association with apoptosis. *Toxicol Appl Pharmacol* 378: 114608.

C.2.1.2. Studies Not Available in HAWC

Bachinskii PP, Gutsalenko OA, Naryzhniuk ND, Sidora VD, Shliakhta AI. 1985. [Action of the body fluorine of healthy persons and thyroidopathy patients on the function of hypophyseal-thyroid the system]. *Probl Endokrinol* 31: 25-29.

Balabolkin MI, Mikhailets ND, Lobovskaia RN, Chernousova NV. 1995. [The interrelationship of the thyroid and immune statuses of workers with long-term fluorine exposure]. *Ter Arkh* 67: 41-42.

Baum K, Boerner W, Reiners C, Moll E. 1981. [Bone density and thyroid function in adolescents in relation to fluoride content of drinking water]. *Fortschr Med* 99: 1470-1472.

Berry WTC, Whittles JH. 1963. Absence of effect of fluoride upon the incidence of thyroid enlargements in Wiltshire schoolgirls. *Mon Bull Minist Health Public Health Lab Serv* 22: 50-52.

Cherkinskii SN, Zaslavskaia RM. 1956. [Significance of fluorides in potable water in the development of endemic goiter]. *Probl Endokrinol Gormonoter* 2: 70-75.

Choubisa SL. 2001. Endemic fluorosis in southern Rajasthan, India. *Fluoride* 34: 61-70.

Chuka A, Zhukovskil V, Mirku I, Postel'Niku D. 1964. Prezhdevremennoe starenie naseleniya v zone rasprostraneniya endemicheskogo zoba. *Vestnik Akad Med Nauk Sssr* 19: 23-27.

Dai HX, Zeng P, Wang KY, Zhang XG, Ma ZJ, Zhou YG, Fan ZX, Guo SH. 2013. [Analysis of a survey results of patients with suspected high iodine goiter in Liuji Town Fuping County of Shaanxi Province]. *Chin J Endemiol* 32: 408-411.

Day T, Powell-Jackson P. 1972. Fluoride, water hardness, and endemic goitre. *Lancet* 299(7761): 1135-1138.

Desai VK, Solanki DM, Bansal RK. 1993. Epidemiological study of goitre in endemic fluorosis district of Gujarat. *Fluoride* 26: 187-190.

Díaz-Cadórniga FJ, Delgado E, Tartón T, Valdés MM, Méndez A, Fernández MT, Rojo C. 2003. Endemic goiter associated with high iodine intake in primary school children in the Saharawi Arab Democratic Republic. *Endocrinol Nutr* 50: 357-362.

Eichner R, Borner W, Henschler D, Kohler W, Moll E. 1981. [Osteoporosis therapy and thyroid function. Influence of 6 months of sodium fluoride treatment on thyroid function and bone density]. *Fortschr Med* 99: 342-348.

Fiorentini S, Galeazzi M, Visintin B. 1947. II fluoro in natura come agente morbigeno II. La fluorosi die Campagnano di Roma. III. Un focolaio di fluorosi umana a Campagnano di Roma. IV. Osservazioni radiologiche sui processi alveolari, sulle ossa mascellari, e sul paradenzio degli abitanti die Campagnano. V. Zona fluorotica intorno a Campagnano di Roma. VI. Frequenza e caratteri clinici della carie dentale in soggetti fluorotici. *Rend Ist Superiore Sanita* 10: 721-804.

Gabovich RD, Verzhikovskaia NV. 1958. [Effect of fluoride compounds on absorption of radioactive iodine by the thyroid gland in humans and in experimental conditions]. *Probl Endokrinol Gormonoter* 4: 49-54.

Galletti P, Held HR, Korrodi H, Wegmann T. 1956. [Investigations on the possible damaging effect of fluorine on the thyroid gland]. *Helv Med Acta* 23: 601-605.

Galletti PM, Joyet G, Jallut O. 1957. [Effect of sodium fluoride on thyroid function in Basedow's Disease]. *Helv Med Acta* 24: 209-215.

Galletti PM, Joyet G. 1958. Effect of fluorine on thyroidal iodine metabolism in hyperthyroidism. *J Clin Endocrinol Metab* 18: 1102-1110.

Gas'kov AI, Savchenkov MF, Iushkov NN. 2005. [The specific features of the development of iodine deficiencies in children living under environmental pollution with fluorine compounds]. *Gig Sanit*: 53-55.

Gedalia I, Brand N. 1963. The relationship of fluoride and iodine in drinking water in the occurrence of goiter. *Arch Int Pharmacodyn Ther* 142: 312-315.

Grimm H. 1973. [The physical development of schoolchildren under the influence of drinking water fluoridation in Karl Marx Stadt]. *Dtsch Gesundheitsw* 28: 2363-2369.

Hasling C, Nielsen HE, Melsen F, Mosekilde L. 1987. Safety of osteoporosis treatment with sodium fluoride, calcium phosphate and vitamin D. *Miner Electrolyte Metab* 13: 96-103.

Hidehiko T. 1958. On the relation between the distribution of endemic goiter and the fluorine content of natural water in Hidaka Province, Hokkaido. *Folia Pharmacol Jpn* 54: 225-229.

Hoffmann-Axthelm W. 1953. [Observations on the influence of fluorine on dental enamel and thyroid gland]. *Dtsch Zahnarztl Z* 8: 757-765.

Jentzer A. 1956. [Effect of fluorine on the iodine content of the human thyroid gland]. *Bull Schweiz Akad Med Wiss* 12: 539-543.

Jooste PL, Weight MJ, Kriek JA, Louw AJ. 1999. Endemic goitre in the absence of iodine deficiency in schoolchildren of the Northern Cape Province of South Africa. *Eur J Clin Nutr* 53: 8-12.

Kolomiitseva MG. 1961. [The content of fluorine in the external environment of the Upper Altai autonomous region and its role in the etiology of endemic goiter]. *Gig Sanit* 26: 101-103.

Korrodi H, Wegmann T, Galleti P, Held HR. 1955. [Caries prophylaxis and the untoward effects of fluor on the thyroid gland]. *Schweiz Med Wochenschr* 85: 1016-1019.

Kutlucan A, Kale Koroglu B, Numan Tamer M, Aydin Y, Baltaci D, Akdogan M, Ozturk M, Vural H, Ermis F. 2013. The investigation of effects of fluorosis on thyroid volume in school-age children. *Med Glas* 10: 93-98.

Latham MC, Grech P. 1967. The effects of excessive fluoride intake. *Am J Public Health* 57: 651-660.

Leone NC, Leatherwood EC, Petrie IM, Lieberman L. 1964. Effect of fluoride on thyroid gland: Clinical study. *J Am Dent Assoc* 69: 179-180.

Levi JE, Silberstein HE. 1955. Lack of effect of fluorine ingestion on uptake of iodine 131 by the thyroid gland. *J Lab Clin Med* 45: 348-351.

McGlashan N, Chelkowska E, Sasananan S. 2010. A survey of goiter morbidity in Ban Mae Toen, northwest Thailand. *Southeast Asian J Trop Med Public Health* 41: 1200-1208.

Rathore S, Meena C, Gonmei Z, Dwivedi S, Toteja GS, Bala K. 2018. Study of excess fluoride ingestion and thyroid hormone derangement in relation with different fluoride levels in drinking water among children of Jodhpur District, Rajasthan, India. *Asian J Microbiol Biotechnol Environ Sci* 20: 327-331.

Reisenauer R, Rezler D, Křemenová J, Preininger Q. 1961. [Fluorization of the waters in Czechoslovakia. IV. Endocrinological control of results of two years' fluorization of drinking-water in school children]. *Cesk Stomatol* 61: 91-97.

Romer TEZ, Kowalczyk B, Kacprzak M, Wiktorowski M. 1976. [Incidence of goiter in pubertal girls of the Piotrkow Region and iodide content in drinking water]. *Endokrynol Pol* 27: 373-380.

Savchenkov MF, Efimova NV, Manueva RS, Nikolaeva LA, Shin NS. 2016. [Thyroid gland pathology in children population exposed to the combination of iodine deficiency and fluoride pollution of environment]. *Gig Sanit* 95: 1201-1205.

Shtifanova AK. 1962. [The fluorine content in water, soil and vegetal products of the Alma-Atinsk District areas and its role in the etiology of dental caries and endemic goiter]. *Zdravookhranenie Kazakhstana*: 60-63.

Siddiqui AH. 1969. Incidence of simple goiter in areas of endemic fluorosis in Nalgonda District, Andhra Pradesh, India. *Fluoride* 2: 200-205.

Sidora VD, Shliakhta AI, Iugov VK, Kas'ianenko AS, Piatenko VG. 1983. [Indices of the pituitary-thyroid system in residents of cities with various fluorine concentrations in drinking water]. *Probl Endokrinol* 29: 32-35.

Sung FC, Chen KP, Chen CY, Tai PW, Yang CF. 1973. Studies of the effect of salt iodization on endemic goiter in Taiwan. IV. A survey of drinking water in relation to endemic goiter. *J Fomosan Med Assoc* 72: 96-103.

Tokar VI, Voroshnin VV, Sherbakov SV. 1989. [Chronic effects of fluorides on the pituitary-thyroid system in industrial workers]. *Gig Tr Prof Zabol*: 19-22.

Wespi HJ. 1954. [Iodized-fluoridized salt for the prevention of goiter and caries]. *Schweiz Med Wochenschr* 84: 885-890.

Yu YN. 1985. [Effects of chronic fluorosis on the thyroid gland]. *Chin Med J* 65: 747-7479.

C.2.2. Studies in Non-human Animals

As described in Figure 2, 339 non-human mammal studies were included; however, full data extraction was conducted only on studies with primary neurological outcomes and/or secondary functional neurological outcomes (e.g., motor activity). Data extraction was completed using HAWC. Data were extracted from a subset of included studies in animals (n = 123) and are available in HAWC based on outcome. The following lists of references are organized as studies that are available in HAWC followed by studies that are not available in HAWC. Specifically, all primary outcomes and functional neurological secondary outcomes (e.g., motor activity) were extracted from animal studies and are available in HAWC, including studies from the NTP (2016) assessment. Studies are also available in HAWC that evaluated mechanistic effects related to oral fluoride exposure at or below 20 ppm fluoride drinking water equivalents for categories of mechanistic endpoints with the largest amount of available data (i.e., biochemistry of the brain or neurons, neurotransmission, oxidative stress, and histopathology [n = 70]); however, these mechanistic data were generally not extracted. Several animal studies assessed primary neurological outcomes and/or functional neurological secondary outcomes and mechanistic effects in the four mechanistic categories listed above (n = 56). In total, 140 animal studies are available in HAWC (70 with primary neurological outcomes and/or secondary functional neurological outcomes without relevant mechanistic data; 15 with relevant mechanistic data only; and 55 with primary and/or secondary functional neurological outcomes with relevant mechanistic data). Studies that evaluated other mechanistic endpoints, as well as studies that assessed only mechanistic effects at fluoride levels above 20 ppm fluoride drinking water equivalents, are not available in HAWC (n = 199). The list below presents the 339 non-human animal studies that were included in the review. An overview of the screening results is outlined in the study selection diagram (Figure 2) that reports numbers of included studies as well as numbers of studies excluded for each reason at the full-text review stage.

C.2.2.1. Studies Available in HAWC

Adedara IA, Abolaji AO, Idris UF, Olabiyi BF, Onibiyo EM, Ojuade TD, Farombi EO. 2017. Neuroprotective influence of taurine on fluoride-induced biochemical and behavioral deficits in rats. *Chem Biol Interact* 261: 1-10.

Adedara IA, Olabiyi BF, Ojuade TD, Idris UF, Onibiyo EM, Farombi EO. 2017. Taurine reverses sodium fluoride-mediated increase in inflammation, caspase-3 activity, and oxidative damage along the brain-pituitary-gonadal axis in male rats. *Can J Phys Pharmacol* 95: 1019-1029.

Agustina F, Sofro ZM, Partadiredja G. 2018. Subchronic administration of high-dose sodium fluoride causes deficits in cerebellar purkinje cells but not motor coordination of rats. *Biol Trace Elem Res* 188(2): 424-433.

Ahmad KR, Noor S, Jabeen S, Nauroze T, Kanwal MA, Raees K, Abbas T. 2017. Amelioration by jambul fruit extract of fluoride-induced hepato-nephronal histopathologies and impaired neuromotor capacity in mice. *Fluoride* 50: 2-14.

Akinrinade ID, Memudu AE, Ogundele OM. 2015. Fluoride and aluminium disturb neuronal morphology, transport functions, cholinesterase, lysosomal and cell cycle activities. *Pathophysiology* 22: 105-115.

Akinrinade ID, Memudu AE, Ogundele OM, Ajetunmobi OI. 2015. Interplay of glia activation and oxidative stress formation in fluoride and aluminium exposure. *Pathophysiology* 22: 39-48.

Baba NA, Raina R, Verma PK, Sultana M. 2014. Alterations in plasma and tissue acetylcholinesterase activity following repeated oral exposure of chlorpyrifos alone and in conjunction with fluoride in Wistar rats. *Proc Indian Natl Sci Acad B Biol Sci* 84: 969-972.

Bagmut I, Kolisnyk I, Titkova A, Babiy L, Filipchenko S. 2018. The antioxidant system enzymes' activity in rats' brain, intoxicated with sodium fluoride in subtoxic doses. *Arch Balkan Med Union* 53(4): 506-511.

Balaji B, Kumar EP, Kumar A. 2015. Evaluation of standardized bacopa monniera extract in sodium fluoride-induced behavioural, biochemical, and histopathological alterations in mice. *Toxicol Ind Health* 31: 18-30.

Balayssac D, Richard D, Authier N, Nicolay A, Jourdan D, Eschalier A, Coudore F. 2002. Absence of painful neuropathy after chronic oral fluoride intake in Sprague-Dawley and Lou/C rats. *Neurosci Lett* 327: 169-172.

Banala RR, Karnati PR. 2015. Vitamin A deficiency: An oxidative stress marker in sodium fluoride (NaF) induced oxidative damage in developing rat brain. *Int J Dev Neurosci* 47: 298-303.

Banala R, Nagapuri K, Mohd K, Reddy M, Karnati P. 2018. Carica Papaya leaf extract as a neuroprotective agent against behavioral and neurotransmitter changes in brain of the rat treated with sodium fluoride in pre- and post-natal periods. *Pharmacogn Mag* 14(55): 123-131.

Banji D, Banji OJ, Pratusha NG, Annamalai AR. 2013. Investigation on the role of spirulina platensis in ameliorating behavioural changes, thyroid dysfunction and oxidative stress in offspring of pregnant rats exposed to fluoride. *Food Chem* 140: 321-331.

Baran-Poesina V, Negres S, Dobrescu D, Dimcevici-Poesina N, Dimcevici-Poesina A, Feghiu A, Soare T, Militaru M. 2013. Experimental pharmacological researches regarding the influence of sodium fluoride in allopathic and homeopathic doses on central nervous system's performances:

A correlation between behavioral response in classic maze test and morphological aspects of cerebral cortex. *Farmacia* 61: 781-799.

Bartos M, Gumilar F, Bras C, Gallegos CE, Giannuzzi L, Cancela LM, Minetti A. 2015. Neurobehavioural effects of exposure to fluoride in the earliest stages of rat development. *Physiol Behav* 147: 205-212.

Bartos M, Gumilar F, Gallegos CE, Bras C, Dominguez S, Monaco N, Esandi MDC, Bouzat C, Cancela LM, Minetti A. 2018. Alterations in the memory of rat offspring exposed to low levels of fluoride during gestation and lactation: Involvement of the alpha7 nicotinic receptor and oxidative stress. *Reprod Toxicol* 81: 108-114.

Basha PM, Rai P, Begum S. 2011. Fluoride toxicity and status of serum thyroid hormones, brain histopathology, and learning memory in rats: A multigenerational assessment. *Biol Trace Elem Res* 144: 1083-1094.

Basha PM, Sujitha NS. 2012. Combined impact of exercise and temperature in learning and memory performance of fluoride toxicated rats. *Biol Trace Elem Res* 150: 306-313.

Bataineh HN, Nusier MK. 2006. Impact of 12-week ingestion of sodium fluoride on aggression, sexual behavior, and fertility in adult male rats. *Fluoride* 39: 293-301.

Bera I, Sabatini R, Auteri P, Flace P, Sisto G, Montagnani M, Potenza MA, Marasciulo FL, Carratu MR, Coluccia A, Borracci P, Tarullo A, Cagiano R. 2007. Neurofunctional effects of developmental sodium fluoride exposure in rats. *Eur Rev Med Pharmacol Sci* 11: 211-224.

Bhatnagar M, Rao P, Sushma J, Bhatnagar R. 2002. Neurotoxicity of fluoride: Neurodegeneration in hippocampus of female mice. *Indian J Exp Biol* 40: 546-554.

Bhatnagar M, Sukhwal P, Suhalka P, Jain A, Joshi C, Sharma D. 2011. Effects of fluoride in drinking water on NADPH-diaphorase neurons in the forebrain of mice: A possible mechanism of fluoride neurotoxicity. *Fluoride* 44: 195-209.

Chen H, Geng D. 2011. [The change of cognition induced by chronic fluoride in rats]. *Acta Academiae Medicinae Xuzhou* 31(5): 319-322.

Chen J, Niu Q, Xia T, Zhou G, Li P, Zhao Q, Xu C, Dong L, Zhang S, Wang A. 2018. ERK1/2-mediated disruption of BDNF-TrkB signaling causes synaptic impairment contributing to fluoride-induced developmental neurotoxicity. *Toxicology* 410: 222-230.

Chinoy NJ, Shah SD. 2004. Biochemical effects of sodium fluoride and arsenic trioxide toxicity and their reversal in the brain of mice. *Fluoride* 37: 80-87.

Chioca LR, Raupp IM, Da Cunha C, Losso EM, Andreatini R. 2008. Subchronic fluoride intake induces impairment in habituation and active avoidance tasks in rats. *Eur J Pharmacol* 579: 196-201.

Chouhan S, Flora SJS. 2008. Effects of fluoride on the tissue oxidative stress and apoptosis in rats: Biochemical assays supported by IR spectroscopy data. *Toxicology* 254: 61-67.

Chouhan S, Lomash V, Flora SJ. 2010. Fluoride-induced changes in haem biosynthesis pathway, neurological variables and tissue histopathology of rats. *J Appl Toxicol* 30: 63-73.

Cui YS, Zhong Q, Li WF, Liu ZH, Wang Y, Hou CC. 2017. [Effects of fluoride exposure on thyroid hormone level and intelligence in rats]. *Chin J Ind Hyg Occup Dis* 35: 888-892.

Dabrowska E. 1997. Effect of different fluorine doses on the supraoptic nucleus of the rat. *Folia Histochem Cytobiol* 35: 115-116.

Dong Y, Wang Y, Wei N, Guan Z. 2015. [Expression levels of brain muscarinic acetylcholine receptor in offspring rats of drinking-water borne fluorosis]. *Chin J Endemiol* 34: 326-330.

Dong YT, Wang Y, Wei N, Zhang QF, Guan ZZ. 2015. Deficit in learning and memory of rats with chronic fluorosis correlates with the decreased expressions of M1 and M3 muscarinic acetylcholine receptors. *Arch Toxicol* 89: 1981-1991.

Dong YT, Wei N, Qi XL, Liu XH, Chen D, Zeng XX, Guan ZZ. 2017. Attenuating effect of vitamin E on the deficit of learning and memory of rats with chronic fluorosis: The mechanism may involve muscarinic acetylcholine receptors. *Fluoride* 50: 354-364.

Dong YW, Y. Wei, N. Guan, Z. 2015. [Expression of muscarinic acetylcholine receptors in the brain of rats with chronic fluorosis]. *Chin J Endemiol* 34(2): 84-88.

Ekambaram P, Paul V. 2001. Calcium preventing locomotor behavioral and dental toxicities of fluoride by decreasing serum fluoride level in rats. *Environ Toxicol Pharmacol* 9: 141-146.

Ekambaram P, Paul V. 2002. Modulation of fluoride toxicity in rats by calcium carbonate and by withdrawal of fluoride exposure. *Pharmacol Toxicol* 90: 53-58.

Ekambaram P, Paul V. 2003. Effect of vitamin D on chronic behavioral and dental toxicities of sodium fluoride in rats. *Fluoride* 36: 189-197.

El-lethey HS, Kamel MM, Shaheed IB. 2010. Neurobehavioral toxicity produced by sodium fluoride in drinking water of laboratory rats. *J Am Sci* 6(5): 54-63.

El-lethey HS, Kamel MM. 2011. Effects of black tea in mitigation of sodium fluoride potency to suppress motor activity and coordination in laboratory rats. *J Am Sci* 7(4): 243-254.

El-lethey HS, Shaheed IB. 2011. Potential health impact of black tea against Na-F-induced alterations in territorial aggression, sexual behaviour and fertility of male rats. *Life Sci J* 8: 828-839.

Elliott L. 1967. Lack of effect of administration of fluoride on the central nervous system of rats. *Acta Pharmacol Toxicol (Copenh)* 25: 323-328.

Flace P, Benagiano V, Vermesan D, Sabatini R, Inchingolo AM, Auteri P, Ambrosi G, Tarullo A, Cagiano R. 2010. Effects of developmental fluoride exposure on rat ultrasonic vocalization, acoustic startle reflex and pre-pulse inhibition. *Eur Rev Med Pharmacol Sci* 14: 507-512.

Gabovich RD. 1962. [On the problem of the effect of fluorine in drinking water on the functional state of the central nervous system]. *Gig Sanit* 27: 10-12.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008. [Level of oxidative stress in rat brains and learning and memory function of rats with chronic fluorosis]. *Chin J Endemiol* 27: 371-373.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008. [Effects of fluoride on learning and memory and cholinesterase activity in rat brains]. *Chin J Endemiol* 27: 128-130.

Gao Q, Liu YJ, Guan ZZ. 2009. Decreased learning and memory ability in rats with fluorosis: Increased oxidative stress and reduced cholinesterase activity in the brain. *Fluoride* 42: 277-285.

Gao Y, Liu L, Young L, Huan L, Jin H. 2009. Effects of learning and memory of fluoride and the antagonism of selenium in rats. *Studies of Trace Elements and Health* 26(2): 1-3.

Ge QD, Tan Y, Luo Y, Wang WJ, Zhang H, Xie C. 2018. MiR-132, miR-204 and BDNF-TrkB signaling pathway may be involved in spatial learning and memory impairment of the offspring rats caused by fluorine and aluminum exposure during the embryonic stage and into adulthood. *Environ Toxicol Pharmacol* 63: 60-68.

Ge Y, Chen L, Yin Z, Song X, Ruan T, Hua L, Liu J, Wang J, Ning H. 2018. Fluoride-induced alterations of synapse-related proteins in the cerebral cortex of ICR offspring mouse brain. *Chemosphere* 201: 874-883.

Gopal K, Saxena R, Gupta GSD, Rana MD, Agrawal D. 2006. Fluoride induced alterations in neurobehavioural and cardiovascular responses in rats. *J Adv Zool* 27: 1-7.

Gui CZ, Ran LY, Wu CX, Long YG, He J, Zhang H, Guan ZZ. 2009. [Changes in learning and memory ability and brain cholinesterase activity in the rats with coal burning fluorosis]. *Chin J Endemiol* 28: 497-500.

Gui CZ, Ran LY, Li JP, Guan ZZ. 2010. Changes of learning and memory ability and brain nicotinic receptors of rat offspring with coal burning fluorosis. *Neurotoxicol Teratol* 32: 536-541.

Gui CZ, Ran LY, Guan ZZ. 2011. [Expression levels of brain nicotinic acetylcholine receptor mRNA and protein in coal-burning type of fluorosis rats]. *Chin J Endemiol* 30: 239-242.

Guner S, Uyar-Bozkurt S, Haznedaroglu E, Mentes A. 2016. Dental fluorosis and catalase immunoreactivity of the brain tissues in rats exposed to high fluoride pre- and postnatally. *Biol Trace Elem Res* 174: 150-157.

Han H, Du W, Zhou B, Zhang W, Xu G, Niu R, Sun Z. 2014. Effects of chronic fluoride exposure on object recognition memory and mRNA expression of SNARE complex in hippocampus of male mice. *Biol Trace Elem Res* 158: 58-64.

Hong JH, Ge YM, Ning HM, Wang JD. 2005. [Effects of High Fluoride and Low Iodine on Learning-Memory and TchE of Brain in Offspring Rats]. *Chin Prev Med* 6: 489-491.

Inkielewicz I, Krechniak J. 2004. Fluoride effects on glutathione peroxidase and lipid peroxidation in rats. *Fluoride* 37: 7-12.

Jain A, Mehta VK, Chittora RA, Mahdi A, Bhatnagar M. 2015. Melatonin ameliorates fluoride induced neurotoxicity in young rats: An in vivo evidence. *Asian J Pharm Clin Res* 8: 164-167.

Jetti R, Raghuveer CV, Mallikarjuna RC. 2016. Protective effect of ascorbic acid and Ginkgo biloba against learning and memory deficits caused by fluoride. *Toxicol Ind Health* 32: 183-187.

Jia B, Zong L, Lee JY, Lei J, Zhu Y, Xie H, Clemens JL, Feller MC, Na Q, Dong J, McLane MW, Jones-Beatty K, Burd I. 2019. Maternal supplementation of low dose fluoride alleviates adverse perinatal outcomes following exposure to intrauterine inflammation. *Sci Rep* 9(1): 2575.

Jiang C, Zhang S, Liu H, Guan Z, Zeng Q, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Zhang X, Cui Y, Yu L, Wang A. 2014. Low glucose utilization and neurodegenerative changes caused by sodium fluoride exposure in rat's developmental brain. *Neuromolecular Med* 16: 94-105.

Jiang S, Su J, Yao S, Zhang Y, Cao F, Wang F, Wang H, Li J, Xi S. 2014. Fluoride and arsenic exposure impairs learning and memory and decreases mGluR5 expression in the hippocampus and cortex in rats. *PLoS One* 9(4): e96041.

Khan AM, Raina R, Dubey N, Verma PK. 2017. Effect of deltamethrin and fluoride co-exposure on the brain antioxidant status and cholinesterase activity in Wistar rats. *Drug Chem Toxicol* 41(2): 1-5.

Kinawy AA, Al-Eidan AA. 2018. Impact of prenatal and postnatal treatment of sodium fluoride and aluminum chloride on some hormonal and sensorimotor aspects in rats. *Biol Trace Elem Res*: 1-8.

Kivrak Y. 2012. Effects of fluoride on anxiety and depression in mice. *Fluoride* 45: 302-306.

Li M, Cui J, Gao YH, Zhang W, Sun LY, Liu XN, Liu Y, Sun DJ. 2015. Pathological changes and effect on the learning and memory ability in rats exposed to fluoride and aluminum. *Toxicol Res* 4: 1366-1373.

Li X, Zhang J, Niu R, Manthari RK, Yang K, Wang J. 2019. Effect of fluoride exposure on anxiety- and depression-like behavior in mouse. *Chemosphere* 215: 454-460.

Liu F, Ma J, Zhang H, Liu P, Liu YP, Xing B, Dang YH. 2014. Fluoride exposure during development affects both cognition and emotion in mice. *Physiol Behav* 124: 1-7.

Liu WX. 1989. [Experimental study of behavior and cerebral morphology of rat pups generated by fluorotic female rat]. *Chin J Pathol* 18: 290-292.

Liu YJ, Gao Q, Wu CX, Long YG, Guan ZZ. 2009. [Modified expression of extracellular signal-regulated protein kinase signal transduction in rat brains and changed capacity of learning and memory of rats with chronic fluorosis]. *Chin J Endemiol* 28: 32-35.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. Alterations of nAChRs and ERK1/2 in the brains of rats with chronic fluorosis and their connections with the decreased capacity of learning and memory. *Toxicol Lett* 192: 324-329.

Liu YJ, Gao Q, Long YG, Yu YN, Guan ZZ. 2011. [Influence of chronic fluorosis on expression of phospho-Elk-1 in rat brains]. *Chin J Endemiol* 30: 251-255.

Lou DD, Guan ZZ, Liu YJ, Liu YF, Zhang KL, Pan JG, Pei JJ. 2013. The influence of chronic fluorosis on mitochondrial dynamics morphology and distribution in cortical neurons of the rat brain. *Arch Toxicol* 87: 449-457.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S, Mentes A, DenBesten PK. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. *Physiol Behav* 206: 76-83.

Ma J, Liu F, Liu P, Dong YY, Chu Z, Hou TZ, Dang YH. 2015. Impact of early developmental fluoride exposure on the peripheral pain sensitivity in mice. *Int J Dev Neurosci* 47: 165-171.

Manusha S, Sudhakar K, Reddy KP. 2019. Protective effects of allium sativum extract against sodium fluoride induced neurotoxicity. *Int J Pharm Sci Res* 10(2): 625-633.

McPherson CA, Zhang G, Gilliam R, Brar SS, Wilson R, Brix A, Picut C, Harry GJ. 2018. An evaluation of neurotoxicity following fluoride exposure from gestational through adult ages in Long-Evans hooded rats. *Neurotoxicol Res*: 1-18.

Mesram N, Nagapuri K, Banala RR, Nalagoni CR, Karnati PR. 2016. Quercetin treatment against NaF induced oxidative stress related neuronal and learning changes in developing rats. *J King Saud Univ Sci* 29: 221-229.

Mullenix PJ, Denbesten PK, Schunior A, Kernan WJ. 1995. Neurotoxicity of sodium fluoride in rats. *Neurotoxicol Teratol* 17: 169-177.

Nageshwar M, Sudhakar K, Reddy NCC, Reddy KP. 2017. Neuroprotective effects of curcumin on sodium fluoride induced behavioural and enzymatic changes in brain and muscles of rat. *J Environ Biol* 38: 675-681.

Nageshwar M, Sudhakar K, Reddy KP. 2018. Quercetin ameliorates oxidative stress, neural damage of brain and behavioral impairment of rat with fluoride exposure. *Int J Pharm Sci Res* 9(8): 3247-3256.

Nian W, Wang X, Shao D, Yu Q, Ouyang W, Zhang Z, Ruan Q. 2018. Effects of subchronic exposure to fluorine on hippocampal injury in mice and its molecular mechanism. *Acta Sci Circumst* 38(11): 4512-4519.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. *Environ Pollut* 233: 889-899.

Niu R, Sun Z, Wang J, Cheng Z. 2008. Effects of fluoride and lead on locomotor behavior and expression of nissl body in brain of adult rats. *Fluoride* 41: 276-282.

Niu R, Sun Z, Cheng Z, Li Z, Wang J. 2009. Decreased learning ability and low hippocampus glutamate in offspring rats exposed to fluoride and lead. *Environ Toxicol Pharmacol* 28: 254-258.

Niu R, Liu S, Wang J, Zhang J, Sun Z. 2014. Proteomic analysis of hippocampus in offspring male mice exposed to fluoride and lead. *Biol Trace Elem Res* 162: 227-233.

Niu R, Xue X, Zhao Y, Sun Z, Yan X, Li X, Feng C, Wang J. 2015. Effects of fluoride on microtubule ultrastructure and expression of Tubalpha1a and Tubbeta2a in mouse hippocampus. *Chemosphere* 139: 422-427.

Niu R, Chen H, Manthari RK, Sun Z, Wang J, Zhang J, Wang J. 2018. Effects of fluoride on synapse morphology and myelin damage in mouse hippocampus. *Chemosphere* 194: 628-633.

Nkpaa KW, Onyeso GI. 2018. Rutin attenuates neurobehavioral deficits, oxidative stress, neuro-inflammation and apoptosis in fluoride treated rats. *Neurosci Lett* 682: 92-99.

Paul V, Ekambaram P, Jayakumar AR. 1998. Effects of sodium fluoride on locomotor behavior and a few biochemical parameters in rats. *Environ Toxicol Pharmacol* 6: 187-191.

Pereira M, Dombrowski PA, Losso EM, Chioca LR, Da Cunha C, Andreatini R. 2011. Memory impairment induced by sodium fluoride is associated with changes in brain monoamine levels. *Neurotoxicol Res* 19: 55-62.

Pulungan ZS, Sofro ZM, Partadiredja G. 2016. Sodium fluoride does not affect the working memory and number of pyramidal cells in rat medial prefrontal cortex. *Anat Sci Int* 93(1): 128-138.

Raghu J, Raghuveer VC, Rao MC, Somayaji NS, Babu PB. 2013. The ameliorative effect of ascorbic acid and Ginkgo biloba on learning and memory deficits associated with fluoride exposure. *Interdiscip Toxicol* 6: 217-221.

Raju S, Sivanesan SK, Gudemalla K. 2019. Cognitive enhancement effect of Ginkgo biloba extract on memory and learning impairments induced by fluoride neurotoxicity. *Int J Res Pharm Sci* 10(1): 129-134.

Reddy MM, Karnati PR. 2015. Protective effects of aqueous extract of fruit pulp of tamarindus indica on motor activity and metabolism of the gastrocnemius muscle of rats treated with fluoride. *Int J Toxicol Pharmacol Res* 7: 241-246.

Reddy YP, Tiwari SK, Shaik AP, Alsaeed A, Sultana A, Reddy PK. 2014. Effect of sodium fluoride on neuroimmunological parameters, oxidative stress and antioxidative defenses. *Toxicol Mech Methods* 24: 31-36.

Rumiantsev GI, Novikov SM, Mel'nikova NN, Levchenko NI, Kozeeva EE. 1988. [Experimental study of the biological effect of salts of hydrofluosilicic acid]. *Gig Sanit*: 80-82.

Sarkozi K, Horvath E, Vezer T, Papp A, Paulik E. 2015. Behavioral and general effects of subacute oral arsenic exposure in rats with and without fluoride. *Int J Environ Health Res* 25: 418-431.

Shah SD, Chinoy NJ. 2004. Adverse effects of fluoride and/or arsenic on the cerebral hemisphere of mice and recovery by some antidotes. *Fluoride* 37: 162-171.

Shalini B, Sharma JD. 2015. Beneficial effects of emblica officinalis on fluoride-induced toxicity on brain biochemical indexes and learning-memory in rats. *Toxicol Int* 22: 35-39.

Shan KR, Qi XL, Long YG, Nordberg A, Guan ZZ. 2004. Decreased nicotinic receptors in PC12 cells and rat brains influenced by fluoride toxicity: A mechanism relating to a damage at the level in post-transcription of the receptor genes. *Toxicology* 200: 169-177.

Sharma C, Suhalka P, Bhatnagar M. 2018. Curcumin and resveratrol rescue cortical-hippocampal system from chronic fluoride-induced neurodegeneration and enhance memory retrieval. *Int J Neurosci*: 1-15.

Shen X, Zhang Z, Xu X. 2004. [Effect of iodine and selenium on learning memory impairment induced by fluorosis and blood biochemical criterion of rats]. *Occupation and Health* 20(1): 6-8.

Sudhakar K, Nageshwar M, Pratap Reddy K. 2017. Seed extract of Abelmoschus moschatus medik reverses NAF-induced behavioral changes through neurodegeneration and oxidative stress in brain of rat. *Asian J Pharm Clin Res* 10: 165-171.

Sudhakar K, Nageshwar M, Reddy KP. 2018. Protective effect of okra, Abelmoschus moschatus seed extract on developing brain of rats during pre- and post-natal fluoride exposure. *Int J Pharm Sci Res* 9: 1519-1528.

Sudhakar K, Nageshwar M, Reddy KP. 2018. Abelmoschus moschatus extract reverses altered pain and neurohistology of a rat with developmental exposure of fluoride. *J Appl Pharm Sci* 8(6): 94-104.

Sudhakar K, Reddy KP. 2018. Protective effects of Abelmoschus moschatus seed extract on neurotransmitter system of developing brain of Wistar rats with gestational and post-natal exposure of sodium fluoride. *Int J Green Pharm* 12: S131-S139.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2000. [Effects of high fluoride drinking water on the cerebral function of mice]. *Chin J Epidemiol* 19: 262-263.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2008. Effects of high fluoride drinking water on the cerebral function of mice. *Fluoride* 41: 148-151.

Sun Z, Zhang Y, Xue X, Niu R, Wang J. 2018. Maternal fluoride exposure during gestation and lactation decreased learning and memory ability, and glutamate receptor mRNA expressions of mouse pups. *Hum Exp Toxicol* 37: 87-93.

Trivedi MH, Verma RJ, Chinoy NJ. 2007. Amelioration by black tea of sodium fluoride-induced changes in protein content of cerebral hemisphere, cerebellum and medulla oblongata in brain region of mice. *Acta Poloniae Pharm* 64: 221-225.

Trivedi MH, Verma RJ, Chinoy NJ. 2009. Mitigation of sodium fluoride induced toxicity in mice brain by black tea infusion. *Fluoride* 42: 29-33.

Trivedi MH, Verma RJ, Sangai NP, Chinoy NJ. 2011. Black tea extract mitigation of NaF-induced lipid peroxidation in different regions of mice brains. *Fluoride* 44: 243-254.

Trivedi MH, Verma RJ, Sangai NP, Chinoy NJ. 2012. Mitigation by black tea extract of sodium fluoride induced histopathological changes in brain of mice. *Fluoride* 45: 13-26.

Tsunoda M, Aizawa Y, Nakano K, Liu Y, Horiuchi T, Itai K, Tsunoda H. 2005. Changes in fluoride levels in the liver, kidney, and brain and in neurotransmitters of mice after subacute administration of fluoride. *Fluoride* 38: 284-292.

Varner JA, Jensen KF, Horvath W, Isaacson RL. 1998. Chronic administration of aluminum-fluoride or sodium-fluoride to rats in drinking water: Alterations in neuronal and cerebrovascular integrity. *Brain Res* 784(1-2): 284-298.

Verma RJ, Trivedi MH, Chinoy NJ. 2007. Black tea amelioration of sodium fluoride-induced alterations of DNA, RNA, and protein contents in the cerebral hemisphere, cerebellum, and medulla oblongata regions of mouse brain. *Fluoride* 40: 7-12.

Wang G, Li J, Zhu H, Zhu J. 2006. Effect of different doses of chronic exposure of fluoride on rat learning and memory behavior. *Studies of Trace Elements and Health* 23(2): 1-2.

Wang JD, Ge YM, Ning HM, Wang SL. 2004. Effects of high fluoride and low iodine on biochemical indexes of the brain and learning-memory of offspring rats. *Fluoride* 37: 201-208.

Wang J, Zhang Y, Guo Z, Li R, Xue X, Sun Z, Niu R. 2018. Effects of perinatal fluoride exposure on the expressions of miR-124 and miR-132 in hippocampus of mouse pups. *Chemosphere* 197: 117-122.

Wei N, Dong Y, Wang Y, Guan Z. 2014. [Effects of chronic fluorosis on neurobehavioral development in offspring of rats and antagonistic effect of vitamin E]. *Chin J Endemiol* 33: 125-128.

Whitford GM, Whitford JL, Hobbs SH. 2009. Appetitive-based learning in rats: Lack of effect of chronic exposure to fluoride. *Neurotoxicol Teratol* 31: 210-215.

Wu CX, Gu XL, Ge YM, Zhang JH, Wang JD. 2006. Effects of high fluoride and arsenic on brain biochemical indexes and learning-memory in rats. *Fluoride* 39: 274-279.

Wu NP, Zhao ZL, Gao WH, Li XQ. 1995. [Behavioral teratology in rats exposed to fluoride.] *Chin J Endemiol* 12(5): 271-273.

Wu NP, Zhao ZL, Gao WH, Li XQ. 2008. Behavioral teratology in rats exposed to fluoride. *Fluoride* 41: 129-133.

Xu X, Shen X, Zhang Z. 2001. Effect of fluorosis on mice learning and memory behaviors and brain SOD activity and MDA content *China Public Health* 17(1): 8-10.

Yang L, Jin P, Wang X, Zhou Q, Lin X, Xi S. 2018. Fluoride activates microglia, secretes inflammatory factors and influences synaptic neuron plasticity in the hippocampus of rats. *Neurotox* 69: 108-120.

Yu Q, Shao D, Zhang R, Ouyang W, Zhang Z. 2019. Effects of drinking water fluorosis on L-type calcium channel of hippocampal neurons in mice. *Chemosphere* 220: 169-175.

Yuan J, Li Q, Niu R, Wang J. 2019. Fluoride exposure decreased learning ability and the expressions of the insulin receptor in male mouse hippocampus and olfactory bulb. *Chemosphere* 224: 71-76.

Zhang C, Ren C, Chen H, Geng R, Fan H, Zhao H, Guo K, Geng D. 2013. The analog of Ginkgo biloba extract 761 is a protective factor of cognitive impairment induced by chronic fluorosis. *Biol Trace Elem Res* 153: 229-236.

Zhang KL, Lou DD, Guan ZZ. 2015. Activation of the AGE/RAGE system in the brains of rats and in SH-SY5Y cells exposed to high level of fluoride might connect to oxidative stress. *Neurotoxicol Teratol* 48: 49-55.

Zhang J, Zhu W, Zhang Z. 2009. [The effect of fluorine exposure of pregnant rats on the learning and memory capabilities of baby rats]. *Chinese Journal of Public Health* 25(11): 1347-1348.

Zhang J, Zhu WJ, Xu XH, Zhang ZG. 2011. Effect of fluoride on calcium ion concentration and expression of nuclear transcription factor kappa-B rho65 in rat hippocampus. *Exp Toxicol Pathol* 63: 407-411.

Zhang J, Zhang Z. 2013. Effects of chronic fluorosis on camkiiα, c-FOS, BAX, and BCL-2 channel signaling in the hippocampus of rats. *Fluoride* 46: 135-141.

Zhang Z, Shen X, Xu X. 2001. [Effects of selenium on the damage of learning-memory ability of mice induced by fluoride]. *J Hyg Res* 30: 144-146.

Zhang Z, Xu X, Shen X, Xua XH. 1999. [Effect of fluoride exposure on synaptic structure of brain areas related to learning-memory in mice]. *J Hyg Res* 28(4): 210-212.

Zhang Z, Xu X, Shen X, Xua XH. 2008. Effect of fluoride exposure on synaptic structure of brain areas related to learning-memory in mice. *Fluoride* 41: 139-143.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao XL, Wu JH. 1998. Actions of sodium fluoride on acetylcholinesterase activities in rats. *Biomed Environ Sci* 11: 1-6.

Zheng X, Sun Y, Ke L, Ouyang W, Zhang Z. 2016. Molecular mechanism of brain impairment caused by drinking-acquired fluorosis and selenium intervention. *Environ Toxicol Pharmacol* 43: 134-139.

Zhu W, Zhang J, Zhang Z. 2011. Effects of fluoride on synaptic membrane fluidity and PSD-95 expression level in rat hippocampus. *Biol Trace Elem Res* 139: 197-203.

Zhu YL, Zheng YJ, LV XM, Ma Y, Zhang J. 2012. Effects of fluoride exposure on performance in water labyrinth and monoamine neurotransmitters of rats. *Journal of Xinjiang Medical University* 3: 014.

Zhu YP, Xi SH, Li MY, Ding TT, Liu N, Cao FY, Zeng Y, Liu XJ, Tong JW, Jiang SF. 2017. Fluoride and arsenic exposure affects spatial memory and activates the ERK/CREB signaling pathway in offspring rats. *Neurotox* 59: 56-64.

C.2.2.2. Studies Not Available in HAWC

Abdelaleem MM, El-Tahawy NFG, Abozaid SMM, Abdel-Hakim SA. 2018. Possible protective effect of curcumin on the thyroid gland changes induced by sodium fluoride in albino rats: Light and electron microscopic study. *Endocr Regul* 52: 59-68.

Abd-Elhakim YM, Mohammed AT, Ali HA. 2018. Impact of subchronic exposure to triclosan and/or fluoride on estrogenic activity in immature female rats: The expression pattern of calbindin-D9k and estrogen receptor alpha genes. *J Biochem Mol Toxicol* 32(2): 22027.

Abdumajidov OR. 2004. [Sex differences in lipid peroxidation and antioxidant defense of the brain tissue in intoxication with low doses of inorganic compounds]. *Uzbekiston Tibbiet Zhurnali*: 58-60.

Adebayo OL, Shallie PD, Salau BA, Ajani EO, Adenuga GA. 2013. Comparative study on the influence of fluoride on lipid peroxidation and antioxidants levels in the different brain regions of well-fed and protein undernourished rats. *J Trace Elem Med Biol* 27: 370-374.

Adedara IA, Ojuade TJD, Olabiyi BF, Idris UF, Onibiyo EM, Ajeigbe OF, Farombi EO. 2016. Taurine ameliorates renal oxidative damage and thyroid dysfunction in rats chronically exposed to fluoride. *Biol Trace Elem Res*: 1-8.

Ahmed SK, Kalleny NK, Attia AAEM, Elkateb LA. 2015. The possible protective role of chromium chloride against sodium fluoride-induced changes in the structure of the cerebellar cortex of the adult male albino rat. *Egypt J Histol* 38: 402-414.

Al Badawi MH, Mahmoud OM, Salem NA. 2016. Therapeutic potential of omega-3 against sodium fluoride toxicity on the cerebellar cortex of adult male albino rats: Histological and immunohistochemical study. *Egypt J Histol* 39: 170-178.

Alhayani A, Elshal EB, Aal IHA, Al-Shammeri E, Kabra H. 2013. Does vitamin E protect against sodium fluoride toxicity on the cerebellar cortex of albino rats? *Middle East J Sci Res* 16: 1019-1026.

Ameeramja J, Raghunath A, Perumal E. 2018. Tamarind seed coat extract restores fluoride-induced hematological and biochemical alterations in rats. *Environ Sci Pollut Res Int* 25(26): 26157-26166.

Antonyan OA. 1980. [Lipid per oxidation in fluorosis and the protective role of dietary factors]. *Zh Eksp Klin Med* 20: 381-388.

Ardelean I, Racoveanu N, Manescu S, Lupulescu A, Diaconescu M, Ghelerter L. 1964. Experimental investigations concerning the effect of fluorine on the thyroid gland. *Rum Med Rev* 8: 17-20.

Atmaca N, Atmaca HT, Kanici A, Anteplioglu T. 2014. Protective effect of resveratrol on sodium fluoride-induced oxidative stress, hepatotoxicity and neurotoxicity in rats. *Food Chem Toxicol* 70: 191-197.

Auskaps AM, Shaw JH. 1955. Hemoglobin concentration, thyroid weight and growth rate in rats during minimum fluoride ingestion. *J Nutr* 55: 611-621.

Bagmut I, Kolisnyk I, Titkova A, Petrenko T, Filipchenko S. 2018. Content of catecholamines in blood serum of rats under fluoride intoxication. *Georgian Med News* (280-281): 125-129.

Bakalyan PH, Antonyan OA. 1981. [Effect of fluorosis on glutathione peroxidase and glutathione reductase activities and sulfhydryl groups]. *Zh Eksp Klin Med* 21: 10-14.

Basha PM, Madhusudhan N. 2010. Pre and post natal exposure of fluoride induced oxidative macromolecular alterations in developing central nervous system of rat and amelioration by antioxidants. *Neurochem Res* 35: 1017-1028.

Basha PM, Madhusudhan N. 2011. Effect of maternal exposure of fluoride on oxidative stress markers and amelioration by selected antioxidants in developing central nervous system of rats. *Biologia* 66: 187-193.

Basha PM, Rai P, Begum S. 2011. Evaluation of fluoride-induced oxidative stress in rat brain: A multigeneration study. *Biol Trace Elem Res* 142: 623-637.

Basha PM, Sujitha NS. 2012. Combined influence of intermittent exercise and temperature stress on the modulation of fluoride toxicity. *Biol Trace Elem Res* 148: 69-75.

Basha PM, Saumya SM. 2013. Suppression of mitochondrial oxidative phosphorylation and TCA enzymes in discrete brain regions of mice exposed to high fluoride: Amelioration by panax ginseng (ginseng) and lagerstroemia speciosa (banaba) extracts. *Cell Mol Neurobiol* 33: 453-464.

Basha MP, Begum S, Madhusudhan N. 2014. Antioxidants in the management of fluoride induced neural oxidative stress in developing rats. *Int J Pharm Sci Res* 5: 201-206.

Benetato G, Giuran AM, Cirmaciu R, Cirje M, Petrescu A, Vacariu A. 1970. [Effect of fluorine in drinking water on the metabolism of Ca and Mg and on neuromuscular excitability: Experimental studies and clinical observations]. *Rev Roum Physiol* 7: 335-352.

Bharti VK, Srivastava RS. 2009. Fluoride-induced oxidative stress in rat's brain and its amelioration by buffalo (Bubalus bubalis) pineal proteins and melatonin. *Biol Trace Elem Res* 130: 131-140.

Bhatnagar M, Rao P, Saxena A, Bhatnagar R, Meena P, Barbar S, Chouhan A, Vimal S. 2006. Biochemical changes in brain and other tissues of young adult female mice from fluoride in their drinking water. *Fluoride* 39: 280-284.

Bilgili A, Akdogan M, Yildiz M, Eraslan G, Cetin N. 2004. The effects of fluoride on thyroid hormones in rabbits. *Indian Vet J* 81: 986-988.

Bobek S, Kahl S, Ewy Z. 1976. Effect of long-term fluoride administration on thyroid hormones level blood in rats. *Endocrinol Exp* 10: 289-295.

Bouaziz H, Ammar E, Ghorbel H, Ketata S, Jamoussi K, Ayadi F, Guermazi F, Zeghal N. 2004. Effect of fluoride ingested by lactating mice on the thyroid function and bone maturation of their suckling pups. *Fluoride* 37: 133-142.

Bouaziz H, Soussia L, Guermazi F, Zeghal N. 2005. Fluoride-induced thyroid proliferative changes and their reversal in female mice and their pups. *Fluoride* 38: 185-192.

Bouaziz HB, Amara I, Essefi M, Croute F, Zeghal N. 2010. Fluoride-induced brain damages in suckling mice. *Pestic Biochem Physiol* 96: 24-29.

Chauhan SS, Ojha S, Mahmood A. 2013. Effects of fluoride and ethanol administration on lipid peroxidation systems in rat brain. *Indian J Exp Biol* 51: 249-255.

Chen J, Chen X, Yang K, Xia T, Xie H. 2002. [Studies on DNA damage and apoptosis in rat brain induced by fluoride]. *Chin J Prev Med* 36: 222-224.

Chirumari K, Reddy PK. 2007. Dose-dependent effects of fluoride on neurochemical milieu in the hippocampus and neocortex of rat brain. *Fluoride* 40: 101-110.

Chouhan S, Yadav A, Kushwah P, Kaul RK, Flora SJS. 2011. Silymarin and quercetin abrogates fluoride induced oxidative stress and toxic effects in rats. *Mol Cell Toxicol* 7: 25-32.

Clay AB, Suttie JW. 1987. Effect of dietary fluoride on dairy cattle: Growth of young heifers. *J Dairy Sci* 70: 1241-1251.

Czechowicz K, Osada A, Slesak B. 1974. Histochemical studies on the effect of sodium fluoride on metabolism in Purkinje's cells. *Folia Histochem Cytochem* 12: 37-44.

Demole V, Lerch P. 1956. [Normality of fixation of radioactive iodine in the thyroid of rats during experimental fluorosis]. *Helv Physiol Pharmacol Acta* 14(4): 62-63.

Dhurvey V, Patil V, Thakare M. 2017. Effect of sodium fluoride on the structure and function of the thyroid and ovary in albino rats (rattus norvegicus). *Fluoride* 50: 235-246.

Domzalska E. 1966. [Influence of sodium fluoride on hypophysis, thyroid gland, parathyroid, and adrenal gland in the white rat]. *Czas Stomatol* 19: 839-844.

El-lethey HS, Kamel MM, Shaheed IB. 2011. Perinatal exposure to sodium fluoride with emphasis on territorial aggression, sexual behaviour and fertility in male rats. *Life Sci J* 8: 686-694.

Flora SJS, Mittal M, Mishra D. 2009. Co-exposure to arsenic and fluoride on oxidative stress, glutathione linked enzymes, biogenic amines and DNA damage in mouse brain. *J Neurol Sci* 285: 198-205.

Flora SJS, Mittal M, Pachauri V, Dwivedi N. 2012. A possible mechanism for combined arsenic and fluoride induced cellular and DNA damage in mice. *Metallomics* 4: 78-90.

Gabovich RD, Verzhikovskaia NV. 1958. [Effect of fluoride compounds on absorption of radioactive iodine by the thyroid gland in humans and in experimental conditions]. *Probl Endokrinol Gormonoter* 4: 49-54.

Galamini-Ligori M, Di Blasi F. 1961. [Action of sodium fluoride on the thyroid of hypophysectomized rats]. *Boll Soc Ital Biol Sper* 37: 1503-1506.

Galletti P, Held HR, Korrodi H, Wegmann T. 1956. [Investigations on the possible damaging effect of fluorine on the thyroid gland]. *Helv Med Acta* 23: 601-605.

Ge Y, Ning H, Feng C, Wang H, Yan X, Wang S, Wang J. 2006. Apoptosis in brain cells of offspring rats exposed to high fluoride and low iodine. *Fluoride* 39: 173-178.

Ge Y, Niu R, Zhang J, Wang J. 2011. Proteomic analysis of brain proteins of rats exposed to high fluoride and low iodine. *Arch Toxicol* 85: 27-33.

Ge YM, Ning HM, Wang SL, Wang JD. 2005. DNA damage in thyroid gland cells of rats exposed to long-term intake of high fluoride and low iodine. *Fluoride* 38: 318-323.

Ge YM, Ning HM, Wang SL, Wang JD. 2005. Comet assay of DNA damage in brain cells of adult rats exposed to high fluoride and low iodine. *Fluoride* 38: 209-214.

Ge YM, Ning HM, Wang SL, Wang JD. 2005. Effects of high fluoride and low iodine on brain histopathology in offspring rats. *Fluoride* 38: 127-132.

Ge YM, Ning HM, Gu XL, Yin M, Yang XF, Qi YH, Wang JD. 2013. Effects of high fluoride and low iodine on thyroid function in offspring rats. *J Integr Agric* 12: 502-508.

Guan ZZ. 1986. [Morphology of the brain of the offspring of rats with chronic fluorosis]. *Chin J Pathol* 15: 297-299.

Guan Z, Wang Y, Xiao K. 1997. [Influence of experimental fluorosis on phospholipid content and fatty acid composition in rat brain]. *Chin Med J* 77: 592-596.

Guan Z-Z, Wang Y-N, Xiao K-Q, Dai D-Y, Chen Y-H, Liu J-L, Sindelar P, Dallner G. 1998. Influence of chronic fluorosis on membrane lipids in rat brain. *Neurotoxicol Teratol* 20: 537-542.

Guan ZZ, Shan KR, Wang YN, Dallner G. 2006. [Changes of lipids and nicotinic receptors in rat brains and pheochromocytoma with fluorosis]. *Chin J Endemiol* 25: 121-124.

Gushchin SK. 1951. [Effect of sodium fluoride on iodine metabolism in rabbit tissue organs; on the etiology of endemic goiter]. *Gig Sanit* 2: 45-48.

Hamza RZ, Al-Harbi MS. 2014. Sodium fluoride induced neurotoxicity and possible antioxidant role of selenium and curcumin in male mice. *Biosci Biotechnol Res Asia* 11: 81-87.

Hamza RZ, El-Shenawy NS, Ismail HAA. 2015. Protective effects of blackberry and quercetin on sodium fluoride-induced oxidative stress and histological changes in the hepatic, renal, testis and brain tissue of male rat. *J Basic Clin Physiol Pharmacol* 26: 237-251.

Haojun Z, Yaoling W, Ke Z, Jin L, Junling W. 2012. Effects of NaF on the expression of intracellular Ca2+ fluxes and apoptosis and the antagonism of taurine in murine neuron. *Toxicol Mech Methods* 22: 305-308.

Hara K. 1980. Studies on fluorosis especially effects of fluoride on thyroid metabolism. *J Dent Health* 30: 42-57.

Harris NO, Hayes RL. 1955. A tracer study of the effect of acute and chronic exposure to sodium fluoride on the thyroid iodine metabolism of rats. *J Dent Res* 34: 470-477.

Hassan HA, Abdel-Aziz AF. 2010. Evaluation of free radical-scavenging and anti-oxidant properties of black berry against fluoride toxicity in rats. *Food Chem Toxicol* 48: 1999-2004.

Hoogstratten B, Leone NCLG, Shupe J, Greenwood DA, Lieberman J. 1965. Effect of fluorides on hematopoietic system, liver, and thyroid gland in cattle. *J Amer Med Assoc* 192: 26-32.

Inkielewicz I, Rogowska M, Krechniak J. 2006. Lipid peroxidation and antioxidant enzyme activity in rats exposed to fluoride and ethanol. *Fluoride* 39: 53-59.

Inkielewicz I, Czarnowski W. 2008. Oxidative stress parameters in rats exposed to fluoride and aspirin. *Fluoride* 41: 76-82.

Inkielewicz-Stepniak I, Czarnowski W. 2010. Oxidative stress parameters in rats exposed to fluoride and caffeine. *Food Chem Toxicol* 48: 1607-1611.

Jiang P, Li G, Zhou X, Wang C, Qiao Y, Liao D, Shi D. 2018. Chronic fluoride exposure induces neuronal apoptosis and impairs neurogenesis and synaptic plasticity: Role of GSK-3beta/beta-catenin pathway. *Chemosphere* 214: 430-435.

Jiang SF, Xi SH, Yao SQ, Tong JW, Zhang YS, Wang Q, Su J, Li MY. 2013. [Effects of fluoride, arsenic and co-exposure on expression of Bcl-2 and Bax in hippocampus and cerebral cortex of rats]. *Chin J Endemiol* 32: 365-369.

Jiang Y, Guo X, Sun Q, Shan Z, Teng W. 2016. Effects of excess fluoride and iodide on thyroid function and morphology. *Biol Trace Elem Res* 170: 382-389.

Jin TX, Guan ZZ, Zhang H. 2011. [The effect of fluoride on α subunit of calcium/calmodulin-dependent protein kinase-II mRNA and protein expression in central nervous system]. *Chin J Endemiol* 30: 247-250.

Jonderko G, Kita K, Pietrzak J, Primus-Slowinska B, Ruranska B, Zylka-Wloszczyk M, Straszecka J. 1983. [Effect of subchronic sodium fluoride poisoning on the thyroid gland of rabbits with normal and increased supply of iodine]. *Endokrynol Pol* 34: 195-203.

Kahl S, Bobek S. 1975. [Effect of fluoride administration on radiothyroxine turnower in rats]. *Endokrynol Pol* 26: 391-396.

Kahl S, Ewy Z. 1975. Effect of single and long term sodium fluoride administration on biosynthesis of the thyroid hormone in rats. *Fluoride* 8: 191-198.

Kapoor V, Prasad T, Paliwal VK. 2001. Blood biochemical constituents in calves following subclinical levels of fluoride toxicosis. *Fluoride* 34: 126-131.

Karawya FS, Zahran NM, Azzam EZ. 2015. Is water fluoridation a hidden cause of obesity? Histological study on thyroid follicular cells of albino rats. *Egypt J Histol* 38: 547-557.

Kaur T, Bijarnia RK, Nehru B. 2009. Effect of concurrent chronic exposure of fluoride and aluminum on rat brain. *Drug Chem Toxicol* 32: 215-221.

Kelimu A, Liu KT, Lian J, Hu HH, Zheng YJ, Wang TM. 2008. [Effects of vitamin C and E on the ultrastructure in liver, kidney and brain of fluorosis rats]. *Chin J Endemiol* 27: 378-381.

Kinawy AA. 2019. Synergistic oxidative impact of aluminum chloride and sodium fluoride exposure during early stages of brain development in the rat. *Environ Sci Pollut Res Int* 26(11): 10951-10960.

Knizhnikov VA. 1959. [Effect of potable water with high fluoride concentration on thyroid function]. *Gig Sanit* 24: 20-25.

Knizhnikov VA, Tsypin AB, Shcherbova EN, Bugryshev PF. 1963. [The effect of drinking water with an increased fluorine content on the bioelectrical activity of the brain and heart under experimental conditions]. *Gig Sanit* 28: 16-19.

Kondo T, Yoshida M, Kasahara K. 1976. [Acute fluorosis in female rats: Time of inhibition and recovery of cholinesterase in serum and salivary glands]. *Jpn J Dent Health* 26: 187-192.

Kowalewska M. 1974. [Biopotentials of the organ of hearing in chronic sodium fluoride poisoning]. *J Pol Otolaryngol* 28: 417-424.

Krechniak J, Inkielewicz I. 2005. Correlations between fluoride concentrations and free radical parameters in soft tissues of rats. *Fluoride* 38: 293-296.

Leonard BE. 1972. Effect of phentolamine on the increase in brain glycolysis following the intraventricular administration of dibutyryl-3,5-cyclic adenosine monophosphate and sodium fluoride to mice. *Biochem Pharmacol* 21: 115-117.

Liu G, Zhang W, Jiang P, Li X, Liu C, Chai C. 2012. Role of nitric oxide and vascular endothelial growth factor in fluoride-induced goitrogenesis in rats. *Environ Toxicol Pharmacol* 34: 209-217.

Li H, Cai Q, Wang D. 2012. [Effect of fluoride on the expression of rat thyroid peroxidase mRNA]. *Chin J Endemiol* 31: 515-517.

Li H, Cai Q, Wang D. 2012. [Effects of fluoride on rat thyroid morphology, thyroid peroxidase activity and the expression of thyroid peroxidase protein]. *Chin J Endemiol* 31: 271-274.

Liu H, Hou C, Zeng Q, Zhao L, Cui Y, Yu L, Wang L, Zhao Y, Nie J, Zhang B, Wang A. 2016. Role of endoplasmic reticulum stress-induced apoptosis in rat thyroid toxicity caused by excess fluoride and/or iodide. *Environ Toxicol Pharmacol* 46: 277-285.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. [Changes of the c-Jun N-terminal kinase in the brains of rats with chronic fluorosis]. *Chin J Endemiol* 29: 608-612.

Liu YJ, Guan ZZ, Gao Q, Pei JJ. 2011. Increased level of apoptosis in rat brains and SH-SY5Y cells exposed to excessive fluoride-A mechanism connected with activating JNK phosphorylation. *Toxicol Lett* 204: 183-189.

Lohakare J, Pattanaik AK. 2013. Effects of addition of fluorine in diets differing in protein content on urinary fluoride excretion, clinical chemistry and thyroid hormones in calves. *Brazilian J Anim Sci* 42: 751-758.

Long YG, Wang YN, Chen J, Jiang SF, Nordberg A, Guan ZZ. 2002. Chronic fluoride toxicity decreases the number of nicotinic acetylcholine receptors in rat brain. *Neurotoxicol Teratol* 24: 751-757.

Lou DD, Liu YF, Zhang KL, Yu YN, Guan ZZ. 2011. [Changes of reactive oxygen species level and mitochondria fission-fusion in cortical neurons of rats with chronic fluorosis]. *Chin J Endemiol* 30: 256-260.

Lou DD, Liu YF, Qin SL, Zhang KL, Yu YN, Guan ZZ. 2012. [Changed transcription level of mitochondrial fission and fusion gene loci in cortical neurons of rats with chronic fluorosis]. *Chin J Endemiol* 31: 125-129.

Lou DD, Zhang KL, Qin SL, Liu YF, Yu YN, Guan ZZ. 2012. [Alteration of mitochondrial distribution and gene expression of fission 1 protein in cortical neurons of rats with chronic fluorosis]. *Chin J Pathol* 41: 243-247.

Lou DD, Pan JG, Zhang KL, Qin SL, Liu YF, Yu YN, Guan ZZ. 2013. [Changed expression of mito-fusion 1 and mitochondrial fragmentation in the cortical neurons of rats with chronic fluorosis]. *Chin J Prev Med* 47: 170-174.

Lou DD, Zhang KL, Pan JG, Qin SL, Liu YF, Yu YN, Guan ZZ. 2013. [Influence of chronic fluorosis on the expression of mitochondrial fission protein dynamin-related 1 in the cortical neurons of rats]. *Chin J Prev Med* 47: 561-564.

Lou DD, Zhang KL, Qin SL, Liu YF, Liu YJ, Guan ZZ. 2013. [Effects of chronic fluorosis on 4.8 kb mitochondrial DNA in liver, kidney and brain of rats]. *Chin J Endemiol* 32: 121-124.

Lou DD, Guan ZZ, Pei JJ. 2014. Alterations of apoptosis and expressions of Bax and Bcl-2 in the cerebral cortices of rats with chronic fluorosis. *Fluoride* 47: 199-207.

Luo GY, Niu RY, Sun ZL, Zhang JH, Wang JM, Wang C, Wang JD. 2011. Reduction of CaMKII expression in the hippocampus of rats from ingestion of fluoride and/or lead. *Fluoride* 44: 63-69.

Ma T, Liu D, Song K. 1999. Cytochemical study of neuron enzyme at anterior horn of spinal cord in rats with experimental fluorosis. *J Chin Med Univ* 28: 81-82.

Ma TX, Yu HT, Song KQ. 2008. [Expression of c-fos and Caspase 8 in cerebral cortex of rats with experimental fluorosis]. *Chin J Endemiol* 27: 131-133.

Mach Z, Zygulska-Machowa H. 1959. O wplywie fluoru na przemiane J131 [Russian and English summ.]. *Endokrynol Pol* 10: 157-162.

Machida H. 1989. [A study on the rabbit thermoregulatory system effects of high dose sodium fluoride]. *Dent Sci Rep* 89: 607-626.

Madan J, Puri JP, Singh JK. 2009. Growth, feed efficiency and blood profile of buffalo calves consuming high levels of fluoride. *Trop Anim Health Prod* 41: 295-298.

Madhusudhan N, Basha PM, Begum S, Ahmed F. 2009. Fluoride-induced neuronal oxidative stress and its amelioration by antioxidants in developing rats. *Fluoride* 42: 179-187.

Madhusudhan N, Basha PM, Rai P, Ahmed F, Prasad GR. 2010. Effect of maternal fluoride exposure on developing CNS of rats: Protective role of Aloe vera, Curcuma longa and Ocimum sanctum. *Indian J Exp Biol* 48: 830-836.

Manocha SL, Warner H, Olkowski ZL. 1975. Cytochemical response of kidney, liver and nervous system of fluoride ions in drinking water. *Histochem J* 7: 343-355.

Mansour HH, Tawfik SS. 2012. Efficacy of lycopene against fluoride toxicity in rats. *Pharm Biol* 50: 707-711.

Mietkiewski K, Walczak M, Trojanowicz R. 1966. [Effect of sodium fluoride on the neurosecretory system in guinea pigs]. *Endokrynol Pol* 17: 121-131.

Mohamed NE. 2016. The role of calcium in ameliorating the oxidative stress of fluoride in rats. *Biol Trace Elem Res* 170: 128-144.

Muhlemann HR, Schneider R. 1956. [Mitotic activity of rat thyroid epithelium after administration of fluoridated drinking water]. *Schweiz Med Wochenschr* 86: 625-627.

Nabavi SF, Eslami S, Moghaddam AH, Nabavi SM. 2011. Protective effects of curcumin against fluoride-induced oxidative stress in the rat brain. *Neurophysiology* 43: 287-291.

Nabavi SF, Moghaddam AH, Nabavi SM, Eslami S. 2011. Protective effect of curcumin and quercetin on thyroid function in sodium fluoride intoxicated rats. *Fluoride* 44: 147-152.

Nabavi SF, Habtemariam S, Jafari M, Sureda A, Nabavi SM. 2012. Protective role of gallic acid on sodium fluoride induced oxidative stress in rat brain. *Bull Environ Contam Toxicol* 89: 73-77.

Nabavi SF, Nabavi SM, Latifi AM, Mirzaei M, Habtemariam S, Moghaddam AH. 2012. Mitigating role of quercetin against sodium fluoride-induced oxidative stress in the rat brain. *Pharm Biol* 50: 1380-1383.

Nabavi SF, Nabavi SM, Habtemariam S, Moghaddam AH, Sureda A, Mirzaei M. 2013. Neuroprotective effects of methyl-3-O-methyl gallate against sodium fluoride-induced oxidative stress in the brain of rats. *Cell Mol Neurobiol* 33: 261-267.

Nabavi SM, Sureda A, Nabavi SF, Latifi AM, Moghaddam AH, Hellio C. 2012. Neuroprotective effects of silymarin on sodium fluoride-induced oxidative stress. *J Fluor Chem* 142: 79-82.

Narayanaswamy M, Piler MB. 2010. Effect of maternal exposure of fluoride on biometals and oxidative stress parameters in developing CNS of rat. *Biol Trace Elem Res* 133: 71-82.

Narbutt B, Romer TE, Grabski J, Szymik N. 1971. [Influence of natrium fluoride on the structure of the rat thyroid]. *Endokrynol Pol* 22: 445-451.

Narbutt B, Romer TE, Grabski J, Szymik N. 1971. [The influence of sodium fluoride on the morphology of the thyroid gland in rats]. *Endokrynol Pol* 22: 361-365.

Niu RY, Sun ZL, Cheng ZT, Liu HT, Chen HC, Wang JD. 2008. Effects of fluoride and lead on N-methyl-D-aspartate receptor 1 expression in the hippocampus of offspring rat pups. *Fluoride* 41: 101-110.

Niu R, Wang J, Sun Z, Xue X, Yan X, Zhang J. 2015. Transcriptional regulatory dynamics of the hypothalamic-pituitary-testicular axis in male mice exposed to fluoride. *Environ Toxicol Pharmacol* 40: 557-562.

Niu R, Zhang Y, Liu S, Liu F, Sun Z, Wang J. 2015. Proteome alterations in cortex of mice exposed to fluoride and lead. *Biol Trace Elem Res* 164: 99-105.

Ogilvie AL. 1952. Histological findings in the kidney, liver, pancreas, adrenal and thyroid gland of the rat following sodium fluoride administration. *J Dent Res* 31: 598-598.

Okayasu I, Tsuchida M, Yanagisawa F. 1985. Hyperplastic nodules of thyroid parafollicular cells (C cells) in rats induced by prolonged low dose ingestion of NaF. *Fluoride* 18: 111-117.

Pal S, Sarkar C. 2014. Protective effect of resveratrol on fluoride induced alteration in protein and nucleic acid metabolism, DNA damage and biogenic amines in rat brain. *Environ Toxicol Pharmacol* 38: 684-699.

Pan Y, Lu P, Yin L, Chen K, He Y. 2015. Effect of fluoride on the proteomic profile of the hippocampus in rats. *Z Naturforsch C* 70: 151-157.

Phillips PH, Lamb AR. 1934. Histology of certain organs and teeth in chronic toxicosis due to fluorin. *Arch Path* 17: 169-176.

Portela ML. 1972. [Biochemical effects in the prolonged ingestion of fluorides in rats]. *Arch Latinoam Nutr* 22: 291-308.

Prestes DS, Filappi A, Schossler DR, Duarte FA, Dressler VL, Flores EMM, Cecim M. 2009. Functional and histological evaluations of thyroid of sheep submitted to sodium fluoride administration. *Arq Bras Med Vet Zootec* 61: 293-298.

Puentes F, Cremer HD. 1966. Experiments on fluoride-iodine antagonism in the thyroid gland. *Adv Fluorine Res* 4: 213-220.

Qian W, Miao K, Li T, Zhang Z. 2013. Effect of selenium on fluoride-induced changes in synaptic plasticity in rat hippocampus. *Biol Trace Elem Res* 155: 253-260.

Qing-Feng S, Ying-Peng X, Tian-Tong X. 2019. Matrix metalloproteinase-9 and p53 involved in chronic fluorosis induced blood-brain barrier damage and neurocyte changes. *Arch Med Sci* 15(2): 457-466.

Qiu YH, Kong DM, Yang Q, Zhao N. 2010. [Influence of high-fluoride on thyroid function and brain damage in rats]. *Chin J Endemiol* 29: 146-149.

Raghavendra M, Ravindra RK, Raghuveer YP, Narasimha JK, Uma MRV, Navakishor P. 2016. Alleviatory effects of hydroalcoholic extract of cauliflower (brassica oleracea var. botrytis) on thyroid function in fluoride intoxicated rats. *Fluoride* 49: 84-90.

Rakhov GM. 1964. [Effect of calcium and fluorine in drinking water on the iodine metabolism and status of the thyroid gland in iodine insufficiency in food]. *Gig Sanit* 29: 12-17.

Ranpariya VL, Parmar SK, Sheth NR, Chandrashekhar VM. 2011. Neuroprotective activity of matricaria recutita against fluoride-induced stress in rats. *Pharm Biol* 49: 696-701.

Reddy KP, Sailaja G, Krishnaiah C. 2009. Protective effects of selenium on fluoride induced alterations in certain enzymes in brain of mice. *J Environ Biol* 30: 859-864.

Rogalska A, Kuter K, Zelazko A, Glogowska-Gruszka A, Swietochowska E, Nowak P. 2017. Fluoride alteration of [3H]glucose uptake in Wistar rat brain and peripheral tissues. *Neurotoxicol Res* 31: 436-443.

Saka O, Hallac P, Urgancioğlu I. 1965. The effect of fluoride on the thyroid of the rat. *New Istanbul Contrib Clin Sci* 8: 87-90.

Samanta A, Chanda S, Bandyopadhyay B, Das N. 2016. Establishment of drug delivery system nanocapsulated with an antioxidant (+)-catechin hydrate and sodium meta borate chelator against sodium fluoride induced oxidative stress in rats. *J Trace Elem Med Biol* 33: 54-67.

Sarkar C, Das N, Pal S, Dinda B. 2014. Oxidative stress induced alteration of protein and nucleic acid metabolism in fluoride-intoxicated rat brain: Protection by 3α-hydroxy olean-12-en-27-oic acid isolated from neanotis wightiana. *Int J Pharm Sci Res* 5: 3047-3066.

Sarkar C, Pal S. 2014. Ameliorative effect of resveratrol against fluoride-induced alteration of thyroid function in male Wistar rats. *Biol Trace Elem Res* 162: 278-287.

Sarkar C, Pal S, Das N, Dinda B. 2014. Ameliorative effects of oleanolic acid on fluoride induced metabolic and oxidative dysfunctions in rat brain: Experimental and biochemical studies. *Food Chem Toxicol* 66: 224-236.

Seffner W, Teubener W, Runde H, Wiedner H, Vogt J, Otto G, Zschau E, Geinitz D, Franke J. 1990. Boron as an antidote to fluorosis? II. Studies on various organs of pigs. *Fluoride* 23: 68-79.

Selim AOA, El-Haleem MR, Ibrahim IH. 2012. Effect of sodium fluoride on the thyroid gland of growing male albino rats: Histological and biochemical study. *Egypt J Histol* 35: 470-482.

Shao Q, Wang Yn, Guan Z. 2000. [Influence of free radical inducer on the level of oxidative stress in brain of rats with fluorosis]. *Chin J Prev Med* 34: 330-332.

Sharma C, Suhalka P, Sukhwal P, Jaiswal N, Bhatnagar M. 2014. Curcumin attenuates neurotoxicity induced by fluoride: An in vivo evidence. *Pharmacogn Mag* 10: 61-65.

Shashi A. 1992. Studies on alterations in brain lipid metabolism following experimental fluorosis. *Fluoride* 25: 77-84.

Shashi A. 1993. Nucleic acid levels in thyroid gland in acute and chronic fluoride intoxication. *Fluoride* 26: 191-196.

Shashi A, Singh JP, Thapar SP. 1994. Effect of long-term administration of fluoride on levels of protein, free amino acids and RNA in rabbit brain. *Fluoride* 27: 155-159.

Shashi A. 2003. Histopathological investigation of fluoride-induced neurotoxicity in rabbits. *Fluoride* 36: 95-105.

Shashi A, Neetika S, Bhardwaj M. 2009. Neuronal DNA damage and apoptosis in brain of rat exposed to fluoride. *Asian J Microbiol Biotechnol Environ Sci* 11: 629-632.

Shayiq RM, Raza H, Kidwai AM. 1986. Fluoride and lipid peroxidation a comparative study in different rat tissues. *Bull Environ Contam Toxicol* 37: 70-76.

Shen QF, Li HN, Xu TT, Xia YP. 2012. [Damage of blood brain barrier of spinal cord in rats with chronic fluorosis]. *Chin Med J* 92: 2357-2361.

Shen Q, Tian R, Li H, Xu T, Xia Y. 2014. [White matter injury of spinal cord in rats with chronic fluorosis and recovery after defluoridation]. *Chin Med J* 94: 1189-1192.

Shen X, Zhang Z, Xu X. 2004. [Influence of combined iodine and fluoride on phospholipid and fatty acid composition in brain cells of rats]. *J Hyg Res* 33: 158-161.

Shivarajashankara YM, Shivashankara AR, Bhat PG, Rao SH. 2001. Effect of fluoride intoxication on lipid peroxidation and antioxidant systems in rats. *Fluoride* 34: 108-113.

Shivarajashankara YM, Shivashankara AR, Bhat PG, Rao SH. 2002. Brain lipid peroxidation and antioxidant systems of young rats in chronic fluoride intoxication. *Fluoride* 35: 197-203.

Shivarajashankara YM, Shivashankara AR, Bhat PG, Rao SM, Rao SH. 2002. Histological changes in the brain of young fluoride-intoxicated rats. *Fluoride* 35: 12-21.

Siebenhuner L, Miloni E, Burgi H. 1984. [Effects of fluoride on thyroid hormone biosynthesis: Studies in a highly sensitive test system]. *Klin Wochenschr* 62: 859-861.

Singh R, Srivastava AK, Gangwar NK. 2017. Clinico-pathological studies on the co-exposure of cypermethrin and fluoride in experimental rats with ameliorative action of Vitamin E. *Vet Pract* 18(2): 207-210.

Soni KK, Shrivastava VK. 2007. Sodium fluoride induced histopathological changes in thyroid gland of male mus musculus. *Biochem Cell Arch* 7: 317-320.

Stee EW. 1968. *Effect of sodium fluoride and AMOX (NF3O) on growth and thyroid function in the rat*. No. AMRL-TR-67-189. Wright-Patterson Air Force Base, OH: pp. 67.

Štolc V, Podoba J. 1960. Effect of fluoride on the biogenesis of thyroid hormones. *Nature* 188: 855-856.

Sugivama Y. 1967. [The effect of sodium fluoride administration on the parathyroid glands]. *Hirosaki Med J* 19: 520-529.

Sun Y, Ke L, Zheng X, Li T, Ouyang W, Zhang Z. 2016. Effects of different levels of calcium intake on brain cell apoptosis in fluorosis rat offspring and its molecular mechanism. *Biol Trace Elem Res*: 1-12.

Takata H. 1958. The effect of fluorine upon the uptake of I131 by the thyroid glands. *Folia Pharmacol Jpn* 54: 230-236.

Teng Y, Zhang J, Zhang Z, Feng J. 2017. The effect of chronic fluorosis on calcium ions and CaMKIIα, and c-fos expression in the rat hippocampus. *Biol Trace Elem Res*: 295-302.

Trabelsi M, Guermazi F, Zeghal N. 2001. Effect of fluoride on thyroid function and cerebellar development in mice. *Fluoride* 34: 165-173.

Tsuchida M, Okayasu I, Kohyama Y, Kurihara H, Tanaka H, Yanagisawa F, Date C, Hayashi M, Mui K, Asada M. 1986. Effects of long term, low dose ingestion of fluoride on the thyroid gland in rats. *Stud Environ Sci* 27: 307-312.

Vani ML, Reddy KP. 2000. Effects of fluoride accumulation on some enzymes of brain and gastrocnemius muscle of mice. *Fluoride* 33: 17-26.

Wang C, Liang C, Ma J, Manthari RK, Niu R, Wang J, Wang J, Zhang J. 2018. Co-exposure to fluoride and sulfur dioxide on histological alteration and DNA damage in rat brain. *J Biochem Mol Toxicol* 32.

Wang H, Yang Z, Zhou B, Gao H, Yan X, Wang J. 2009. Fluoride-induced thyroid dysfunction in rats: Roles of dietary protein and calcium level. *Toxicol Ind Health* 25: 49-57.

Wang J, Niu R, Sun Z, Lv L, Smith GW. 2008. Effects of protein and calcium supplementation on bone metabolism and thyroid function in protein and calcium deficient rabbits exposed to fluoride. *Fluoride* 41: 283-291.

Wang JD, Ge YM, Ning HM, Wang SL. 2004. Effects of high fluoride and low iodine on oxidative stress and antioxidant defense of the brain in offspring rats. *Fluoride* 37: 264-270.

Wang JL. 2007. [Effect of fluoride on the intracellular Ca2+ in neurons of mice]. *Chin J Endemiol* 26: 505-507.

Wang Y, Guan Z, Xiao K. 1997. [Changes of coenzyme Q content in brain tissues of rats with fluorosis]. *Chin J Prev Med* 31: 330-333.

Wang Y, Dong Y, Wei N, Guan Z. 2015. [Influence of chronic fluorosis on expression of quinone oxidoreductase-1 and heme oxygenase-1 in rat brains]. *Chin J Endemiol* 34: 250-253.

Wedzisz A, Cieciura J. 1988. Effect of small sodium fluoride feed supplements on the serum thyroid hormone content of rats. *Bromatol Chem Toksykol* 21: 174-175.

Wei N, Dong YT, Deng J, Wang Y, Qi XL, Yu WF, Xiao Y, Zhou JJ, Guan ZZ. 2018. Changed expressions of N-methyl-d-aspartate receptors in the brains of rats and primary neurons exposed to high level of fluoride. *J Trace Elem Med Biol* 45: 31-40.

Yan N, Liu Y, Liu S, Cao S, Wang F, Wang Z, Xi S. 2016. Fluoride-induced neuron apoptosis and expressions of inflammatory factors by activating microglia in rat brain. *Mol Neurobiol* 53: 4449-4460.

Yang H, Xing R, Liu S, Yu H, Li P. 2016. Gamma-Aminobutyric acid ameliorates fluoride-induced hypothyroidism in male Kunming mice. *Life Sci* 146: 1-7.

Yang H, Xing R, Liu S, Yu H, Li P. 2019. Analysis of the protective effects of gamma-aminobutyric acid during fluoride-induced hypothyroidism in male Kunming mice. *Pharm Biol* 57(1): 29-37.

Yang M, Ren Z, Zhou B, Guan Z, Yu W. 2017. [Expression of endonuclease G in the brain tissue of rats with chronic fluorosis]. *Chin J Endemiol* 36: 327-332.

Yuan SD, Xie QW, Lu FY. 1993. Changes of serotonin content and turnover rate in hypothalamus of female rat during fluorosis. *Fluoride* 26: 57-60.

Zhai JX, Guo ZY, Hu CL, Wang QN, Zhu QX. 2003. [Studies on fluoride concentration and cholinesterase activity in rat hippocampus]. *Chin J Ind Hyg Occup Dis* 21: 102-104.

Zhan CW, Huo DJ. 1988. Ultrastructural findings in liver, kidneys, thyroid-gland and cardiac-muscle of rabbits following sodium-fluoride administration. *Fluoride* 21: 32-38.

Zhan XA, Xu ZR, Li JX, Wang M. 2005. Effects of fluorosis on lipid peroxidation and antioxidant systems in young pigs. *Fluoride* 38: 157-161.

Zhan XA, Li JX, Wang M, Xu ZR. 2006. Effects of fluoride on growth and thyroid function in young pigs. *Fluoride* 39: 95-100.

Zhang KL, Lou DD, Liu YF, Qin SL, Guan ZZ. 2012. [Changes of P-glycoprotein and nuclear factor κB in the cerebral cortex of rat with chronic fluorosis]. *Chin J Endemiol* 31: 613-616.

Zhang KL, Lou DD, Guan ZZ. 2013. [Expression of receptor for advanced glycation endproducts and nuclear factor κB in brain hippocampus of rat with chronic fluorosis]. *Chin J Endemiol* 32: 625-628.

Zhang WD, Zhang Y, Liu GY, Jiang P, Chai CY. 2008. [Effects of fluoride on ultrastructure of thyroids in rats]. *Chin J Endemiol* 27: 622-624.

Zhang ZG, Wang XY, Nian WW, Liao QX, Zhang R, Ouyang W. 2017. Effects of calcium on drinking fluorosis-induced hippocampal syntaptic plasticity impairment in the offspring of rats. *Transl Neurosci* 8: 191-200.

Zhao W, Zhu H, Yu Z, Aoki K, Misumi J, Zhang X. 1998. Long-term effects of various iodine and fluorine doses on the thyroid and fluorosis in mice. *Endocr Regul* 32: 63-70.

Zhao WY. 1988. [A preliminary study of the interaction of iodine and fluoride in experimental iodine goiter and fluorosis]. *Chin J Prev Med* 22: 146-148.

Zhao XL, Gao WH, Zhao ZL. 1994. [Effects of sodium fluoride on the activity of Ca2+Mg(2+)-ATPase in synaptic membrane in rat brain]. *Chin J Prev Med* 28: 264-266.

Zhavoronkov AA, Polyakova GA. 1973. Morphological and functional state of the hypothalamo-hypophyseal neurosecretory system in experimental fluorosis. *Bull Exp Biol Med* 75: 194-196.

Zhou B, Luo G, Wang C, Niu R, Wang J. 2014. Effects of fluoride on expression of cytokines in the hippocampus of adult rats. *Fluoride* 47: 191-198.

C.2.3. In Vitro Experimental Studies

As described in Figure 2, 60 in vitro experimental studies were included; however, data extraction was not conducted on in vitro studies. Therefore, in vitro experimental studies are not available in HAWC with the exception of in vitro studies that also reported in vivo non-human animal data that met the relevant criteria for being made available in HAWC. The following lists of references are organized as studies that are available in HAWC (n = 6) followed by studies that are not available in HAWC (n = 54).

C.2.3.1. Studies Available in HAWC

Chen J, Niu Q, Xia T, Zhou G, Li P, Zhao Q, Xu C, Dong L, Zhang S, Wang A. 2018. ERK1/2-mediated disruption of BDNF-TrkB signaling causes synaptic impairment contributing to fluoride-induced developmental neurotoxicity. *Toxicology* 410: 222-230.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. *Environ Pollut* 233: 889-899.

Shan KR, Qi XL, Long YG, Nordberg A, Guan ZZ. 2004. Decreased nicotinic receptors in PC12 cells and rat brains influenced by fluoride toxicity: A mechanism relating to a damage at the level in post-transcription of the receptor genes. *Toxicology* 200: 169-177.

Zhang KL, Lou DD, Guan ZZ. 2015. Activation of the AGE/RAGE system in the brains of rats and in SH-SY5Y cells exposed to high level of fluoride might connect to oxidative stress. *Neurotoxicol Teratol* 48: 49-55.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao XL, Wu JH. 1998. Actions of sodium fluoride on acetylcholinesterase activities in rats. *Biomed Environ Sci* 11: 1-6.

C.2.3.2. Studies Not Available in HAWC

Ardelean I, Racoveanu N, Manescu S, Lupulescu A, Diaconescu M, Ghelerter L. 1964. Experimental investigations concerning the effect of fluorine on the thyroid gland. *Rum Med Rev* 8: 17-20.

Chen J, Chen X, Yang K. 2000. [Effects of selenium and zinc on the DNA damage caused by fluoride in pallium neural cells of rats]. *J Hyg Res* 29: 216-217.

Chen L, Ning H, Yin Z, Song X, Feng Y, Qin H, Li Y, Wang J, Ge Y, Wang W. 2017. The effects of fluoride on neuronal function occurs via cytoskeleton damage and decreased signal transmission. *Chemosphere* 185: 589-594.

Chen R, Zhao LD, Liu H, Li HH, Ren C, Zhang P, Guo KT, Zhang HX, Geng DQ, Zhang CY. 2017. Fluoride induces neuroinflammation and alters Wnt signaling pathway in BV2 microglial cells. *Inflammation* 40: 1123-1130.

Cheng TJ, Chen TM, Chen CH, Lai YK. 1998. Induction of stress response and differential expression of 70 kDa stress proteins by sodium fluoride in HeLa and rat brain tumor 9L cells. *J Cell Biochem* 69: 221-231.

Deng MF, Zhu D, Liu YP, He WW, Gui CZ, Guan ZZ. 2018. Attenuation by 7-nitroindazole of fluoride-induced toxicity in SH-SY5Y cells exposed to high fluoride: Effects on nitric oxide, nitric oxide synthetase activity, nNOS, and apoptosis. *Fluoride* 51(4): 328-339.

Flores-Mendez M, Ramirez D, Alamillo N, Hernandez-Kelly LC, Del Razo LM, Ortega A. 2014. Fluoride exposure regulates the elongation phase of protein synthesis in cultured Bergmann glia cells. *Toxicol Lett* 229: 126-133.

Gao Q, Liu YH, Guan ZZ. 2008. Oxidative stress might be a mechanism connected with the decreased alpha 7 nicotinic receptor influenced by high-concentration of fluoride in SH-SY5Y neuroblastoma cells. *Toxicol In Vitro* 22: 837-843.

Goschorska M, Gutowska I, Baranowska-Bosiacka I, Piotrowska K, Metryka E, Safranow K, Chlubek D. 2018. Influence of acetylcholinesterase inhibitors used in Alzheimer's Disease treatment on the activity of antioxidant enzymes and the concentration of glutathione in THP-1 macrophages under fluoride-induced oxidative stress. *Int J Environ Res Pub Health* 16(1).

Guan ZZ, Shan KR, Xiu J, Long YG. 2005. [Fluorosis on expression of nicotinic acetylcholine receptors in protein and gene levels in human SH-SY5Y neuroblastoma cells]. *Chin J Prev Med* 39: 26-29.

Guan ZZ, Shan KR, Wang YN, Dallner G. 2006. [Changes of lipids and nicotinic receptors in rat brains and pheochromocytoma with fluorosis]. *Chin J Endemiol* 25: 121-124.

Haojun Z, Yaoling W, Ke Z, Jin L, Junling W. 2012. Effects of NaF on the expression of intracellular Ca2+ fluxes and apoptosis and the antagonism of taurine in murine neuron. *Toxicol Mech Methods* 22: 305-308.

Hong-Liang L, Qiang Z, Yu-Shan C, Lei Z, Gang F, Chang-Chun H, Liang Z, Aiguo W. 2014. Fluoride-induced thyroid cell apoptosis. *Fluoride* 47: 161-169.

Inkielewicz-Stepniak I, Radomski MW, Wozniak M. 2012. Fisetin prevents fluoride- and dexamethasone-induced oxidative damage in osteoblast and hippocampal cells. *Food Chem Toxicol* 50: 583-589.

Jin TX, Guan ZZ, Zhang H. 2011. [The effect of fluoride on α subunit of calcium/calmodulin-dependent protein kinase-II mRNA and protein expression in central nervous system]. *Chin J Endemiol* 30: 247-250.

Kariya T, Kotani M, Field JB. 1974. Effects of sodium fluoride and other metabolic inhibitors on basal and TSH stimulated cyclic AMP and thyroid metabolism. *Metab Clin Exper* 23: 967-973.

Ke L, Zheng X, Sun Y, Ouyang W, Zhang Z. 2016. Effects of sodium fluoride on lipid peroxidation and PARP, XBP-1 expression in PC12 cell. *Biol Trace Elem Res* 173: 161-167.

Lee J, Han YE, Favorov O, Tommerdahl M, Whitsel B, Lee CJ. 2016. Fluoride induces a volume reduction in CA1 hippocampal slices via MAP kinase pathway through volume regulated anion channels. *Exp Neurobiol* 25: 72-78.

Levesque L, Mizzen CA, McLachlan DR, Fraser PE. 2000. Ligand specific effects on aluminum incorporation and toxicity in neurons and astrocytes. *Brain Res* 877: 191-202.

Li H, Gao MT, Xu KY, Wang CY. 2007. Effect of sodium fluoride on the primary porcine thyroid cells and thyroid peroxidase activity. *J Clin Rehabil Tissue Eng Res* 11: 7425-7428.

Li H, Gao MT, Xu KY, Cui MY, Dai X. 2008. [Effect of fluoride on thyroid functioning in primary porcine thyrocyte]. *Chin J Endemiol* 27: 38-40.

Li H, Huang H, Xu Y, Gao Y, Liu Z. 2010. [Toxic effects of fluoride on rat cerebral cortex astrocytes in vitro]. *J Hyg Res* 39: 86-88.

Liu H, Zeng Q, Cui Y, Yu L, Zhao L, Hou C, Zhang S, Zhang L, Fu G, Liu Y, Jiang C, Chen X, Wang A. 2014. The effects and underlying mechanism of excessive iodide on excessive fluoride-induced thyroid cytotoxicity. *Environ Toxicol Pharmacol* 38: 332-340.

Liu HL, Zeng Q, Cui YS, Zhao L, Zhang L, Fu G, Hou CC, Zhang S, Yu LY, Jiang CY, Wang ZL, Chen XM, Wang AG. 2014. The role of the IRE1 pathway in excessive iodide- and/or fluoride-induced apoptosis in Nthy-ori 3-1 cells in vitro. *Toxicol Lett* 224: 341-348.

Liu YJ, Guan ZZ, Gao Q, Pei JJ. 2011. Increased level of apoptosis in rat brains and SH-SY5Y cells exposed to excessive fluoride-A mechanism connected with activating JNK phosphorylation. *Toxicol Lett* 204: 183-189.

Liu Y, Gao Q, Tang Z, Zhang X, Guan Z. 2015. [The expression and correlation between neural nicotinic acetylcholine receptor subunit α3 and mitogen-activated protein kinase cell signaling transduction pathway in human neuroblastoma cell line SH-SY5Y overexposed to fluoride]. *Chin J Endemiol* 34: 553-558.

Madaoui S, Rappaport L, Nunez J. 1974. Prostaglandins and in vitro TSH-dependent iodide binding by rat thyroid glands. *Biochimie* 56: 109-113.

Nakagawa-Yagi Y, Saito Y, Kitoh N, Ogane N, Fujisawa E, Nakamura H. 1993. Fluoride causes suppression of neurite outgrowth in human neuroblastoma via an influx of extracellular calcium. *Biochem Biophys Res Commun* 191: 727-736.

Ong J, Kerr DIB. 1995. Interactions of N-ethylmaleimide and aluminium fluoride with GABA(B) receptor function in rat neocortical slices. *Eur J Pharmacol* 287: 197-200.

Pastan I, Macchia V, Katzen R. 1968. Effect of fluoride on the metabolic activity of thyroid slices. *Endocrinology* 83: 157-160.

Rubakhova VM. 1977. [Effect of serotonin and sodium fluoride on visceral nerve conductors]. *Vyestsi Akademii Navuk BSSR Syeryya Biyalahichnykh Navuk* 1: 117-119.

Shayiq RM, Raza H, Kidwai AM. 1986. Fluoride and lipid peroxidation a comparative study in different rat tissues. *Bull Environ Contam Toxicol* 37: 70-76.

Shuhua X, Ziyou L, Ling Y, Fei W, Sun G. 2012. A role of fluoride on free radical generation and oxidative stress in BV-2 microglia cells. *Mediators Inflamm* 2012: 1-8.

Singh P, Das TK. 2019. Ultrastructural localization of 4-hydroxynonenal adducts in fluoride-exposed cells: Protective role of dietary antioxidants. *Fluoride* 52(1): 49-58.

Taylor P. 1972. Comparison of the effects of various agents on thyroidal adenyl cyclase activity with their effects on thyroid hormone release. *J Endocrinol* 54: 137-145.

Tu W, Zhang Q, Liu Y, Han LY, Wang Q, Chen PP, Zhang S, Wang AG, Zhou X. 2018. Fluoride induces apoptosis via inhibiting SIRT1 activity to activate mitochondrial p53 pathway in human neuroblastoma SH-SY5Y cells. *Toxicol Appl Pharmacol* 347: 60-69.

van der Voet GB, Schijns O, de Wolff FA. 1999. Fluoride enhances the effect of aluminium chloride on interconnections between aggregates of hippocampal neurons. *Arch Physiol Biochem* 107: 15-21.

Wang JL. 2007. [Effect of fluoride on the intracellular Ca2+ in neurons of mice]. *Chin J Endemiol* 26: 505-507.

Wang J, Gao Y, Cheng X, Yang J, Zhao Y, Xu H, Zhu Y, Yan Z, Manthari RK, Mehdi OM, Wang J. 2019. GSTO1 acts as a mediator in sodium fluoride-induced alterations of learning and memory related factors expressions in the hippocampus cell line. *Chemosphere* 226: 201-209.

Wei N, Dong YT, Deng J, Wang Y, Qi XL, Yu WF, Xiao Y, Zhou JJ, Guan ZZ. 2018. Changed expressions of N-methyl-d-aspartate receptors in the brains of rats and primary neurons exposed to high level of fluoride. *J Trace Elem Med Biol* 45: 31-40.

Willems CB-V, Sande J, Dumont JE. 1972. Inhibition of thyroid secretion by sodium fluoride in vitro. *Biochim Biophys Acta* 264: 197-204.

Woodward JJ, Harms J. 1992. Potentiation of N-methyl-D-aspartate-stimulated dopamine release from rat brain slices by aluminum fluoride and carbachol. *J Neurochem* 58: 1547-1554.

Wu J, Cheng M, Liu Q, Yang J, Wu S, Lu X, Jin C, Ma H, Cai Y. 2015. Protective role of tert-butylhydroquinone against sodium fluoride-induced oxidative stress and apoptosis in PC12 cells. *Cell Mol Neurobiol* 35: 1017-1025.

Xia T, Zhang M, He WH, He P, Wang AG. 2007. [Effects of fluoride on neural cell adhesion molecules mRNA and protein expression levels in primary rat hippocampal neurons]. *Chin J Prev Med* 41: 475-478.

Xu B, Xu Z, Xia T, He P, Gao P, He W, Zhang M, Guo L, Niu Q, Wang A. 2011. Effects of the Fas/Fas-L pathway on fluoride-induced apoptosis in SH-SY5Y cells. *Environ Toxicol* 26: 86-92.

Xu Z, Xu B, Xia T, He W, Gao P, Guo L, Wang Z, Niu Q, Wang A. 2013. Relationship between intracellular Ca2+ and ROS during fluoride-induced injury in SH-SY5Y cells. *Environ Toxicol* 28: 307-312.

Yamashita K, Field JB. 1972. Elevation of cyclic guanosine 3,5; monophosphate levels in dog thyroid slices caused by acetylcholine and sodium fluoride. *J Biol Chem* 247: 7062-7066.

Yan L, Liu S, Wang C, Wang F, Song Y, Yan N, Xi S, Liu Z, Sun G. 2013. JNK and NADPH oxidase involved in fluoride-induced oxidative stress in BV-2 microglia cells. *Mediators Inflamm* 2013: 895-975.

Zhang CY, Chen R, Wang F, Ren C, Zhang P, Li Q, Li HH, Guo KT, Geng DQ, Liu CF. 2016. EGb-761 attenuates the anti-proliferative activity of fluoride via DDK1 in PC-12 cells. *Neurochem Res* 42(2): 606-614.

Zhang M, Wang A, He W, He P, Xu B, Xia T, Chen X, Yang K. 2007. Effects of fluoride on the expression of NCAM, oxidative stress, and apoptosis in primary cultured hippocampal neurons. *Toxicology* 236: 208-216.

Zhang M, Wang A, Xia T, He P. 2008. Effects of fluoride on DNA damage, S-phase cell-cycle arrest and the expression of NF-kappaB in primary cultured rat hippocampal neurons. *Toxicol Lett* 179: 1-5.

Zhang S, Zheng X, Sun Y, Wang Y, Zhang Z. 2015. Alterations in oxidative stress and apoptosis in cultured PC12 cells exposed to fluoride. *Fluoride* 48: 213-222.

Zhao L, Xiao Y, Deng CM, Tan LC, Guan ZZ. 2016. Protective effect of lovastatin on neurotoxicity of excessive fluoride in primary hippocampal neurons. *Fluoride* 49: 36-46.

Zhao XL, Gao WH, Zhao ZL. 1994. [Effects of sodium fluoride on the activity of Ca2+Mg(2+)-ATPase in synaptic membrane in rat brain]. *Chin J Prev Med* 28: 264-266.

Appendix D. Risk-of-bias Figures

Figures

D.1. Studies in Humans

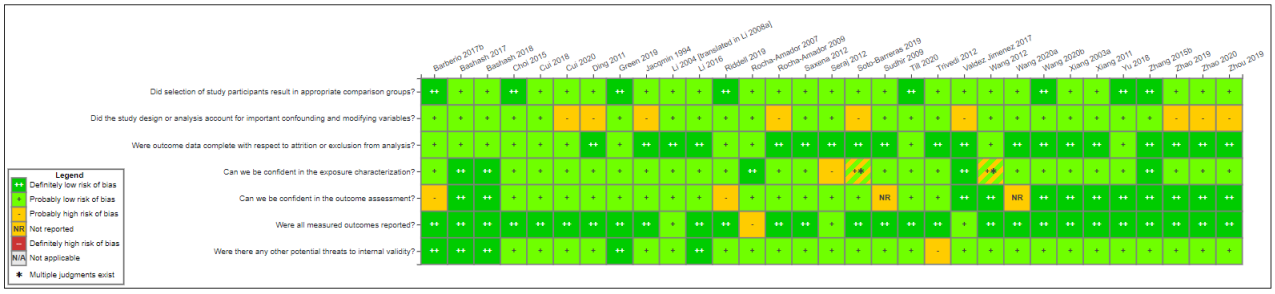


Figure D-1. Risk-of-bias Heatmap for Low Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-1 and additional study details in HAWC here.

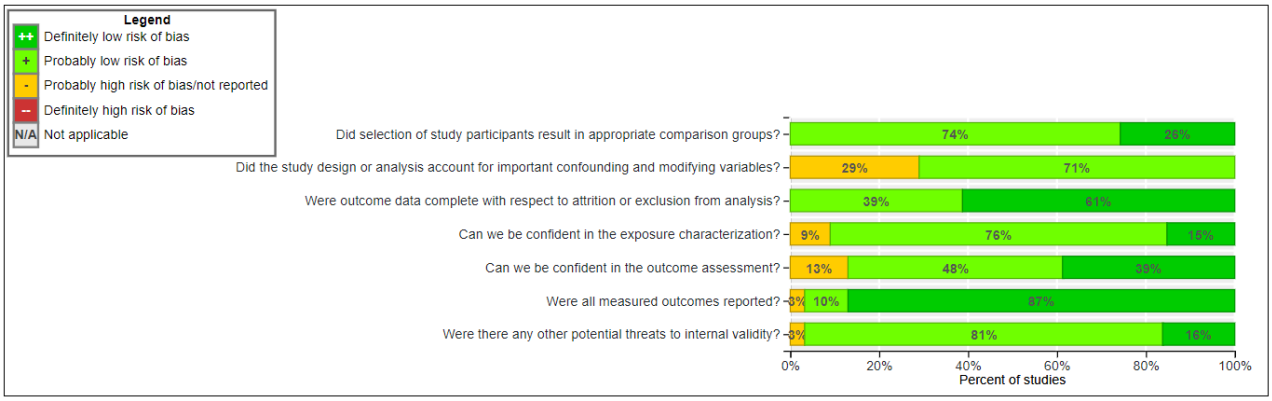


Figure D-2. Risk-of-bias Bar Chart for Low Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-2 and additional study details in HAWC here.

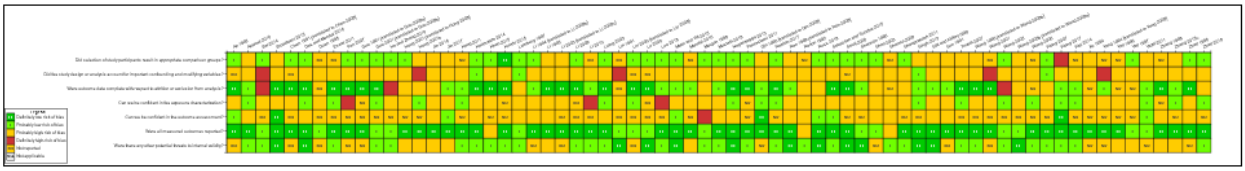


Figure D-3. Risk-of-bias Heatmap for High Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-3 and additional study details in HAWC here.

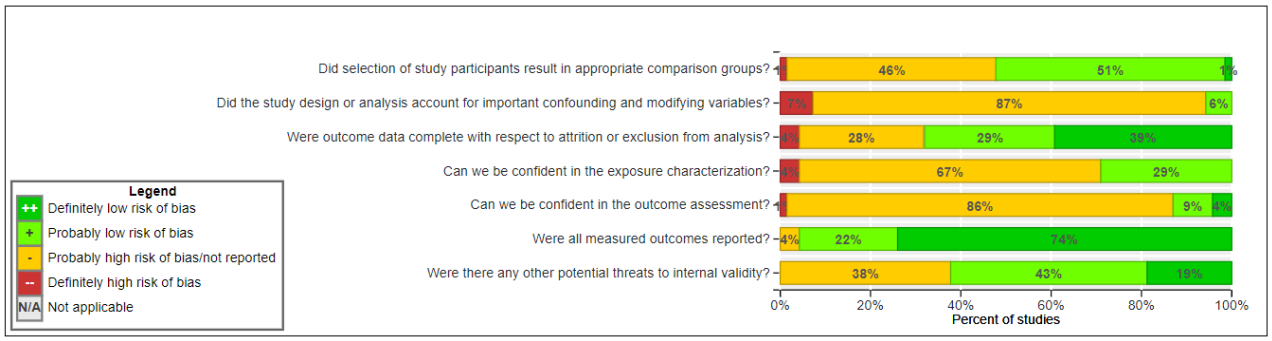


Figure D-4. Risk-of-bias Bar Chart for High Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-4 and additional study details in HAWC here.

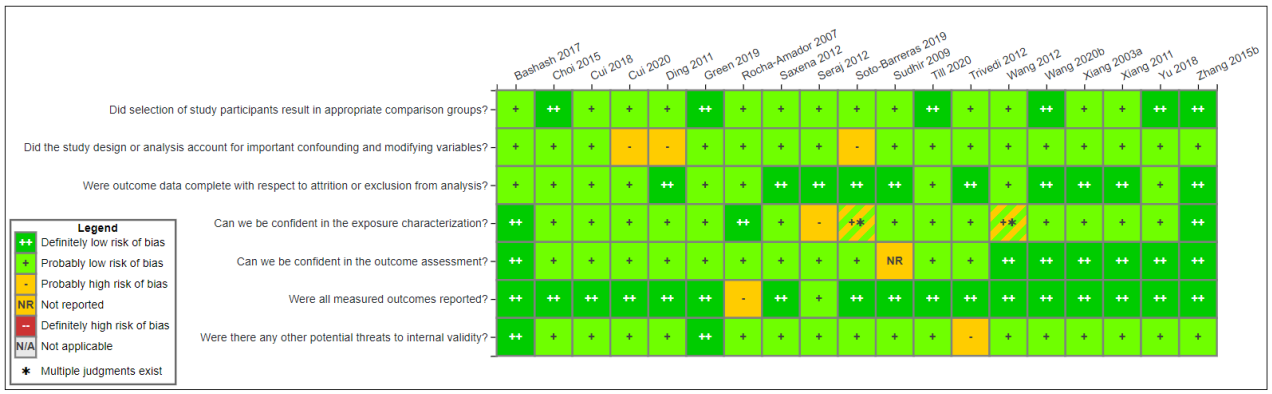


Figure D-5. Risk-of-bias Heatmap for Low Risk-of-bias Children's IQ Studies Following Fluoride Exposure

An interactive version of Figure D-5 and additional study details in HAWC here.

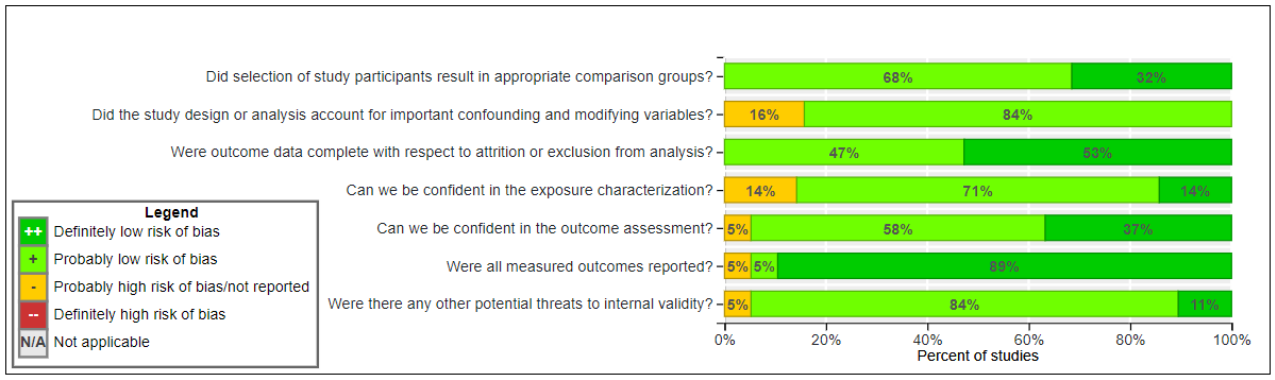


Figure D-6. Risk-of-bias Bar Chart for Low Risk-of-bias Children's IQ Studies Following Fluoride Exposure

An interactive version of Figure D-6 and additional study details in HAWC here.

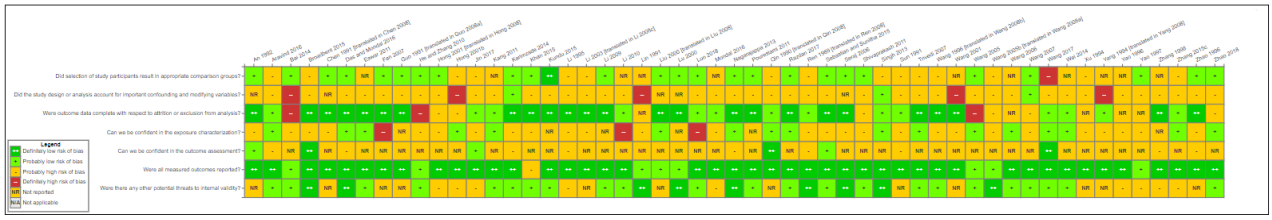


Figure D-7. Risk-of-bias Heatmap for High Risk-of-bias Children's IQ Studies Following Fluoride Exposure

An interactive version of Figure D-7 and additional study details in HAWC here.

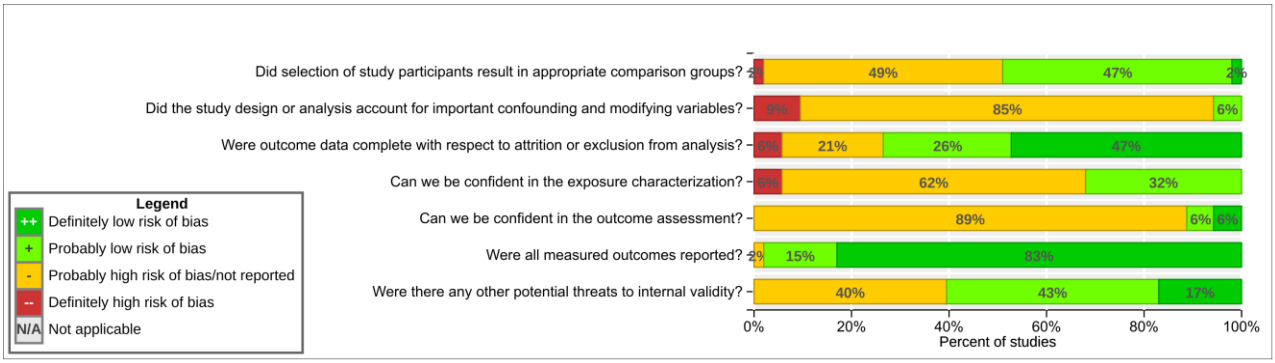


Figure D-8. Risk-of-bias Bar Chart for High Risk-of-bias Children's IQ Studies Following Fluoride Exposure

An interactive version of Figure D-8 and additional study details in HAWC here.

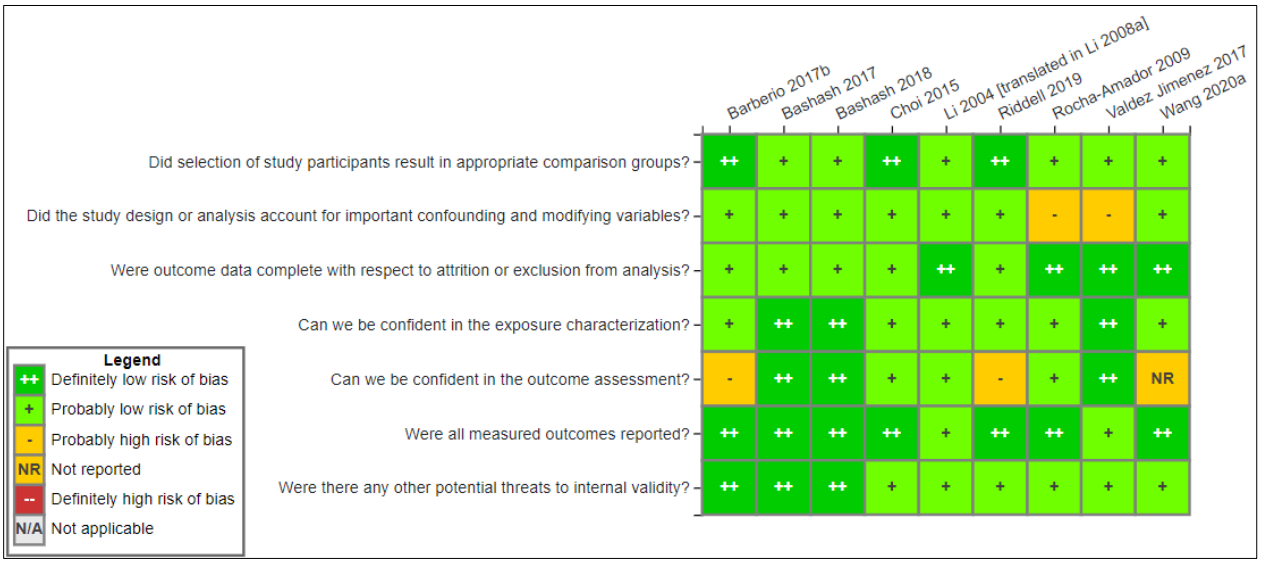


Figure D-9. Risk-of-bias Heatmap for Low Risk-of-bias Children's Other Neurodevelopmental Effect Studies Following Fluoride Exposure

An interactive version of Figure D-9 and additional study details in HAWC here.

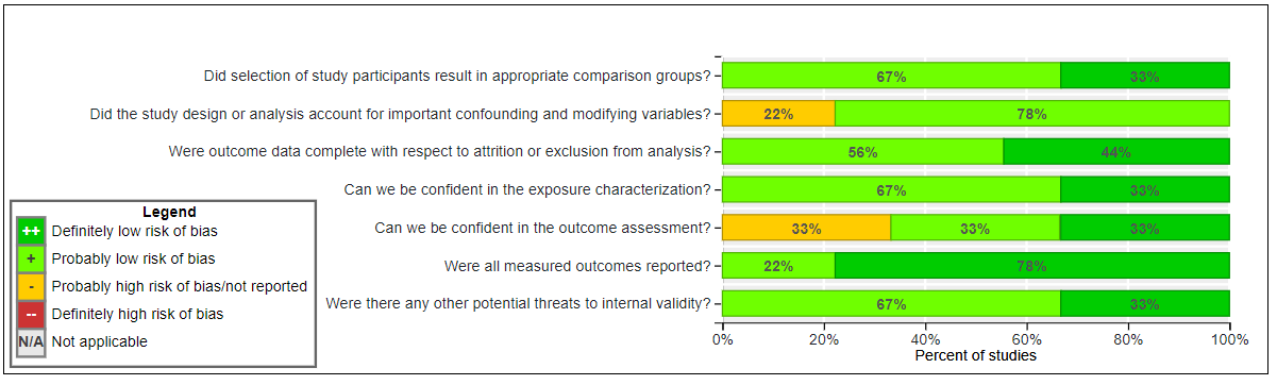


Figure D-10. Risk-of-bias Bar Chart for Low Risk-of-bias Children's Other Neurodevelopmental Effect Studies Following Fluoride Exposure

An interactive version of Figure D-10 and additional study details in HAWC here.

Figure D-11. Risk-of-bias Heatmap for High Risk-of-bias Children's Other Neurodevelopmental Effect Studies Following Fluoride Exposure

An interactive version of Figure D-11 and additional study details in HAWC here.

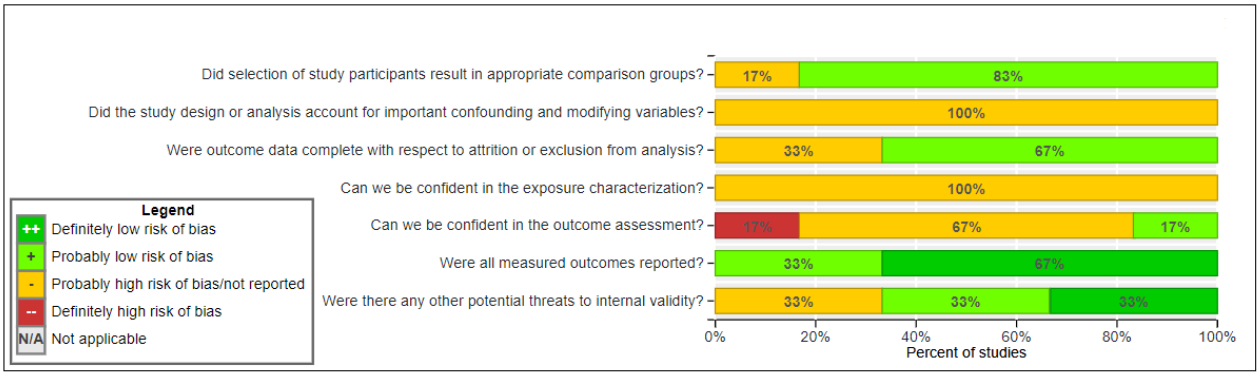


Figure D-12. Risk-of-bias Bar Chart for High Risk-of-bias Children's Other Neurodevelopmental Effect Studies Following Fluoride Exposure

An interactive version of Figure D-12 and additional study details in HAWC here.

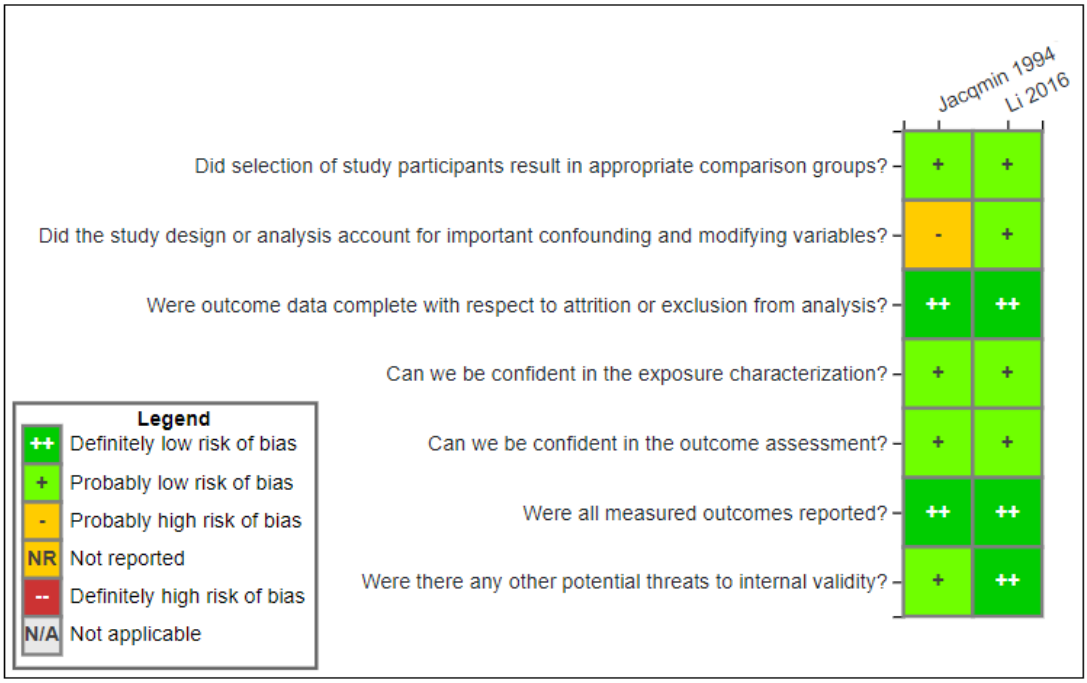


Figure D-13. Risk-of-bias Heatmap for Low Risk-of-bias Adult Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-13 and additional study details in HAWC here.

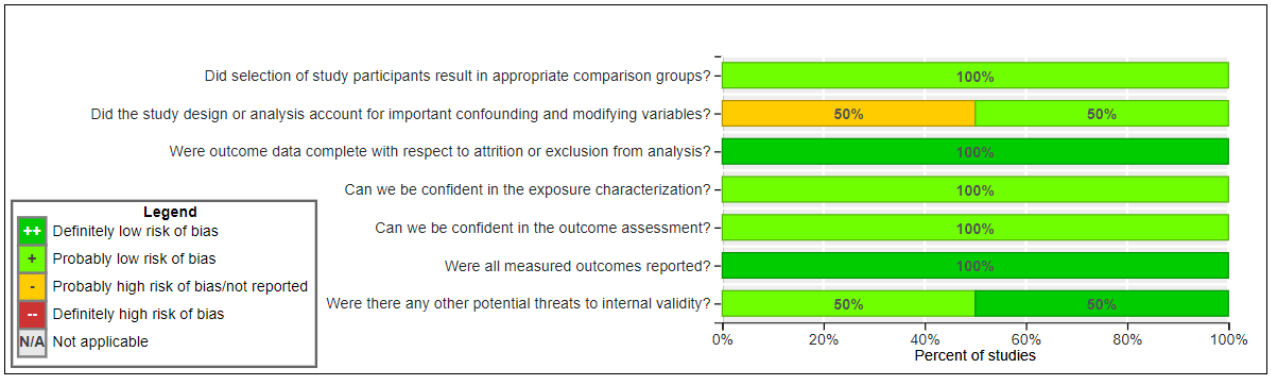


Figure D-14. Risk-of-bias Bar Chart for Low Risk-of-bias Adult Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-14 and additional study details in HAWC here.

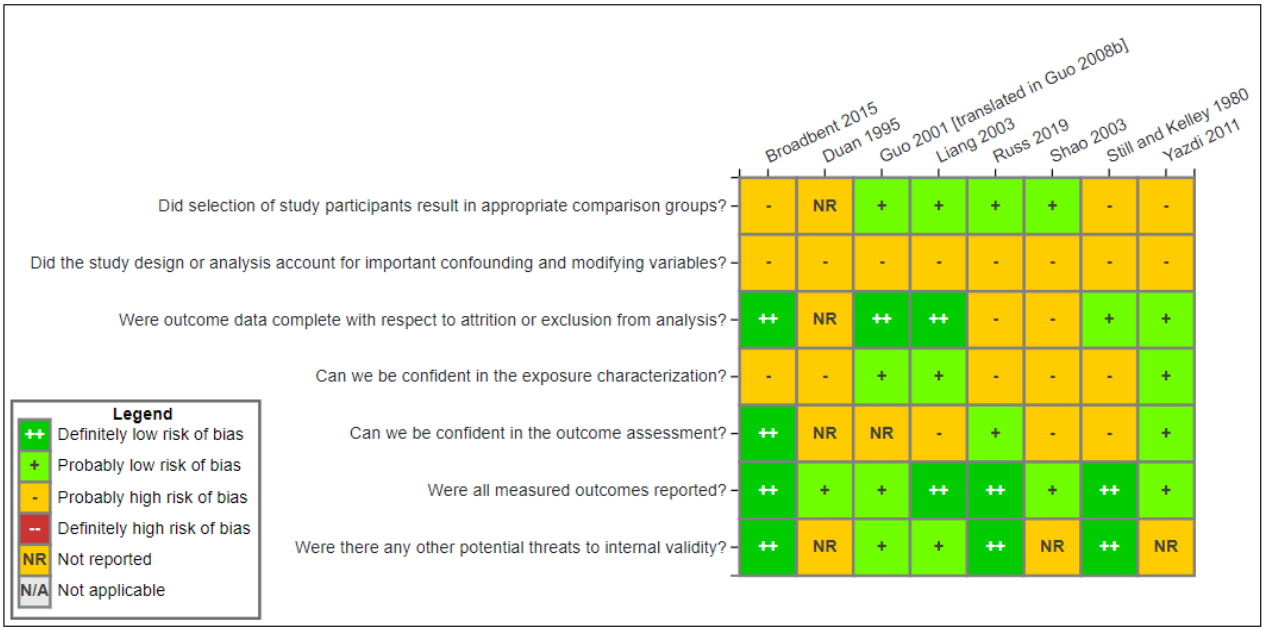


Figure D-15. Risk-of-bias Heatmap for High Risk-of-bias Adult Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-15 and additional study details in HAWC here.

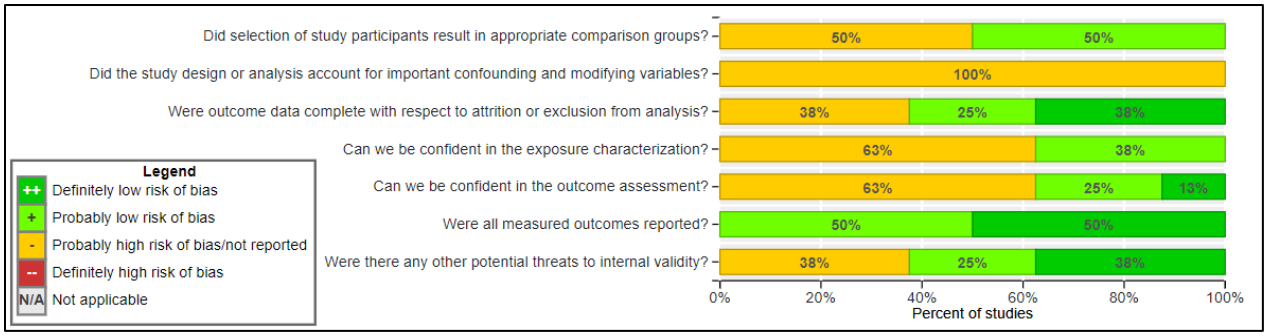


Figure D-16. Risk-of-bias Bar Chart for High Risk-of-bias Adult Cognitive Studies Following Fluoride Exposure

An interactive version of Figure D-16 and additional study details in HAWC here.

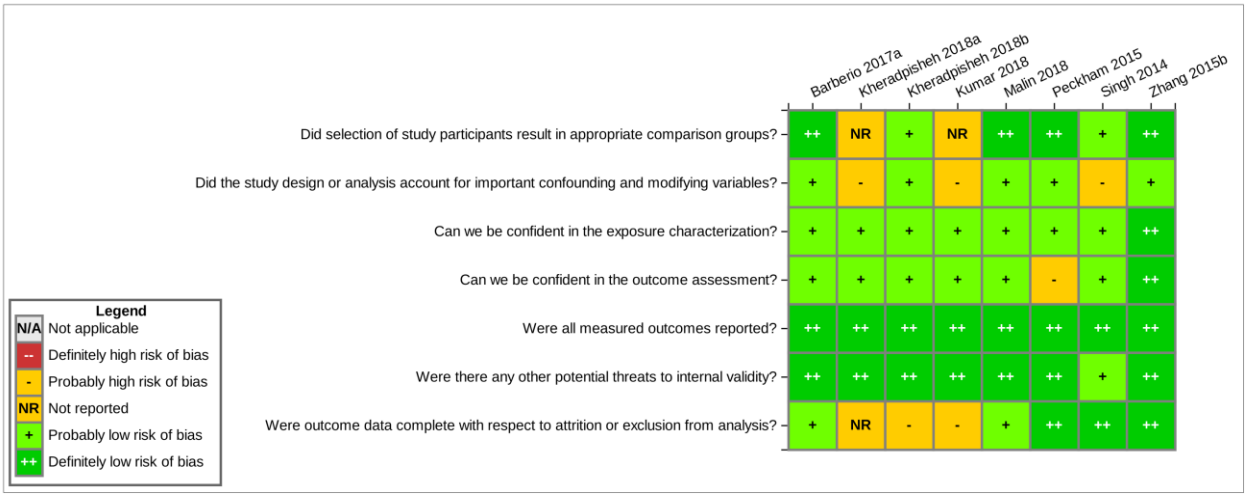


Figure D-17. Risk-of-bias Heatmap for Low Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure

An interactive version of Figure D-17 and additional study details in HAWC here.

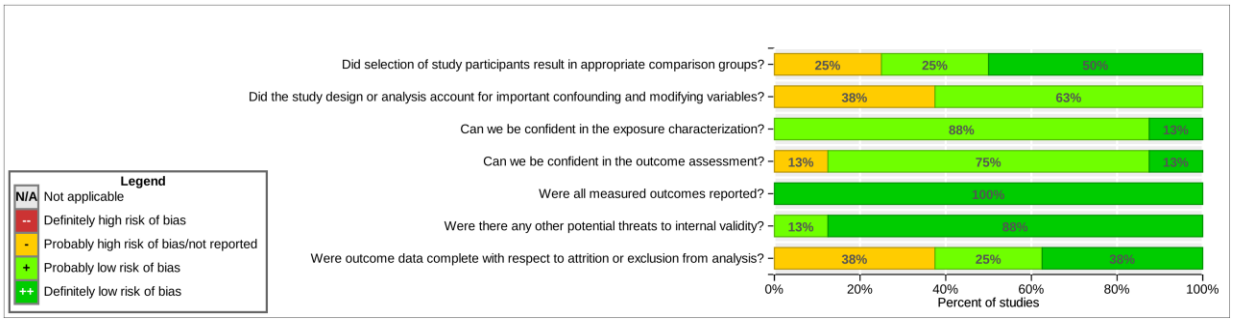


Figure D-18. Risk-of-bias Bar Chart for Low Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure

An interactive version of Figure D-18 and additional study details in HAWC here.

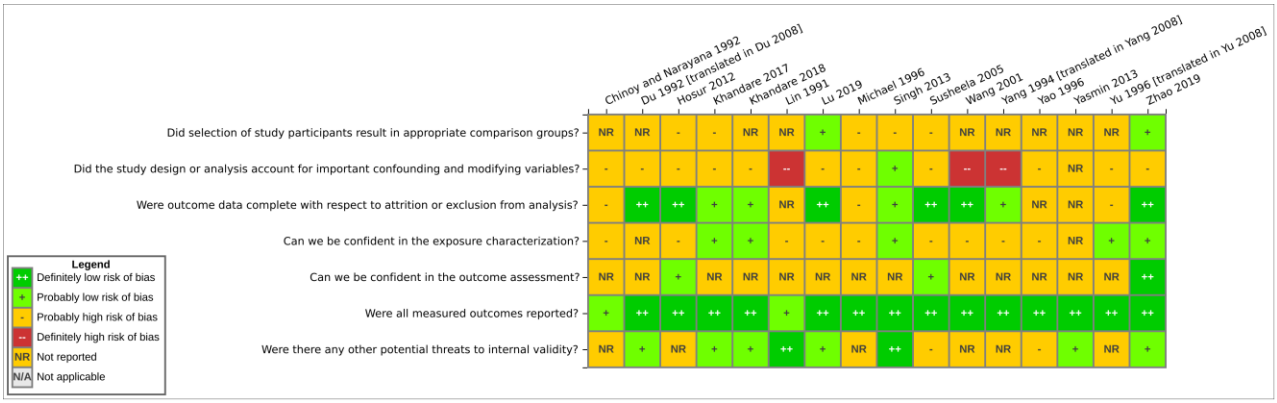


Figure D-19. Risk-of-bias Heatmap for High Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure

An interactive version of Figure D-19 and additional study details in HAWC here.

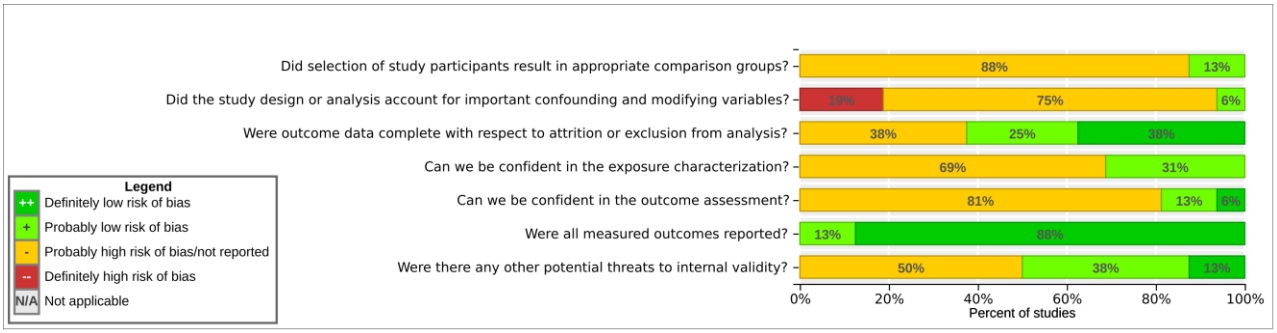


Figure D-20. Risk-of-bias Bar Chart for High Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure

An interactive version of Figure D-20 and additional study details in HAWC here.

D.2. Studies in Non-human Animals

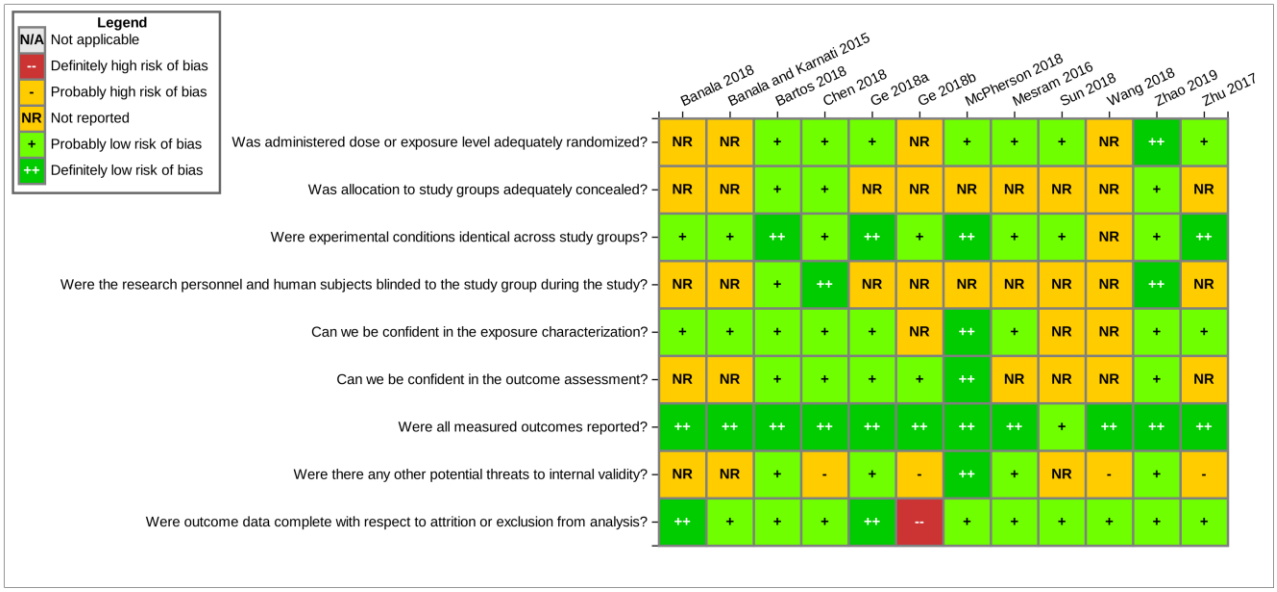


Figure D-21. Risk-of-bias Heatmap for New Developmental Animal Learning and Memory Studies Following Fluoride Exposure

An interactive version of Figure D-21 and additional study details in HAWC here.

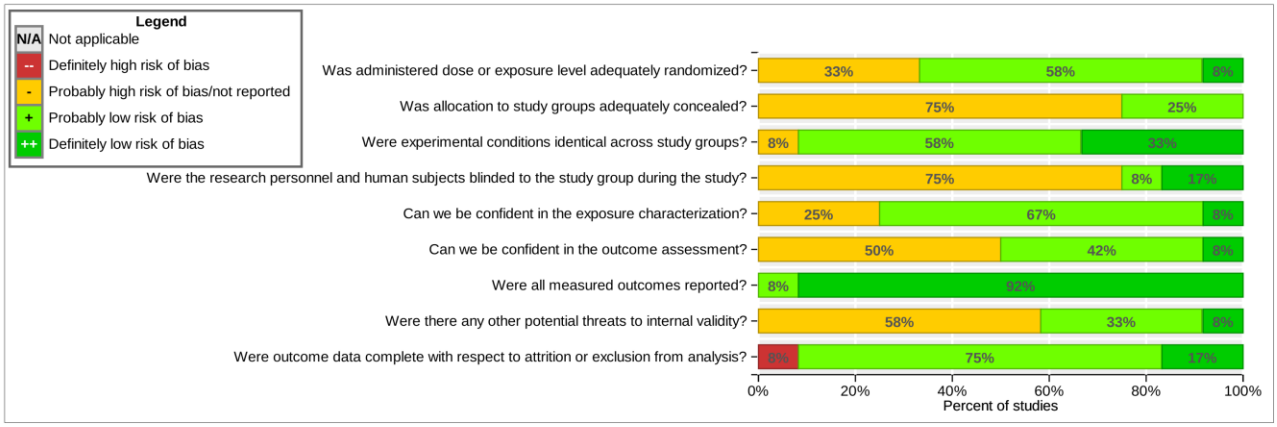


Figure D-22. Risk-of-bias Bar Chart for New Developmental Animal Learning and Memory Studies Following Fluoride Exposure

An interactive version of Figure D-22 and additional study details in HAWC here.

Figure D-23. Risk-of-bias Heatmap for New Adult Animal Learning and Memory Studies Following Fluoride Exposure

An interactive version of Figure D-23 and additional study details in HAWC here.

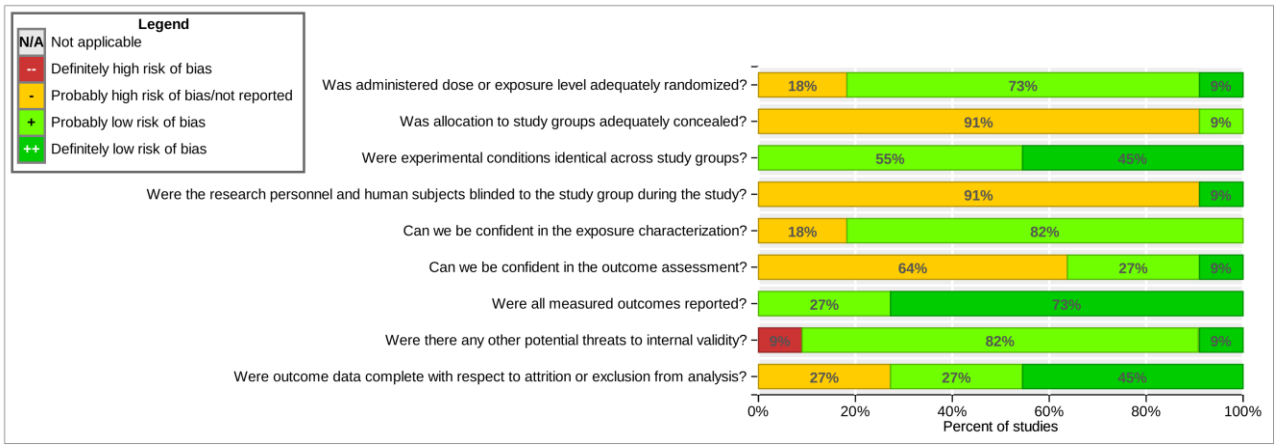


Figure D-24. Risk-of-bias Bar Chart for New Adult Animal Learning and Memory Studies Following Fluoride Exposure

An interactive version of Figure D-24 and additional study details in HAWC here.

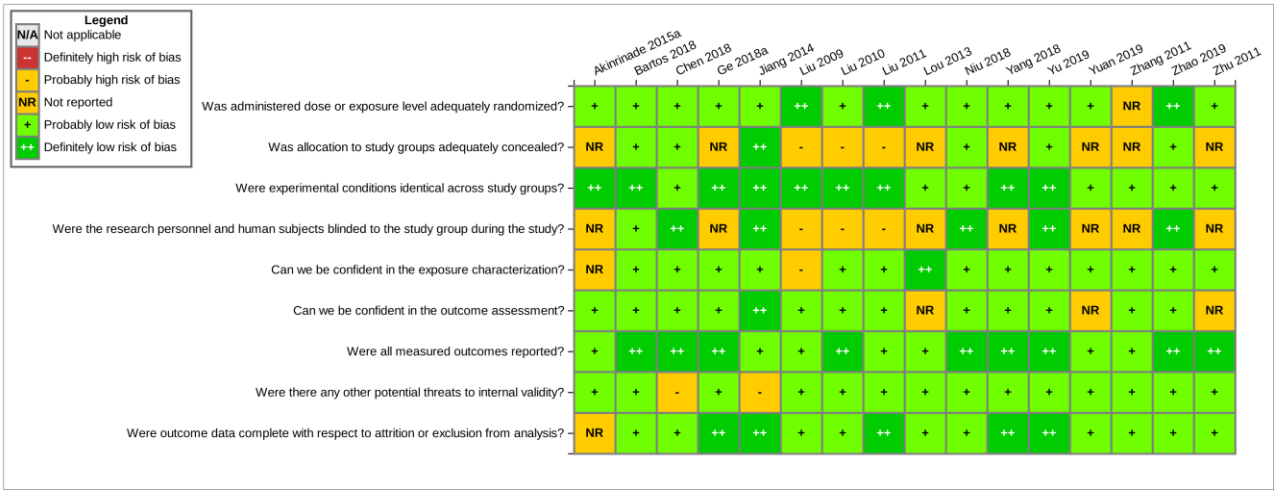


Figure D-25. Risk-of-bias Heatmap for Low Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure

An interactive version of Figure D-25 and additional study details in HAWC here.

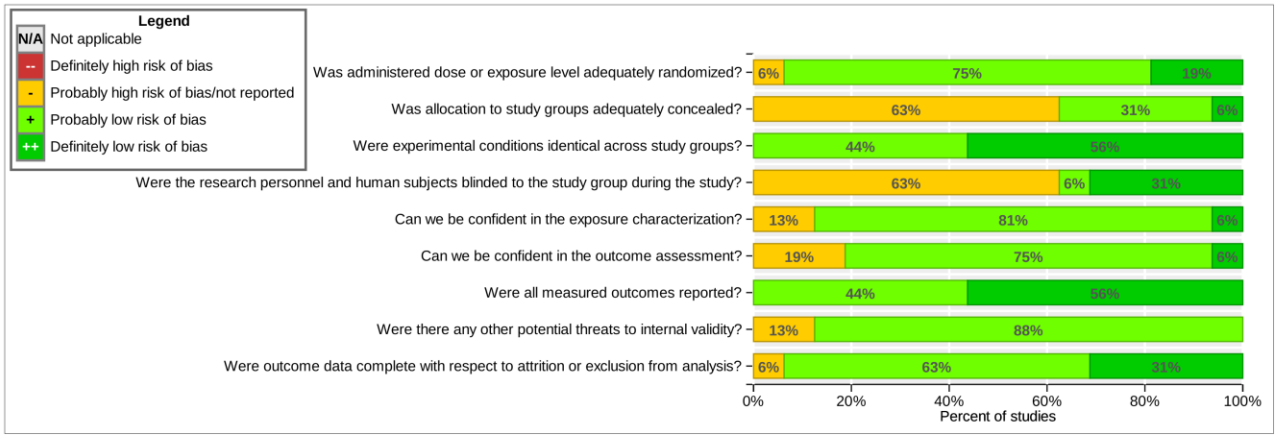


Figure D-26. Risk-of-bias Bar Chart for Low Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure

An interactive version of Figure D-26 and additional study details in HAWC here.

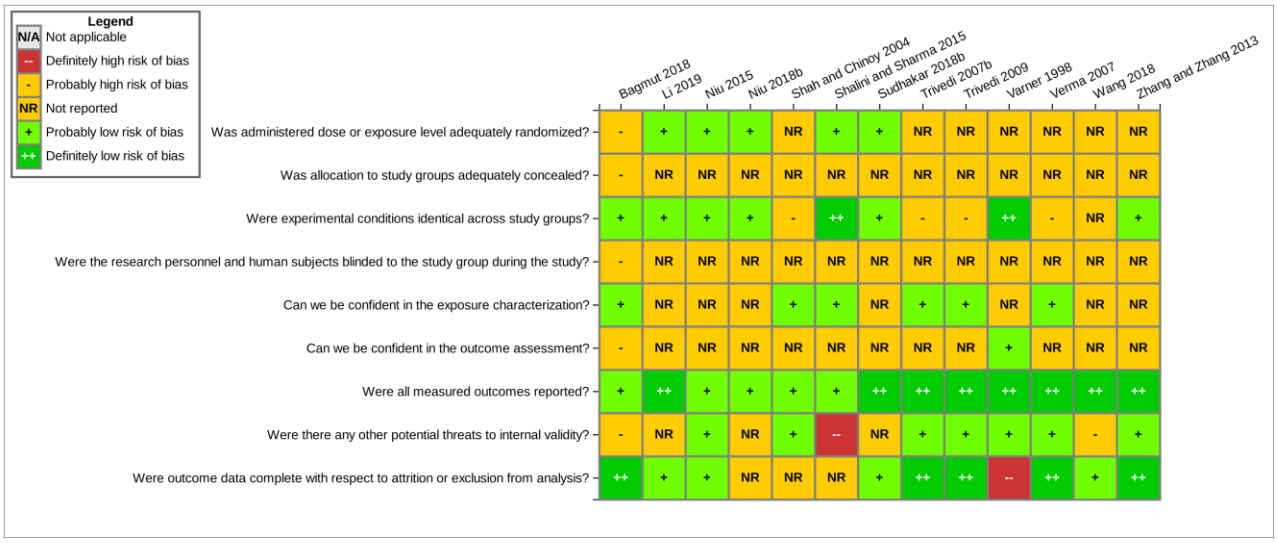


Figure D-27. Risk-of-bias Heatmap for High Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure

An interactive version of Figure D-27 and additional study details in HAWC here.

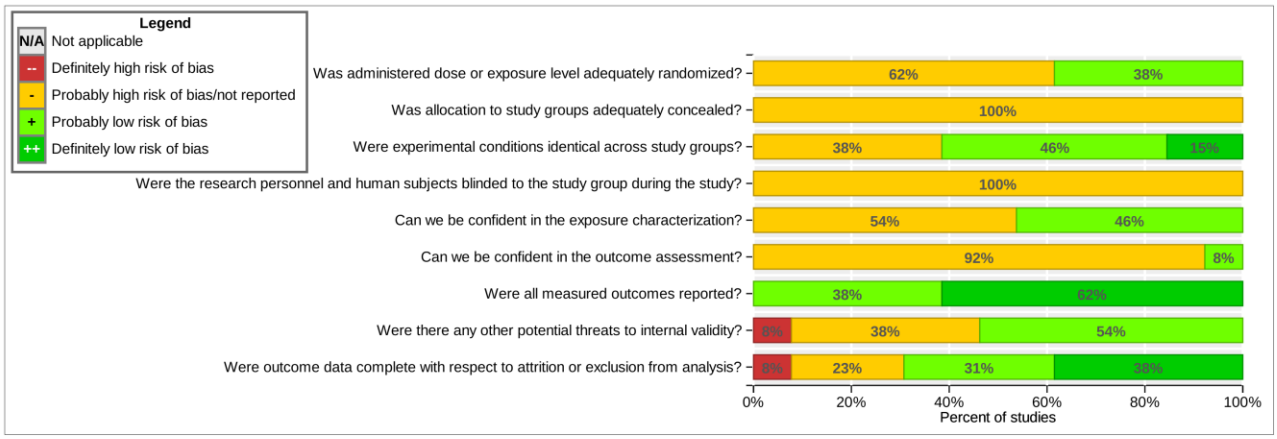


Figure D-28. Risk-of-bias Bar Chart for High Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure

An interactive version of Figure D-28 and additional study details in HAWC here.

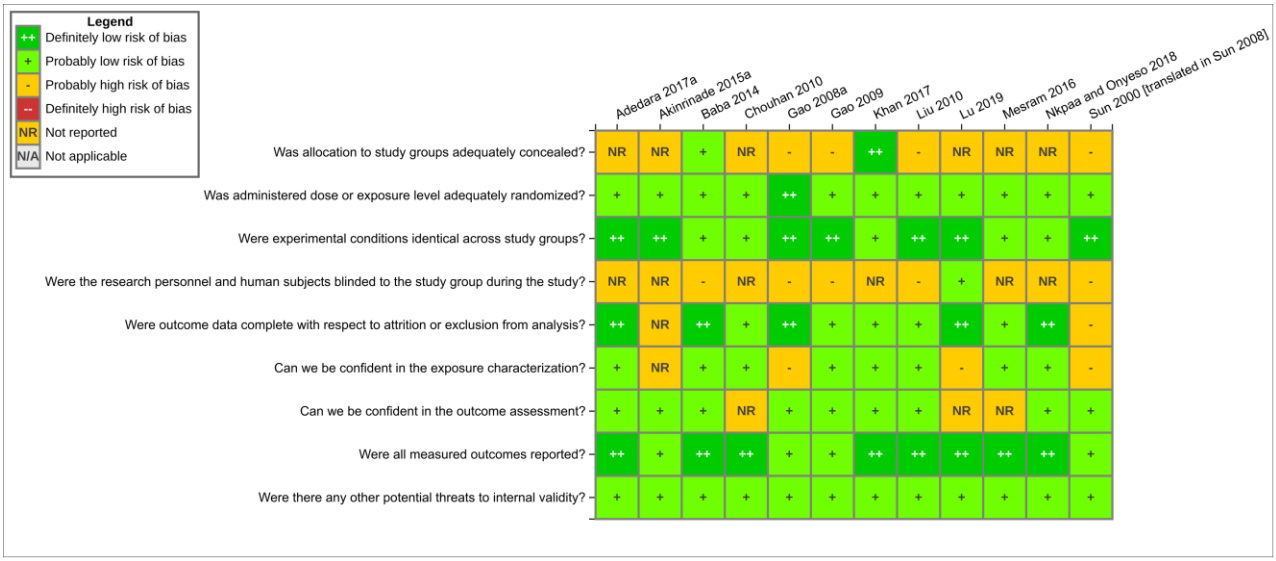


Figure D-29. Risk-of-bias Heatmap for Low Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure

An interactive version of Figure D-29 and additional study details in HAWC here.

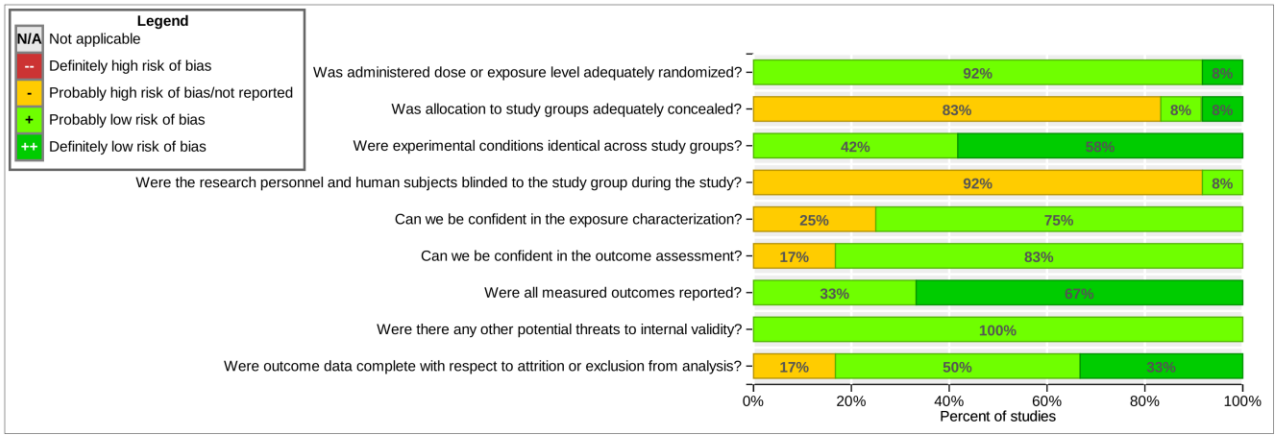


Figure D-30. Risk-of-bias Bar Chart for Low Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure

An interactive version of Figure D-30 and additional study details in HAWC here.

Figure D-31. Risk-of-bias Heatmap for High Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure

An interactive version of Figure D-31 and additional study details in HAWC here.

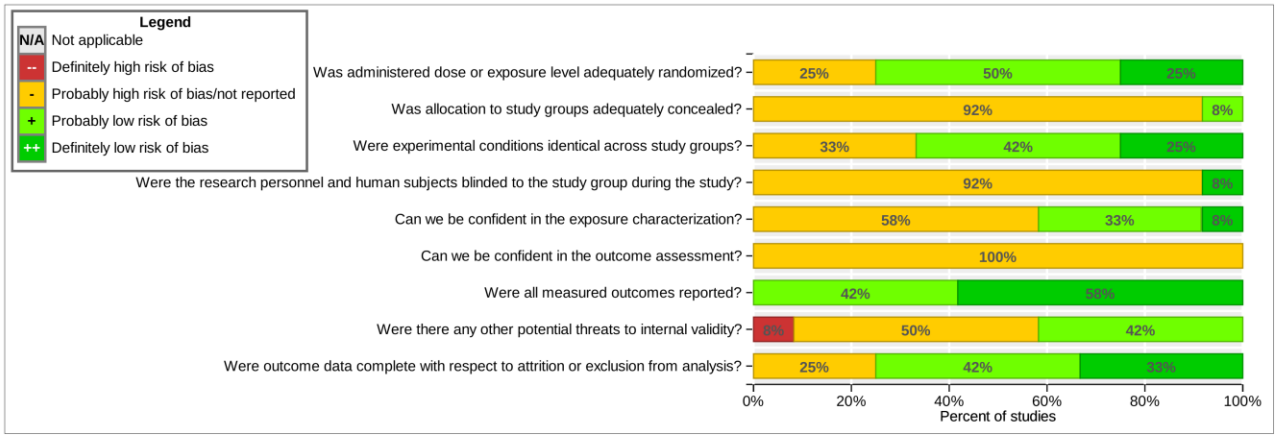


Figure D-32. Risk-of-bias Bar Chart for High Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure

An interactive version of Figure D-32 and additional study details in HAWC here.

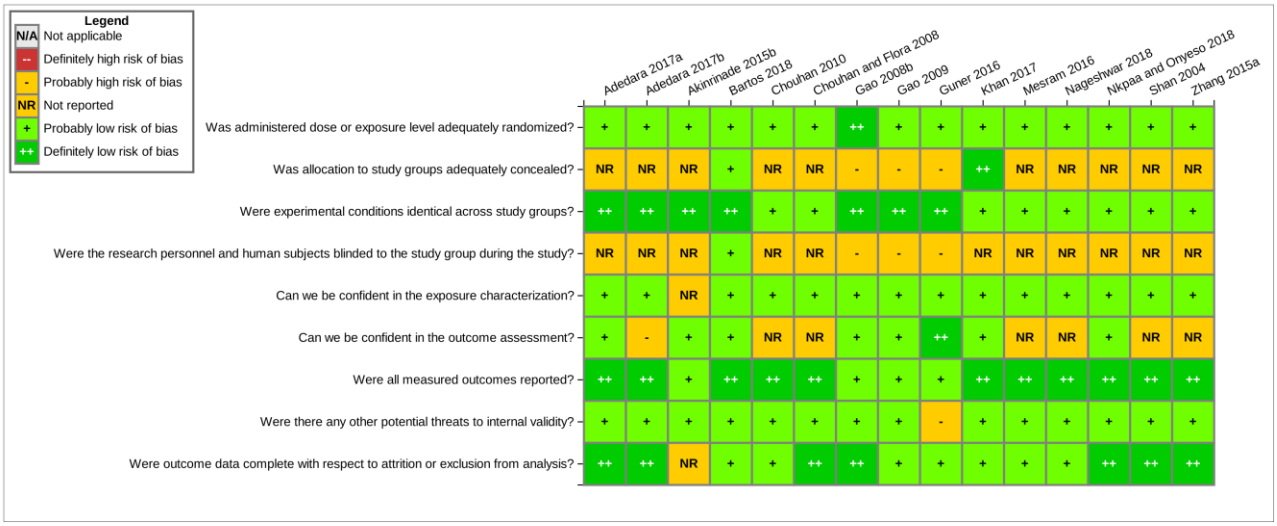


Figure D-33. Risk-of-bias Heatmap for Low Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure

An interactive version of Figure D-33 and additional study details in HAWC here.

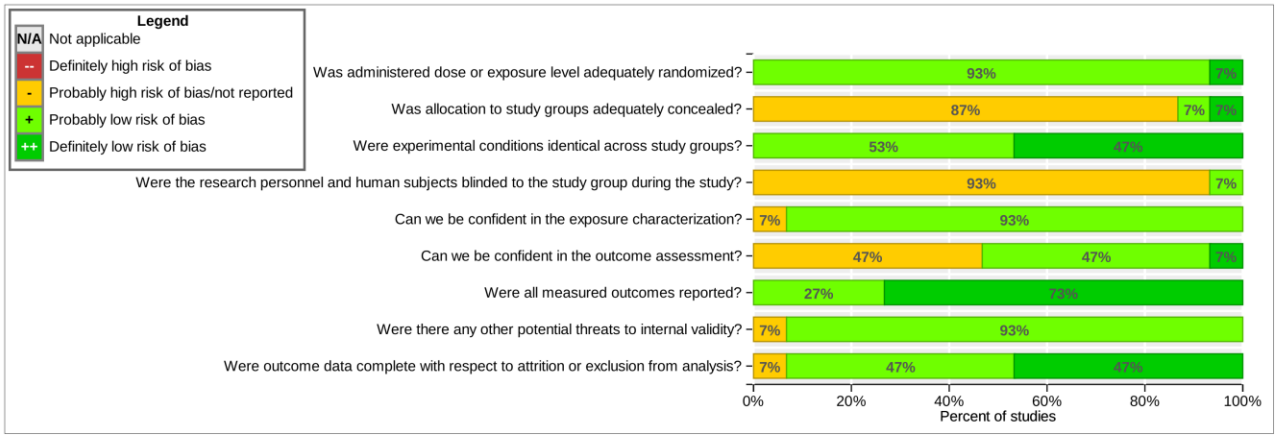


Figure D-34. Risk-of-bias Bar Chart for Low Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure

An interactive version of Figure D-34 and additional study details in HAWC here.

Figure D-35. Risk-of-bias Heatmap for High Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure

An interactive version of Figure D-35 and additional study details in HAWC here.

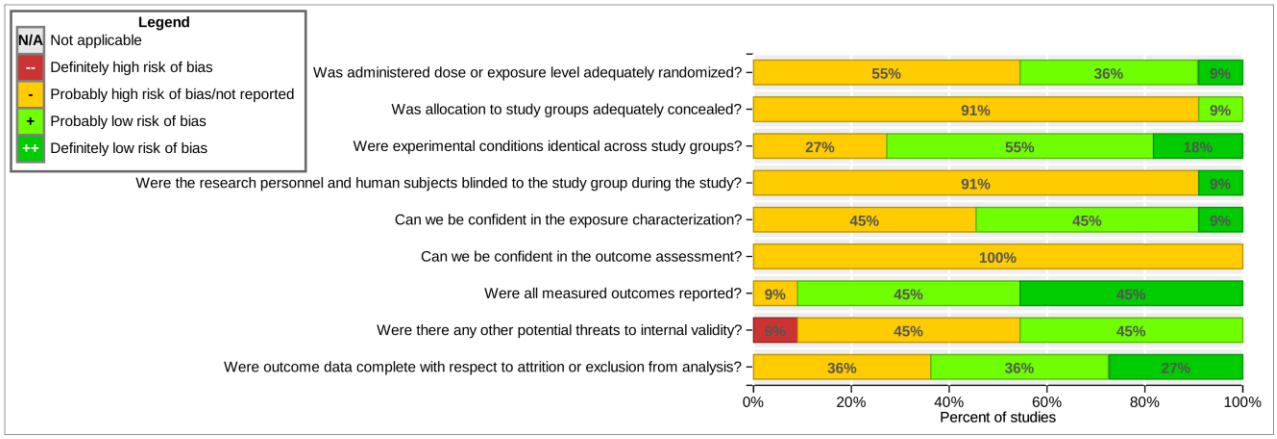


Figure D-36. Risk-of-bias Bar Chart for High Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure

An interactive version of Figure D-36 and additional study details in HAWC here.

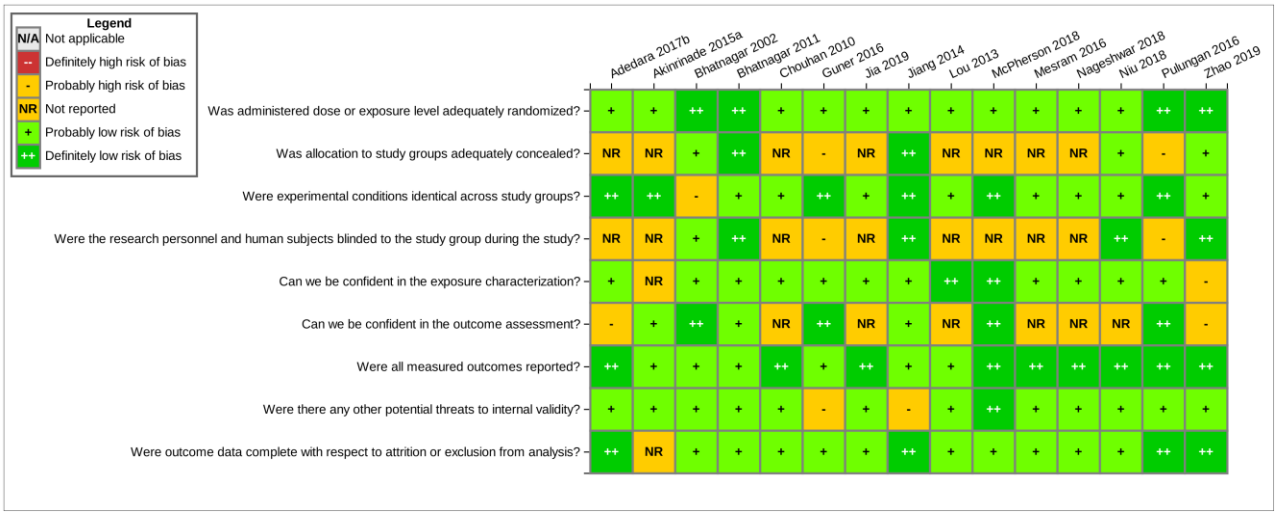


Figure D-37. Risk-of-bias Heatmap for Low Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure

An interactive version of Figure D-37 and additional study details in HAWC here.

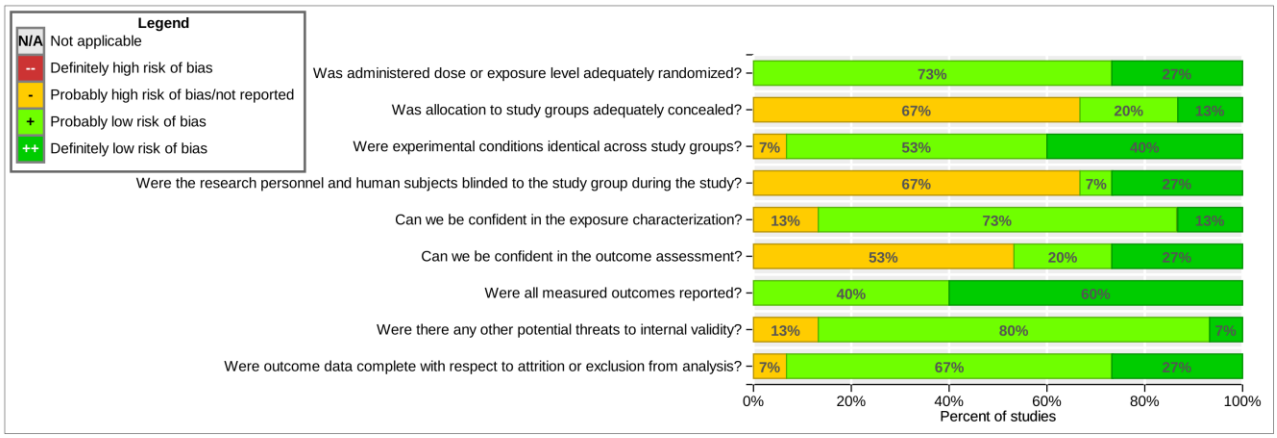


Figure D-38. Risk-of-bias Bar Chart for Low Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure

An interactive version of Figure D-38 and additional study details in HAWC here.

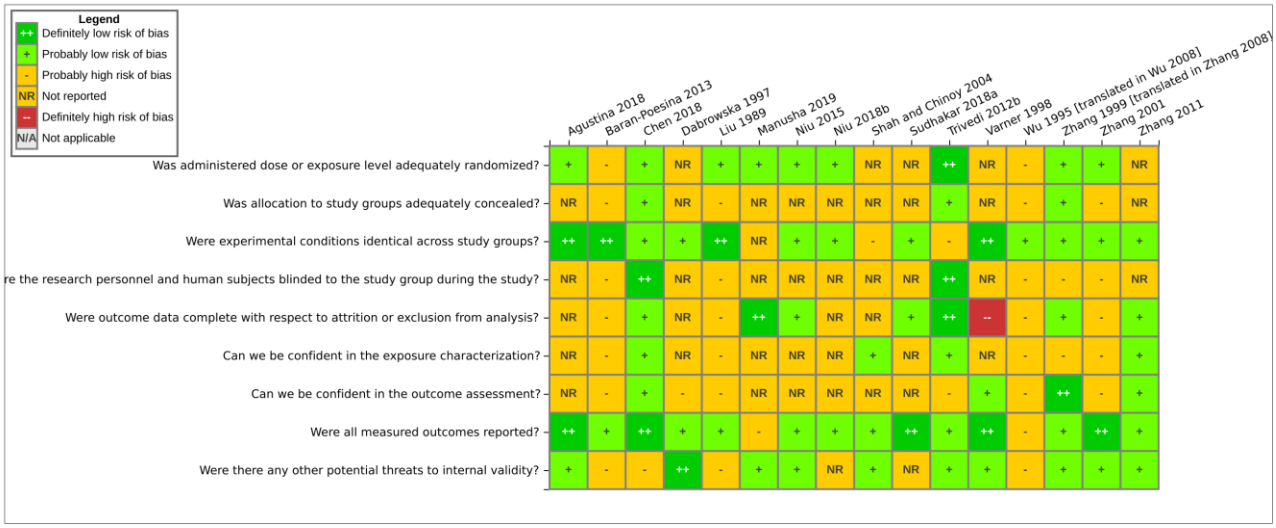


Figure D-39. Risk-of-bias Heatmap for High Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure

An interactive version of Figure D-39 and additional study details in HAWC here.

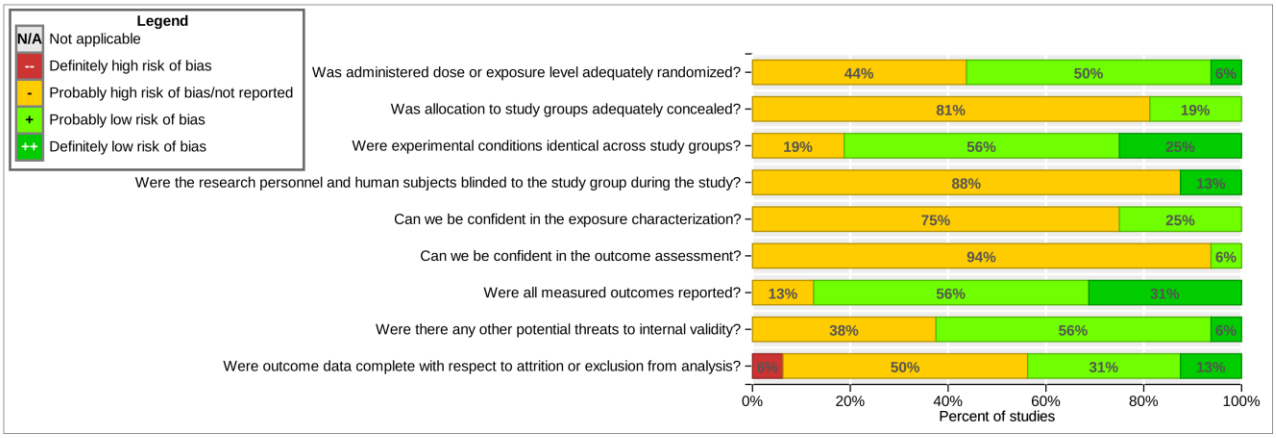


Figure D-40. Risk-of-bias Bar Chart for High Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure

An interactive version of Figure D-40 and additional study details in HAWC here.

Appendix E. Details for Low Risk-of-bias Studies

Table of Contents

E.1. IQ Studies

E.1.1. Bashash et al. (2017)

E.1.1.1. Study Details

- ***Study design:*** Prospective cohort
- ***Population:*** Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants (pregnant mothers and their children aged 4 or 6–12 years).
- ***Study area:*** Mexico City, Mexico
- ***Sample size:*** 299 mother-child pairs, of whom 211 had data for the IQ analyses.
- ***Data relevant to the review:*** Adjusted and unadjusted associations between IQ scores and maternal or child's urinary fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant association between maternal urinary fluoride and IQ score (adjusted β = −2.50 per 0.5 mg/L increase; 95% CI: −4.12, −0.59). No significant associations with children's urinary fluoride.

E.1.1.2. Risk of Bias

- **Author contacts:**
  - Authors were contacted for additional information on whether clustering was addressed. The authors provided results from additional models with cohort as a random effect.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Study participants were selected from two different cohorts from three hospitals in Mexico City that serve low-to-moderate income populations. One cohort was from an observational study of prenatal lead exposure and neurodevelopment outcomes, and the other was from a randomized trial of the effect of calcium on maternal blood lead levels. The authors state that participants had no history of psychiatric disorders, high-risk pregnancies, gestational diabetes, illegal drug use, or continuous prescription drugs, but no information on smoking habits was considered. Study populations appear to be similar, but there may be some differences because subjects were selected from two different cohorts that were recruited from slightly different time periods.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar despite the subjects coming from different original study populations wherein different methods were used for recruitment.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Data were collected via questionnaire on maternal age, education, marital status at first prenatal visit, birth order, birth weight, gestational age at delivery, maternal smoking, maternal IQ, and HOME scores. All models were adjusted for gestational age at birth, sex, birth weight, birth order, age at testing,

maternal marital status, smoking history, age at delivery, maternal IQ, education, and cohort, with additional testing for children's urinary fluoride, mercury, lead, and calcium. Sensitivity analyses additionally adjusted for HOME score. Important covariates not considered included BMI, iodine deficiency, arsenic, and maternal mental health and nutrition. Arsenic is assumed not to be a potential co-exposure in this population because the study authors did not discuss it as an issue, but did consider other co-exposures. Arsenic is included in the water quality control program in Mexico City and is not considered a concern in this population.

- *Potentially important study-specific covariates:* All key covariates were addressed.
  - *Direction/magnitude of effect size:* Not applicable.
- *Basis for rating:* Probably low risk of bias based on direct evidence that key covariates, including other potential co-exposures, were addressed and indirect evidence that the methods used to collect the information were valid and reliable and that arsenic is not likely to be an issue in this study population.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although there was a large amount of attrition, the study authors clearly describe all reasons for attrition and also provide characteristics to compare those participants included to those excluded. There were some slight differences between those included and those excluded, but there is nothing to indicate that the attrition would potentially bias the results.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Urinary fluoride concentrations were determined in spot urine samples (2nd morning void) collected from mothers (during at least one trimester) and children ages 6–12 years. Fluoride content was measured using ion-selective electrode-based assays. QC methods were described including between laboratory correlations. All samples were measured in duplicate. Extreme outliers were excluded. Urinary dilution was addressed by using creatinine-adjusted levels.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)

- *Summary:* Outcome was assessed using the McCarthy Scales of Children's Abilities (MSCA) in 4-year-old children (translated into Spanish) and the Wechsler Abbreviated Scale of Intelligence (WASI) in 6–12-year-olds. The WASI is a well-established test, and the validity of both tests is well documented by the authors. Inter-examiner reliability was evaluated and reported with a correlation of 0.99 (++ for methods). The study report stated that psychologists were blind to the children's fluoride exposure (++ for blinding). Overall rating for methods and blinding = ++.
- *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study. Statistical tests of bivariate associations (using Chi-square tests for categorical variables and analysis of variance [ANOVA]) were used to compare the means of the outcomes or exposure within groups based on the distribution of each covariate. Generalized additive models (GAMs) were used to estimate the adjusted association between fluoride exposure and measures of children's intelligence. Residual diagnostics were used to examine model assumptions and identify any potentially influential observations. Results are reported as adjusted effects and 95% CIs. In sensitivity analyses, regression models accounted for clustering at the cohort level by using cohort as a fixed effect in the models. Although using cohort as a random effect would be more appropriate, using individual-level exposure data and accounting for numerous important covariates in the models likely captured the cohort effect. Additional models with cohort as a random effect were also subsequently made available via personal communication with the study authors and showed similar results to the main model.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk if bias is based on direct evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include

individual exposure measurements, blinding of outcome assessor to participants' fluoride exposure, and the prospective cohort study design.

E.1.2. Choi et al. (2015)

E.1.2.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** First-grade children (ages 6–8 years)
- ***Study area:*** Mianning County in southern Sichuan, China
- ***Sample size:*** 51 first-grade children
- ***Data relevant to the review:*** Associations between IQ (digit span for auditory span and working memory and block design for visual organization and reasoning components of WISC-IV only) with continuous urine or drinking water fluoride levels. Study also had information based on dental fluorosis score.
- ***Reported association with fluoride exposure:*** Yes: Compared to the normal/questionable dental fluorosis, the moderate/severe dental fluorosis group was associated with significantly lower total (adjusted β = −4.28; 95% CI: −8.22, −0.33) and backward (adjusted β = −2.13; 95% CI: −4.24, −0.02) digit span scores. Linear associations between total digit span and log-transformed fluoride in urine (adjusted β = −1.67; 95% CI: −5.46, 2.12) and in drinking water (adjusted β = −1.39; 95% CI: −6.76, 3.98) were observed but not significant. Other outcomes not significantly associated with fluoride exposure.

E.1.2.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were selected during the same time frame using the same methods. Fifty-one first-grade children residing in Mianning County in southern Sichuan, China were included in this pilot study. It is not specified whether the 51 children represented all the first-grade children from this area or whether some refused to participate. Children who did not speak Chinese, were not students at the Primary School of Sunshui Village in Mianning County, or those with chronic or acute disease that might affect neurobehavioral function tests were excluded. Demographic characteristics are presented in Table 1 of the study, which indicates that subjects were similar. Important covariates are adjusted for in the statistical analyses.
  - *Basis for Rating:* Definitely low risk of bias based on direct evidence that the exposure groups were similar and were recruited within the same time frame using the same methods with no evidence of differences in participation/response rates.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The parents or guardians completed a questionnaire on demographic and personal characteristics of the children (sex, age at testing, parity, illnesses before age 3, and past medical history) and caretakers (age, parity, education and occupational histories, residential history, and household income). A 20-µL capillary blood sample was collected at the school by a Mianning County Center for Disease Control (CDC) health practitioner and tested for possible iron deficiency, which could have been used as a covariate of neurodevelopmental performance. Important covariates that were not assessed include maternal BMI, parental mental health, maternal smoking status, maternal reproductive factors, parental IQ, and HOME score. However, the study authors noted that confounding bias appeared to be limited due to the minimal diversity in the social characteristics of the subjects. The study authors indicated that CDC records documented that levels of other contaminants, including arsenic and lead, were very low in the area. Iodine differences were not specifically addressed, but there is no indication from the information provided that this might have been a concern.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Probably low risk of bias because there is direct evidence that the key covariates were considered and indirect evidence that co-exposure to arsenic was likely not an issue in this area and that methods used for collecting the information were valid and reliable.
- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The majority of results were reported for the 51 children stated to be included in the pilot study. In Table 5 of the study, the N for each dental fluorosis category totals only 43, but the text indicates 8 children did not have a Dean Index because permanent teeth had not erupted.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study used three different measurements of fluoride exposure: well water fluoride concentrations from the residence during pregnancy and onwards, fluoride concentrations from children's first morning urine samples, and degree of children's dental fluorosis. Fluoride concentrations in community well water were measured and recorded by Mianning County CDC; specific analytic methods were not reported, but it is likely that standard methods were used because the

analyses were conducted by the CDC and were likely the same as those used to measure the fluoride in urine. Migration of subjects was noted to be limited. Well water fluoride concentrations of the mother's residence during pregnancy and onward were used to characterize a child's lifetime exposure. To provide a measure of the accumulated body burden, each child was given a 330-mL (11.2-oz) bottle of Robust© distilled water (free from fluoride and other contaminants) to drink the night before the clinical examinations, after emptying the bladder and before bedtime. The first urine sample the following morning was collected at home, and the fluoride concentration was determined on a 5-mL sample using an ion-specific electrode at the Mianning CDC. There is no indication that urinary fluoride levels accounted for dilution, nor was it clear that the method of administering water to the children and collection methods sufficiently controlled for differences in dilution. One of the investigators, a dentist, performed a blinded dental examination on each child's permanent teeth to rate the degree of dental fluorosis using the Dean Index. The Dean Index is a commonly used index in epidemiological studies and remains the gold standard in the dentistry armamentarium. The Index has the following classifications: normal, questionable, very mild, mild, moderate, and severe. Quality control (QC) procedures are not reported but were likely appropriate.

        - *Direction/magnitude of effect size:* Current levels were used to assess lifetime exposure. This is likely to be a non-differential exposure misclassification, and direction of bias is unknown. Because subject migration appears to be limited, it is likely that the current fluoride levels are adequate reflections of past exposure. Dental fluorosis would be an indicator that exposure occurred in the past, and there was a fair correlation between degree of dental fluorosis and current urine and water fluoride levels, with both increasing with increasing levels of dental fluorosis.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.

- ***Outcome:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* The study authors adopted culture-independent tests considered feasible for children aged 6 to 8 years. The Wide Range Assessment of Memory and Learning (WRAML) was used for the assessment of memory and learning. Three subtests were also used. The Finger Windows subtest assesses sequential visual memory. The Design Memory subtest assesses the ability to reproduce designs from memory following a brief exposure. The Visual Learning subtest assesses the ability to learn the locations of pictured objects over repeated exposures. The Wechsler Intelligence Scale for Children-Revised (WISC-IV) includes digit span for auditory span and working memory and block design for visual organization and reasoning. The grooved pegboard test assesses manual dexterity. The tests used have been validated on a Western population. Although there is no information provided to indicate that the tests were validated on the study population, the study authors indicated that the tests were culture-

independent (+ for methods). Blinding of the outcome assessors to participants' fluoride exposure, or steps to minimize potential bias were not reported. However, it is unlikely that the assessors had knowledge of the individual exposure as children all came from the same area, and water and urine levels were tested at the CDC. (+ for blinding). Overall = +.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that all outcomes were assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses are appropriate. Multiple regression models evaluate the associations between exposure indicators and test scores after adjusting for covariates. Specific regression models are not described or refenced, just stated to be "standard regression analysis with confounder adjustment." The distributions of fluoride concentrations in urine and water are skewed and log10-transformed to approximate a Gaussian distribution (test not specified). Results are reported as adjusted effects and 95% CIs. There is no evidence that residual diagnostics were used to examine model assumptions; however, the impact on the effect estimates is expected to be minimal.
    - *Other potential concerns:* It should be noted that this study was a pilot study and, therefore, had a relatively small sample size (i.e., 51 children).
  - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in the confounding, exposure, and outcome domains. Study strengths include individual fluoride measurements with blinding at outcome assessment likely. All key covariates and many other important covariates were considered in the study design or analysis.

E.1.3. Cui et al. (2018)

E.1.3.1. Study Details

- ***Study design:*** Cross-sectional

- ***Population:*** School children aged 7–12 years from four schools in two districts in China with different fluoride levels
- ***Study area:*** Jinghai and Dagang in Tianjin City, China
- ***Sample size:*** 323 school children
- ***Data relevant to the review:*** IQ scores by urine fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant association between IQ score and log-transformed urinary fluoride (adjusted β = −2.47; 95% CI: −4.93, −0.01).

E.1.3.2. Risk of Bias

- **Author contacts:**
  - Authors were contacted in June 2019 to obtain additional information for risk-of-bias evaluation.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Four schools were selected from the same district in China. The schools were selected based on levels of fluoride in the local drinking water and the degree of school cooperation. No details were provided on the number of schools in given areas or the difficulty in getting school cooperation. It was noted that the residents in the four areas had similar living habits, economic situations, and educational standards. Although authors do not provide the specific data to support this, fluoride levels and IQ scores were provided by different subject characteristics. The areas were classified as historically endemic fluorosis and non-fluorosis. Cluster sampling was used to select the grades in each school according to previously set child ages, and classroom was randomly selected with all students within a selected classroom included. Reasons for exclusion do not appear to be related to exposure or outcome.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and recruited within the same time frame using the same methods, with no evidence of differences in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The measurements of all covariates were obtained by structured questionnaires that were completed by children with the help of their parents. Covariates that were assessed include: sex, age, child's ethnicity, child's BMI, birth (normal vs. abnormal), mother's age at delivery, mother's education, income per family member, mother's smoking/alcohol during pregnancy, family member smoking, environmental noise, iodine region (non-endemic vs. iodine-excess-endemic area), factory within 30 m of residence, iodine salt, diet supplements, seafood/pickled food/tea consumption, surface water consumption, physical activity, stress, anger, anxiety/depression, trauma, having a cold 5 times a year, thyroid disease in relatives, mental retardation in relatives, and cancer in relatives. Covariates not considered include parity, maternal and paternal IQ, and quantity

and quality of caregiving environment (e.g., HOME score). The authors report that there were no other environmentally toxic substances that might have affected intelligence, such as high arsenic or iodine deficiency according to the Tianjin Centers for Disease Prevention and Control.

- *Potentially important study-specific covariates:* All key covariates were considered in this study.
  - *Direction/magnitude of effect size:* Not applicable.
- *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key covariates were considered, methods for collecting the information were valid and reliable, and co-exposure to arsenic was likely not an issue in this area.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of the 400 children enrolled, 35 were excluded because they did not have informed consent signed by a guardian or they moved out of the area. Forty-two children were excluded because they did not have a DRD2 genotyping measurement. It is unclear whether these children were from the same schools or whether they were evenly distributed throughout the study area. It is also unclear whether the excluded subjects were similar to those included in the study. In the study, some analyses had fewer than the 323 subjects, but this seems reasonable given the subgroups that were being evaluated.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although children were selected based on area fluoride levels, fluoride in the urine was used in the analysis. Urine was collected from each child during the morning of enrollment and analyzed within a week. Fluoride levels were measured using an ion-selective electrode according to the China standard. A brief description of the method was provided, but no QC methods were reported. The study authors did not account for urinary dilution in the spot samples.
    - *Direction/magnitude of effect size:* Not accounting for dilution could cause some exposure misclassification. The direction and magnitude would depend on where the differences occurred.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* IQ was measured by professionals using the Combined Raven's Test–The Rural in China method, which is the appropriate test for the study population

(++ for methods). Blinding or other methods to reduce bias were not reported. Although it was unlikely that the outcome assessor would have knowledge of the child's urine fluoride levels, there was potential that they would know whether the child was from an endemic or non-endemic area if the IQ tests were conducted at the child's school, and there was no information provided on how the IQ tests were administered. Correspondence with the study author noted the cross-sectional nature of the study with outcome and exposure assessed at the same time, making the outcome assessors blind to the exposure status of participants. However, there was still potential for knowledge of the area (+ for blinding).

  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses were appropriate. Multiple linear regression models were applied to evaluate the relationship between urine fluoride levels and IQ scores, accounting for numerous important covariates. The urinary fluoride levels were log-transformed due to a skewed distribution. Residual diagnostics were used to examine model assumptions. Model robustness was tested through bootstrap, sensitivity analysis after excluding potential outliers, and cross-validation techniques. Results are reported as adjusted effects and 95% CIs. The analysis did not account for clustering at the school level or at the grade level (students were from four schools in grades selected via a clustered sampling method). There is no evidence that the sampling strategy was otherwise accounted for via sampling weights. The impact of these factors on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for several important covariates.
      - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate, and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements but is limited by the cross-sectional study design and lack of

accounting for urine dilution. All key covariates were considered in the study design or analysis.

E.1.4. Cui et al. (2020)

E.1.4.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** School children aged 7–12 years
- ***Study area:*** Tianjin City, China (one randomly selected school from each district based on iodine levels in the water), presumably was an expansion of the Cui et al. (2018) study
- ***Sample size:*** 498 school children
- ***Data relevant to the review:*** IQ scores by urine fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: A 2-point decrease in IQ was observed in the highest urinary fluoride group compared to the lowest urinary fluoride group (i.e., 110.00 in ≥2.5-mg/L group versus 112.16 in <1.6-mg/L group); however, the results did not achieve statistical significance based on a one-way ANOVA comparing the three different urinary fluoride categories only.

E.1.4.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for the 2020 publication. Authors were contacted in June 2019 for additional information on the Cui et al. (2018) publication. Information obtained from that correspondence may have been used for additional information in the 2020 publication.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were recruited from 2014 to 2018. One school was selected from each district where water concentrations of water iodine were <10, 10–100, 100–150, 150–300 and >300 µg/L. In each school, classes were randomly sampled for the appropriate age group of 7–12 years old. A table of subject characteristics was provided by IQ. A total of 620 children were recruited, and 122 children who did not have complete information or enough blood sample were excluded. Reasons for exclusion do not appear to be related to exposure or outcome. The characteristics of the 498 included children are presented.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited within the same time frame using the same methods, with no evidence of differences in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably high risk of bias (−)

- *Summary:* It was noted by the study authors that there were no other environmental poisons except water fluoride. Other studies also conducted in this area of China noted specifically that arsenic was not a concern. Iodine was addressed as that was one of the main points of the study. Twenty-one factors (provided in Table 1 of the study) were selected as covariates, and a homemade questionnaire of unspecified validity was used for obtaining the information. It was noted that child age, stress, and anger were significantly associated with IQ although it is unclear whether these varied by fluoride level. However, Cui et al. (2018) indicate that stress and anger were not significantly associated with fluoride, and it was assumed that results would be similar for this study even though more children were included.
- *Potentially important study-specific covariates:* Age (children 7–12 years old)
  - *Direction/magnitude of effect size:* Age is a key covariate for IQ, even in the narrow age range evaluated in this study. The direction of the association may depend on the number of children in each age group within the different urinary fluoride categories; however, these data were not provided. In general, there were fewer subjects ≤9 years of age (i.e., 111) compared to >9 years of age (i.e., 387) with a significantly higher IQ in the ≤9-year-old age group. Therefore, if exposure were higher in the older subjects, this could likely bias the association away from the null.
- *Basis for rating:* Probably high risk of bias because there is indirect evidence that age was not addressed as a key covariate and it may be related to both IQ and exposure.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of the 620 children recruited, 122 (20%) were excluded due to incomplete information or inadequate blood sample. No information was provided to indicate whether there were similarities or differences in the children included versus the children excluded, but exclusion is unlikely to be related to either outcome or exposure.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children's morning urine was collected with a clean polyethylene tube, and fluoride was measured using a fluoride ion-selective electrode following Chinese standard WS/T 89-2015. A brief description was provided, but no QC methods were reported. The study authors do not account for urinary dilution in the spot samples.

        - *Direction/magnitude of effect size:* Not accounting for dilution could cause some exposure misclassification. The direction and magnitude would depend on where the differences occurred.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.
- ***Outcome:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* IQ was measured using the Combined Raven's Test, which is an appropriate test for the study population (++ for methods). Blinding was not mentioned; however, the outcome assessors would not likely have had knowledge of the child's urinary fluoride. Subjects appear to have been recruited based on iodine levels; therefore, it is unlikely that there would have been any knowledge of potential fluoride exposure. Correspondence with the study authors for the Cui et al. (2018) study also indicated that the outcome assessors would have been blind.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:*
        - *Statistical analyses:* One-way ANOVA was used to make comparisons between mean IQ by urinary fluoride levels. Consideration of heterogeneity of variances was not reported. There is no adjustment for covariates or for clustering of children at the school level. There is no evidence that the sampling strategy was otherwise accounted for (i.e., via sampling weights). The impact of these factors on the effect estimates is expected to be minimal given the use of individual-level data. The primary focus of the study was to evaluate associations between IQ and thyroid hormone or dopamine levels (not between IQ and fluoride levels). It should also be noted that more advanced analyses used for thyroid hormone- and dopamine-IQ associations still lacked adjustment for school and accounting for clustering of children from the same school.
        - *Other potential concerns:* None identified.

  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate, and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements, but the study is limited by the cross-sectional study design, lack of accounting for urine dilution, and lack of addressing age as a key covariate.

E.1.5. Ding et al. (2011)

E.1.5.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Elementary school children aged 7–14 years old
- ***Study area:*** Hulunbuir City, Inner Mongolia, China
- ***Sample size:*** 331 school children
- ***Data relevant to the review:*** IQ mean difference based on 10 categories of urine fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant association between urinary fluoride and IQ score (each 1-mg/L increase in urinary fluoride was associated with a decrease in IQ score of 0.59 points; 95% CI: −1.09, −0.08).

E.1.5.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study randomly selected 340 7–14-year-olds from four nearby elementary schools in Hulunbuir. Authors stated that the four elementary schools appeared to be very similar in teaching quality. The study authors noted that they followed the principles of matching social and natural factors like economic situation, educational standards, and geological environments as much as possible; however, how this was done is unclear and no table of study subject characteristics by group was provided.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited within the same time frame using the same methods, with no evidence of differences in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably high risk of bias (−)
  - *Summary:* It was noted that none of the four sites had other potential neurotoxins, including arsenic, in their drinking water. Details were not provided, except for a

reference supporting the statement. In addition, iodine deficiency was noted as not being issue in any of the four areas. Age was the only key covariate adjusted for in the regression model. Although dental fluorosis severity by % female was reported, not enough data were provided to determine whether sex should have been considered in the regression model. The study authors note that future studies will include covariates such as parents' educational attainment, mother's age at delivery, and household income.

- *Potentially important study-specific covariates:* Sex
  - *Direction/magnitude of effect size:* There is not enough information to determine whether there was an effect from sex. There were some differences in dental fluorosis level by sex, but it is unclear how this might impact the results or whether the distribution of sex differed by age.
- *Basis for rating:* Probably high risk of bias based on indirect evidence that there were differences in sex that were not considered in the study design or analyses.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data were relatively complete (i.e., <5% loss). Of the 340 subjects selected for inclusion, 5 were excluded because they lived in the area for less than a year with an additional 4 not consenting to participate.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analysis was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Spot urine samples were collected and measured using China CDC standards. All samples were analyzed twice using a fluoride ion-selective electrode. Recovery rates were specified as 95%–105% with an LOD of 0.05 mg/L. Water samples were collected from small-scale central water supply systems and tube wells with handy pumps and were processed using standard methods, similar to the urine samples. Quality assurance validation was reported. A blind professional examiner evaluated the children for dental fluorosis using Dean's Index. All urine and water samples were above the LOD. Urine levels were the primary exposure measure used in the analysis. The study authors did not account for urinary dilution in the spot samples. The mean urine fluoride concentration was correlated with the dental fluorosis levels.
    - *Direction/magnitude of effect size:* Spot urine samples that did not account for dilution could have exposure misclassification. The misclassification is likely non-differential, and the potential direction of bias is unknown.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* IQ was determined using the Combined Raven's Test–The Rural in China (CRT-RC3) (++ for methods). Although blinding was not reported, it is unlikely that the IQ assessors had knowledge of the children's urine levels or even of the water levels from the four sites, as these were sent to a separate lab for testing (+ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses were reasonable (ANOVA and multiple linear regression), but consideration of homogeneity of variance was not reported. The NASEM committee's review (NASEM 2021) pointed out a potential concern regarding the lack of accounting for clustering at the school level because children were selected from four elementary schools. However, as outlined in the ***Selection*** domain, the authors stated that they followed the principles of matching social and natural factors like economic situation, educational standards, and geological environments to the extent possible and that the four elementary schools appeared to be very similar in teaching quality. There is no evidence that the sampling strategy was otherwise accounted for (i.e., via sampling weights). The impact of these factors on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for age as a key covariate.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and that there were no other potential threats of risk of bias.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements, but the study is limited by the cross-sectional study design, lack of accounting for urine dilution, and lack of consideration of sex as a key covariate.

E.1.6. Green et al. (2019)

E.1.6.1. Study Details

- ***Study design:*** Prospective cohort
- ***Population:*** Maternal-Infant Research on Environmental Chemicals (MIREC) participants (pregnant mothers and their children aged 3–4 years)
- ***Study area:*** 10 cities, Canada
- ***Sample size:*** 512 mother-child pairs (238 from non-fluoridated areas, 162 from fluoridated areas; 264 females, 248 males)
- ***Data relevant to the review:*** Adjusted linear regression models evaluating associations between IQ in both sexes together and separately, with maternal urinary fluoride across all three trimesters or with estimated maternal fluoride intake.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower full-scale IQ per 1-mg/L increase in maternal urinary fluoride in boys (adjusted $\beta = -4.49$) but not girls (adjusted $\beta = 2.40$) and not in both sexes combined (adjusted $\beta = -1.95$); significantly lower full-scale IQ per 1-mg increase in maternal intake in both sexes combined (adjusted $\beta = -3.66$ [no sex interaction]); significantly lower full-scale IQ per 1-mg/L increase in drinking water fluoride in both sexes combined (adjusted $\beta = -5.29$ [no sex interaction]).

E.1.6.2. Risk of Bias

- **Author contacts:**
  - Authors were contacted in June 2019 for additional information for the risk-of-bias evaluation.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Pregnant women were recruited from the same population during the same time frame and using the same methods as the MIREC program. Methods were reported in detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were similar and were recruited with the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study considered several possible covariates, including maternal age, pre-pregnancy BMI, marriage status, birth country, race, maternal education, employment, income, HOME score, smoking during pregnancy, secondhand smoke in the home, alcohol consumption during pregnancy, parity, sex, age at testing, gestational age, birth weight, time of void, and time since last void. The study also conducted secondary analyses to test for lead, mercury, arsenic, and PFOA. There is no indication of any other potential co-exposures in this study population. Iodine deficiency or excess could not be assessed but is not expected

to differentially occur. The study was not able to assess parental IQ or mental health disorders. Methods used to obtain the information included questionnaires and laboratory tests.

- *Potentially important study-specific covariates:* All key covariates were addressed.
  - *Direction/magnitude of effect size:* Not applicable.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and direct evidence that key covariates, including potential co-exposures, were addressed.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of the 610 recruited children, 601 (98.5%) completed testing. Of the 601 mother-child pairs, 512 (85.2%) had all three maternal urine samples and complete covariate data, and 400 (66.6%) had data available to estimate fluoride intake.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Spot urine samples from all three trimesters of pregnancy were evaluated using appropriate methods, and results were adjusted for creatinine and specific gravity. Fluoride intake was estimated based on fluoride water levels, and information on consumption of tap water and other water-based beverages (e.g., tea, coffee) was obtained via questionnaire.
    - *Direction/magnitude of effect size:* There is not any specific direction or magnitude of bias expected. Urinary fluoride levels are reflective of a recent exposure. Having measurements from all three trimesters of pregnancy provides a better representation of actual exposure than a single measurement, although the potential for missed high exposure is possible. However, the possibility of the occurrence of missed high exposure would be similar in all females and would be non-differential. For the fluoride intake, exposure was based on the fluoride levels in the water at the residence. If women worked outside the home and the majority of intake occurred from areas outside the home (and were different from levels in the home), there is potential to bias toward the null.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)

  - *Summary:* The Wechsler Preschool and Primary Scale of Intelligence was normalized for ages 2.5–<4.0 and sex using the U.S population-based norms. Blinding was not reported, but it is unlikely that the outcome assessors had knowledge of the maternal fluoride level or were aware of whether the city had fluoridated water.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes were reported.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Multivariate linear regression analyses were used to evaluate the associations between maternal urinary fluoride and fluoride intake and children's IQ scores. Regression diagnostics were used to test assumptions for linearity, normality, and homogeneity. There were no potential influential observations (based on Cook's distance). Sensitivity analyses showed that the effects of maternal urinary fluoride (MUF), fluoride intake, and water fluoride were robust to the exclusion of two very low IQ scores in males (<70). City was accounted for as a covariate in the regression models published. Additional models with city as a random effect were also subsequently made publicly available and showed similar results to the main model.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, prospective cohort design, and the consideration of key covariates.

E.1.7. Rocha-Amador et al. (2007)

E.1.7.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 6–10 years

- ***Study area:*** Moctezuma (low fluoride, low arsenic) and Salitral (high fluoride, high arsenic) of San Luis Potosí State and 5 de Febrero (high fluoride, high arsenic) of Durango State, Mexico
- ***Sample size:*** 132 children
- ***Data relevant to the review:*** Associations between full-scale IQ, performance IQ, verbal IQ, and child's urine or water fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant associations between log-transformed fluoride and IQ scores (full-scale IQ adjusted βs of −10.2 [water] and −16.9 [urine]; CIs not reported); arsenic also present, but the effect from arsenic was smaller (full-scale IQ adjusted βs of −6.15 [water] and −5.72 [urine]; CIs not reported).

E.1.7.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* All children in 1st through 3rd grades in three rural areas in Mexico (n = 480) were screened for study eligibility, including age, time at residence, and address. Authors report that the three selected communities were similar in population and general demographic characteristics. Children who had lived in the area since birth and were 6–10 years old were eligible to participate (n = 308). Of the 308 children, 155 were randomly selected and the response rate was 85%, but participation was not reported by area. It was noted, however, that no significant differences in age, sex, or time of residence were observed between participants and non-participants. Time frame for selection was not mentioned but appears to be similar. Sociodemographic characteristics of subjects was provided in Table 1 of the study. There was a significant difference in SES and transferrin saturation, but these were considered in the analysis.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the populations were similar, and differences were noted and addressed in the analysis.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study design or analysis accounted for age, sex, SES, transferrin saturation, weight, height, blood lead levels, and mother's education. Arsenic levels were highly correlated with fluoride levels; however, arsenic and fluoride were evaluated alone, and arsenic was found to have less of an effect on IQ than fluoride. This provides evidence that arsenic had been addressed as a co-exposure and cannot explain the association between fluoride exposure and decreased IQ. Smoking was not addressed and methods for measuring many of the covariates were not reported.

- *Potentially important study-specific covariates:* Arsenic
  - *Direction/magnitude of effect size:* The presence of arsenic in this study, which also demonstrated an association, would likely bias the association away from the null. Although arsenic may contribute to some of the magnitude of the observed effect of fluoride (the exact impact of arsenic on the magnitude cannot be assessed), the presence of arsenic does not fully explain the observed association between fluoride exposure and IQ. The presence of arsenic may affect the magnitude of the association between fluoride and IQ, but it has no impact on the direction of the association.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and direct evidence that key covariates were addressed.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of 155 children randomly selected for study participation, 85% responded to enroll. According to the authors, there were no significant differences in age, sex, or time of residence between responders and non-responders. However, no data were provided to support this, and no breakdown of responders/non-responders by region was provided. Data were provided for the 132 children agreeing to participate.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Urine samples were collected on the same day as psychological evaluations and were analyzed for fluoride according to NIOSH Method 8308 (Fluoride in Urine). For QC, a reference standard was also used (NIST SRM 2671a). Urine samples were also analyzed for arsenic by using the Atomic Absorption Spectrophotometer with hydride system and a reference standard for QC. Levels were adjusted for urinary creatinine levels to account for dilution in the spot samples. Tap water samples were collected from each child's home on the day of biological monitoring. Fluoride was measured with a sensitive, specific ion electrode. Detailed methods are provided including internal quality controls. It was noted that in the high fluoride group, it was common to drink bottled water low in fluoride and to use the tap water only for cooking; therefore, urine was considered the most appropriate measure of exposure. Only children who had lived at the same residence since birth were included.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Neuropsychological profiles were assessed through the WISC-RM (revised for Mexico). This is a well-established test appropriately adjusted for the study population. However, no additional validation was provided (+ for methods). The study report stated that the test assessors were masked to both arsenic and fluoride water levels (++ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- ***Selective Reporting:***
  - *Rating:* Probably high risk of bias (−)
  - *Summary:* It was reported that an interaction between fluoride and arsenic was measured, but it was noted only in the discussion that the study design precluded testing statistical interaction between fluoride and arsenic. This provides indirect evidence of selective reporting.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that there was selective reporting.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study. Multivariate linear analyses were used to evaluate the associations between fluoride in water and urine and children's IQ scores. Exposures were natural log-transformed, but the rationale was not provided. Regression diagnostics were not used to test model assumptions for linearity, normality, and homogeneity. The analyses did not account for clustering at the community level. The three selected communities were similar in population and general demographic characteristics. Although the analysis used individual-level exposures rather than area-level exposures, if the exposure levels within a certain area were highly correlated (which might be expected), then the results might still be biased. However, the overall impact on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for multiple important covariates.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include

individual exposure measurements and blinding of outcome assessors to participants' fluoride exposure, but it is limited by the cross-sectional study design and the inability to completely rule out the influence of arsenic in the results.

E.1.8. Saxena et al. (2012)

E.1.8.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 12 years
- ***Study area:*** Madhya Pradesh, India
- ***Sample size:*** 170 children
- ***Data relevant to the review:*** Mean IQ grade (not standard scores; higher IQ grades are associated with lower intelligence) by water fluoride quartiles, continuous water fluoride, or continuous urinary fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant correlations between IQ score and water (r = 0.534) and urinary (r = 0.542) fluoride levels. Significant increase in mean IQ grade (i.e., increase in proportion of children with intellectual impairment) with increasing urinary fluoride in adjusted analyses.

E.1.8.2. Risk of Bias

- **Author contacts:**
  - Authors were contacted in August of 2017 to obtain additional information for risk-of-bias evaluation.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There was indirect evidence that subjects were similar and were recruited using the same methods during the same time frame. The study participants were selected from a stratified cluster of geographic areas based on fluoride concentration in groundwater. According to the authors, the selected villages were similar in population and demographic characteristics. Data are provided to show the breakdown in SES, parental education, height/age, and weight/height, and no significant differences were noted. Participation was stated to be voluntary, but participation rates were not provided. It is unclear whether the 170 subjects were selected with 100% participation or whether the 170 subjects were all who were asked to participate, but it appears that all subjects participated. Timing of the recruitment was not provided but is assumed to occur during the same time frame.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and recruited using the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)

- *Summary:* There was indirect evidence that key covariates, including potential co-exposures, were addressed using reasonable methods. A questionnaire, completed with the assistance of parents, was used to collect information on child characteristics (age, sex, height, weight), residential history, medical history (including illness affecting the nervous system and head trauma), educational level of the head of the family (in years), and SES of the family. The SES was recorded according to the Pareek and Trivedi classification. The nutritional status of the children was calculated using Waterlow's classification, which defines two groups for malnutrition using height-for-age ratio (chronic condition) and weight for height ratio (acute condition). Within both groups, it categorizes the malnutrition as normal, mildly impaired, moderately impaired, or severely impaired. Urinary lead and arsenic were analyzed using the atomic absorption spectrophotometer. Urinary iodine was measured using the Dunn method. Authors do not report which covariates were included in the multivariate regression models; however, there was no difference in reported demographic characteristics. All subjects were the same age, and there was no difference in iodine, lead, or arsenic between the groups. Mean urinary arsenic levels increased with increasing fluoride even though there was no significant difference by group.
- *Potentially important study-specific covariates:* All key covariates were considered in this study.
  - *Direction/magnitude of effect size:* Not applicable.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and that key covariates, including potential co-exposures, were addressed.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results were provided for all 170 children stated to be included in the study.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence of no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* A sample of 200 mL of drinking water was collected at each child's home. The fluoride levels were analyzed by a fluoride ion-selective electrode. Each subject was also asked to collect a sample of his/her first morning urine. The fluoride content in the urine was determined using a fluoride ion-selective electrode. QA/QC and LOD were not reported, and urinary dilution was not assessed. Although only current levels were measured, children who had changed their water source since birth were excluded.
    - *Direction/magnitude of effect size:* Spot urine samples that did not account for dilution could have exposure misclassification. The misclassification is likely non-differential and not likely to bias in any specific direction. Children who had changed water source since birth were excluded, but it was not

specifically noted that the fluoride in the water source was stable over the years.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Intelligence was assessed using Raven's Standard Progressive Matrices and categorized into five grade levels. Although it was not noted that the test was validated to the study population, the test is visual and would be applicable to most populations (+ for methods). There is no mention of blinding by test administrators or evaluators, and the exposure groups come from different geographic areas. It was also not reported who measured the levels of fluoride from the home or urine samples. Correspondence with the study authors indicated that the outcome assessors were blind to the children's fluoride status (++ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* One-way analysis of variance (ANOVA), simple linear regression, and multiple linear regression were used to compare mean intelligence grades by water fluoride levels and to assess the association between grades and urinary fluoride. Consideration of heterogeneity of variance (for ANOVA) was not reported. Regression diagnostics were not used to test model assumptions for linearity, normality, and homogeneity. Given the ordinal nature of the intelligence grade variable (score from 1 to 5), ordinal logistic regression would have been a more appropriate method. There was no adjustment for area-level clustering in multivariate analyses (although subjects were selected via stratified cluster sampling from two areas). Although the analysis used individual-level exposures rather than area-level exposures, if the exposure levels within a certain area were highly correlated (which might be expected), then the results might still be biased. However, the

overall impact on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for important covariates.

- *Other potential concerns:* None identified.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate, and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and the consideration of key covariates, but it was limited by the cross-sectional study design and lack of addressing dilution in the urine samples.

E.1.9. Seraj et al. (2012)

E.1.9.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 6–11 years
- ***Study area:*** five villages, Makoo, Iran
- ***Sample size:*** 293 children
- ***Data relevant to the review:*** IQ (mean and distribution) assessed by Raven's Colored Progressive Matrices and presented by fluoride area; beta was also provided for water fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant association between water fluoride and IQ score (adjusted β per 1-mg/L increase in water fluoride = −3.865; CIs not reported); significantly higher IQ score in normal area (97.77 ± 18.91) compared with medium (89.03 ± 12.99) and high (88.58 ± 16.01) areas.

E.1.9.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were selected from five villages in Makoo. The villages were stated to all be rural with similar general demographic and geographic characteristics and were comparable in terms of SES and parental occupations. Children were 6–11 years old. Age, sex, and education were taken into account in the analysis. No other characteristics were provided or discussed. Participation rates were not reported. There is indirect evidence that the populations were similar, and some possible differences were addressed.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and recruited using the same methods during the same time frame.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Age, sex, dental fluorosis intensity, and educational levels (child's and parents') were evaluated as important covariates. Other covariates such as smoking were not discussed. Information was obtained from a detailed questionnaire. Lead was measured but found only in low levels in the drinking water throughout the study regions. Iodine in the water was also stated to be measured, and residents were receiving iodine-enriched salt. Arsenic was not addressed, but there is no evidence that arsenic levels would vary across villages in this area. Based on water quality maps, co-exposure to arsenic is likely not a major concern in this area.
  - *Potentially important study-specific covariates:* Arsenic.
    - *Direction/magnitude of effect size:* Conceptually, if there were differential amounts of arsenic in the different villages, co-exposure to arsenic could bias the association, with the direction of the bias dependent on where the arsenic was present; however, arsenic was not expected to be a major concern in this study area based on water quality maps.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and that key covariates, including potential co-exposures, were addressed or were not likely to be an issue in the study area.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Attrition was low if it occurred. It was noted that 293 out of 314 children living in the villages were recruited. It is not clear whether 21 children were excluded based on exclusion criteria or whether they refused to participate; however, this accounts for less than 10% of the population, and results were available for all 293 subjects.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was minimal, adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***
  - *Rating:* Probably high risk of bias (−)
  - *Summary:* Exposure was primarily based on area of residence. Fluoride in the groundwater was analyzed by the SPADNS (Sulfophenylazo dihydroxynaphthalene-disulfonate) method, utilizing the 4000 UV-Vis spectrophotometer in the environmental health engineering laboratory of the Public Health School of the Tehran University of Medical Sciences. Specific details were not provided on methods of collection or sample locations or whether

these locations represented the primary sources of drinking water for the subjects. Villages were categorized into normal (0.5–1 ppm), moderate (3.1 ± 0.9 ppm), and high (5.2 ± 1.1 ppm) fluoride based on the mean fluoride content of all seasons presumably for the stated 12-year time period. Subjects were stated to be long-life residents of the village. Dental fluorosis was also measured and increased in severity with fluoride levels; however, all areas had some degree of dental fluorosis. Although authors used an average fluoride level in varying seasons over presumably 12 years, they used a less-established method without reporting reliability or validity, and they did not provide data to indicate that the mean was truly representative of the fluoride levels over time and throughout the village. Although dental fluorosis severity increased with increasing fluoride levels, the data could also indicate potential exposure misclassification.

        - *Direction/magnitude of effect size:* The presence of dental fluorosis in all groups indicates that there may have been different exposures in some children at a younger age. Although there were only about 20 children in the "normal" fluoride group with very mild to mild dental fluorosis, this could bias the results toward the null because those children may have experienced a higher level of fluoride at some point. The other two fluoride groups were exposed to fluoride levels that likely exceeded those in the "normal" fluoride group.
    - *Basis for rating:* Probably high risk of bias based on indirect evidence that exposure was assessed using insensitive methods.

- ***Outcome:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Intelligence was evaluated using Raven's Color Progressive Matrices. This is a well-established method. Although the study authors did not provide data to indicate that the methods were valid in this study population, the test is designed to be culturally diverse (+ for methods). The study report stated that test administrators were blinded to subjects' exposure status (++ for blinding). Overall rating for methods and blinding = +.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that outcomes were assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- ***Selective Reporting:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported. However, because the study author did not report the method for obtaining the betas in Table 4 of the study, it is not clear whether these were adjusted or unadjusted regression coefficients.
    - *Basis for rating:* Probably low risk of bias based on direct evidence that all the study's measured outcomes were reported, but the results were not sufficiently reported.

- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical methods for comparisons of IQ level by exposure groups were reasonable (ANOVA, post hoc test, and Kruskal-Wallis test), but consideration of heterogeneity of variance was not reported. Clustering at the village levels was not accounted for in multivariate analyses, which used area-level water fluoride levels. Because the exposure levels within a certain area are highly correlated (which might be expected), the results are likely to be biased. There was adjustment for some individual-level important covariates, and the children were from five rural areas with similar general demographic and geographic characteristics and were comparable in terms of SES and parental occupations. These factors are expected to mitigate some of the impact of lack of accounting for clustering, and the overall impact on the effect estimates is expected to be minimal.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and outcome. Study strengths include addressing potential key covariates, but it was limited by the cross-sectional study design and the group-level exposure data.

E.1.10. Soto-Barreras et al. (2019)

E.1.10.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 9–10 years
- ***Study area:*** Chihuahua, Mexico
- ***Sample size:*** 161 children
- ***Data relevant to the review:*** Water fluoride, urinary fluoride, exposure dose, and dental fluorosis index by IQ grade.
- ***Reported association with fluoride exposure:*** No: Results were not presented to evaluate an association between fluoride exposure and IQ but to compare fluoride levels within IQ grades. For this reason, the results of this study are not comparable to other studies that evaluated IQ scores by fluoride exposure levels. No significant differences in measured fluoride levels across IQ grades were observed.

E.1.10.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.

- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were selected using a multistage cluster sampling. During the first stage, 13 public elementary schools were randomly selected from a pool of 73 using a cluster sample design. Secondly, only fourth-grade students were included. Authors stated that they wanted to keep the same grade level, but there were no specific details as to why fourth graders were selected as opposed to any other grade. Lastly, only children whose parents or guardians attended and responded to the survey were included. There is no information provided on how the 13 schools selected may have been similar to or different from the 60 schools not selected. There is no information provided on the number of children in the fourth grade to know participant rates. It was only noted that 245 children were examined, but 161 were included after the exclusion rules were applied. Inclusion and exclusion criteria are presented. Reasons for exclusion do not appear to be related to exposure or outcome. Characteristics of participants and non-participants are not compared; however, characteristics of the 161 included children were provided, and any differences were taken into account in the analysis.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposed groups were similar and were recruited using similar methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably high risk of bias (−)
  - *Summary:* No covariates were considered when evaluating associations between fluoride exposure and intelligence; covariates were considered only when evaluating associations between fluoride levels and dental caries. According to Table 4 of the study, there was no significant association between IQ grade and age, sex, parental education, or SES status. No other information was reported or considered. There is no information on potential co-exposures. According to water quality maps, the arsenic prediction indicates a greater than 50% probability of exceeding the WHO guidelines for arsenic of 10 µg/L in areas of Chihuahua, Mexico.
  - *Potentially important study-specific covariates:* Arsenic.
    - *Direction/magnitude of effect size:* The impact on the direction and magnitude of effect size is unknown. There is potential for arsenic to occur in the study area, but it is not known how it relates to fluoride exposure. If they occur together in the water, it would likely bias the association away from the null; however, if they occur in different areas, there is potential to bias the association toward the null.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that there is potential for exposure to arsenic that was not sufficiently addressed.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)

- *Summary:* A total of 161 of 245 children were included in the study. Exclusion criteria are presented and are unrelated to outcome or exposure. For the 161 children, there are no missing outcome data.
- *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***
  - *Rating:* Probably low risk of bias (+); Probably high risk of bias (−)
  - *Summary:* **Urinary Fluoride (probably low risk of bias):** First morning void urine samples were collected based on NIOSH methods. Water samples were also stated to be collected, but it does not appear that methods followed any particular standard, and there is no indication that subjects were provided with collection containers. Analysis was based on a calibration curve using fluoride ion-selective electrode. QC methods were mentioned. Based on results, there were values below detection limits, but LODs or % below LOD were not reported.

    Daily fluoride exposure (probably high risk of bias): Daily fluoride exposure was based on the water fluoride level, drinking water consumption (based on parental report of how many glasses of water consumed), and body weight.

    - *Direction/magnitude of effect size:* Spot urine samples that did not account for dilution could have exposure misclassification. The misclassification is likely non-differential and is not likely to bias in any specific direction. Daily exposure was based partially on parental report of water consumption. The direction and magnitude of effect is unknown.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure. The daily fluoride exposure is probably high risk of bias because there is indirect evidence that the exposure was assessed using methods of unknown validity.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Intellectual ability was evaluated using Raven's Colored Progressive Matrices by an independent examiner. Some details were provided, but it was not stated that the tests were assessed blind; however, there is no indication that subjects were from high fluoride areas, and the assessor would not have knowledge of the urine or water fluoride levels. Results for children were converted into a percentile according to age (details not provided), and overall scores were assigned an intellectual grade of I to V as described in the report.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* The Kolmogorov-Smirnov test was used to determine variable distribution. The Kruskal-Wallis test was used to compare exposure levels between IQ grades with Dunn's post hoc test. Multivariate logistic regression was used to estimate the association between presence of dental caries and various risk factors. Fluoride levels in drinking water and urine and fluoride exposure dose were compared across intellectual grades. Children were from 13 schools selected via stratified cluster sample design. There was no adjustment for clustering at the school level or for the sampling design. Although the analysis used individual-level exposures rather than area-level exposures, if the exposure levels within a certain school were highly correlated (which might be expected), then the results might still be biased. The large number of clusters (13 schools) makes clustering less of a concern, and the impact on the effect estimates is expected to be minimal.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements and blinding of outcome assessor to participants' fluoride exposure, but it is limited by the cross-sectional study design, lack of accounting for urine dilution, and lack of consideration for potential exposures to arsenic in the study area. Although the study is considered to have low potential for bias overall, the focus of the study was to evaluate the relationship between fluoride exposure and lower rates of dental caries. In terms of evaluating an association between fluoride exposure and IQ scores, the study is limited by the way the data were reported.

E.1.11. Sudhir et al. (2009)

E.1.11.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 13–15 years
- ***Study area:*** Nalgonda district (Andhra Pradesh), India

- ***Sample size:*** 1,000 children
- ***Data relevant to the review:*** Mean IQ grade (not standard scores) or IQ distribution by water fluoride strata (<0.7, 0.7–1.2, 1.3–4.0, and >4.0 ppm).
- ***Reported association with fluoride exposure:*** Yes: Significant increase in mean and distributions of IQ grades (i.e., increase in proportion of children with intellectual impairment) with increasing drinking water fluoride levels.

E.1.11.2. Risk of Bias

- **Author contacts:**
  - Authors were contacted in September of 2017 for additional information related to risk-of-bias evaluation, but no response was received.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children were selected from the same general population during the same time frame and were then broken down into nearly equal exposure groups. A cross-sectional study was conducted among 13–15-year-old school children of Nalgonda district, Andhra Pradesh, between August and October 2006. Data were collected from the school children who were lifelong residents of Nalgonda district, Andhra Pradesh, and who consumed drinking water from the same source during the first 10 years of life. A stratified random sampling technique was used. The entire geographical area of Nalgonda district was divided into four strata based on different levels of naturally occurring fluoride in the drinking water supply. Children were randomly selected from schools in the different strata. It was noted that the 1,000 selected children were equally divided among all four strata; however, each group did not have 250 children (rather, each had 243–267). Participation rates were not reported. Exclusion criteria included children who had a history of brain disease and head injuries, children whose intelligence had been affected by congenital or acquired disease, children who had migrated or were not permanent residents, children with orthodontic brackets, and children with severe extrinsic stains on their teeth. Age and sex data are presented in Table 1 of the study, but this information is not presented by the different fluoride groups.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and were recruited using the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Data were collected using a self-administered questionnaire and clinical examination. The questionnaire requested information on demographic data (appears to cover age and sex), permanent residential address, staple food consumed, liquids routinely consumed, and aids used for oral hygiene maintenance (fluoridated or non-fluoridated). SES was measured using the Kakkar socioeconomic status scale (KSESS) with eight closed-ended questions

related to parental education, family income, father's occupation, and other factors. All children were asked to fill out the form, and the answers obtained were scored using Kakkar socioeconomic status scoring keys. Based on this scoring, children were divided into three groups: lower class, middle class, or upper class. Age, sex, and SES were not found to be significantly associated with IQ. Other covariates, including smoking, were not addressed. Co-exposures such as arsenic and lead were not addressed; however, there is no indication that lead is a co-exposure in this population, and arsenic is not likely a major concern in this area based on water quality maps.

- *Potentially important study-specific covariates:* Key covariates age, sex, and measures of SES were similar between exposure groups; however, arsenic was not considered. Arsenic often occurs in the drinking water along with fluoride in some Indian populations; however, based on water quality maps, this does not appear to be an issue in the Nalgonda district of Andhra Pradesh. Iodine deficiencies are not mentioned.
  - *Direction/magnitude of effect size:* Conceptually, the presence of arsenic would potentially bias the association away from the null if present with fluoride. Deficiencies in iodine would likely bias the association away from the null if present in areas of high fluoride but toward the null if present in areas of non-high fluoride. Neither of these were considered issues in this study for reasons noted above.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the key covariates were considered, co-exposure to arsenic was likely not an issue in this area, and methods used for collecting the information were valid and reliable.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results were available for the 1,000 children selected to participate.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence of no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children were placed into one of four strata based on the level of fluoride in drinking water. Collection of water samples was done in the districts. The placement into strata was based on fluoride levels obtained from documented records of the District Rural Water Works Department. Once the children were assigned to strata, it was confirmed that the fluoride level of their drinking water was within the strata assigned. This was done using the methodology followed in the National Oral Health Survey and Fluoride Mapping 2002–2003. During the initial visits to the schools, the children were interviewed regarding their history of residence and source of drinking water from birth to 10 years. The first child meeting the criteria was given a bottle for water collection, and the next child was given a bottle for collection only if the water source was different from that of a previous child. Children were asked to collect a water sample from the source that

was used in the initial 10 years of their life (and that sample was collected the next day). It was not reported whether all bottles were returned. The water samples collected were subjected to water fluoride analysis using an ion-specific electrode, Orion 720A fluoride meter at District Water Works, Nalgonda to confirm the fluoride levels in the water before commencement of clinical examination. LOD and QA/QC details were not reported.

        - *Direction/magnitude of effect size:* There is some potential for exposure misclassification based on recall of the children on the source of water used in their first 10 years of life. The misclassification is likely non-differential and not likely to bias in any specific direction. Children who had changed water since birth were excluded, but it was not specifically noted that the fluoride in the water source was stable over the years.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
    - *Rating:* Probably high risk of bias (NR)
    - *Summary:* Raven's standard progressive matrices (1992 edition) was used to assess IQ. Raven's test is a standard test; although there is no information provided to indicate that the methods were reliable and valid in this study population, the test was created to be culturally fair (+ for methods). Blinding or other methods to reduce potential bias were not reported (NR for blinding). No response was received to an email request for clarification in September 2017. Overall rating for methods and blinding = NR.
    - *Basis for rating:* Probably high risk of bias based on indirect evidence that the outcome assessors were not blind to participants' fluoride exposure and could bias the results.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:*
        - *Statistical analyses:* Chi-square test and Spearman rank correlation were used to assess the association between four different fluoride levels and IQ grades. Area-level exposures were used. Clustering of children within the four areas was not accounted for in the analysis; however, because multiple villages were included in each fluoride exposure level, clustering was less of a concern and the impact on the effect estimates was expected to be minimal.

    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include verification of exposure measurements and consideration of key covariates, but it was limited by the cross-sectional study design and lack of information on blinding during outcome assessment.

E.1.12. Till et al. (2020)

E.1.12.1. Study Details

- ***Study design:*** Prospective cohort
- ***Population:*** MIREC participants (pregnant mothers and their children aged 3–4 years)
- ***Study area:*** 10 cities, Canada
- ***Sample size:*** 398 mother-child pairs (247 from non-fluoridated areas, 151 from fluoridated areas; 200 breastfed as infants, 198 formula-fed as infants)
- ***Data relevant to the review:*** Adjusted linear regression models evaluating associations between IQ and water fluoride concentration (with or without adjusting for maternal urine) in formula-fed or breastfed infants or fluoride intake from formula.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower performance IQ with water fluoride per 0.5-mg/L increase by breastfeeding status (adjusted βs = −9.26 formula-fed, −6.19 breastfed) and fluoride intake from formula (adjusted β = −8.76); significantly lower full-scale IQ with water fluoride per 0.5-mg/L increase in formula-fed children (adjusted β = −4.40); no significant changes in full-scale IQ for water fluoride in breastfed children or fluoride intake from formula-fed children; no significant changes in verbal IQ scores with fluoride exposure.

E.1.12.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Pregnant women were recruited between 2008 and 2011 by the MIREC program from 10 cities across Canada. Inclusion and exclusion criteria were provided. Additional details were stated to be available in Arbuckle et al. (2013). A total of 610 children were recruited to participate in the developmental follow-up with 601 children completing all testing. The demographic characteristics of women included in the current analyses (n = 398) were not

substantially different from the original MIREC cohort (n = 1,945) or the subset without complete water fluoride and covariate data (n = 203). A table of characteristics of the study population was provided. Approximately half of the children lived in non-fluoridated cities and half lived in fluoridated cities.

- *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were similar and were recruited with the same methods during the same time frame.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariates were selected a priori that have been associated with fluoride, breast feeding, and children's intellectual ability. Final covariates included sex and age at testing, maternal education, maternal race, secondhand smoke in the home, and HOME score. City was considered but excluded from the models. Covariates that were not assessed include parental mental health, iodine deficiency/excess, parental IQ, and co-exposure to arsenic and lead. Co-exposure to arsenic is less likely an issue in this Canadian population because it receives water mainly from municipal water supplies that monitor for lead and arsenic, and the lack of information is not considered to appreciably bias the results. In addition, a previous study on this population (Green et al. 2019) conducted sensitivity analyses on co-exposures to lead and arsenic. Results from these sensitivity analyses support the conclusion that co-exposures to lead and arsenic are not likely a major concern in this study population.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect:* Not applicable.
  - *Basis for rating:* Probably low risk of bias based on direct evidence that key covariates were considered and indirect evidence that the methods used to collect the information were valid and reliable and co-exposures were not an issue.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of 610 children, 601 (98.5%) in the MIREC developmental study who were ages 3–4 years completed the neurodevelopment testing. Of the 601 children who completed the neurodevelopmental testing, 591 (99%) completed the infant feeding questionnaire and 398 (67.3%) reported drinking tap water. It was noted that the demographic characteristics were not substantially different from the original MIREC cohort or the 203 subjects without complete water fluoride or covariate data.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)

  - *Summary:* Information on breastfeeding was obtained via questionnaire at 30–48 months. Fluoride concentration in the drinking water was assessed by daily or monthly reports provided by water treatment plants. Water reports were first linked with mothers' postal codes, and the daily or weekly amounts were averaged over the first 6 months of each child's life. Additional details can be found in Till et al. (2018). Maternal urinary exposure was used to assess fetal fluoride exposure. Procedures can be found in Green et al. (2019).
    - *Direction/magnitude of effect size:* There is not any specific direction or magnitude of bias expected. Urinary fluoride levels are reflective of recent exposure. The possibility of exposure misclassification would be similar in all subjects and would be non-differential. For the fluoride intake from formula, exposure was based on the fluoride levels in the water at the residence and the proportion of time that the infant was not exclusively breastfed. This exposure misclassification would also be non-differential.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Intelligence was tested using the Wechsler Preschool and Primary Scale of Intelligence III, which is considered a gold standard test. It is appropriate for both the study population and age group. It was not reported whether the evaluators were blind to the child's fluoride exposure status during the assessment. Although it is unlikely that the assessors had knowledge of the specific drinking water levels or maternal urine levels, there is potential that the outcome assessors had knowledge of the city the child lived in and whether the city was fluoridated or non-fluoridated. Correspondence with the study authors on the outcome assessment for Green et al. (2019) indicated that it was unlikely that the testers had knowledge of the city's fluoridation. The same is assumed here. Specific measurements included were identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)

- *Summary:*
  - *Statistical analyses:* Regression diagnostics were used to test assumptions for linearity, normality, and homogeneity. There were two potential influential observations (based on Cook's distance), and sensitivity analyses re-estimated the models without these two variables. Effect modification by breastfeeding status was evaluated. Interestingly, all regression coefficients were divided by 2 to represent change in IQ per 0.5-mg/L change in fluoride. One concern is posed by the lack of accounting for city in the regression models, ideally as a random effect. The authors explored including city as a covariate in the models; however, city was not included either because it was strongly multi-collinear with water fluoride concentration (VIF > 20) (model 1, with water fluoride concentration) or because fluoride intake from formula is a function of water fluoride concentration (assessed at the city level) and was therefore deemed redundant (model 2). However, the models use city-level water fluoride concentrations—and, in sensitivity analyses, adjust for maternal urinary fluoride—which warrants exploration of city as a random effect rather than a fixed effect (as would be the case by having it included as a covariate). Even including individual-level maternal urinary fluoride might not fully account for lack of a city effect, given that the subjects were from six different cities, with half of them fully on fluoridated water. Hence, even individual-level exposures are likely to be correlated at the city level. Based on a previous analysis (Green et al. 2019), it is unlikely that exclusion of city from models (as a fixed or random effect) would significantly impact the effect estimates.
  - *Other potential concerns:* None identified.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, prospective cohort design, and consideration of key covariates.

E.1.13. Trivedi et al. (2012)

E.1.13.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 12–13 years
- ***Study area:*** Kachchh, Gujarat, India
- ***Sample size:*** 84 children
- ***Data relevant to the review:*** Mean IQ scores and distribution by low and high fluoride villages.

- ***Reported association with fluoride exposure:*** Yes: Significantly lower mean IQ score in the high fluoride villages (92.53 ± 3.13) compared with the low-fluoride villages (97.17 ± 2.54) in boys and girls combined (and by sex).

E.1.13.2. Risk of Bias

- **Author contacts:**
  - Authors were contacted in September of 2017 to obtain additional information for risk-of-bias evaluation.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There is insufficient information provided on the sampling methods to determine whether the populations were similar. Although it was noted that samples were obtained for groundwater quality from March to May of 2011, there is no indication that the children were selected at the same time or during a similar time frame. Correspondence with the author indicates that children were selected within a week of the water collection based on random selection of a school in the village. Study participants were selected from six different villages of the Mundra region of Gujarat, India. Subjects were grouped into high and low villages based on the level of fluoride in the drinking water of those villages. The number of subjects per village was not reported, but it was noted that there were 50 children in the low-fluoride group and 34 children in the high fluoride group. It is not clear whether the differences in numbers were based on different participation rates or whether there were fewer children in the high fluoride villages. Recruitment methods, including any exclusion criteria and participation rates, were not provided. SES was stated to be low and equal based on questionnaire information, but the results were not provided. It should also be noted that only regular students (having attendance more than 80%) of standard 6th and 7th grades were selected, but it was not noted whether attendance varied by village. Correspondence with the study author indicated that there was an average of 20 students per class with an average of 40 students per village. It appears that keeping the requirement of 80% attendance was a limiting factor that resulted in different numbers of children by area; however, this was applied similarly to both groups.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and recruited using the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children were stated to be students of the 6th and 7th standard grades. Age was not addressed, but the children would all be of similar ages based on the grades included. Results were reported for males and females separately as well as combined. SES and iodine consumption were stated to be analyzed via a questionnaire and were standardized on the basis of the 2011 census of India. Although it was noted in the abstract that the SES was equal (no data provided),

the study report did not mention the iodine results. Although arsenic and lead were not considered, the study authors provided physicochemical analyses for the water samples from the six different villages. While the authors did not specifically analyze lead or arsenic in the water samples, these physicochemical analyses suggest that differential lead or arsenic exposure was unlikely. Moreover, based on water quality maps, arsenic was not expected to be a major concern in this study area. According to the information from the water quality maps and the physiochemical analysis of the water provided, there is indirect evidence that neither arsenic nor lead were a concern in this study population.

- *Potentially important study-specific covariates:* Key covariates age, sex, and measures of SES were similar between exposure groups; however, arsenic was not considered. Arsenic often occurs in the drinking water along with fluoride in some Indian populations; however, based on water quality maps, arsenic does not appear to be an issue in the study area.
  - *Direction/magnitude of effect size:* Conceptually, the presence of arsenic would potentially bias the association away from the null if present with fluoride, or toward the null if present in the reference group; however, for reasons noted above, arsenic is not considered a concern in this study population.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable, that potential co-exposures were not an issue, and that key covariates were addressed.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results were provided for 84 children, but the methods do not indicate how many children were initially selected to participate, nor were any exclusion criteria provided. It was noted in the results that 84 children had their groundwater and urine tested, but it was not noted whether analyses were restricted to these children or whether exposures were assessed in all the children who had IQ measurements. Correspondence with the study author indicated that the main reason for exclusion was a <80% attendance rate, with fluoride and IQ measured on all 84 children who met the criteria.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence of no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children in villages were grouped based on fluoride levels that were assessed in groundwater (low fluoride villages versus high fluoride villages). The average concentration of these levels was considered to be the levels in the drinking water with confirmation using urinary fluoride levels. The groundwater samples were selected to cover major parts of the taluka and represent overall groundwater quality. Ten samples were obtained from each village. Fluoride was measured in the groundwater using ion exchange chromatography. Although urine

levels were also significantly higher in the high fluoride village, no information was provided on how or when the urinary samples were obtained or how they were measured. However, correspondence with the study author indicated that the groundwater and urine fluoride levels were available for all 84 children, indicating that the urine measures were available for the children that had IQ measures. The urine samples were stated to be collected at the same time the second water sample was collected.

    - *Direction/magnitude of effect size:* Fluoride levels were measured in both the drinking water and urine. Although there is some variability in the measurements, there is no overlap between the two groups, and the urine and drinking water levels in the children support each other. Any potential exposure misclassification would be non-differential, and the impact on the direction and magnitude of the effect size is unknown.

  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Outcome methods were only noted to be reported in Trivedi et al. (2007), which was scored as follows: IQ was measured in the children of both areas using a questionnaire prepared by Professor JH Shah, copyrighted by Akash Manomapan Kendra, Ahmedabad, India, and standardized on the Gujarati population with a 97% reliability rate in relation to the Stanford-Binet Intelligence Scale (+ for methods). Blinding or other methods to reduce bias were not reported, but correspondence with the study author indicated that the teachers were blind to the status of fluoride. The teachers administered the tests in the presence of a research fellow. It is not completely clear who scored the tests, but it is assumed the teachers (+ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcomes were assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably high risk of bias (−)
  - *Summary:*

        - *Statistical analyses:* Mean IQ scores in low and high fluoride villages were compared using a t-test. Consideration of heterogeneity of variances was not reported. Results are reported as means and standard errors of the means, with p-values for significant differences. Area-level exposures were used. There was no accounting for clustering of children within the villages, and comparative analyses did not account for covariates. Urinary fluoride was not considered in the comparative analyses. The lack of individual exposure levels and the lack of accounting for clustering are likely to bias the standard error of the difference in mean IQ levels between the high- and low-fluoride villages and make the differences appear stronger than they actually are.
    - *Basis for rating:* Probably high risk of bias based on indirect evidence that the statistical analyses did not account for clustering, and this lack of accounting could bias the association. There were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and the addressing of potential key covariates, but the study was limited by the cross-sectional study design. Another limitation was the lack of accounting for clustering, which may bias the standard error of the differences, making the effect appear stronger than it actually is; however, this does not change the nearly 5-point difference in IQ scores between the two villages.

E.1.14. Wang et al. (2012)

E.1.14.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 8–13 years [possibly the same study population as Xiang et al. (2003a)]
- ***Study area:*** Wamiao and Xinhuai villages located in Sihong County, Jiangsu Province, China
- ***Sample size:*** 526 school children
- ***Data relevant to the review:*** Mean IQ and % low IQ (<80) by total fluoride intake.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower mean IQ in the endemic versus non-endemic regions, as reported in Xiang et al. (2003a); when the high-exposure group was broken into four exposure groups based on fluoride intake, a dose-dependent decrease in IQ and increase in % with low IQ was observed; significant correlation between total fluoride intake and IQ ($r = -0.332$); for IQ <80, adjusted OR of total fluoride intake per 1 mg/(person/day) was 1.106 (95% CI: 1.052, 1.163).

E.1.14.2. Risk of Bias

- **Author contacts:**
    - Authors were not contacted for additional information because it was not necessary.

- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study appears to have the same study population as Xiang et al. (2003a) and Xiang et al. (2011); however, it does not cite these studies as providing additional information, and the numbers of children differ; therefore, it may be a separate analysis on the same villages. The years of testing were not provided, so it cannot be determined whether study subjects were the same. Two villages, Wamiao and Xinhuai, located 64 km apart in Sihong County, Jiangsu Province, were selected for the study. Wamiao is a village in a region with severe endemic fluorosis, and Xinhuai is a village in a non-endemic fluorosis region. Neither village has fluoride pollution from coal or industrial sources. Villages were stated to be similar in terms of annual per capita income, transportation, education, medical conditions, natural environment, and lifestyle. All primary students ages 8–13 years currently in school in either village were surveyed with exclusions noted. Of 243 children from Wamiao, 236 (97.12%) were included, and of 305 children from Xinhuai, 290 (95.08%) were included. No table of subject characteristics was provided.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited using the same methods within the same time frame, with direct evidence that there was no difference in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Logistic regression of low IQ rate and total fluoride intake adjusted for age and sex. Both villages had hand-pumped well water for drinking water, but the authors do not mention whether arsenic was also present in the drinking water. However, a publication by Xiang et al. (2013) in the same study areas indicates that Xinhuai (the low-fluoride area) had significantly higher arsenic levels compared with Wamiao (the endemic fluorosis area), which would bias the association toward the null. Areas were stated to be similar in annual per capita income, transportation, education, medical conditions, natural environment, and lifestyle; however, no details were provided. This study did not address other co-exposures, but other studies on populations in these villages (Xiang et al. 2003a; Xiang et al. 2011) indicate that iodine and lead are not concerns.
  - *Potentially important study-specific covariates:* Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on information provided in Xiang et al. (2013), arsenic concentrations were higher in the low-fluoride area compared with the high fluoride area. Because there were significant effects on IQ observed in the high fluoride areas, the impact of co-exposure to arsenic is less of a concern. The presence of arsenic in the control village may cause an underestimation of the effect of fluoride, but despite this potential impact, a significant association between fluoride exposure and IQ was reported.

        - *Direction/magnitude of effect size:* Presence of arsenic in this study population would potentially bias the association toward the null.
    - *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key covariates were considered, methods used for collecting the information were valid and reliable, and co-exposures to arsenic and lead and iodine deficiency were not attributing to the association observed in this study. The potential for bias toward the null combined with the reported significant association increases confidence in the observed effect.

- ***Attrition:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Data are reported for all 526 children noted to be included in the study. There is a slight discrepancy in the reported total number of children from the high-fluoride village and the number of participants from the high-fluoride village between this paper (236 participated of 243 total children) and the 2003 and 2011 publications on the same study population (222 of 238). This discrepancy is not explained but is not expected to appreciably bias the results.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***
    - *Rating:* Probably low risk of bias (+); Probably high risk of bias (−)
    - *Summary:* **Water fluoride (+ probably low risk of bias):** Exposure was based on drinking water levels and fluoride intake. Residents in the Wamiao village were divided into five groups based on fluoride levels in the drinking water. Clean, dry polyethylene bottles were used to collect 50 mL of drinking water from each student's household, and fluoride content was measured.

      Total fluoride intake (− probably high risk of bias): Six families from each of the five Wamiao groups were randomly selected as dietary survey households. Intakes of various foods by each person at each meal and intakes of unboiled water, boiled water, and tea were surveyed for four consecutive days. Methods for food collection were described. Five representative households from each village were selected based on geographic location, population distribution, housing structure, and other conditions. Indoor air samples were collected once daily for five consecutive days; outdoor air was sampled at two points once daily for five days. Methods for determining fluoride content in samples were noted to follow specific guidelines. Calculation of total fluoride intake was stated to follow Appendix A of the People's Republic of China Health Industry Standard with some details provided. Although it is assumed the method is valid, it was not detailed how each fluoride determination was made for each subject, and it appears that total fluoride intake was determined based on data from select subjects and not all subjects.

        - *Direction/magnitude of effect size:* There is potential for exposure misclassification based on calculating fluoride intake based on measurements from a few select subjects rather than all subjects. The potential impact on the direction and magnitude of effect size cannot be assessed based on the information provided.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure. The total fluoride intake is probably high risk of bias because there is indirect evidence that the exposure was assessed using methods of unknown validity.
- ***Outcome:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* The IQ of each child was measured with the Combined Raven's Test for Rural China (CRT-RC) (++ for methods). The test was stated to be administered to the children independently in a school classroom under the supervision of three exam proctors. Testing methods, testing language, and testing conditions were all in strict accordance with the CRT-RC guidebook. Major testing personnel received necessary training by the Psychology Department of East China Normal University. The children undergoing IQ testing and the test scorers were kept double-blind throughout the testing process (++ for blinding). Overall rating = ++.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:*
        - *Statistical analyses:* Logistic regression analysis was used to determine the odds of having low IQ with increasing fluoride intake. Analyses and methods are not well described. There is no mention of what tests were used for the mean IQ comparison by village; however, statistical software (SPSS) was used, suggesting appropriate tests were applied. Simple linear regression analyses were conducted to evaluate associations between total fluoride intake and children's IQ or low IQ rate. There is no evidence that regression diagnostics were used to test model assumptions for linearity, normality, and homogeneity. Clustering at the village level was not accounted for in the

analyses. The overall impact of these factors on effect estimates is expected to be minimal given the use of individual-level data and adjustment for important covariates.

- *Other potential concerns:* None identified.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements with blinding at outcome assessment, but is limited by the cross-sectional study design and lack of individual measurements to calculate fluoride intake. All key covariates were accounted for in the study design or analysis, but there is potential for the presence of arsenic to bias the association toward the null.

E.1.15. Wang et al. (2020b)

E.1.15.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** School children aged 7–13 years
- ***Study area:*** Tianjin City, China [possibly a subset of the children from Yu et al. (2018)]
- ***Sample size:*** 571 school children
- ***Data relevant to the review:*** IQ scores by urine and water fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant associations between IQ score and water fluoride (adjusted $\beta = -1.587$ per 1-mg/L increase) and urinary fluoride (adjusted $\beta = -1.214$ per 1-mg/L increase) in boys and girls combined based on both quartiles and continuous measures. No significant modification effect of sex.

E.1.15.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were from a cross-sectional study conducted in 2015, but no citation was provided on this cohort [presumably the Yu et al. (2018) cohort]. It was noted that the subjects in that cohort were from districts with historically high or normal fluoride levels. Subjects for this study were selected by using a stratified and multistage random sampling approach. Brief description was provided. The study area consisted of three historically high fluoride areas and four non-endemic areas. A flow diagram was provided for inclusion and exclusion, but this detail was given for all children and not by area. Therefore, it

cannot be determined whether the participation differed by area. However, there was a 93% recruitment rate, and the 13 excluded due to missing data were not likely excluded due to exposure. Detailed characteristics of the study population are provided. Exclusion criteria included: "children who had congenital or acquired diseases affecting intelligence, or a history of cerebral trauma and neurological disorders, or those with a positive screening test history (like hepatitis B virus infection, Treponema palladium infection and Down's syndrome) and adverse exposures (smoking and drinking) during maternal pregnancy, prior diagnosis of thyroid disease, and children who had had missing values of significant factors (2.2%) were also excluded."

- *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were recruited using similar methods during the same time frame and that any differences between the exposed groups were accounted for in the statistical analyses.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Study authors noted that the study areas were not exposed to other neurotoxins such as lead, arsenic, or mercury nor were they iodine-deficient. Final models included age, sex, child's BMI, maternal and paternal education, household income, and low birth weight. The other covariates that were considered are unclear as the authors only noted that the covariates were selected based on current literature. Reasons for exclusion included history of disease affecting intelligence, history of trauma or neurological disorders, positive screening test history, or exposures such as smoking or drinking during pregnancy. Information was obtained by questionnaire or measurements. Covariates such as parental BMI, behavioral and mental health disorders, IQ, and quantity and quality of the caregiving environment were not considered.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Probably low risk of bias because there is direct evidence that the key covariates were considered and indirect evidence that the methods for collecting the information were valid and reliable and that co-exposure to arsenic was not an issue in this area.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* A detailed chart of the recruitment process is presented. The study had a 93% recruitment rate, and only 2.2% of subjects with missing data for certain covariates were excluded.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children provided spot urine samples, presumably at the time of examination. Water samples were randomly collected from public water supplies in each village. Fluoride concentrations were analyzed using fluoride ion-selective electrode according to the national standardized method in China. There is no indication of whether the urine samples accounted for dilution.
    - *Direction/magnitude of effect size:* Not accounting for dilution could cause some exposure misclassification. The impact on the direction and magnitude of effect size would depend on where the differences occurred.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Assessments of IQ scores were conducted by graduate students at the School of Public Health, Tongji Medical College, Huazhong University of Science and Technology. Each team member was assigned a single task, meaning that only one person would have conducted the IQ tests. A Combined Raven's Test for Rural China was used. Therefore, the test was appropriate for the study population (++ for method). It was noted that the examiner was trained and blind to the exposure (++ for blinding). Overall = ++
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Logistic and multivariate regression models accounting for covariates were used. Results are presented as betas or ORs and 95% CIs. Regression diagnostics were conducted for all models, including examination of multicollinearity, heteroscedasticity, and influential observations. Mediation and interaction analyses were appropriate. There is no evidence that the stratified and multistage random sampling approach for subject selection was accounted for in the analyses by using sampling weights or

accounting for clustering using random effect models; however, selected villages were similar in population and general demographic characteristics. Given the use of individual-level data and adjustment for important covariates, the impact on the regression coefficients is likely to be minimal.

- *Other potential concerns:* None identified.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and no other potential threats of risk of bias were identified.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key covariates were considered in the study design or analysis.

E.1.16. Xiang et al. (2003a)

E.1.16.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 8–13 years
- ***Study area:*** Wamiao and Xinhuai villages located in Sihong County, Jiangsu Province, China
- ***Sample size:*** 512 school children
- ***Data relevant to the review:*** Comparison of IQ (mean and distribution) between Wamiao County (a severe endemic fluorosis area) and Xinhuai County (a non-endemic fluorosis area); additional breakdown of the Wamiao area into five water fluoride exposure groups.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower IQ scores observed with water fluoride levels of 1.53 mg/L or higher. Percentage of subjects with IQ scores below 80 was significantly increased at water fluoride levels of 2.46 mg/L or higher. Significant inverse correlation between IQ and urinary fluoride ($r = -0.164$). Mean IQ scores for children in the non-endemic region (100.41 ± 13.21) were significantly higher than the endemic region (92.02 ± 13.00).

E.1.16.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Two villages, Wamiao and Xinhuai, located 64 km apart in Sihong County, Jiangsu Province, were selected for this study, which was conducted between September and December 2002. Wamiao is located in a severe fluorosis endemic area, and Xinhuai is located in a non-endemic fluorosis area. Neither

village has fluoride pollution from burning coal or other industrial sources. All eligible children in each village were included; children who had been absent from either village for 2 years or longer or who had a history of brain disease or head injury were excluded. In Wamiao, 93% of the children (222 out of 238) were included in the study; in Xinhuai, 95% were included (290 out of 305). The children in Wamiao were divided into five subgroups according to the level of fluoride in their drinking water: <1.0 mg/L (group A), 1.0–1.9 mg/L (group B), 2.0–2.9 mg/L (group C), 3.0–3.9 mg/L (group D), and >3.9 mg/L (group E). Children in Xinhuai (0.18–0.76 mg/L in the drinking water) served as a control group (group F). Demographic characteristics are not presented, and statistical analyses are not adjusted, but mean IQ scores are stratified by age, sex, family income, and parental education.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited using the same methods within the same time frame, with direct evidence that there was no difference in participation/response rates.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although information was stated to be collected on personal characteristics, medical history, education levels of the children and parents, family SES, and lifestyle, only sex, age, family income, and parental education were considered. Potential co-exposures, such as arsenic, were not addressed. A separate publication in 2003 [(Xiang et al. 2003b), letter to the editor] indicated that blood lead levels were not significantly different between the two areas. Although arsenic was not addressed specifically in this publication, Xiang et al. (2013) measured both fluoride and arsenic in the Wamiao and Xinhuai areas. Xinhuai (the low-fluoride area) had significantly higher arsenic levels compared with Wamiao (the endemic fluorosis area). This is likely to bias the association toward the null; however, the study observed a significantly lower IQ score in the endemic fluorosis area. Iodine was tested in a subset of the children and found not to be significantly different between the two groups.
  - *Potentially important study-specific covariates:* Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on information provided in Xiang et al. (2013), arsenic concentrations were higher in the low-fluoride area compared with the high fluoride area. Because there were significant effects on IQ observed in the high fluoride areas, the impact of co-exposure to arsenic is less of a concern. The presence of arsenic in the control village may cause an underestimation of the effect of fluoride, but despite this potential impact, there was still a significant association between fluoride exposure and IQ.
    - *Direction/magnitude of effect size:* Presence of arsenic in this study population would potentially bias the association toward the null.
  - *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key covariates were taken into account, methods used for collecting the

information were valid and reliable, and co-exposures to arsenic and lead and iodine deficiency were not attributing to the effect observed in this area. The potential for bias toward the null, combined with the reported significant association increases confidence in the observed effect.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are complete. IQ results were reported for all 512 children included in the study (222 in the endemic area and 290 in the nonendemic area).
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Exposure was based on drinking water and urinary levels of fluoride. The two study areas were selected to reflect a severe endemic area and a non-endemic area. Drinking water was collected from wells, and early-morning spot urine samples were collected from a randomly selected subsample of children. Both water and urine samples were measured using fluoride ion-selective electrode, but no quality control was discussed. Both absolute and creatinine-adjusted urine results were reported.
    - *Direction/magnitude of effect size:* There is potential for exposure misclassification because only current levels were assessed. Migration of subjects in or out of the area was not assessed, but the study authors noted that if the children had been absent from the village for 2 or more years, they were excluded. Misclassification would likely be non-differential, which could likely bias the association in either direction.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* The IQ of each child was measured with the Combined Raven's Test for Rural China (CRT-RC) (++ for methods). The test was stated to be administered to the children independently in a school classroom, in a double-blind manner, under the supervision of an examiner and two assistants, and in accordance with the directions of the CRT-RC manual regarding test administration conditions, instructions to be given, and test environment (++ for blinding). Overall rating = ++
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.
- ***Selective Reporting:***

- *Rating:* Definitely low risk of bias (++)
- *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
- *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.

- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* There is no mention of the tests conducted, but data were stated to be analyzed using SAS, suggesting appropriate tests were applied. Results provided in the tables indicate that t-tests comparing IQ values between the villages (overall and by sex) were conducted, but it was not reported that heterogeneity of variance was assessed. In addition, correlations between IQ and age, family income, and parents' education level were tested with Pearson's correlation. There is no evidence that a test for trend was conducted to evaluate the stated "significant inverse concentration-response relationship between the fluoride level in drinking water and the IQ of children."
    - A potential concern raised by the NASEM (2020) committee's review was the lack of accounting for relationships in exposure between persons from the same village. Given only two villages were included and the analyses consisted of village-level comparisons (no use of individual-level covariate data), it is likely that the standard error of the difference in mean IQ between fluoride in water exposure groups will be biased, making differences appear stronger than they actually are. Without controlling for village effects and given the large differences in fluoride concentrations and IQ levels between villages, the apparent dose-response relationship could be due to a village effect in addition to a fluoride effect. However, a dose-response relationship is apparent within the "exposed" village, diminishing the concern for a village-only effect and likely minimizing the impact on the effect estimates.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that statistical analyses were appropriate and that there were no other threats of risk of bias.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and blinding of outcome assessor to exposure but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key covariates were considered in the study design or analysis, but there is potential for the presence of arsenic to bias the association toward the null.

E.1.17. Xiang et al. (2011)

E.1.17.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 8–13 years [same study population as Xiang et al. (2003a)]
- ***Study area:*** Wamiao and Xinhuai villages located in Sihong County, Jiangsu Province, China
- ***Sample size:*** 512 school children
- ***Data relevant to the review:*** Mean IQ scores and odds ratio for having an IQ <80 presented by serum fluoride quartiles.
- ***Reported association with fluoride exposure:*** Yes: Significant linear trend across quartiles of serum fluoride and children's IQ score <80 (adjusted ORs for Q1 and Q2; Q1 and Q3; and Q1 and Q4, respectively: 1; 2.22 [95% CI: 1.42, 3.47]; and 2.48 [95% CI: 1.85, 3.32]); significant effects observed at ≥0.05 mg/L serum fluoride.

E.1.17.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study population was the same as that used in the Xiang et al. (2003a) study, but a few more measurements were available and different analyses were conducted. The comparison population was considered the same based on the study populations being recruited from similar populations, using similar methods, during the same time frame. Demographic characteristics were not provided.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited using the same methods within the same time frame, with direct evidence that there was no difference in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* As was noted in the 2003 publication (Xiang et al. 2003a), information was collected on personal characteristics, medical history, education levels in the children and parents, family SES, and lifestyle. In the logistic regression model age and sex were adjusted for in the analysis. In the previous report, no significant associations were observed between groups for family income and parents' education (Xiang et al. 2003a). Urinary iodine and blood lead levels were also stated to be measured and were noted not to be significantly different between the groups. Although the iodine levels were reported in the previous publication, the

lead levels were not and neither were the methods. Lead information is reported in a letter to the editor (Xiang et al. 2003b) and was not significantly different between the areas. Although arsenic was not addressed specifically in this publication, Xiang et al. (2013) measured both fluoride and arsenic in the Wamiao and Xinhuai areas. Xinhuai (the low-fluoride area) had significantly higher arsenic levels compared with Wamiao (the endemic fluorosis area). This is likely to bias the association toward the null; however, the study observed a significantly lower IQ score in the endemic fluorosis area and with increasing serum fluoride.

- *Potentially important study-specific covariates:* Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on information provided in Xiang et al. (2013), arsenic concentrations were higher in the low-fluoride area compared to the high fluoride area. Because there were significant effects on IQ observed in the high fluoride areas, the impact of co-exposure to arsenic is less of a concern. The presence of arsenic in the control village may cause an underestimation of the effect of fluoride, but despite this potential impact, there was still a significant association between fluoride exposure and IQ.
  - *Direction/magnitude of effect size:* Presence of arsenic in this study population would potentially bias the association toward the null.
- *Basis for rating:* Probably low of risk bias because there is indirect evidence that the key covariates were considered, methods used for collecting the information were valid and reliable, and co-exposures to arsenic and lead and iodine deficiency were not attributing to the effects observed in this area. The potential bias toward the null, combined with the reported significant association increases confidence in the observed effect.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are reported for all 512 children noted to be included in the study.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Fluoride levels were measured in serum with a fluoride ion-selective electrode. A fasting venous blood sample was used. No details are provided on validation (including correlation with drinking water levels) or QA. Children who did not reside in their village for at least 2 years were excluded. Results were provided in quartiles, but the authors combined the lower two quartiles. After combining the two lower quartiles, the exposure levels ranged from <0.05 mg/L (Q1 + Q2) to >0.08 mg/L (Q4).
    - *Direction/magnitude of effect size:* Serum fluoride may not be the best estimate for exposure. There is potential for exposure misclassification because only current levels were assessed. Migration of subjects in or out of

the area was not assessed, but the study authors noted that if the children had been absent from the village for 2 or more years, they were excluded. Misclassification would likely be non-differential, which could bias results in either direction.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* IQ was assessed as part of the 2003 evaluation. IQ was measured with the Combined Raven's Test for Rural China, which is appropriate for this population (++ for methods). Although this study does not provide details, the original study article from 2003 provides specific details. The study authors indicate in the 2003 publication that the tests were conducted in a double-blind manner, and these are the same results and population (++ for methods). Overall rating = ++
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses conducted were appropriate for the study. Chi-square tests were used to compare categorical variables, and multiple logistic regression was used to evaluate the association between serum fluoride levels and risk of low IQ. A potential concern raised by the NASEM (2020) peer review was the lack of accounting for relationships in exposure between persons from the same village. Although only two villages were included, in the analyses that consisted of village-level comparisons, it is likely that the standard error of the difference in mean IQ between villages is biased. This is less of a concern for the mean IQ comparisons across quartiles of serum fluoride levels and for the logistic regression analyses of risk of low IQ and individual-level serum fluoride levels. Without controlling for village effects and given the large differences in fluoride concentrations and IQ between villages, the apparent dose-response relationship could be due to a village effect in addition to a fluoride effect. However, the dose-response

relationship is still present within the "exposed" village, diminishing the concern for a village-only effect and likely minimizing the impact on the effect estimates.

- *Other potential concerns:* None identified.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements with blinding at outcome assessment but is limited by the cross-sectional study design and use of serum concentrations. All key covariates were considered in the study design or analysis, but there is potential for the presence of arsenic to bias the association toward the null.

E.1.18. Yu et al. (2018)

E.1.18.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 7–13 years
- ***Study area:*** Tianjin City, China
- ***Sample size:*** 2,886 school children
- ***Data relevant to the review:*** IQ for normal (≤1 mg/L) versus high (>1 mg/L) water fluoride; betas for IQ score by water and urine fluoride groupings; ORs by IQ category using water and urine fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant difference in mean IQ scores in high water fluoride areas (>1.0 mg/L; 106.4 ± 12.3 IQ) compared to the normal water fluoride areas (≤1.0 mg/L; 107.4 ± 13.0). Distribution of IQ scores was also significantly different (p = 0.003). Every 0.5-mg/L increase in water fluoride (between 3.40 and 3.90 mg/L) was associated with a 4.29 decrease in IQ score (95% CI: −8.09, −0.48).

E.1.18.2. Risk of Bias

- **Author contacts:**
  - Authors were contacted in September 2018 to obtain additional information for the risk-of-bias evaluation.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* School children (2,886), aged 7–13 years, were recruited from the rural areas of Tianjin City, China. After exclusion, 1,636 children were assigned to the "normal-fluoride" exposure group, and 1,250 were assigned to the "high-fluoride" exposure group based on a cut-off water fluoride level of 1.0 mg/L. A multistage random sampling technique, stratified by area, was performed to select representative samples among local children who were permanent residents since

birth. Detailed characteristics of the study population were provided. Exclusion criteria included: 1) children who had congenital or acquired diseases affecting intelligence, 2) children with a history of cerebral trauma and neurological disorders, 3) children with a positive screening test history (like hepatitis B virus infection, Treponema palladium infection and Down's syndrome), and 4) children with adverse exposures (smoking and drinking) during maternal pregnancy. A table of characteristics was provided by fluoride level with differences adjusted in the analysis.

- *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were recruited using similar methods during the same time frame and that any differences between the exposed groups were considered in the statistical analyses.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Demographic data were collected by trained investigators during a face-to-face interview with the recruited children and their parents. Questionnaires were not stated to be validated. The developmental status of the children was further assessed by calculation of BMI, and all measurements were conducted by nurses based on recommended standard methods. Variables that presented differential distribution between the normal-fluoride and high-fluoride exposure groups were adjusted in the linear regression analysis of IQ data and included age, sex, paternal and maternal education levels, and low birth weight. Children exposed to smoking in utero were excluded from the study. Sensitivity analyses were conducted by modifying covariates adjusted in multivariable models among demographics (age and sex); development (BMI); socioeconomics (maternal education, paternal education, and household income); history of maternal disease during pregnancy (gestational diabetes, malnutrition, and anemia); and delivery conditions (hypoxia, dystocia, premature birth, post-term birth, and low birth weight). None of the study sites selected were in areas endemic for iodine deficiency disorders, nor were other potential neurotoxins like lead, arsenic, and mercury present. Variables such as parental BMI and behavioral and mental health disorders were not addressed.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that methods of obtaining the information were valid and reliable and direct evidence that all key covariates and co-exposures were considered.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There were 1,636 children assigned to the “normal-fluoride” exposure group based on water fluoride and 1,250 children assigned to the “high-fluoride” exposure group. Exclusion from the original group of 2,886 children was

adequately described. A total of 2,380 children provided urine samples. There is no indication that the data presented excludes any additional children or urine samples, but results do not indicate a sample size for all results.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* According to the annual surveillance data from the CDC, the drinking water sources and water fluoride concentrations in each village had remained at stable levels over the past decade. During the investigation, water samples were collected randomly from the public water supplies in each village. Spot (early-morning) urine samples from every child and water samples from each village were collected in pre-cleaned, labeled polythene tubes and transported to the lab within 24 hours while frozen. Samples were stored at −80°C until analysis. Concentrations of fluoride ions (mg/L) were analyzed using the national standardized ion-selective electrode method in China; the detection limit was 0.01 mg/L. Samples were diluted with an equal volume of total ionic strength adjusted buffer (TISAB) of pH 5–5.5 for optimal analysis. Double-distilled deionized water was used throughout the experiment. There is no reporting of any QC methods.
    - *Direction/magnitude of effect size:* Spot urine samples may lead to non-differential exposure misclassification. The large population size likely dilutes any potential effects of occasional misclassification. Because the drinking water sources of fluoride had been noted to be stable for the past decade and the children were 13 years or younger, there would only be exposure misclassification if there was a lot of migration between areas.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* IQ scores were measured using the second edition of the Combined Raven's Test–The Rural in China (CRT-RC2) for children aged 7–13 years (++ for methods). The test was completed by each participant within 40 minutes, according to the instruction manual. For each test, 40 children were randomly allocated to one classroom to take the test independently under the supervision of four trained professionals. There is no mention of whether the evaluators were blinded to the fluoride group of each child (normal vs. high fluoride) or whether there were steps taken to ensure consistency in scoring across the evaluators. It is also not clear whether the 40 children randomly assigned to the classroom were specific to the village or whether a local center was used. Correspondence with the study authors indicated that the four professionals worked together throughout

the examination without knowledge of the child's fluoride exposure (++ for blinding).

- *Basis for rating:* Definitely low risk of bias based on the direct evidence that the outcome was assessed using instruments that were valid and reliable, and that the outcome assessors were blind to participants' fluoride exposure.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study. Univariate and multivariable piecewise linear regression models were used to estimate the associations between water fluoride or urinary fluoride levels and IQ scores. Multiple logistic regression analysis was used to evaluate the association between water or urinary fluoride levels and IQ degree using the normal intelligence group as the control. Sensitivity analyses were conducted. There is no evidence that residual diagnostics were used to examine model assumptions or that the complex sampling design (stratified multistage random sampling) was accounted for in the analysis using sampling weights and adjustment for clustering. The impact of these factors on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for numerous important covariates.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements with blinding at outcome assessment but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key covariates, including potential co-exposures, were considered in the study design or analysis.

E.1.19. Zhang et al. (2015b)

E.1.19.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 10–12 years

- ***Study area:*** Tianjin City, China
- ***Sample size:*** 180 children
- ***Data relevant to the review:*** IQ by control and high fluoride groups; IQ correlations with water, serum, or urinary fluoride levels; betas for IQ with urinary fluoride levels (by genotypes)
- ***Reported association with fluoride exposure:*** Yes: Significant correlation between IQ score and children's serum fluoride ($r = -0.47$) and urinary fluoride ($r = -0.45$); significant difference in mean IQ score for high-fluoride area (defined as >1 mg/L in drinking water; 102.33 ± 13.46) compared with control area (<1 mg/L; 109.42 ± 13.30).

E.1.19.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were similar and recruited during the same time frame using the same methods. Authors recruited schoolchildren from a high fluoride area (1.40 mg/L) and a control area (0.63 mg/L) in Tianjin City, China. In accordance with the principles of matching social and natural factors such as educational standard, economic situation, and geological environments as much as possible, two areas with different fluoride concentrations in the groundwater were selected by a stratified cluster random sampling of this region. A total of 180 5th grade children aged 10 to 12 years from two primary schools located 18 km apart in the Jinnan District were recruited—Gegu Second Primary School (from an endemic fluorosis area) and Shuanggang Experimental Primary School (from a non-endemic fluorosis area). The areas are not affected by other drinking water contaminants, such as arsenic or iodine. All subjects were unrelated ethnic Han Chinese and residents in Tianjin with similar physical and mental health status. The authors excluded subjects with known neurological conditions, including pervasive developmental disorders and epilepsy. Descriptive statistics of the study population are presented by exposure group in Table 1 of the study. A number of potential differences were considered in the statistical analyses.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposure groups were similar and recruited using similar methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariates included in the statistical models were age, sex, educational levels of parents, drinking water fluoride (mg/L), and levels of thyroid hormones (T3, T4, and TSH). Authors report that the study areas were not affected by other contaminants such as arsenic or iodine, and residents were of similar physical and

mental health status. Other important covariates (maternal demographics, smoking, reproductive health) were not considered. Covariate data were obtained from a study questionnaire.

- *Potentially important study-specific covariates:* All key covariates were considered in this study.
  - *Direction/magnitude of effect size:* Not applicable.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and direct evidence that key covariates, including potential co-exposures, were considered.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results are complete for the 180 children selected for the study.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Drinking water samples (10 mL) were collected from the tube wells of each child's household. Three fasting venous blood samples were also collected. Urine samples were collected in the early morning before breakfast. Fluoride content in drinking water (W-F), serum (S-F), and urine (U-F) was measured using an ion analyzer EA940 with a fluoride ion-selective electrode (Shanghai Constant Magnetic Electronic Technology Co, Ltd, China), according to the China standard GB 7484-87. All reference solutions for the fluoride determinations were double-deionized water. Parallel samples were set for determination, and averages were taken. The quantitation limits of this method for W-F, S-F, and U-F were 0.2, 0.012, and 0.5 mg/L, respectively. Recovery rates for this method were in the range of 94.3%–106.4%. The intra- and inter-assay coefficients of variation for fluoride were 2.7% and 6.7%, respectively. Dilution of the urinary fluoride was not addressed.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* A Combined Raven's Test for Rural China (CRT-RC) was taken to evaluate the IQ of each child (++ for methods). The study report stated that all tests were administered at school by a trained examiner who was masked to participants' drinking water fluoride levels (++ for blinding). Overall rating for methods and blinding = ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study

population, and that the outcome assessor was blind to participants' fluoride exposure.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All results outlined in the abstract, introduction, and methods sections were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Associations between serum and urinary fluoride levels and IQ score were estimated using general linear models and multivariate linear regression by COMT polymorphism. Normality (Kolmogorov-Smirnov test) was evaluated for all continuous variables. There is no evidence that residual diagnostics were used to examine model assumptions or that the complex sampling design (stratified multistage random sampling) was accounted for in the analysis using sampling weights and adjustment for clustering. The impact of these factors on the regression effect estimates is expected to be minimal given the use of individual-level data and adjustment for numerous covariates.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, blinding of outcome assessor to participants' fluoride exposure, and consideration of key covariates including potential co-exposures.

E.2. Other Neurodevelopmental Studies

E.2.1. Barberio et al. (2017b)

E.2.1.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Canadian Health Measures Survey (cycles 2 and 3) participants (children aged 3–12 years)
- ***Study area:*** general population of Canada
- ***Sample size:*** 2,221 children (1,120 from Cycle 2, 1,101 from Cycle 3)

- ***Data relevant to the review:*** Associations between learning disability or ADHD (Cycle 2 only) assessed by parent or child self-report and urinary fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant increase in adjusted OR for learning disability with unadjusted urinary fluoride per 1-µmol/L increase (1.02; 95% CI: 1.00, 1.03) when Cycles 2 and 3 were combined. No significant associations with creatinine-adjusted or specific gravity-adjusted urinary fluoride. No significant association between urinary fluoride and ADHD.

E.2.1.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* The comparison groups were selected from Cycles 2 and 3 of the Canadian Health Measures Survey. This is a nationally representative sample of residents living in 10 provinces, with clear exclusion criteria provided. Exclusion represented only about 4% of the target population (all Canadian residents 3–79 years old living in 10 provinces). A table of characteristics of the study population is provided.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the subjects were recruited from the same population using the same methods during the same time frame, and exposure groups were similar.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study adjusted for sex, age (3–12 years old), household education, and household income adequacy. Variables to discern fluoride source, including drinking water and dental products, were also considered. Cycle 2 data also included adjustments for: 1) children for whom tap water (vs. bottled or other) was the primary source of drinking water at home or away from home and 2) children who had lived in their current home for 3 or more years. Covariates such as parental behavioral and mental health disorders, smoking, and nutrition were not discussed. The study used data from the Canadian Health Measures Survey, which consists of a nationally representative sample of Canadians. Most Canadians (~89%) receive water from municipal water supplies, which monitor for levels of lead and arsenic. Therefore, co-exposure to lead and arsenic are less likely an issue in this population and the lack of information is not considered to appreciably bias the results.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect size:* Not applicable.

- *Basis for rating:* Probably low risk of bias based on direct evidence that key covariates were addressed and indirect evidence that the methods used to collect the information were valid and reliable and that co-exposures were not an issue.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariate data were missing for less than 5% of all analyses, apart from household income; household income was reported for only 71%–77% of participants and was imputed for the remainder.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Estimates of urinary fluoride (µmol/L) from spot urine were available for a subsample of respondents. Analysis was performed under standardized operating procedures at the Human Toxicology Laboratory of the Institut National de Santé Publique du Québec (accredited under ISO 17025). Fluoride content of urine samples was analyzed using an Orion pH meter with a fluoride ion-selective electrode with limits of detection of 20 µg/L (Cycle 2) and 10 µg/L (Cycle 3). Urinary dilution was addressed by using creatinine-adjusted levels as well as specific gravity-adjusted levels. In Cycle 3 only, estimates of the fluoride concentration of tap water samples collected from randomly selected households were available. The subsample of households selected for tap water sample collection corresponded to the person-level urine fluoride subsample. Analysis of the fluoride concentration of tap water was performed using a basic anion exchange chromatography procedure, with a limit of detection of 0.006 mg/L. QC methods were not addressed.
    - *Direction/magnitude of effect size:* There is not any specific impact on the direction or magnitude of effect size expected. Urinary fluoride levels are reflective of a recent exposure. Having a single concurrent measurement may not be reflective of the exposure associated with the outcome, but if subjects lived in the same area throughout life, the exposure may be an adequate representation. Although there is possible exposure misclassification, it would likely be non-differential.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Probably high risk of bias (−)
  - *Summary:* The primary outcome variable, diagnosis of a learning disability by a health professional, was based on a single item from a household survey asked to all respondents: “Do you have a learning disability?” Answer options were: “yes,”

"no," "don't know," or the participant refused to answer. For Cycle 2, those who indicated having a learning disability were also asked what kind, with the answer options of: "ADD," "ADHD," "dyslexia," or "other." This question was omitted in Cycle 3, and the reason for omission was not described. Parents or guardians answered all questions for children aged 3–11 years, while children 12 years and older answered questions themselves. The self-reporting of a learning disability did not appear to have been confirmed by medical records or a health professional (− for methods based on self-report of diagnosis by a health care professional; also, in Cycle 3, no specific disabilities were described). Blinding was not a concern as spot urine samples were sent to a separate lab, and self-reports would not have knowledge of their urine or tap water exposure level (+ for blinding). Overall rating = −.

- *Basis for rating:* Probably high risk of bias based on indirect evidence that the outcome was measured using an insensitive method in the study population.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods sections were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Logistic regression analyses, adjusted and unadjusted for covariates, examined the associations between fluoride exposure and diagnosis of learning disability. Analyses were performed for Cycle 2 only (urinary fluoride and type of learning disability diagnosis), Cycle 3 only (urinary fluoride, water fluoride, and learning disability diagnosis), and Cycles 2 and 3 combined. Analyses used survey weights and bootstrapped weights to ensure proper computation of variance estimates. Results are reported as unadjusted and adjusted ORs with 95% CIs.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include individual exposure measurements and the consideration of key covariates, but was limited by the cross-sectional study design and insensitive outcome measures.

E.2.2. Bashash et al. (2017)

E.2.2.1. Study Details

- ***Study design:*** Prospective cohort
- ***Population:*** Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants (pregnant mothers and their children aged 4 or 6–12 years).
- ***Study area:*** Mexico City, Mexico
- ***Sample size:*** 299 mother-child pairs, of whom 287 had data for the general cognitive index (GCI).
- ***Data relevant to the review:*** Adjusted and unadjusted associations between GCI and maternal or child's urinary fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant association between maternal urinary fluoride and GCI score (adjusted β per 0.5 mg/L increase = −3.15; 95% CI: −5.42, −0.87). No significant associations with children's urinary fluoride.

E.2.2.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Study participants were selected from two different cohorts from three hospitals in Mexico City that serve low-to-moderate income populations. One cohort was from an observational study of prenatal lead exposure and neurodevelopmental outcomes, and the other was from a randomized trial of the effect of calcium on maternal blood lead levels. The authors state that participants had no history of psychiatric disorders, high-risk pregnancies, gestational diabetes, illegal drug use, or continuous prescription drugs, but information on smoking habits was not included. Study populations appear to be similar, but there may be some differences because subjects were selected from two different cohorts that were recruited during slightly different time periods.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar despite the subjects coming from different original study populations for whom different methods were used for recruitment.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Data were collected via questionnaire on maternal age, education, marital status at first prenatal visit, birth order, birth weight, gestational age at delivery, maternal smoking, maternal IQ, and HOME scores. All models were adjusted for gestational age at birth, sex, birth weight, birth order, age at testing, maternal marital status, smoking history, maternal age at delivery, maternal IQ, education, and cohort, with additional testing for children's urinary fluoride,

mercury, lead, and calcium. Sensitivity analyses were additionally adjusted for HOME score. Covariates not considered included BMI, iodine deficiency, arsenic, and maternal mental health and nutrition. Arsenic is assumed not to be a potential co-exposure in this population as the study authors did not discuss it as an issue but did discuss other co-exposures. Arsenic is included in the water quality control program in Mexico City and is not considered a concern in this population.

- *Potentially important study-specific covariates:* All key covariates were addressed.
  - *Direction/magnitude of effect size:* Not applicable.
- *Basis for rating:* Probably low risk of bias based on direct evidence that key covariates, including other potential co-exposures, were considered, and indirect evidence that the methods used to collect the information were valid and reliable and that arsenic was not likely to be an issue in this study population.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although there was a large amount of attrition, the study authors clearly describe all reasons for attrition and also provide characteristics to compare those participants included to those excluded. There were some slight differences between those included and those excluded, but there is nothing to indicate that the attrition would potentially bias the results.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Urinary fluoride concentrations were determined in spot urine samples (2nd morning void) collected from mothers (during at least one trimester) and children ages 6–12 years. Fluoride content was measured using ion-selective electrode-based assays. QC methods were described including between laboratory correlations. All samples were measured in duplicate. Extreme outliers were excluded. Urinary dilution was addressed by using creatinine-adjusted levels.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Outcome was assessed using the McCarthy Scales of Children's Abilities (MSCA) in 4-year-old children (translated into Spanish) and the Wechsler Abbreviated Scale of Intelligence (WASI) in 6–12-year-olds. The

WASI is a well-established test, and the validity of both tests is well documented by the authors. Inter-examiner reliability was evaluated and reported with a correlation of 0.99 (++ for methods). The study report stated that psychologists were blind to the children's fluoride exposure (++ for blinding). Overall rating for methods and blinding = ++.

- *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study. Statistical tests of bivariate associations (using Chi-square tests for categorical variables and analysis of variance [ANOVA]) were used to compare the means of the outcomes or exposures within groups based on the distribution of each covariate. Generalized additive models (GAMs) were used to estimate the adjusted association between fluoride exposure and measures of children's intelligence. Residual diagnostics were used to examine model assumptions and identify any potentially influential observations. Results are reported as adjusted effects and 95% CIs. In sensitivity analyses, regression models accounted for clustering at the cohort level by using cohort as a fixed effect in the models. Although using cohort as a random effect would be more appropriate, using individual-level exposure data and accounting for numerous covariates in the models likely captured the cohort effect.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, blinding of outcome assessor to participants' fluoride exposure, and the prospective cohort study design.

E.2.3. Bashash et al. (2018)

E.2.3.1. Study Details

- ***Study design:*** Prospective cohort
- ***Population:*** ELEMENT participants (pregnant mothers and their children aged 6–12 years)
- ***Study area:*** Mexico City, Mexico
- ***Sample size:*** 210 mother-child pairs
- ***Data relevant to the review:*** Associations between ADHD and other attention/impulsivity scores and maternal urinary fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant associations between maternal urinary fluoride (per 0.5-mg/L increase) and Conners' Rating Scales-Revised (CRS-R) scores, including Cognitive Problems and Inattention Index (adjusted $\beta = 2.54$; 95% CI: 0.44, 4.63), DSM-IV Inattention Index (adjusted $\beta = 2.84$; 95% CI: 0.84, 4.84), DSM-IV ADHD Total Index (adjusted $\beta = 2.38$; 95% CI: 0.42, 4.34), and ADHD Index (adjusted $\beta = 2.47$; 95% CI: 0.43, 4.50).

E.2.3.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Participants were a subset of mother-child dyads enrolled in various longitudinal birth cohort studies of the Early Life Exposure in Mexico to Environmental Toxicants (ELEMENT) project. Subjects were included from two of the four cohorts for which maternal urinary samples were available. Participants in cohort 2A were recruited between 1997 and 1999, and participants in cohort 3 were recruited from 2001 to 2003. Inclusion and exclusion criteria were applied consistently across the two cohorts. A table of subject characteristics was provided in the study, and any differences were considered in the analysis. Study populations appear to be similar, but there may be some differences because subjects were selected from two different cohorts: one from an observational study on prenatal lead exposure and the other from a randomized trial on the effects of calcium on blood lead levels. In addition, they were recruited from slightly different time periods.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposed groups were similar, and any differences were considered in the analysis.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Questionnaires were used to collect information on maternal age, maternal education, history of smoking, and marital status during the first

pregnancy visit. Child information at birth included birth weight, sex, birth order, and gestational age as calculated by the nurse. Mothers also responded to an SES questionnaire during the visit when the psychometric tests were administered. The Home Observation for Measurement of the Environment (HOME) score was evaluated in a subset of participants. Covariates were selected a priori. Models were adjusted for maternal age at delivery, years of education, marital status, smoking history, gestational age at birth, age at outcome assessment, sex, birth order, SES, cohort, and calcium intervention. Arsenic is included in the water quality control program in Mexico City and is not considered a concern in this population.

- *Potentially important study-specific covariates:* None identified, although this study did not specifically address arsenic or other co-exposures. Bashash et al. (2017) addressed potential co-exposure to lead and mercury but did not address arsenic. Arsenic was potentially addressed as part of the water quality program in Mexico City.
  - *Direction/magnitude of effect size:* Not applicable.
- *Basis for rating:* Probably low risk of bias based on direct evidence that key covariates were addressed, and indirect evidence that the methods used to collect the information were valid and reliable and that arsenic and other potential co-exposures were not likely to be an issue in this study population.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although there was a large amount of attrition from the original cohorts, it was unlikely related to outcome or exposure, and there were very little missing data from those included in the study. Of the 231 mothers with a minimum of one maternal urine fluoride measurement and matching outcome identified for the project, only 17 were excluded based on incomplete demographic and outcome information.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Mothers provided at least one spot urine sample during pregnancy. As described in Bashash et al. (2017), urinary concentrations were determined on second morning void. Fluoride content was measured using ion-selective electrode-based assay. Bashash et al. (2017) describes QC methods. All samples were measured in duplicate, and extreme outliers were excluded. Urinary dilution was addressed by using creatinine-adjusted levels.
    - *Direction/magnitude of effect:* N/A

  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Behaviors associated with ADHD were assessed using the Spanish version of Conners' Rating Scales-Revised, which has been validated for the evaluation of ADHD. Mothers completed the CRS-R at the same follow-up visit in which the child completed the CPT-II tests. All tests were applied under the supervision of an experienced psychologist (++ for methods). Use of only parent reports and not teacher reports was noted by the authors as a study limitation because there is considerable variation between the two sources in terms of identifying ADHD-associated behaviors. Blinding was not reported, but it is unlikely that the mothers were aware of their urinary fluoride levels. Although mothers may have had knowledge that they were receiving fluoride through fluoridated salt or naturally occurring fluoride in their water, they would not have knowledge that this was relevant to the study purpose as the ADHD tests were conducted for the original cohort (as was acknowledged by the study authors in the discussion) (++ for blinding). Overall rating = ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Bivariate analyses included Chi-square tests for categorical variables and ANOVA for continuous outcomes. Appropriate univariate statistics and transformations were performed before bivariate analyses. Residuals from fully adjusted linear regressions were checked and suggested skewness. Gamma regression with an identity link was used to examine the adjusted association between prenatal fluoride and each neurobehavioral outcome (instead of using log transformation). Generalized additive models were used to visually examine potential non-linearity. Sensitivity analyses examined impact of other covariates. Diagnostics tests were used to assess violations of the model assumptions and to identify

remaining influential observations. The Benjamini-Hochberg false discovery rate (FDR) procedure was used to correct for multiple testing.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, blinding of outcome assessor to participants' fluoride exposure, and the prospective cohort study design.

E.2.4. Choi et al. (2015)

E.2.4.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** First-grade children (ages 6–8 years)
- ***Study area:*** Mianning County in southern Sichuan, China
- ***Sample size:*** 51 first-grade children
- ***Data relevant to the review:*** Associations between learning, memory, visual motor ability, motor ability, and manual dexterity with continuous urine or drinking water fluoride levels. Study also had information based on dental fluorosis score.
- ***Reported association with fluoride exposure:*** No: None of the outcomes were significantly associated with fluoride exposure.

E.2.4.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were selected during the same time frame using the same methods. Fifty-one first-grade children residing in Mianning County in southern Sichuan, China were included in this pilot study. It is not specified whether the 51 children represented all first-grade children from this area or whether some refused to participate. Children who did not speak Chinese, were not students at the Primary School of Sunshui Village in Mianning County, or those with chronic or acute disease that might affect neurobehavioral function tests were excluded. Demographic characteristics are presented in Table 1 of the study, which indicates that subjects were similar. Covariates were adjusted for in the statistical analyses.
  - *Basis for Rating:* Definitely low risk of bias based on direct evidence that the exposure groups were similar and were recruited within the same time frame

using the same methods with no evidence of differences in participation/response rates.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The parents or guardians completed a questionnaire on demographic and personal characteristics of the children (sex, age at testing, parity, illnesses before age 3, and past medical history) and caretakers (age, parity, education and occupational histories, residential history, and household income). A 20-µL capillary blood sample was collected at the school by a Mianning County Center for Disease Control (CDC) health practitioner and tested for possible iron deficiency, which could be used as a covariate of neurodevelopmental performance. Covariates that were not assessed include maternal BMI, parental mental health, maternal smoking status, maternal reproductive factors, parental IQ, and HOME score. However, the study authors noted that confounding bias appeared to be limited due to the minimal diversity in the social characteristics of the subjects. The study authors indicated that CDC records documented that levels of other contaminants, including arsenic and lead, were very low in the area. Iodine differences were not specifically addressed, but there is no indication from the information provided that this might have been a concern.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Probably low risk of bias because there is direct evidence that the key covariates were considered and indirect evidence that co-exposure to arsenic was likely not an issue in this area and that methods used for collecting the information were valid and reliable.
- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The majority of results were reported for the 51 children stated to be included in the pilot study. In Table 5 of the study, the N for each dental fluorosis category totals only 43, but the text indicates 8 children did not have a Dean Index because permanent teeth had not erupted.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study used three different measurements of fluoride exposure: well water fluoride concentrations from the residence during pregnancy and onwards, fluoride concentrations from children's first morning urine samples, and degree of children's dental fluorosis. Fluoride concentrations in community well water were measured and recorded by Mianning County CDC; specific methods were not

reported, but standard methods were likely used because analyses were conducted by the CDC and were likely the same as those used to measure the fluoride in urine. Migration of subjects was noted to be limited. Well water fluoride concentrations of the mother's residence during pregnancy and onward were used to characterize a child's lifetime exposure. To provide a measure of the accumulated body burden, each child was given a 330-mL (11.2-oz) bottle of Robust© distilled water (free from fluoride and other contaminants) to drink the night before the clinical examinations, after emptying the bladder and before bedtime. The first urine sample was collected at home the following morning, and the fluoride concentration was determined on a 5-mL sample using an ion-specific electrode at the Mianning CDC. There is no indication that urinary fluoride levels accounted for dilution, nor was it clear that the method of administering water to the children and collection methods sufficiently controlled for differences in dilution. One of the investigators, a dentist, performed a blinded dental examination on each child's permanent teeth to rate the degree of dental fluorosis using the Dean Index. The Dean Index is commonly used in epidemiological studies and remains the gold standard in the dentistry armamentarium. The Index has the following classifications: normal, questionable, very mild, mild, moderate, and severe. Quality control (QC) procedures are not reported but were likely appropriate.

        - *Direction/magnitude of effect size:* Current levels were used to assess lifetime exposure. This is likely to be a non-differential exposure misclassification, and direction of bias is unknown. Because subject migration appears to be limited, it is likely that the current fluoride levels are adequate reflections of past exposure. Dental fluorosis would be an indicator that exposure occurred in the past, and there was a fair correlation between degree of dental fluorosis and current urine and water fluoride levels, with both increasing with increasing levels of dental fluorosis.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.

- ***Outcome:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* The study authors adopted culture-independent tests considered feasible for children aged 6 to 8 years. The Wide Range Assessment of Memory and Learning (WRAML) was used for the assessment of memory and learning. Three subtests were also used. The Finger Windows subtest assesses sequential visual memory. The Design Memory subtest assesses the ability to reproduce designs from memory following a brief exposure. The Visual Learning subtest assesses the ability to learn the locations of pictured objects over repeated exposures. The Wechsler Intelligence Scale for Children-Revised (WISC-IV) included digit span for auditory span and working memory and block design for visual organization and reasoning. The grooved pegboard test assesses manual dexterity. The tests used have been validated on a Western population. Although there is no information provided to indicate that the tests were validated on the

study population, the study authors indicated that the tests were culture-independent (+ for methods). Blinding of the outcome assessors or steps to minimize potential bias was not reported. However, it is unlikely that the assessors had knowledge of the individual exposure as children all came from the same area, and water and urine levels were tested at the CDC (+ for blinding). Overall = +.

  - *Basis for rating:* Probably low risk of bias based on indirect evidence that all outcomes were assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses were appropriate. Multiple regression models evaluated the associations between exposure indicators and test scores after adjusting for covariates. Specific regression models are not described or referenced, just stated to be "standard regression analysis with confounder adjustment." The distributions of fluoride concentrations in urine and water were skewed and were log10-transformed to approximate a Gaussian distribution (test not specified). Results were reported as adjusted effects and 95% Cis. There was no evidence that residual diagnostics were used to examine model assumptions; however, the impact on the effect estimates is expected to be minimal.
    - *Other potential concerns:* It should be noted that this study was a pilot study and, therefore, had a relatively small sample size (i.e., 51 children).
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in the confounding, exposure, and outcome domains. Study strengths include individual fluoride measurements with blinding at outcome assessment likely. All key covariates and many other covariates were considered in the study design or analysis.

E.2.5. Li et al. (2004) [translated in Li et al. 2008a]

E.2.5.1. Study Details

- ***Study design:*** Cross-sectional

- ***Population:*** Full-term, normal neonates 24–72 hours old from healthy mothers
- ***Study area:*** Zhaozhou County, Heilongiang Province, China
- ***Sample size:*** 91 neonates (46 males and 45 females)
- ***Data relevant to the review:*** Comparison of neurobehavioral capacity between children in the high-fluoride area compared to the control area.
- ***Reported association with fluoride exposure:*** Yes: Significant differences in neurobehavioral assessment total scores between high-fluoride (36.48 ± 1.09) and control (38.28 ± 1.10) groups; significant differences in total neurobehavioral capacity scores as measured by non-biological visual orientation reaction and biological visual and auditory orientation reaction between the two groups (11.34 ± 0.56 in controls compared to 10.05 ± 0.94 in high-fluoride group).

E.2.5.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There is indirect evidence that the exposure groups were similar. Participants were recruited during the same time frame using the same methods. From 2002 to 2003, 273 neonates were born in a hospital in Zhaozhou County, China. Ninety-one of 273 full-term neonates (46 males, 45 females) were randomly selected. Mothers ranged in age from 20 to 31 years, met multiple health criteria, and had not changed residence during pregnancy. Authors report that the two study groups were located in the same area with similar climate, living habits, economic and nutritional conditions, and cultural backgrounds, but do not provide these data in the manuscript. There is no statistically significant difference in the mode of delivery, birth weight, infant length, or sex. Subjects were separated into exposure groups after random selection.
  - *Basis for Rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited within the same time frame using the same methods with no evidence of differences in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* No covariates were specifically considered in the analysis. The study authors note similarities in characteristics in the two populations (i.e., living habits, economic and nutritional conditions, and cultural backgrounds) but do not provide these data nor do they indicate which specific characteristics were considered. There were no significant differences in infant sex, birth method, gestational age, or infant weight and length. All tests were conducted when children were 1–3 days old. No potential co-exposures were discussed. Although arsenic is considered a potential issue in China, water quality maps indicate that

there is a 25%–50% probability that the drinking water in that area exceeds the WHO guideline for arsenic of 10 µg/L.

- *Potentially important study-specific covariates:* Key covariates, including age, sex, and measures of socioeconomic status (SES), were similar between exposure groups; however, arsenic was not considered. Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on water quality maps, arsenic does not appear to be an issue in Zhaozhou County of the Heilongjiang Province. Iodine deficiencies are not mentioned.
  - *Direction/magnitude of effect size:* Conceptually, the presence of arsenic would potentially bias the association away from the null if it were present with fluoride. Deficiencies in iodine would potentially bias the association away from the null if it were present in areas of higher fluoride but toward the null if it were present in areas of lower fluoride. Neither of these are considered a concern in this study for reasons detailed above.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the key covariates were considered, co-exposure to arsenic was likely not an issue in this area, and methods used for collecting the information were valid and reliable.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Although authors did not discuss why only 91 of the 273 neonates available were randomly selected, results were available for all 91 subjects.
  - *Basis for rating:* Definitely low risk of bias based on results being available for all subjects.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were split into control and high-fluoride groups based on fluoride levels in their places of residence. Although the levels were provided (1.7–6.0 mg/L for the high-fluoride group compared to 0.5–1.0 mg/L for the control group), it was not reported how or when these levels were measured. Urine was collected when women were hospitalized but before labor began. Urine samples were sent to a specific lab for measurement using fluoride ion-selective electrode. It was noted that this procedure strictly followed the internal controls of the laboratory, indicating quality control. Level of detection (LOD) was not provided. Urinary fluoride levels were significantly higher in the high-fluoride mothers (3.58 ± 1.47 mg/L) compared to the control-group mothers (1.74 ± 0.96 mg/L). There was indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure. Although results were mainly based on exposure area, they were supported by urine data, making exposure misclassification less of a concern.
    - *Direction/magnitude of effect size:* There is high variability in both water fluoride and urine fluoride in the subjects from the high-exposure area. Although there is no overlap in the water fluoride levels in the exposure areas, there is some overlap in the urine concentrations in the mothers from the two

areas. This may reflect the single measurement and pose no specific bias, or it could indicate that some mothers in the high-fluoride area have lower fluoride exposure, which could bias the association toward the null.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* A standard neonatal behavioral neurological assessment method was carried out by professionals in the pediatric department working in a neonatal section trained specifically for these programs and passing the training exams (+ for methods). The examinations were carried out 1 to 3 days after delivery. Because urine samples were collected on the day of delivery and sent to a separate laboratory, it is likely that the outcome assessors were blind. Although the subjects were separated by fluoride exposure area, it is not likely that the professionals were aware of the exposure as the tests were conducted in the hospital (+ for blinding).
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessors were blind to participants' fluoride exposure.
- ***Selective Reporting:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study authors reported numerous outcomes in sufficient detail; however, because a list of outcomes tested was not provided, there is no direct evidence that all were reported.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that all the study's measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses are described only as a t-test. Consideration of heterogeneity of variance was not reported. Results are reported as mean and standard deviations of neurological scores. Maternal urinary fluoride levels were used only to compare exposures between exposed and control groups. Infants in the control group were from four villages, and those in the exposed group were from five villages within the same district. Infants were randomly selected before they were assigned to exposed or control groups. In the comparisons, there was no accounting for clustering at the village level. It is likely that the standard error of the difference in mean neurobehavioral assessment scores between the high fluoride group and control group will be biased, making differences appear stronger than they actually are. However, the use of multiple villages per exposure group is

likely to mitigate some of the impact of this lack of accounting for clustering, and the overall impact on effect estimates is expected to be minimal.

- *Other potential concerns:* It should be noted that although the study states that subjects were randomly selected, it is unclear why only 91 subjects were included and whether they were randomly selected to obtain equal numbers in the high-fluoride and control groups.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that statistical analyses were appropriate and that there were no other potential threats of risk of bias.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in the confounding, exposure, and outcome domains. Study strengths include individual fluoride measurements to support the differences in the two areas. Tests were noted to be conducted at the hospital, providing indirect evidence that blinding was not a concern during the outcome evaluation. Although there was some potential for bias due to the lack of accounting for arsenic or iodine deficiencies, co-exposure to arsenic was likely not a major concern according to groundwater quality maps.

E.2.6. Riddell et al. (2019)

E.2.6.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Canadian Health Measures Survey (Cycles 2 and 3) participants (children aged 6–17 years)
- ***Study area:*** General population, Canada
- ***Sample size:*** 3,745 children
- ***Data relevant to the review:*** Adjusted odds ratios for ADHD and attention symptoms per 1 unit increase in urinary fluoride by water fluoride in the tap water or community fluoridation status.
- ***Reported association with fluoride exposure:*** Yes: Significantly increased odds of ADHD diagnosis (adjusted OR = 6.10; 95% CI: 1.60, 22.8) or hyperactivity/inattentive symptoms (adjusted β = 0.31; 95% CI: 0.04, 0.58) per 1-mg/L increase in tap water fluoride. In addition, a significant association between ADHD diagnosis (adjusted OR = 1.21; 95% CI: 1.03, 1.42) or hyperactivity/inattentive symptoms (adjusted β = 0.11; 95% CI: 0.02, 0.58) and community water fluoridation status. No significant associations with urinary fluoride levels.

E.2.6.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)

  - *Summary:* Subjects were part of Cycles 2 and 3 of the Canadian Health Measures Survey. This is a nationally representative sample of residents living in 10 provinces. Specific inclusion criteria were provided. This study was restricted to children 6–17 years of age with different fluoride measurements that consisted of three participant samples. One of the samples was available only in Cycle 3.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were similar and were recruited with the same methods during the same time frame.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariates included in all models included age at testing, sex, ethnicity, BMI, parents' education, total household income, exposure to cigarette smoke inside the home, and log-transformed concurrent blood lead levels. Covariates such as parental behavioral and mental health disorders, quantity and quality of caregiving environment, and co-exposure to arsenic were not discussed. The study used data from the Canadian Health Measures Survey, which consists of a nationally representative sample of Canadians. Most Canadians (~89%) receive water from municipal water supplies, which monitor for levels of arsenic. Therefore, co-exposure to arsenic is not likely an issue in this population. Rationale for selection of covariates was based on relationship to ADHD diagnosis and to fluoride metabolism based on literature review and consultation with an ADHD expert. There is no information of the source of data for covariates, but it is likely the questionnaires from the Canadian Health Measures Survey, which are considered standardized and validated.
  - *Potentially important study-specific covariates:* All key covariates were considered in this study.
    - *Direction/magnitude of effect size:* Not applicable.
  - *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key covariates were considered, co-exposure to arsenic was likely not an issue, and methods used for collecting the information were valid and reliable.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There is no information indicating that there were any data excluded due to missing covariates. All exclusions of children were described and reasonable (i.e., drinking bottled water when considering city fluoridation as a measure of fluoride exposure). Outliers were stated to be excluded, but methods for determining this were provided, and it was noted that the outliers were 0.27% of the values.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***

- *Rating:* Probably low risk of bias (+)
- *Summary:* **Urinary Fluoride:** Spot urine samples were collected under normal non-fasting conditions and analyzed using an Orion pH meter with a fluoride ion-selective electrode after being diluted with an ionic adjustment buffer. Analysis was performed at the Human Toxicology Laboratory of the Institut National de Santé Publique du Québec. The precision and accuracy of the fluoride analyses, including quality control and quality assurance, were described by Health Canada (2015). The limits of detection were 20 µg/L for Cycle 2 and 10 µg/L for Cycle 3 with no values below detection. Fluoride levels were adjusted for specific gravity.

  Water Fluoride in Tap Water: Tap water was collected at the subjects' homes in Cycle 3 only. Samples were analyzed for fluoride concentrations using anion exchange chromatography procedure with an LOD of 0.006 mg/L. Values below the LOD were imputed with LOD/square root(2). Of the 980 samples, 150 (15%) were below detection.

  Chlorinated Water Fluoride Status: This was determined by viewing reports on each city's website or contacting the water treatment plant (provided in supplemental material). Children were excluded if they drank bottled water, had a well, had a home filtration system, lived in the current residence for 2 years or less, or lived in an area with mixed city fluoridation.

  - *Direction/magnitude of effect size:* There is not any specific impact on the direction or magnitude of effect size expected. Urinary fluoride levels are reflective of a recent exposure, but the study authors adjusted to account for dilution. The possibility of exposure misclassification would be similar in all subjects and would be non-differential. There is less potential for exposure misclassification due to tap water or chlorinated water fluoride status, since children who drank bottled water were excluded and children who had a home filtration system were excluded from the chlorinated water status.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Probably high risk of bias (−)
  - *Summary:*

    Strengths and Difficulties Questionnaire (SDQ): The questionnaire was administered to youths under 18 years. Children aged 6–11 years had SDQ ratings provided by parents and guardians, but youths aged 12–17 years completed the questionnaire themselves. Tests consist of 25 items with a 3-point scale. Items were divided into five subscales: emotional problems, conduct problems, hyperactivity-inattention, peer problems, and prosocial behavior. The current study used only the hyperactivity-inattention subscale. Validation of this method was not reported (− for methods).

ADHD: Ninety percent of youths with ADHD are diagnosed after age 6. For children aged 6–11 years, ADHD diagnosis was provided by parents, but youths aged 12–17 years completed the questionnaire themselves. Cycle 2 asked "Do you have a learning disability?"; if the subject answered "yes," he/she was asked to specify the type (four options were available and described). In Cycle 3, parents were asked directly whether they had ADHD, and children 12 years and older were asked whether they had a physician diagnosis of ADHD and, if so, what subtype (− for methods because different methods were used, and only the children 12 years and older in Cycle 3 were asked specifically about a doctor's diagnosis). Both were measured in both cycles. Blinding is likely not an issue as subjects would not have knowledge of the urine or tap water fluoride levels. However, they would likely have knowledge of the city.

- *Basis for rating:* Probably high risk of bias based on indirect evidence that the outcome was assessed using insensitive methods that varied based subject age.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods sections were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Robust logistic regression was used to examine the association between fluoride exposure and ADHD diagnosis, adjusting for covariates. Box-Tidewell tests were used to check the linearity of the relationship with the continuous predictors. Linear regression was used for the SDQ scores using Huber-White standard errors. Multicollinearity was evaluated using variance inflation factor (VIF) statistics. Outliers with high studentized residuals, high leverage, or large Cook's distance values were removed from all analyses with urinary fluoride. All regressions were tested for interactions between fluoride exposure and age and between fluoride exposure and sex. Sensitivity analyses were conducted to test the different survey cycles. There is no mention of adjustment for the complex survey design using survey weights or bootstrapped weights to ensure appropriate calculation of the estimated variances; however, the overall impact on effect estimates is expected to be minimal.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include individual exposure measurements and the addressing of key covariates, but was limited by the cross-sectional study design and insensitive outcome measures.

E.2.7. Rocha-Amador et al. (2009)

E.2.7.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** Children aged 6–11 years
- ***Study area:*** Durango, Mexico
- ***Sample size:*** 80 children
- ***Data relevant to the review:*** Associations between visuospatial organization and visual memory (using the Rey-Osterrieth Complex Figure Test, children's version) and urinary fluoride levels in the children.
- ***Reported association with fluoride exposure:*** Yes: Significant correlation between urinary fluoride and visuospatial organization ($r = -0.29$) and visual memory ($r = -0.27$) scores. No significant correlations with arsenic.

E.2.7.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were from the same population and were recruited during the same time frame using the same methods. Although this study compared three sites with antecedents of environmental pollution to mixtures of either F-As, Pb-As, or DDT-PCBs, authors evaluated each contaminant separately. The only area of interest with F and As contamination is in Durango state (5 de Febrero) where drinking water is polluted naturally with F and As at levels exceeding 6 and 19 times, respectively, the World Health Organization (WHO) limits (WHO 2008). Children attending public schools were screened through personal interviews for study eligibility. Inclusion criteria were children between 6 and 11 years old, living in the study area since birth, whose parents signed the agreement to participate. Children with a neurological disease diagnosed by a physician and reported by the mother were excluded from the study. The final sample for the F-As group was 80. Participation rates were not reported. Selected demographic characteristics are presented in Table 1 of the study.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the populations were similar and recruited during the same time frame using the same methods.
- ***Confounding:***

- *Rating:* Probably high risk of bias (−)
- *Summary:* Covariates included blood lead (PbB), age, sex, and height-for-age z-scores; only age had significant associations and was included in the final analysis. Arsenic was also assessed and analyzed separately from fluoride. Arsenic in urine was analyzed by atomic absorption spectrophotometer coupled to a hydride system (Perkin-Elmer model AAnalyst 100). Although the model did not adjust for arsenic, arsenic in the F-As group was not associated with either outcome; therefore, arsenic co-exposure is not considered a major concern in this study. PbB was analyzed with a Perkin-Elmer 3110 atomic absorption spectrophotometer using a graphite furnace. Authors note that the mean blood lead level in the F-As study area was 5.2 µg/dL, and 8% of the children had values above the reference value of 10 µg/dL. PbB was stated not to affect results and was not included in the final analysis. Other covariate data were obtained during the study interview. Father's education was provided and, in the F-As group, was stated to range from 0–16 years, but this was not considered. Maternal education, smoking, and SES were also not considered. The authors provide an SES score of 5.9 ± 1.4 for the 5 de Febrero region (the fluoride region). It is not clear whether this would vary by fluoride or arsenic levels.
- *Potentially important study-specific covariates:* SES.
  - *Direction/magnitude of effect size:* There are insufficient data to determine the impact on the magnitude or direction of effect size. The impact on the direction of the association would likely depend on the association between fluoride exposure and SES.
- *Basis for rating:* Probably high risk of bias based on indirect evidence that the SES was not considered in the study design or analysis and may have varied by fluoride levels.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are complete. All 80 participants stated to be the final sample for the site of interest (F-As) were included in all analyses.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Fluoride in urine (FU) was analyzed according to method 8308 (''fluoride in urine'') from the National Institute for Occupational Safety and Health (NIOSH 1984) with a sensitive specific ion electrode. As a quality control check, reference standard ''fluoride in freeze dried urine'' (NIST SRM 2671a) was analyzed. The accuracy was 97.0% ± 6.0%. Levels of FU and AsU were adjusted for urinary creatinine, which was analyzed by a colorimetric method (Bayer Diagnostic Kit, Sera-Pak1 Plus). However, details on the collection methods were not reported.

        - *Direction/magnitude of effect size:* Spot urine samples in a small sample size (i.e., 80 children) may have some exposure misclassification. Adjusting for dilution reduces the potential for misclassification based on differences in dilution. Exposure misclassification would likely be non-differential.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* IQ was assessed through the Rey-Osterrieth Complex Figure Test (ROCF). This is a less well-established method, although the authors provide citations suggesting it has been validated and standardized for the Mexican population (+ for methods). According to the study report, the neuropsychologist who administered the test was blinded to all exposure types and levels (++ for blinding). Overall rating for methods and blinding = +.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:*
        - *Statistical analyses:* Statistical analyses used log-transformed exposure variables (although rationale was not provided). Crude and partial correlations were calculated to evaluate associations between serum fluoride levels and TOCF scores. There is no other description of the regression model, and regression diagnostics to evaluate model assumptions are not presented; however, the overall impact on effect estimates is expected to be minimal.
        - *Other potential concerns:* None identified.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements and blinding of outcome assessor to participants' fluoride exposure, but it is limited by the cross-sectional study design, lack of consideration of SES in

the study population, co-exposure with arsenic, and use of spot samples in a small population.

E.2.8. Valdez Jimenez et al. (2017)

E.2.8.1. Study Details

- ***Study design:*** Prospective cohort
- ***Population:*** Infants aged 3–15 months
- ***Study area:*** Durango City and Lagos de Moreno, Jalisco, Mexico
- ***Sample size:*** 65 infants
- ***Data relevant to the review:*** The Bayley Scales of Infant Development II was used to assess Mental Development Index scale and the Psychomotor Development Index scale in children aged 3 to 15 months and evaluated for associations with first and second trimester maternal urine fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant association between log10-mg/L maternal urinary fluoride and MDI score during first trimester (adjusted $\beta = -19.05$; SE = 8.9) and second trimester (adjusted $\beta = -19.34$; SE = 7.46). No association between maternal fluoride during any trimester and Psychomotor Developmental Index (PDI).

E.2.8.2. Risk of Bias

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were recruited from two endemic areas in Mexico. The study authors do not provide information on the similarities or differences between the two areas, nor do they indicate whether there were different participation rates. However, recruitment methods were the same. Women receiving prenatal care in health centers located in Durango City and Lagos de Moreno, Jalisco, Mexico were recruited in 2013–2014. Participation rates are not likely to be an issue as characteristics were similar between those who participated and those who did not. Although the authors did not provide characteristics by area, the characteristics provided do not indicate any differences that may be biased by the selection. Considering the age range for the non-participants, the mean age for non-participants appears to be incorrect (or the age range is incorrect); however, there does not appear to be a difference that would potentially indicate selection bias.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited with the same methods in the same time frame, with no evidence of differences or issues with participation/response rates.

- ***Confounding:***
  - *Rating:* Probably high risk of bias (−)
  - *Summary:* Questionnaires were used to obtain information about sociodemographic factors, prenatal history, mother's health status before pregnancy (e.g., use of drugs, vaccines, diseases), and the type of water for drinking and cooking. The marginalization index (MI) was obtained from the National Population Council (CONAPO). Two additional surveys were conducted during the second and third trimester of pregnancy to get information about the mother's health, pregnancy evolution, and sources of water consumption. A survey was also conducted to get information about childbirth (type of birth, week of birth, weight and length of the baby at birth, Apgar score and health conditions of the baby during the first month of life). This information was corroborated with the birth certificate. Linear regression models included gestational age, children's age, marginality index, and type of drinking water. Bivariate analyses were conducted on the other factors, including sex, prior to conducting multivariable regression models. Some important covariates were not considered, including parental mental health, IQ, smoking, and potential co-exposures. Water quality maps indicate a potential for arsenic to be present in the study area.
  - *Potentially important study-specific covariates:* Arsenic is a potential co-exposure in this area of Mexico.
    - *Direction/magnitude of effect size:* If arsenic were present as a co-exposure, it would likely bias the association away from the null.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that there is a potential for co-exposure with arsenic that was not addressed.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Out of the 90 women selected for inclusion in the study, 65 approved the participation of their infants. The authors provide a table of characteristics between women who consented to their children's cognitive evaluation and those who participated only in biological monitoring. There were no significant differences between the groups. There were fewer women who provided urine during the second and third trimesters. All specified children are included in the relevant analyses.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Fluoride exposure was assessed through morning urine samples and water fluoride levels collected from the children's homes. Sampling methodology was appropriately documented, and water levels were quantified through specific

ion-sensitive electrode assays. QC was described, and accuracy was >90%. Urinary fluoride was corrected by specific gravity.

- *Direction/magnitude of effect size:* Not applicable.

- *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Neurodevelopment was assessed with the Bayley Scales of Infant Development II (BSDI-II) that was noted to be reliable and valid for evaluating children from 3 months to 5 years of age. The average age of children assessed was 8 months, with a range of 3–15 months) (++ for methods). The study report stated that a trained psychologist who was blinded about the mother's fluoride exposure evaluated the infants at home (++ for blinding). Overall rating for methods and blinding = ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was assessed using instruments that were valid and reliable in the study population, and that the outcome assessor was blind to participants' fluoride exposure.
- ***Selective Reporting:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported. Table 4 of the study displays only data for trimesters 1 and 2. Although third trimester data were collected, they were not reported, likely because they were available for only 29 subjects. No discussion of this was provided.
  - *Basis for rating:* Probably low risk of bias because, although it appears some data were not reported, it is likely because there were insufficient data and not because the authors were selectively reporting the results.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used log10-transformed exposure variables. Normality, homoscedasticity, and linearity assumptions were tested and satisfied for MDI and PDI scores. Bivariate analyses included correlations, t-tests, and ANOVA. Multiple linear regression models by the first and second trimester of pregnancy were used to evaluate the association between maternal fluoride exposure and MDI and PDI scores. The best-fit model was selected using a "stepwise method," and the best-fit line was evaluated using "the curve fitting method." It is not further specified or cited what these methods entailed. Best-fit or goodness-of-fit statistics are not reported. It is unclear how a best-fit model could be selected when the authors state that all models adjusted for the same set of covariates regardless of

significance, and these covariates also appear in the final model—presumably the best-fit model. It is unlikely that a stepwise method would retain all those covariates unless they were forced in the model. Residual analysis was conducted to assess model validity; however, there is no description of the results of the residual analysis. Nonetheless, the impact on effect estimates is expected to be minimal.

        - *Other potential concerns:* No other potential concerns were identified. In the peer-review report, NASEM (2020) cited the following as potential concerns: "the large difference in numbers of males and females in the offspring (20 males, 45 females), and apparently incorrect probabilities were reported for age differences between participants and nonparticipants, high rates of cesarean deliveries and premature births among participants (degree of overlap not reported), and incorrect comparisons of observed prematurity rates with national expected rates." However, these concerns were taken into consideration in other domains (***Selection, Confounding***).
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats of risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements and blinding of outcome assessor to participants' fluoride exposure, but it is limited by the cross-sectional study design and lack of accounting for potential co-exposures to arsenic.

E.2.9. Wang et al. (2020a)

E.2.9.1. Study Details

- ***Study design:*** Cross-sectional
- ***Population:*** School children aged 7–13 years
- ***Study area:*** Tongxu County, China
- ***Sample size:*** 325 school children
- ***Data relevant to the review:*** Associations between ADHD and other measures of learning disability with urine fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant association between psychosomatic problems and urinary fluoride (per 1-mg/L increase; adjusted $\beta = 4.01$ [95% CI: 2.74, 5.28]) and increased risk of a T-score >70 with urinary fluoride (per 1-mg/L increase; adjusted OR = 1.97 [95% CI: 1.19, 3.27]). No significant associations with ADHD or other measures of learning disability.

E.2.9.2. Risk of Bias

- **Author contacts:**
    - Authors were contacted in July of 2020 to obtain additional information for risk-of-bias evaluation. No response was received.

- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were recruited in 2017 from Tongxu County, China. Children were selected from four randomly selected primary schools in the area. Selection was based on specified inclusion rules. It was noted that the living habits and diets of the participants from the four schools were well matched, but details were not provided. The area did not have industrial pollution within 1 km of the living environment of the children, and it was noted that the children were not exposed to other neurodevelopmental toxicants (lead, cadmium, arsenic, or mercury). A table of subject characteristics was provided in the study but not by school or exposure. This was a pilot study, and it was not explicitly stated whether all eligible subjects participated in the study. There is no information on participation rates or whether they varied by school.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposed groups were recruited using similar methods during the same time frame and that any differences between the exposed groups were considered in the statistical analyses.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* It was noted that subjects were well matched in terms of living habits and diets, but there were no specifics provided. It was noted that there was no industrial exposure or exposure to other neurotoxins such as lead, cadmium, arsenic, or mercury. Covariates were collected using a standardized and structured questionnaire completed by the children and their guardians under the direction of investigators, but reliability or validity of the questionnaire was not reported. Information collected included age, sex, weight, height, parental education level, and parental migration (or work as migrant workers). IQ scores evaluated by the Combined Raven's Test–the Rural in China were used to represent basic cognitive function. Models were adjusted for age, BMI, sex, mother and father migration, and urinary creatinine. Adjustments were not made for parental education, race/ethnicity, maternal demographics (e.g., maternal age, BMI), parental behavioral and mental health disorders (e.g., ADHD, depression), smoking (e.g., maternal smoking status, secondhand tobacco smoke exposure), reproductive factors (e.g., parity), iodine deficiency/excess, maternal (and paternal) IQ, quantity and quality of caregiving environment (e.g., HOME score), or SES other than parental migration. There is no evidence to suggest that SES would differ substantially among the four rural schools in the same area of China that were randomly selected.
  - *Potentially important study-specific covariates:* SES.
    - *Direction/magnitude of effect size:* The impact on the direction and magnitude of effect size are unknown. It was noted that the subjects were matched in terms of living habits and diet, and this could be an indication that SES was not different among the groups, but details were not provided.

  - *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key covariates were considered, that the methods for collecting the information were valid and reliable, and that co-exposure to arsenic was not an issue in this area.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are complete. It was noted that there were 325 subjects included, and results were available on all subjects.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Spot urine samples were collected from each child in the early morning into cleaned polyethylene tubes. Fluoride concentrations were measured using fluoride ion-selective electrode [with reference to Ma et al. (2017); however, that reference cites Zhou et al. (2012)]. Therefore, no QC methods or LODs were available. Fluoride concentrations were creatinine-adjusted.
    - *Direction/magnitude of effect size:* Spot urine samples account for only recent exposure. Although this could cause some exposure misclassification, the number of subjects should help dilute any issues with the non-differential misclassification.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.
- ***Outcome:***
  - *Rating:* Probably high risk of bias (NR)
  - *Summary:* Children's behavior was assessed by the Chinese version of Conners' Parent Rating Scale-Revised (CPRS-48). The homogeneity reliability of Cronbach α in the Chinese version of CPRS-48 was 0.932, the correlation of Spearman-brown split-half was 0.900, and the retest reliability of total score was 0.594. Raw scores for each subscale were converted into sex- and age-adjusted T-scores within a mean ± standard deviation (SD) of 50 ± 10. The guardians independently completed the CPRS-48 according to the instruction manual under the direction of trained investigators (++ for methods). Blinding is not reported. Although it is unlikely that the outcome assessors were aware of the fluoride levels in the urine, it is unclear whether subjects were selected based on areas with endemic fluoride or whether parents were aware of fluoride concentrations in the areas (NR for blinding). Overall rating for methods and blinding = NR.
  - *Basis for rating:* Probably high risk of bias based on no information provided to indicate that the outcome assessors were blind to the participants' fluoride exposure.
- ***Selective Reporting:***

  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Multiple linear regression models were used to assess the association between urinary fluoride exposure and each behavioral outcome. Logistic regression was used to assess the risk of behavioral problems (T-scores >70) due to fluoride exposure. Sensitivity analyses were performed, with models adjusting for combinations of age, BMI, sex, mother migrated, father migrated, and urinary creatinine levels. Regression diagnostics to evaluate model assumptions are not described; however, the overall impact on effect estimates is expected to be minimal.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and no other potential threats of risk of bias were identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include individual exposure measurements, but it is limited by the cross-sectional study design and lack of details on blinding of the outcome assessment. All key covariates were considered in the study design or analysis.

Appendix F. Mechanistic Data from Animal Studies

Table of Contents

Figures

A number of animal studies were available that presented mechanistic data in several effect categories (see Figure F-1). Limiting the data to studies with at least one exposure at or below 20 ppm fluoride drinking water equivalents (gavage and dietary exposures were backcalculated into equivalent drinking water concentrations for comparison) still provided a sufficient number of studies for evaluation of several mechanistic endpoints while allowing for a more focused look at exposure levels most relevant to human exposures. The following sections summarize the mechanistic data by effect category. Although there is some evidence of consistency in mechanistic effects, overall these data are insufficient to increase confidence in the assessment of findings from human epidemiological studies.

	Dose Level	
Mechanism	**All**	**≤20 ppm**
Biochemical (brain/neurons)	66	25
Neurotransmitters	53	23
Oxidative stress	78	25
Histopathology	67	30
Apoptosis/cell death	19	7
Inflammation	7	5
Thyroid	50	17
Other	3	2

Figure F-1. Number of Animal Mechanistic Studies for Fluoride by Mechanistic Category and Exposure Level

An interactive version of Figure F-1 and additional study details in Tableau® (https://public.tableau.com/app/profile/ntp.visuals/viz/Animal_Mechanisms_2021/FigureA5-1). The number of studies that evaluated mechanistic effects associated with at least one exposure at or below 20 ppm fluoride is tabulated in the "≤20 ppm" column. The total number of studies per mechanistic category is summarized in the "All" column.

F.1. Neurotransmitters

Neurotransmitter and biochemical changes in the brain and neurons were considered the mechanistic areas with the greatest potential to demonstrate effects of fluoride on the brain of animals in the lower dose range and provide evidence of changes in the brain that may relate to lower IQ in children (see Figure F-2). Twenty of 23 neurotransmitter studies assessed changes in brain cholinesterase activity associated with fluoride exposure at or below 20 ppm fluoride. Acetylcholine is a major neurotransmitter involved in learning, memory, and intelligence (Chen 2012; Gais and Schonauer 2017). AChE is responsible for the breakdown of acetylcholine in the synapses of nerve cells. Changes in cholinesterase, acetylcholine, or AChE could be related to effects on memory. Evidence of an effect varied among the low risk-of-bias studies that assessed changes in cholinesterase or acetylcholine (n = 11 drinking water studies) (Adedara et al. 2017a; Akinrinade et al. 2015a; Baba et al. 2014; Chouhan et al. 2010; Gao et al. 2008b; Gao et al. 2009; Khan et al. 2017; Liu et al. 2010; Mesram et al. 2016; Nkpaa and Onyeso 2018; Sun et al. 2000 [translated in Sun et al. 2008]), with the majority reporting evidence of an effect that is considered inconsistent with the phenotypic outcome (see Quality Assessment of Individual

Studies section for methods on determining which studies pose low risk of bias). Decreases in cholinesterase will cause increases in acetylcholine, which can have a positive effect on learning and memory; however, long-term decreases in cholinesterase can lead to secondary neuronal damage occurring in the cholinergic region of the brain (Chen 2012).

Five of the 11 studies with low risk of bias (Adedara et al. 2017a; Baba et al. 2014; Gao et al. 2009; Khan et al. 2017; Nkpaa and Onyeso 2018) found statistically significant decreases in cholinesterase or AChE in brain homogenates (with some brains dissected into specific regions prior to homogenizing) with fluoride concentrations in drinking water at or below 20 ppm, and four of the five studies found statistically significant decreases in cholinesterase or AChE below 10 ppm. The five studies were conducted in rats (Wistar or Sprague-Dawley) with exposure ranging from 28 days to 6 months. An additional 2 out of 11 studies (Akinrinade et al. 2015a; Gao et al. 2008b) reported decreases in brain homogenate AChE at concentrations at or below 20 ppm fluoride in drinking water, but statistical significance was not reached. These studies were also conducted in rats with exposure for 30 days or 3 months. Gao et al. (2008b) reported a dose-dependent decrease in brain homogenate AChE in the low (5 ppm fluoride) and high (50 ppm fluoride) treatment groups compared with the control group, but the decrease was statistically significant only in the high-dose group. Similarly, Akinrinade et al. (2015a) observed a dose-dependent decrease in percent intensity of AChE immunohistochemistry in the prefrontal cortex associated with 2.1 and 10 ppm sodium fluoride in drinking water, but neither result was statistically significant. Gao et al. (2009) found lower brain homogenate AChE levels in the 5-ppm animals compared with the 50-ppm animals; therefore, the results were not always dose-dependent.

Relative to the above-mentioned studies, 2 of the 11 low risk-of-bias studies observed opposite effects on brain cholinesterase levels. Sun et al. (2000) [translated in Sun et al. (2008)] observed a significant increase in brain cholinesterase in Kunming mice associated with fluoride drinking water concentrations from 10 to 100 mg/L but did not observe a dose response. Chouhan et al. (2010) did observe a dose-related increase in AChE levels in brain homogenate of Wistar rats with sodium fluoride concentrations of 1 to 100 ppm for 12 weeks and noted statistically significant results at 1, 50, and 100 ppm but not at 10 ppm.

Mesram et al. (2016) did not assess changes in AChE but observed a significant decrease in acetylcholine levels in cerebral cortex homogenate through 30 days of age in rats treated in utero with 20 ppm sodium fluoride, which may suggest an increase in AChE levels. Likewise, Liu et al. (2010) did not assess changes in AChE but measured nicotinic acetylcholine receptors (nAChRs) in brain homogenate of rats following drinking water fluoride exposure, which the authors stated could modulate physiological and pharmacological functions that are involved in learning- and memory-related behaviors. Significant decreases in the protein expressions of nAChR subunits at 2.26 ppm fluoride were observed; however, the corresponding receptor subunit mRNAs did not exhibit any changes (Liu et al. 2010).

The studies that assessed other neurotransmitters of the brain and neurons were too heterogeneous or limited in number to make any determination on mechanism, even before limiting the review of the data to low risk-of-bias studies. There were only five studies that evaluated dopamine and/or metabolites (Banala et al. 2018; Chouhan et al. 2010; Reddy et al. 2014; Sudhakar and Reddy 2018; Tsunoda et al. 2005). Four of the studies observed decreases in dopamine levels in the brain with exposures of less than 20 ppm fluoride (Banala et al. 2018;

Chouhan et al. 2010; Reddy et al. 2014; Sudhakar and Reddy 2018); however, the fifth study (Tsunoda et al. 2005) observed increased dopamine and metabolites at fluoride exposures below 20 ppm (with statistical significance achieved only for the metabolite homovanillic acid in one brain region). No differences from the control group were observed at levels above 20 ppm fluoride. Other neurotransmitters were evaluated at or below 20 ppm fluoride exposure, but generally only in a couple of studies.

F.2. Biochemistry (Brain/Neurons)

Similar to the above, the endpoints measured in brain biochemistry studies were too heterogeneous or limited in number to make any determination on potential relevance of mechanism, even before limiting the review of the data to low risk-of-bias studies (see Figure F-2). Endpoints related to biochemical changes in the brain or neurons included carbohydrate or lipid changes, RNA or DNA changes, changes in gene expression, or changes in protein expression. For the most part, only a single study was available for any given endpoint. The largest body of evidence on biochemistry was on protein level in various brain regions. Eleven low risk-of-bias studies were identified that evaluated protein levels; however, few studies evaluated the same proteins or areas of the brain. In the few cases in which the same protein was evaluated, results were not always consistent. These data are insufficient to increase confidence or support a change to hazard conclusions.

F.3. Histopathology

Histological data can be useful in determining whether effects are occurring in the brain at lower fluoride concentrations; however, author descriptions of these effects may be limited, thereby making it difficult to directly link histological changes in the brain to learning and memory effects. Histopathology of the brain was evaluated in 31 studies with concentrations at or below 20 ppm fluoride, of which 15 were considered low risk-of-bias studies (Adedara et al. 2017b; Akinrinade et al. 2015a; Bhatnagar et al. 2002; Bhatnagar et al. 2011; Chouhan et al. 2010; Guner et al. 2016; Jia et al. 2019; Jiang et al. 2014; Lou et al. 2013; McPherson et al. 2018; Mesram et al. 2016; Nageshwar et al. 2018; Niu et al. 2018; Pulungan et al. 2016; Zhao et al. 2019). In all but one low risk-of-bias study [Pulungan et al. (2016); gavage], animals were exposed to fluoride via drinking water. All low risk-of-bias studies were conducted in rodents, and all but three were conducted in rats (Wistar [seven studies], Sprague-Dawley [four studies], Long-Evans hooded [one study]). Overall, the low risk-of-bias studies that evaluated histopathology in the brain had low potential for bias for key questions regarding randomization and exposure characterization; however, eight studies were rated as probably high risk of bias for the key risk-of-bias question regarding outcome assessment based on lack of reporting of blinding of outcome assessors and/or inadequate description of outcome measures or lesions. Moreover, low image quality in some of the studies hampered the ability to verify the quality of the data. Further technical review of the 15 low risk-of-bias studies was conducted by a board-certified pathologist. Based on confidence in the results for each study, the technical reviewer further categorized the low risk-of-bias studies as studies with higher or lower confidence in the outcome assessment, which is reflected in the following summary of the brain histopathology results. Main limitations of the histopathology data identified by the pathologist included lack of information on methods of euthanasia and fixation. Perfusion fixation is generally considered the

best practice for lesions of the central nervous system in addition to complete fixation of the brain prior to its removal from the skull (Garman et al. 2016). Four of the low risk-of-bias studies reported that they used this method (Bhatnagar et al. 2002; Bhatnagar et al. 2011; McPherson et al. 2018; Pulungan et al. 2016). Two of the low risk-of-bias studies handled the brains before fixation was complete, which can produce artifacts that can resemble dead neurons (Nageshwar et al. 2018; Zhao et al. 2019). Fixation and brain removal details were inadequately described in the remaining low risk-of-bias studies.

Although there was heterogeneity in the endpoints reported (e.g., cell size, shape, and counts; nuclei fragmentation; increased vacuolar spaces) and some variation in the consistency of the evidence based on the area of the brain evaluated, the majority of the low risk-of-bias studies (11 of 14 drinking water studies) found some histological change in the brain of rats or mice treated with fluoride at concentrations at or below 20 ppm, of which 8 studies reported histological changes in the brain at or below 10 ppm. Histological changes in the hippocampus (one of the areas of the brain most evaluated for histological changes) associated with fluoride exposure at or below 20 ppm were reported in three of four low risk-of-bias studies with higher confidence in the outcome assessment (Bhatnagar et al. 2002; Bhatnagar et al. 2011; Guner et al. 2016) and in three of four low risk-of-bias studies with lower confidence in the outcome assessment (Jiang et al. 2014; Nageshwar et al. 2018; Niu et al. 2018). McPherson et al. (2018) was the only drinking water study (with higher confidence in the histopathology outcome assessment) that did not observe any histological changes in hippocampus at 10 or 20 ppm fluoride in male Long-Evans hooded rats exposed in utero through adulthood (>PND 80). Although there are too few studies to definitively explain the inconsistency in results, McPherson et al. (2018) also did not observe any associations between fluoride exposure and impairments to learning and memory, which is inconsistent with the majority of developmental exposure studies that observed learning and impairments associated with fluoride exposure for other strains of rats. Similarly, histological changes in the cortex were reported in three of the four low risk-of-bias drinking water studies with higher confidence in the outcome assessment (Akinrinade et al. 2015a; Bhatnagar et al. 2011; Chouhan et al. 2010) and in three of four low risk-of-bias studies with lower confidence in the outcome assessment (Lou et al. 2013; Mesram et al. 2016; Nageshwar et al. 2018).

Histological changes were also consistently reported in other areas of the brain in studies with higher confidence in the outcome assessment, including the amygdala, caudate putamen, cerebellum, and hypothalamus, although each of these areas of the brain was evaluated in only one low risk-of-bias study (Bhatnagar et al. 2011; Guner et al. 2016). Pulungan et al. (2016), one of two low risk-of-bias studies with higher confidence in the outcome assessment that did not report histological changes in the brain, observed a decreasing trend in the number of pyramidal cells in the prefrontal cortex with increasing dose, but this was not changed at concentrations below 20 ppm (the study administered sodium fluoride via gavage; the 5-mg/kg/day dose was considered equivalent to 15.3 ppm fluoride in drinking water), nor were any of the results statistically significant.

F.4. Oxidative Stress

Oxidative stress is considered a general mechanistic endpoint that cannot be specifically linked to neurodevelopmental or cognitive effects in humans; however, like histopathology, it may help in identifying changes in the brain occurring at lower concentrations of fluoride. Oxidative stress

in the brain was evaluated in 25 studies that examined concentrations at or below 20 ppm fluoride, of which 15 studies had low potential for bias (Adedara et al. 2017a; Adedara et al. 2017b; Akinrinade et al. 2015b; Bartos et al. 2018; Chouhan and Flora 2008; Chouhan et al. 2010; Gao et al. 2008a; Gao et al. 2009; Guner et al. 2016; Khan et al. 2017; Mesram et al. 2016; Nageshwar et al. 2018; Nkpaa and Onyeso 2018; Shan et al. 2004; Zhang et al. 2015a). All of the low risk-of-bias studies were conducted in rats (mainly Wistar or Sprague-Dawley) and administered fluoride via drinking water with exposure durations ranging from 28 days to 7 months. Although there was heterogeneity in the endpoints reported (i.e., varying measures of protein oxidation, antioxidant activity, lipid peroxidation, and reactive oxygen species [ROS]) and some variation in the consistency of the evidence based on the endpoint, the majority of the studies (13 of 15) (Adedara et al. 2017a; Adedara et al. 2017b; Akinrinade et al. 2015b; Bartos et al. 2018; Gao et al. 2008a; Gao et al. 2009; Guner et al. 2016; Khan et al. 2017; Mesram et al. 2016; Nageshwar et al. 2018; Nkpaa and Onyeso 2018; Shan et al. 2004; Zhang et al. 2015a) found evidence of oxidative stress in the brains of rats treated with fluoride at concentrations at or below 20 ppm, of which 10 studies reported oxidative stress in the brain below 10 ppm fluoride. The most consistent evidence of oxidative stress in the brain was reported through changes in antioxidant activity. Eleven of the 12 low risk-of-bias studies that evaluated antioxidant activity reported an effect at concentrations at or below 20 ppm (Adedara et al. 2017a; Adedara et al. 2017b; Akinrinade et al. 2015b; Bartos et al. 2018; Gao et al. 2008a; Gao et al. 2009; Guner et al. 2016; Khan et al. 2017; Mesram et al. 2016; Nageshwar et al. 2018; Nkpaa and Onyeso 2018). Decreases in antioxidant activity using measures of superoxide dismutase (SOD) activity were reported in seven of eight low risk-of-bias studies (Adedara et al. 2017a; Adedara et al. 2017b; Akinrinade et al. 2015b; Khan et al. 2017; Mesram et al. 2016; Nageshwar et al. 2018; Nkpaa and Onyeso 2018), and, among these seven studies, all that also measured changes in catalase (CAT) activity (n = 6 studies) also reported decreased activity (Adedara et al. 2017a; Adedara et al. 2017b; Khan et al. 2017; Mesram et al. 2016; Nageshwar et al. 2018; Nkpaa and Onyeso 2018). A decrease in total antioxidant capacity (T-AOC) as a measure of antioxidant activity was also consistently reported in two low risk-of-bias studies (Gao et al. 2008a; Gao et al. 2009), and a decrease in glutathione peroxidase (GPx) activity was reported in two of three low risk-of-bias studies (Adedara et al. 2017b; Nkpaa and Onyeso 2018).

Relative to the above-mentioned studies, 2 of the 15 low risk-of-bias studies (Chouhan and Flora 2008; Chouhan et al. 2010) did not observe statistically significant effects on oxidative stress in the brain with concentrations at or below 20 ppm fluoride; however, the measure of oxidative stress evaluated in Chouhan and Flora (2008) and Chouhan et al. (2010) (glutathione [GSH] to oxidized glutathione [GSSG] ratio as an indication of antioxidant activity and ROS levels) were not evaluated in any other low risk-of-bias study. Chouhan and Flora (2008) observed a dose-dependent increase in ROS levels associated with 10, 50, and 100 mg/L sodium fluoride in drinking water; however, results were not statistically significant at any dose. In Chouhan et al. (2010), the levels of ROS were significantly higher at 50 ppm sodium fluoride in drinking water, but statistical significance was not met at doses below 20 ppm fluoride (1 and 10 ppm sodium fluoride) or at 100 ppm sodium fluoride; yet, hydrogen peroxide levels as a measure of ROS were found to be significantly increased at 15 ppm sodium fluoride in drinking water in studies conducted by another group of authors (Adedara et al. 2017a; Adedara et al. 2017b).

F.5. Apoptosis/Cell Death

Seven low risk-of-bias studies were identified that evaluated apoptosis with concentrations at or below 20 ppm fluoride. Results from these studies were inconsistent and were insufficient for evaluating fluoride-induced apoptosis. These data are insufficient to increase confidence or support a change to hazard conclusions.

F.6. Inflammation

Five low risk-of-bias studies were identified that evaluated potential effects of fluoride on inflammation with concentrations at or below 20 ppm. The inflammation markers were too heterogeneous or limited in number to make any determination on potential relevance of mechanism, even before limiting the review of the data to low risk-of-bias studies. These data are insufficient to increase confidence or support a change to hazard conclusions.

F.7. Thyroid

Seventeen studies were identified that evaluated potential effects of fluoride on the thyroid with concentrations at or below 20 ppm (see Figure F-1). These animal thyroid data are not further described because this endpoint has been directly evaluated in a number of human studies that have failed to identify consistent evidence to suggest that thyroid effects are a requisite mechanism by which fluoride causes neurodevelopmental or cognitive effects in humans.

Mechanism	Mechanism Subcategory	Direction of Effect ↑	↓	NS	Grand Total
Biochemistry (brain/neurons)	Carbohydrate/lipid-related		1	1	2
	Gene expression		1		1
	Protein levels associated with brain function	8	7	7	11
	Other biochemical	2			2
Neurotransmitters	Cholinesterase	2	7	3	11
	Dopamine and metabolites		1		1
	Other neurotransmitters	2	2	1	2
Oxidative stress	Antioxidant activity	1	10	4	12
	Lipid peroxidation byproduct	7		4	11
	Protein oxidation	2		1	2
	ROS	2		2	4

Figure F-2. Number of Low Risk-of-bias Animal Studies That Evaluated Biochemical, Neurotransmission, and Oxidative Stress Effects at or below 20 ppm by Mechanism Subcategory and Direction of Effect

An interactive version of Figure F-2 and additional study details in Tableau® (https://public.tableau.com/app/profile/ntp.visuals/viz/Fluoride_Animal_SelectMechanisms_2021/FigureA5-2). This figure displays study counts for low risk-of-bias studies, as these counts are most relevant to the text in this section. Counts for high risk-of bias studies or all studies combined can be accessed in the interactive figure in Tableau®. Study counts are tabulated by significance—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with at least one endpoint in the mechanistic subcategory with significantly increasing results at fluoride exposure levels of ≤20 ppm. These columns are not mutually exclusive (i.e., a study may report on multiple endpoints with varying results within a single mechanistic subcategory and therefore may be reflected in the counts for the "↑", "↓", and NS columns but would be counted only once in the Grand Total column). Endpoints, species, strain, sex, and exposure duration are available for each study in the interactive figure in Tableau®.

Appendix G. Protocol History and Revisions

Date	Activity or Revision
December 14, 2016	**Draft evaluation protocol reviewed:** sent to technical advisors for peer review
April 10, 2017	**Draft human risk-of-bias protocol reviewed:** sent to technical advisors for peer review
May 2, 2017	**Draft animal risk-of-bias protocol reviewed:** sent to technical advisors for peer review
June 2017	**Evaluation protocol finalized:** Review protocol finalized for use and posting
May 29, 2019	**Revised protocol:** Revised review protocol posted
September 16, 2020	**Revised protocol:** Revised review protocol posted

National Toxicology Program
U.S. Department of Health and Human Services

DRAFT NTP MONOGRAPH ON THE

STATE OF THE SCIENCE CONCERNING FLUORIDE EXPOSURE AND NEURODEVELOPMENTAL AND COGNITIVE HEALTH EFFECTS: A SYSTEMATIC REVIEW

October 5, 2021

Division of the National Toxicology Program
National Institute of Environmental Health Sciences
National Institutes of Health

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOREWORD

The National Toxicology Program (NTP), established in 1978, is an interagency program within the Public Health Service of the U.S. Department of Health and Human Services. Its activities are executed through a partnership of the National Institute for Occupational Safety and Health (part of the Centers for Disease Control and Prevention), the Food and Drug Administration (primarily at the National Center for Toxicological Research), and the National Institute of Environmental Health Sciences (part of the National Institutes of Health), where the program is administratively located. NTP offers a unique venue for the testing, research, and analysis of agents of concern to identify toxic and biological effects, provide information that strengthens the science base, and inform decisions by health regulatory and research agencies to safeguard public health. NTP also works to develop and apply new and improved methods and approaches that advance toxicology and better assess health effects from environmental exposures.

NTP conducts literature-based evaluations to determine whether exposure to environmental substances (e.g., chemicals, physical agents, and mixtures) may be associated with adverse health effects. These evaluations result in hazard conclusions or characterize the extent of the evidence and are published in the NTP Monograph series, which began in 2011. NTP Monographs serve as an environmental health resource to provide information that can used to make informed decisions about whether exposure to a substance may be of concern for human health.

NTP conducts these health effects evaluations following pre-specified protocols that apply the general methods outlined in the "Handbook for Conducting a Literature-Based Health Assessment Using the OHAT Approach for Systematic Review and Evidence Integration."[1] The protocol describes project-specific procedures tailored to each systematic review in a process that facilitates evaluation and integration of scientific evidence from published human, experimental animal, and mechanistic studies.

Systematic review procedures are not algorithms and the methods require scientific judgments. The key feature of the systematic review approach is the application of a transparent framework to document the evaluation methods and the basis for scientific judgements. This process includes steps to comprehensively search for studies, select relevant evidence, assess individual study quality, rate confidence in bodies of evidence across studies, and then integrate evidence to develop conclusions for the specific research question. Draft monographs undergo external peer review prior to being finalized and published.

NTP Monographs are available free of charge on the NTP website and cataloged in PubMed, a free resource developed and maintained by the National Library of Medicine (part of the National Institutes of Health). Data for these evaluations are included in the Health Assessment and Workspace Collaborative.

For questions about the monographs, please email NTP or call 984-287-3211.

[1]OHAT is the abbreviation for Office of Health Assessment and Translation, which has become the Health Assessment and Translation group in the Integrative Health Assessment Branch of the Division of the National Toxicology Program at the National Institute of Environmental Health Sciences.

PREFACE

The National Toxicology Program (NTP) conducted a systematic review of the published scientific literature because of public concern for the potential association between fluoride exposure and adverse neurodevelopmental and cognitive health effects.

NTP initially published a systematic review of the experimental animal literature in 2016 that was subsequently expanded to include human epidemiology studies, mechanistic studies, and newer experimental animal literature. Because of the high public interest in fluoride's benefits and potential risks, the NTP asked the National Academies of Science, Engineering, and Medicine (NASEM) to conduct an independent evaluation of the *draft NTP Monograph on Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects*" (September 6, 2019) and the revised draft (September 16, 2020), which addressed the NASEM committee's recommendations for improvement. The NASEM committee determined that, "Overall the revised monograph seems to include a wealth of evidence and a number of evaluations that support its main conclusion, but the monograph falls short of providing a clear and convincing argument that supports its assessments…" Thus, the NTP has removed the hazard assessment step and retitled this systematic review of fluoride exposure and neurodevelopmental and cognitive health effects as a "state-of-the-science" document to indicate the change. This state-of-the-science document does not include the meta-analysis of epidemiology studies or hazard conclusions found in previous draft monographs; however, it provides a comprehensive and current assessment of the scientific literature on fluoride as an important resource to inform safe and appropriate use.

The NTP has responded to the NASEM committee's comments on the revised draft (September 16, 2020) in a separate document (place holder for URL) and revised relevant sections of this monograph.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

TABLES:

FIGURES:

DATA FIGURES:

RISK-OF-BIAS FIGURES:

ANIMAL MECHANISTIC DATA FIGURES:

ABSTRACT

Background: A 2006 evaluation by the National Research Council (NRC) found support for an association between consumption of high levels of naturally occurring fluoride in drinking water and adverse neurological effects in humans and recommended further investigation.[2] The evidence reviewed at the time was from dental and skeletal fluorosis-endemic regions of China. Since the NRC review, the number and location of studies examining cognitive and neurobehavioral effects of fluoride in humans has grown considerably, including several recent North American prospective cohort studies evaluating prenatal fluoride exposures.

NTP previously published a systematic review of the evidence from experimental animal studies of the effects of fluoride on learning and memory in 2016. The previous systematic review found a low-to-moderate level of evidence that learning and memory deficits occur in non-human mammals exposed to fluoride.

Objective: To conduct a systematic review of the human, experimental animal, and mechanistic literature to evaluate the extent and quality of the evidence linking fluoride exposure to neurodevelopmental and cognitive effects in humans.

Method: A systematic review protocol was developed and utilized following the Office of Health Assessment and Translation's (OHAT's) standardized systematic review approach for conducting literature-based health assessments. This review addresses the state of the science with regard to whether exposure to fluoride could present a potential hazard (i.e., has the potential to cause harm at any exposure level). Benefits of fluoride with respect to oral health are not addressed in this monograph.

Results: Examination of newer experimental animal literature, including studies carried out at the NTP, did not provide information that adds clarity to the findings of the 2016 review. Eight low risk-of-bias studies evaluated fluoride exposure and mechanistic data in humans, including thyroid hormone levels in adults and/or children, thyroid conditions in children and/or adults, and thyroid diseases in adults. The findings of these studies and other human studies of brain histopathology or other biochemical changes do not provide evidence of a consistent mechanism by which fluoride may cause adverse neurological effects.

This systematic review identified studies of cognitive or neurodevelopmental effects in both adults and children, which were evaluated separately. In adults, only two cross-sectional studies examining cognitive effects were available. The literature in children was more extensive and was separated into studies assessing IQ and studies assessing other cognitive neurodevelopmental outcomes. Eight of nine high quality studies examining other cognitive neurodevelopmental outcomes reported lower performance in a neurological outcome associated with fluoride exposure. Sixty-six studies assessed the association between fluoride exposure and IQ in children. Of the 19 high quality IQ studies in children, 18 reported an inverse association with fluoride exposure. These 18 studies, which include 3 prospective cohort studies and 15 cross-sectional studies, were conducted in 5 different countries. Forty-one of the

[2]NRC (National Research Council). 2006. *Fluoride in Drinking Water: A Scientific Review of EPA's Standards*. Washington, DC: The National Academies Press. Available at: http://www.nap.edu/catalog/11571/fluoride-in-drinking-water-a-scientific-review-of-epas-standards.

47 low quality studies in children also found evidence of an inverse relationship between fluoride exposure and IQ.

Discussion: The animal studies provide little insight into the question of whether fluoride exposure affects IQ. Human mechanistic studies were too heterogenous and limited in number to make any determination on biological plausibility. The body of evidence from studies on adults is also limited and provides low confidence that fluoride exposures are associated with adverse effects on adult cognition. There is, however, extensive literature on IQ effects in children. There is also some evidence that fluoride exposure is associated with other neurodevelopmental and cognitive effects although, because of the heterogeneity of the outcomes, there is low confidence in the literature for these other effects. The body of evidence from epidemiological studies that assess IQ in children is large, presents a consistent pattern of effects, and provides moderate confidence that fluoride exposures are associated with lower IQ in children.

INTRODUCTION

The NTP's Office of Health Assessment and Translation (OHAT) conducted a systematic review to evaluate the evidence that exposure to fluoride is associated with neurodevelopmental or cognitive effects. There are numerous human and animal studies reporting neurodevelopmental and cognitive health effects of exposure to excess fluoride. As noted by the National Research Council (NRC) in their 2006 report, although the studies lacked sufficient detail to fully assess their quality and relevance to the U.S. populations, the consistency of the results suggesting that fluoride may be neurotoxic warrants additional research (NRC 2006).

Fluoride salts are added to community water systems and dental products in the United States (e.g., toothpaste, mouth rinses, and supplements) for the prevention of dental caries. Approximately 67% of the U.S. population receives fluoridated water through a community drinking water system (CDC 2013). In other countries fluoride supplementation has been achieved by fluoridating food products such as salt, or milk. Fluoride supplementation has been recommended to prevent bone fractures (Jones *et al.* 2005). Fluoride also can occur naturally in drinking water. Other sources of human exposure include other foods and beverages, industrial emissions, pharmaceuticals, and pesticides (e.g., cryolite, sulfuryl fluoride). Soil ingestion is another source of fluoride exposure in young children (US EPA 2010).

The U.S. Public Health Service (PHS) first recommended communities add fluoride to drinking water in 1962. PHS guidance is advisory, not regulatory, which means that while PHS recommends community water fluoridation as a public health intervention, the decision to fluoridate water systems is made by state and local governments.[3] For community water systems that add fluoride, PHS now recommends a fluoride concentration of 0.7 milligrams/liter (mg/L) (equal to 0.7 parts per million [ppm]). Under the Safe Drinking Water Act, the U.S. Environmental Protection Agency (EPA) sets maximum exposure level standards for drinking water quality. The current enforceable drinking water standard for fluoride, or the maximum contaminant level (MCL), is 4.0 mg/L. This level is the maximum amount of fluoride contamination (naturally occurring not from water fluoridation) that is allowed in water from public water systems and is set to protect against increased risk of skeletal fluorosis, a condition characterized by pain and tenderness of the major joints. EPA also has a non-enforceable secondary drinking water standard of 2.0 mg/L, which is recommended to protect children against the tooth discoloration and/or pitting that can be caused by severe dental fluorosis during the formative period prior to eruption of the teeth. Although the secondary standard is not enforceable, EPA requires that public water systems notify the public if the average levels exceed 2.0 mg/L (NRC 2006). The World Health Organization (WHO) set a safe water guideline of 1.5 mg/L (first established in 1984 and reaffirmed in 1993 and 2011), which is recommended to protect against increasing risk of dental and skeletal fluorosis (WHO 2011).

As of April 2020, 1.08% of persons living in the United States (~ 3.5 million people) were served by community water systems (CWS) containing ≥ 1.1 mg/L naturally occurring fluoride. CWS supplying water with ≥ 1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population (~ 1.9 million

[3]For many years, most fluoridated community water systems used fluoride concentrations ranging from 0.8 to 1.2 mg/L (US DHHS 2015).

people), and systems supplying water with ≥ 2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people) (https://www.cdc.gov/fluoridation/data-tools/reporting-system.html).

Commonly cited health concerns related to fluoride are bone fractures and skeletal fluorosis, lower intelligence quotient (IQ) and other neurological effects, cancer, and endocrine disruption. Effects on neurological function, endocrine function (e.g., thyroid, parathyroid, pineal), metabolic function (e.g., glucose metabolism), and carcinogenicity were assessed in the 2006 NRC report Fluoride in Drinking Water: A Scientific Review of EPA's Standards (NRC 2006). The NRC review considered adverse effects of water fluoride, focusing on a range of concentrations (2–4 mg/L) above the current 0.7-mg/L recommendation for community water fluoridation. The NRC report concluded that the Maximum Contaminant Level Goal (MCLG), also 4 mg/L should be lowered to protect against severe enamel fluorosis and to reduce the risk of bone fractures associated with skeletal fluorosis (NRC 2006). Other than severe fluorosis, the NRC did not find sufficient evidence of negative health effects at fluoride levels below 4.0 mg/L; however, the NRC concluded that the consistency of the results of IQ deficits in children exposed to fluoride at 2.5 to 4 mg/L in drinking water from a few epidemiological studies of Chinese populations appeared significant enough to warrant additional research on the effects of fluoride on intelligence. The NRC report noted several challenges to evaluating the literature, citing deficiencies in reporting quality, lack of consideration of all sources of fluoride exposure, incomplete consideration of potential confounding, selection of inappropriate control subject populations in epidemiological studies, absence of demonstrated clinical significance of reported endocrine effects, and incomplete understanding of the biological relationship between histological, biochemical, and molecular alterations with behavioral effects (NRC 2006).

In 2016, NTP conducted a systematic review of the evidence from experimental animal studies on the potential effects of fluoride exposure on learning and memory (NTP 2016). The NTP (2016) systematic review found a low-to-moderate level of evidence that learning and memory deficits occur in experimental animals exposed to fluoride. Based on the findings in NTP (2016), NTP decided to conduct additional animal studies before carrying out a full systematic review to incorporate human, animal, and potentially relevant mechanistic evidence in order to reach hazard identification conclusions for fluoride and learning and memory effects. As the NTP (2016) report on the experimental animal evidence focused on learning and memory and developed confidence ratings for bodies of evidence by life stage of exposure (i.e., exposure during development or adulthood), this report also evaluates two different age groups in humans (i.e., children and adults) with a focus on cognitive neurodevelopmental effects in children and cognitive effects in adults in order to address potential differences in the health impact based on timeframe of exposure (i.e., during development or during adulthood). The evaluation of experimental animal studies in this report has been conducted separately from the 2016 experimental animal assessment, but like the 2016 assessment, it has assessed mainly learning and memory effects in experimental animal studies to determine whether the findings inform the assessment of cognitive neurodevelopmental effects in children and cognitive effects in adults.

A committee convened by the National Academy of Sciences, Engineering, and Medicine (NASEM) reviewed earlier drafts of this monograph (September 6, 2019 and September 16, 2020) (NASEM 2020, 2021). The current document incorporates changes in regard to those reviews, and responses to the 2020 review are available at (placeholder to cite NTP 2021 Response to NASEM comments).

OBJECTIVE AND SPECIFIC AIMS

Objective

The overall objective of this evaluation was to undertake a systematic review to develop NTP hazard identification conclusions on the association between exposure to fluoride and neurodevelopmental and cognitive effects based on assessing levels of evidence from human and non-human animal studies in consideration of the degree of support from mechanistic data. However, the NASEM Committee reviews (NASEM 2020, 2021) of the 2019 and 2020 drafts of the monograph indicated that, "Overall the revised monograph seems to include a wealth of evidence and a number of evaluations that support its main conclusion, but the monograph falls short of providing a clear and convincing argument that supports its assessments…" For this reason, the methods were revised to remove the hazard assessment step (i.e., the section to Integrate Evidence to Develop Hazard Identification Conclusions). In addition, a meta-analysis of the epidemiology studies examining children's IQ in relation to fluoride exposure added to the 2020 draft in response to NASEM comments (NASEM 2020) will be submitted as a separate publication and is not part of this document.

Therefore, the objective of this revised document is to undertake a systematic review of the literature concerning the association of fluoride exposure and neurodevelopmental and cognitive effects based on evaluating evidence from human and non-human animal studies with consideration of mechanistic understanding.

Specific Aims

- Identify literature that assessed neurodevelopmental and cognitive health effects, especially outcomes related to learning, memory, and intelligence, following exposure to fluoride in human, animal, and relevant in vitro/mechanistic studies.
- Extract data on potential neurodevelopmental and cognitive health effects from relevant studies.
- Assess the internal validity (risk of bias) of individual studies using pre-defined criteria.
- Assess effects on thyroid function to help evaluate potential mechanisms of impaired neurological function.
- Summarize the extent and types of health effects evidence available.
- Describe limitations of the systematic review, strengths and limitations of the evidence base, identify areas of uncertainty, as well as data gaps and research needs for neurodevelopmental and cognitive health effects of fluoride.

Dependent on the extent and nature of the available evidence:

- Synthesize the evidence using a narrative approach.
- Rate confidence in the body of evidence for human and animal studies separately according to one of four statements: High, Moderate, Low, or Very Low/No Evidence Available.

METHODS

Problem Formulation and Protocol Development

The research question and specific aims stated above were developed and refined through a series of problem formulation steps including:

(1) receipt of nomination from the public in June 2015 to conduct analyses of fluoride and developmental neurobehavioral toxicity;

(2) analysis of the extent of evidence available and the merit of pursuing systematic reviews, given factors such as the extent of new research published since previous evaluations and whether these new reports address or correct the deficiencies noted in the literature (NRC 2006, OEHHA 2011, SCHER 2011);

(3) request for information in a Federal Register notice (dated October 7, 2015);

(4) consideration of comments providing a list of studies to review through Federal Register notice and public comment period from October 7, 2015 to November 6, 2015;

(5) release of draft concept titled *Proposed NTP Evaluation on Fluoride Exposure and Potential for Developmental Neurobehavioral Effects* in November 2015;

(6) presentation of draft concept at the NTP Board of Scientific Counselors (BSC) meeting on December 1−2, 2015;

(7) consideration of comments on NTP's draft concept from the NTP BSC meeting in December 2015; and

(8) consideration of input on the draft protocol from review by technical advisors.

The protocol used to conduct this systematic review was posted in June 2017 with updates posted in May 2019 and September 2020 (https://ntp.niehs.nih.gov/go/785076).[4] The protocol served as the complete methods followed for the conduct of this systematic review. The OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673) is a source of general systematic review methods that were selected and tailored in developing this protocol. Options in the OHAT handbook that were not specifically referred to in the protocol were not part of the methods for the systematic review.

A brief summary of the methods is presented below. Although the methods were revised to remove the hazard assessment step and meta-analysis from this document, the protocol was not further revised.

PECO Statements

PECO (**P**opulation, **E**xposure, **C**omparators and **O**utcomes) statements were developed as an aid to identify search terms and inclusion/exclusion criteria as appropriate for addressing the overall research question (effects on neurodevelopmental or cognitive function and thyroid associated with fluoride

[4]NTP conducts systematic reviews following pre-specified protocols that describe the review procedures selected and applied from the general methods outlined in the OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673). The protocol describes project-specific procedures tailored to each systematic review that supersede the methods in the OHAT Handbook.

exposure) for the systematic review (Higgins and Green 2011).The PECO statements are listed below for human, animal, and in vitro/mechanistic studies (see **Table 1**, **Table 2**, and **Table 3**).

Using the PECO statements, the evaluation searched for evidence of neurodevelopmental or cognitive function, and thyroid effects associated with fluoride exposure from human studies, controlled exposure animal studies, and mechanistic/in vitro studies. Mechanistic data can come from a wide variety of studies that are not intended to identify a disease phenotype. This source of experimental data includes in vitro and in vivo laboratory studies directed at cellular, biochemical, and molecular mechanisms that attempt to explain how a substance produces particular adverse health effects. The mechanistic data were first organized by general categories (e.g., biochemical effects in the brain and neurons, neurotransmitters, oxidative stress, etc.) to evaluate the information available. Categories focused on were those with more robust data at levels of fluoride more relevant to human exposure. The intent was not to develop a mechanism for fluoride induction of learning and memory effects, but to evaluate whether a plausible series of mechanistic events exists to support effects observed in the low-dose region (below approximate drinking water equivalent concentrations of 20 ppm for animal studies) that may strengthen the hazard conclusion.

Table 1. Human PECO (Population, Exposure, Comparator and Outcome) Statement

PECO Element	Evidence
Population	Humans without restriction as to age or sex, geographic location, or life stage at exposure or outcome assessment
Exposure	Exposure to fluoride based on administered dose or concentration, biomonitoring data (e.g., urine, blood, other specimens), environmental measures (e.g., air, water levels), or job title or residence. Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; Chemical Abstracts Service Registry Number [CASRN] 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable populations not exposed to fluoride or exposed to lower levels of fluoride (e.g., exposure below detection levels)
Outcomes	Neurodevelopmental outcomes including learning, memory, intelligence, other forms of cognitive behavior, other neurological outcomes (e.g., anxiety, aggression, motor activity), and biochemical changes in the brain or nervous system tissue; or measures of thyroid function, biochemical changes, or thyroid tissue

Table 2. Animal PECO Statement	
PECO Element	**Evidence**
Population	Non-human mammalian animal species (whole organism)
Exposure	Exposure to fluoride based on administered dose or concentration, and biomonitoring data (e.g., urine, blood, other specimens). Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; CASRN 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable animals that were untreated or exposed to vehicle-only treatment
Outcomes	Neurodevelopmental outcomes including learning, memory, intelligence, other forms of cognitive behavior, other neurological outcomes (e.g., anxiety, aggression, motor activity), and biochemical changes in the brain or nervous system tissue; or measures of thyroid function, biochemical changes, or thyroid tissue

Table 3. In Vitro/Mechanistic PECO Statement	
PECO Element	**Evidence**
Population	Human or animal cells, tissues, or biochemical reactions (e.g., ligand binding assays)
Exposure	Exposure to fluoride based on administered dose or concentration. Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; CASRN 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable cells or tissues that were untreated or exposed to vehicle-only treatment
Outcomes	Endpoints related to neurological and thyroid function, including neuronal electrophysiology; mRNA, gene, or protein expression; cell proliferation or death in brain or thyroid tissue/cells; neuronal signaling; synaptogenesis, etc.

Literature Search

Main Literature Search

Search terms were developed to identify all relevant published evidence on developmental neurobehavioral toxicity or thyroid-related health effects potentially associated with exposure to fluoride by reviewing Medical Subject Headings for relevant and appropriate neurobehavioral and thyroid-related terms, and by extracting key neurological and thyroid-related health effects and

developmental neurobehavioral terminology from reviews and a sample of relevant studies[5]. A combination of relevant subject headings and keywords were subsequently identified. A test set of relevant studies was used to ensure the search terms retrieve 100% of the test set. Six electronic databases were searched (see **Main Literature Database Search**) using a search strategy tailored for each database (specific search terms used for the PubMed search are presented in **Appendix 1**; the search strategy for other databases are available in the protocol (https://ntp.niehs.nih.gov/go/785076). A search of PubChem indicated that sodium fluoride was not found in either the Tox21 or ToxCast databases; therefore, these databases were not included in the search. No language restrictions or publication year limits were imposed. These six databases were searched in December 2016 and the search was regularly updated during the review process through April 1, 2019.

An additional search was conducted on May 1, 2020, where human epidemiology studies with primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence) were prioritized during screening. The review of the 2020 search results focused only on the human studies because they formed the basis of the conclusions in the September 6, 2019, draft. A supplemental literature search of Chinese-language databases (described below) was also conducted.

Publications identified in these searches are categorized as "references identified through database searches" in **Figure 2.** Studies identified from other sources or manual review that might impact conclusions were considered under "references identified through other sources" in **Figure 2**. Literature searches for this systematic review were conducted independently from the literature search conducted for NTP (2016). The current literature search strategy was based on the search terms used for NTP (2016) and refined for the current evaluation, including the addition of search terms to identify human studies. Although the review process identified studies prior to 2015, the current assessment did not evaluate the studies published prior to 2015 and relied on the NTP (2016) assessment. The focus of the literature searches for this systematic review was to identify and evaluate relevant animal studies that were published since completion of the literature searches for the NTP (2016) assessment in addition to the human and mechanistic data that were not previously evaluated.

Supplemental Chinese Database Literature Search

In order to identify non-English-language studies that might not appear in databases for the main literature search, additional searches were developed for non-English-language databases. We were unable to find definitive guidance on the most comprehensive, highest quality, or otherwise most appropriate non-English-language databases for health studies of fluoride. Therefore, we chose databases that identified non-English-language studies that we were aware of – those previously identified from other resources (e.g., Chinese-language studies from the Fluoride Action Network website). Multiple non-English language databases were explored before finding two databases (CNKI and Wanfang) that covered studies previously identified from other sources. These two Chinese electronic databases were searched in May 2020 with no language restrictions or publication year limits. Search terms from the main literature search were refined to focus on human epidemiology studies. The CNKI and Wanfang databases have character limits in the search strings; therefore, key terms were prioritized using text analytics to identify the most prevalent terms from neurodevelopmental or cognitive human epidemiology studies previously identified as relevant. Search strings were designed to capture known relevant studies that were previously identified from searching other resources without identifying large numbers of non-relevant studies [the search strategy for both databases is available in

[5]The terms "study" and "publication" are used interchangeably in this document to refer to a published work drawn from an original body of research conducted on a defined population.

the protocol (https://ntp.niehs.nih.gov/go/785076)]. Publications retrieved were compared to publications retrieved from the main literature search and duplicates were removed. The remaining relevant publications are categorized as "references identified through database searches" in **Figure 2**.

New animal and mechanistic references retrieved were scanned for evidence that might extend the information currently in the September 6, 2019 draft. Although additional studies were identified, data that would materially advance the animal and mechanistic findings were not identified; therefore, these studies were not extracted nor were they added to the draft. A primary goal of the screening of the newly-retrieved human references in the supplemental search of Chinese databases was to identify null, or no-effect, studies that evaluated primary neurodevelopmental or cognitive outcomes (i.e., learning, memory, and intelligence) that may have been missed in previous searches that did not include the Chinese databases. NTP also wanted to examine whether the non-English studies on the Fluoride Action website had been selectively presented to only list studies reporting effects of fluoride; therefore, identifying null, or no-effect, studies was of particular interest. Newly-retrieved human references were reviewed to identify studies that might impact conclusions with priority given to identifying and translating null studies that may have been missed using previous approaches. Null studies that were identified were translated and included.

Databases Searched

Main Literature Database Search

- BIOSIS (Thomson Reuters)
- EMBASE
- PsycINFO (APA PsycNet)
- PubMed (NLM)
- Scopus (Elsevier)
- Web of Science (Thomson Reuters, Web of Science indexes the journal Fluoride)

Supplemental Chinese Database Literature Search

- CNKI
- Wanfang

Searching Other Resources

The reference lists of all included studies; relevant reviews, editorials and commentaries; and the Fluoride Action Network website (http://fluoridealert.org/) were manually searched for additional relevant publications.

Unpublished Data

Unpublished data were eligible for inclusion provided the owner of the data was willing to have the data made public and peer reviewed (see protocol for more details https://ntp.niehs.nih.gov/go/785076).

Study Selection

Evidence Selection Criteria

In order to be eligible for inclusion, studies had to satisfy eligibility criteria that reflect the PECO statement in **Table 1**, **Table 2**, and **Table 3**. The following additional exclusion criteria were applied (see protocol for additional details; https://ntp.niehs.nih.gov/go/785076):

(1) Case studies and case reports.

(2) Articles without original data (e.g., reviews, editorials, or commentaries). Reference lists from these materials, however, were reviewed to identify potentially relevant studies not identified from the database searches. New studies identified were assessed for eligibility for inclusion.

(3) Conference abstracts, theses, dissertations, and other non-peer reviewed reports.

Screening Process

References retrieved from the literature search were independently screened by two trained screeners at the title and abstract level to determine whether a reference met the evidence selection criteria. Screening procedures following the evidence selection criteria in the protocol were pilot-tested with experienced contract staff overseen by NTP. For citations with no abstract or non-English abstracts, articles were screened based on title relevance (title would need to indicate clear relevance); number of pages (articles ≤ 2 pages were assumed to be conference reports, editorials, or letters unlikely to contain original data); and/or PubMed Medical Subject Headings (MeSH). Using this approach, literature was manually screened for relevance and eligibility against the evidence selection criteria using a structured form in SWIFT-Active Screener. While the human screeners review studies, SWIFT-Active Screener aids in the process by employing a machine-learning software program used to priority-rank studies for screening (Howard *et al.* 2020). SWIFT-Active Screener also refines a statistical model that continually ranks the remaining studies according to their likelihood for inclusion. In addition, SWIFT-Active Screener employs active learning to continually incorporate user feedback during title and abstract screening to predict the total number of included studies, thus providing a statistical basis for a decision about when to stop screening (Miller *et al.* 2016). Title and abstract screening was stopped once the statistical algorithm in SWIFT-Active Screener estimated that 98% of the predicted number of relevant studies were identified.

Studies that were not excluded during the title and abstract screening were further screened for inclusion with a full-text review by two independent reviewers using DistillerSR® by Evidence Partners, a web-based, systematic-review software program with structured forms and procedures to ensure standardization of the process. Screening conflicts were resolved through discussion and consultation with technical advisor(s), if necessary. During full-text review, studies that were considered relevant were tagged to the appropriate evidence streams (i.e., human, animal, and/or in vitro). Studies tagged to human or animal evidence streams were also categorized by outcome as primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence); secondary neurobehavioral outcomes (anxiety, aggression, motor activity, or biochemical); or related to thyroid effects. In vitro data were tagged as being related to neurological effects or thyroid effects. Translation assistance was sought to assess the relevance of non-English studies. Following full-text review, the remaining studies were "included" and used for the evaluation.

Evaluation of SWIFT-Active Screener Results

During the initial title and abstract screening of 20,883 references using SWIFT-Active Screener, approximately 38%[6] of the 20,883 studies were manually screened in duplicate to identify an estimated

[6]Howard *et al.* (2020) evaluated the performance of the SWIFT-Active Screener methods for estimating total number of relevant studies using 26 diverse systematic review datasets that were previously screened manually by reviewers. The authors found that on average, 95% of the relevant articles were identified after screening 40% of the total reference list when using SWIFT-Active Screener. In the document sets with 5,000 or more references, 95% of the relevant articles were identified after screening 34% of the available references, on average, using SWIFT-Active Screener.

98.6% of the predicted number of relevant studies using the statistical algorithm in SWIFT-Active Screener (13,023 references were not screened). SWIFT-Active Screener predicted that there were 739 relevant studies during the initial title and abstract screening, of which 729 studies were identified and moved to full-text review. The SWIFT-Active statistical algorithm predicted that 10 relevant studies at the title and abstract level (10 represents 1.4% × 739 predicted relevant studies; or 739 predicted relevant studies minus 729 identified relevant studies during screening) were not identified by not screening the remaining 13,023 studies.

To further consider the impact of using SWIFT-Active Screener for this systematic review, NTP evaluated the SWIFT-Active screening results to gain a better understanding of the relevance of the last group of studies that were screened before 98% predicted recall (i.e., 98% of the predicted number of relevant studies were identified). The goal was to determine the likelihood of having missed important studies by not screening all of the literature. To do this, NTP evaluated subsets of studies screened in SWIFT-Active for trends and followed those studies through to full-text review for a final determination of relevance and potential impact (i.e., whether the studies had data on primary outcomes). Based on this evaluation, NTP estimates that the use of SWIFT-Active Screener may have resulted in missing 1–2 relevant human studies and 1–2 relevant animal studies with primary neurodevelopmental or cognitive outcomes. Therefore, the use of SWIFT-Active Screener saved considerable time and resources and is expected to miss very few potentially relevant publications.

Screening of the May 2020 Literature Search Update

For the May 1, 2020, literature search, only primary human epidemiology studies were identified for data extraction. The study screening and selection process was focused on the human studies with primary outcomes for the evaluation because they form the basis of the conclusions. Animal in vivo, human secondary outcome-only, and human and animal mechanistic references were identified as part of the screening process. These studies were then scanned for evidence that might extend the information in the September 6, 2019 draft. All included studies from the May 2020 literature search update appear in **Appendix 2**; however, other than the primary human epidemiology studies, data from the new studies were not extracted unless it was believed they would materially advance the findings.

Note that NTP is aware of a conference abstract by Santa-Marina *et al.* on a Spanish cohort study that looked at fluoride exposure and neuropsychological development in children (Santa-Marina *et al.* 2019). NTP conducted a targeted literature search in April 2021 to see if the data from this study had been published. When no publication was found, NTP contacted the study authors to inquire about the publication of their data. The response from the study authors indicated that the study report was being finalized but had not yet been sent to a journal for review; therefore, it was not considered here.

Supplemental Chinese Database Searches and Human Epidemiology Studies

Supplemental searches were conducted in non-English-language databases (CNKI and Wanfang). One focus of the screening of these supplemental search results was to identify null or no-effect studies that evaluated primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence) that may have been missed in previous approaches or may have been absent from the Fluoride Action Network website. Of the 906 references that were identified in the supplemental Chinese database searches, 13 relevant studies published in Chinese with primary neurological outcomes were identified during title and abstract screening (which were not identified through the main literature searches). Based on information in the titles and abstracts of these 13 studies, Kang *et al.* (2011) was the only null study with primary neurological outcomes that was identified through the supplemental Chinese database searches. NTP had this study translated to English, and the study was included. Note that Kang

et al. (2011) is also identified by the Fluoride Action Network as a null study, but their website does not include an English translation of the study. Full texts were not found for four studies after an extensive search. Among the eight studies for which full texts were retrieved, one study evaluated adults, and the remaining seven studies contained results that would likely add to the body of evidence showing an inverse association between fluoride exposure and IQ in children. An epidemiologist fluent in Chinese evaluated the key risk-of-bias questions for observational human studies (i.e., confounding, exposure characterization, and outcome assessment) for the seven studies considered likely to add to the body of evidence in children. The review indicated that all seven studies would fall into the high risk-of-bias category. Author inquiries were conducted in Chinese to obtain missing information relevant to the assessment of the key risk-of-bias questions, but these inquiries did not result in additional information that would alleviate the risk-of-bias concerns. Because the body of evidence is already large, and because time was a factor in the revision of the monograph, these studies were not translated or included as this information on additional high risk-of-bias studies would likely not materially advance the human findings.

Data Extraction

Data were collected (i.e., extracted) from included studies by one member of the evaluation team and checked by a second member of the team for completeness and accuracy. Any discrepancies in data extraction were resolved by discussion or consultation with a third member of the evaluation team.

Data extraction was completed using the Health Assessment Workspace Collaborative (HAWC), an open source and freely available web-based interface application.[7] Data extraction elements are listed separately for human, animal, and in vitro studies in the protocol (https://ntp.niehs.nih.gov/go/785076). Data for primary and secondary outcomes as well as thyroid hormone level data were extracted from human studies. Studies evaluating only goiters or thyroid size were not extracted. All primary outcomes and functional neurological secondary outcomes (e.g., motor activity) were extracted from animal studies identified since the NTP (2016) report. For animal mechanistic data, studies were tiered based on exposure dose (with preference given to fluoride drinking water equivalent exposures, which were calculated using the method described in the NTP (2016) report) of 20 ppm or less as deemed most relevant to exposures in humans), exposure duration or relevant time window (i.e., developmental), exposure route (with preference given to oral exposures over injection exposures), and commonality of mechanism (e.g., inflammation, oxidative stress, changes in neurotransmitters, and histopathological changes were considered pockets of mechanistic data). Data were not extracted from in vitro studies; however, these studies were evaluated for data that could inform the biological plausibility of the human and animal results. Thyroid data were also reviewed but not extracted. The data extraction results for included studies are publicly available and can be downloaded in Excel format through HAWC (https://hawcproject.org/assessment/405/). Methods for transforming and standardizing dose levels and results from behavioral tests in experimental animals are detailed in the protocol (https://ntp.niehs.nih.gov/go/785076).

In 2016, NTP conducted a systematic review of the evidence from experimental animal studies on the potential effects of fluoride exposure on learning and memory (NTP 2016). The literature searches for the current assessment identified and evaluated relevant animal studies published since the 2016

[7]HAWC (Health Assessment Workspace Collaborative): A Modular Web-based Interface to Facilitate Development of Human Health Assessments of Chemicals (https://hawcproject.org/portal/).

assessment and also included human and mechanistic data that were not previously evaluated. Although literature search activities for the current assessment identified experimental animal studies prior to 2015, the current assessment did not re-evaluate studies published prior to 2015 and relied on the NTP (2016) assessment.

Quality Assessment of Individual Studies

Risk of bias was assessed for individual studies using a tool developed by OHAT that outlines a parallel approach to evaluating risk of bias from human, animal, and mechanistic studies to facilitate consideration of risk of bias across evidence streams with common terms and categories. The risk-of-bias tool is comprised of a common set of 11 questions that are answered based on the specific details of individual studies to develop risk-of-bias ratings for each question. Study design determines the subset of questions used to assess risk of bias for an individual study (see **Table 4**). When evaluating the risk of bias for an individual study, the direction and magnitude of effect of any specific bias is considered.

Assessors were trained with an initial pilot phase undertaken to improve clarity of rating criteria and to improve consistency among assessors. Studies were independently evaluated by two trained assessors who answered all applicable risk-of-bias questions with one of four options in **Table 5** following pre-specified criteria detailed in the protocol (https://ntp.niehs.nih.gov/go/785076). The criteria describe aspects of study design, conduct, and reporting required to reach risk-of-bias ratings for each question and specify factors that can distinguish among ratings (e.g., what separates “definitely low” from “probably low” risk of bias).

Key Risk-of-bias Questions

In the OHAT approach, some risk-of-bias questions or elements are considered potentially more important when assessing studies because these issues are generally considered to have a greater impact on estimates of the effect size or on the credibility of study results in environmental health studies. There are three Key Questions for observational human studies: confounding, exposure characterization, and outcome assessment. Based on the complexity of the possible responses for these questions in epidemiology studies, considerations made and methods used for evaluating the Key Questions are provided below. There are also three Key Questions for experimental animal studies: randomization, exposure characterization, and outcome assessment. In addition, for animal developmental studies, failure to consider the litter as the unit of analysis was also a key risk-of-bias concern. When there was not enough information to assess the potential bias for a risk-of-bias question and authors did not respond to an inquiry for further information, a conservative approach was followed, and the studies were rated probably high risk of bias for that question.

Risk-of-bias Considerations for Human Studies

The risk of bias of individual studies in the body of evidence was considered in developing confidence ratings. The key risk-of-bias questions (i.e., confounding, exposure characterization, and outcome assessment for human studies) are discussed in the consideration of the body of evidence. For this assessment, the key risk-of-bias questions, if not addressed appropriately, are considered to potentially have the greatest impact on the results. The other risk-of-bias questions were also taken into consideration and were used to identify any other risk-of-bias concerns that may indicate serious issues with the studies. No study was excluded based on concerns for risk of bias; however, the low risk-of-bias studies generally drive conclusions on confidence in the results across the body of evidence. Human

evidence was evaluated with and without high risk-of-bias studies to assess the impact of these studies on confidence in the association.

High risk-of-bias studies: Studies rated probably high risk of bias for at least two key risk-of-bias questions or definitely high for any single question are considered studies with higher potential for bias (i.e., high risk-of-bias studies) and to be of low quality.

Low risk-of-bias studies: The remaining studies (i.e., other than the high risk-of-bias studies) were considered to have lower potential for bias (i.e., low risk of bias) and to be of high quality. **Appendix 4** describes strengths and limitations of the low risk-of-bias/high quality studies identified during the assessment and clarifies why they are considered to pose low risk of bias. Details on the statistical analyses are provided in the "Other potential threats" domain in order to evaluate the adequacy of the statistical approach for individual studies.

Given the number non-English-language studies in this assessment, the potential for the translation to introduce bias was examined as described below, and it was determined that translation of non-English studies did not impact evaluation of risk of bias. Thirty-two of 92 studies included in the entire human body of evidence on neurodevelopmental and cognitive effects were initially published in a foreign language (mainly Chinese) and were either translated and published in volume 41 of the journal *Fluoride* (n = 19) or were translated by the Fluoride Action Network (n = 13) (http://fluoridealert.org/researchers/translations/complete_archive/). Most of these studies were considered to have high potential for bias due to lack of information across the key risk-of-bias questions. Therefore, in order to assess if the lack of information relevant to key risk-of-bias concerns was the result of a loss in translation, the original Chinese publications and the translated versions of the five studies that had the most potential for being included in the low risk-of-bias group of studies were reviewed to determine if any of the risk-of-bias concerns could be addressed (An *et al.* 1992, Chen *et al.* 1991 [translated in Chen *et al.* 2008], Du *et al.* 1992 [translated in Du *et al.* 2008], Guo *et al.* 1991 [translated in Guo *et al.* 2008a], Li *et al.* 2009). For all five studies, the translations were determined to be accurate, and there was no impact of the translations on the key risk-of-bias concerns.

Confounding

Potential confounding variables and/or effect modifiers that were considered key for all studies, populations, and outcomes included child's age, child's sex, and socioeconomic status (e.g., maternal education, household income, marital status, crowding). Additional potential confounding variables and/or effect modifiers considered important for this evaluation depending on the study population and outcome included race/ethnicity; maternal demographics (e.g., maternal age, body mass index [BMI]); parental behavioral and mental health disorders (e.g., attention deficit hyperactivity disorder [ADHD], depression); smoking (e.g., maternal smoking status, secondhand tobacco smoke exposure); reproductive factors (e.g., parity); nutrition (e.g., BMI, growth, anemia); iodine deficiency/excess; minerals and other chemicals in water associated with neurotoxicity (e.g., arsenic, lead); maternal and paternal IQ; and quantity and quality of caregiving environment (e.g., Home Observation Measurement of the Environment [HOME] score). To be assigned a rating of probably low risk of bias for the key risk-of-bias question regarding confounding, studies were not required to address every potential confounder listed; however, studies were required to address the three key covariates for all studies, the potential for co-exposures if applicable (e.g., arsenic and lead, both of which could affect cognitive function), and any other potential confounders considered important for the specific study population and outcome. For example, studies of populations in China, India, and Mexico, where there is concern for exposures to high fluoride and high arsenic, were required to address arsenic. If the authors did not

directly specify that arsenic exposures were evaluated, groundwater quality maps were evaluated (https://www.gapmaps.org/Home/Public#) in order to identify areas of China, India, and Mexico where arsenic is a concern (Podgorski and Berg 2020). If no arsenic measurements were available for the area, the arsenic groundwater quality predictions from the global arsenic 2020 map were used (Podgorski and Berg 2020). If an area had less than 50% probability of having arsenic levels greater than 10 µg/L (the WHO guideline concentration), the area was considered not to have an issue with arsenic that needed to be addressed by the study authors.

Exposure

Exposure was assessed using a variety of methods in the human body of evidence. Studies provided varying levels of details on the methods used and employed different exposure characterization methods to group study subjects into exposed and reference groups. Exposure metrics included spot urine (from children or mothers during at least one trimester of gestation), serum, individual drinking water, intake from infant formula, estimated total exposure dose, municipal drinking water (with residence information), evidence of dental or skeletal fluorosis, area of residence (endemic versus a non-endemic fluorosis area with or without individual validation of exposure), burning coal (with or without fluoride), and occupation type.

Urinary fluoride levels measured during pregnancy and in children include all ingested fluoride and are considered a valid measure to estimate total fluoride exposure (Villa *et al.* 2010, Watanabe *et al.* 1995); however, the type and timing of urinary sample collection is important to consider. Urinary fluoride is thought to reflect recent exposure but can be influenced by the timing of exposure (e.g., when water was last consumed, when teeth were last brushed). When compared to 24-hour urine samples, spot urine samples are more prone to the influence of timing of exposure and can also be affected by differences in dilution; however, many studies attempted to account for dilution either using urinary creatinine or specific gravity. Good correlations between 24-hour samples and urinary fluoride concentrations from spot samples adjusted for urinary dilution have been described (Zohouri *et al.* 2006). Despite potential issues with spot urine samples, if authors made appropriate efforts to reduce the concern for bias (e.g., accounting for dilution), studies that used this metric were generally considered to have probably low risk of bias for exposure.

Analytical methods to measure fluoride in biological or water samples also varied, some of which included atomic absorption, ion selective electrode methods, colorimetric methods, or the hexamethyldisiloxane microdiffusion method. Individual-level measures of exposure were generally considered more accurate than group-level measures; however, using group-level measures (e.g., endemic versus non-endemic area) in an analysis was less of a concern if the study provided water or urine fluoride levels from some individuals to verify that there were differences in the fluoride exposure between groups. Studies that provided results by area but also reported individual urinary or serum fluoride concentrations or other biochemical measures, including dental fluorosis in the children or urinary levels in mothers during pregnancy, were considered to have probably low risk of bias.

Outcome

Studies included in this evaluation used a wide variety of methods to measure IQ and other cognitive effects. Measures of IQ were generally standardized tests of IQ; however, for these standardized methods to be considered low potential for bias they needed to be conducted in the appropriate population or modified for the study population. Because results of many of the tests to measure neurodevelopment and cognitive function can be subjective, it was important that the outcome assessors were blind to the fluoride exposure when evaluating the results of the tests. If the study reported that the assessor was blind to the exposure, this was assumed to mean that the outcome assessor did not have any knowledge of the exposure, including whether the study subjects were from high-fluoride communities. If cross-sectional studies collected biomarker measurements at the time of an IQ assessment, this was considered indirect evidence that the outcome assessor would not have knowledge of the fluoride exposure unless there was also potential for the outcome assessor to have knowledge of varying levels of fluoride by study area. In cases where the study did not specify that the outcome assessors were blind, the study authors were contacted and asked if the outcome assessors were blind to exposure. When authors responded and indicated that outcome assessors were blind to exposure or that it was not likely that they would have had knowledge of exposure, this was considered direct or indirect evidence, respectively, that blinding was not a concern for those studies.

Any discrepancies in ratings between assessors were resolved by a senior technical specialist and through discussion when necessary to reach the final recorded risk-of-bias rating for each question along with a statement of the basis for that rating. Members of the evaluation team were consulted for assistance if additional expertise was necessary to reach final risk-of-bias ratings based on specific aspects of study design or performance reported for individual studies. Study procedures that were not reported were assumed not to have been conducted, resulting in an assessment of “probably high” risk of bias. Authors were queried by email to obtain missing information and responses received were used to update risk-of-bias ratings.

Table 4. OHAT Risk-of-bias Questions and Applicability by Study Design						
Risk-of-bias Questions	**Experimental Animal***	**Human Controlled Trials****	**Cohort**	**Case-Control**	**Cross-Sectional*****	**Case Series**
1. Was administered dose or exposure level adequately randomized?	X	X				
2. Was allocation to study groups adequately concealed?	X	X				
3. Did selection of study participants result in the appropriate comparison groups?			X	X	X	
4. Did study design or analysis account for important confounding and modifying variables?			X	X	X	X
5. Were experimental conditions identical across study groups?	X					
6. Were research personnel blinded to the study group during the study?	X	X				
7. Were outcome data complete without attrition or exclusion from analysis?	X	X	X	X	X	
8. Can we be confident in the exposure characterization?	X	X	X	X	X	X
9. Can we be confident in the outcome assessment (including blinding of outcome assessors)?	X	X	X	X	X	X
10. Were all measured outcomes reported?	X	X	X	X	X	X
11. Were there no other potential threats to internal validity?	X	X	X	X	X	X

*Experimental animal studies are controlled exposure studies. Non-human animal observational studies can be evaluated using the design features of observational human studies such as cross-sectional study design.

**Human Controlled Trials are studies in humans with controlled exposure (e.g., Randomized Controlled Trials, non-randomized experimental studies)

***Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).

Table 5. The Four Risk-of-bias Rating Options	
Answers to the risk-of-bias questions result in one of the following four risk-of-bias ratings	
++	***Definitely Low*** *risk of bias*: There is direct evidence of low risk-of-bias practices
+	***Probably Low*** *risk of bias*: There is indirect evidence of low risk-of-bias practices **OR** it is deemed that deviations from low risk-of-bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias
– NR	***Probably High*** *risk of bias*: There is indirect evidence of high risk-of-bias practices (indicated with "-") **OR** there is insufficient information provided about relevant risk-of-bias practices (indicated with "NR" for not reported). Both symbols indicate probably high risk of bias.
––	***Definitely High*** *risk of bias*: There is direct evidence of high risk-of-bias practices

Organizing and Rating Confidence in Bodies of Evidence

Health Outcome Categories for Neurodevelopmental and Cognitive Effects

After data were extracted from all studies, the health effects results within the category of neurodevelopmental or cognitive effects were grouped across studies to develop bodies of evidence or collections of studies with data on the same or related outcomes. The grouping of health effect results was not planned a priori. The vast majority of the human studies evaluated IQ in children as the single outcome; therefore, the discussion of cognitive neurodevelopmental effects in children focuses on IQ studies with supporting information from data on other endpoints. Cognitive function in adults was evaluated separately. Consistent with the NTP (2016) assessment, the primary focus within the animal study body of evidence was on animal studies with endpoints related to learning and memory.

Considerations for Pursuing a Narrative or Quantitative Evidence Synthesis

This evaluation provides only a narrative review of the data; however, heterogeneity within the available evidence was evaluated to determine if a quantitative synthesis (i.e., meta-analysis) is appropriate. Choi *et al.* (2012) and Duan *et al.* (2018) conducted meta-analyses and found that high fluoride exposure was associated with lower IQ scores. Choi *et al.* (2012) was able to determine a risk ratio for living in an endemic fluorosis area but was unable to develop a dose-response relationship. Duan *et al.* (2018) suggested a significant non-linear dose-response relationship between fluoride dose and intelligence with the relationship stated to be most evident with exposures from drinking water containing above 4 mg/L (or 4 ppm). Duan *et al.* (2018) found similar results as Choi *et al.* (2012) for the standardized mean difference; however, the majority of the available studies in both analyses compare populations with high fluoride exposure to those with lower fluoride exposure (with the lower exposure levels frequently in the range of drinking water fluoridation in the United States). NTP conducted a meta-analysis and is preparing it as a separate report for publication.

Confidence Rating: Assessment of Body of Evidence

The quality of evidence for neurodevelopmental and cognitive function outcomes was evaluated using the GRADE system for rating the confidence in the body of evidence (Guyatt *et al.* 2011, Rooney *et al.* 2014). More detailed guidance on reaching confidence ratings in the body of evidence as "high," "moderate," "low," or "very low" is provided in the protocol (https://ntp.niehs.nih.gov/go/785076). In brief, available human and animal studies on a particular health outcome were initially grouped by key study design features, and each grouping of studies was given an initial confidence rating by those features. Starting at this initial rating (see column 1 of **Figure 1**), potential downgrading of the confidence rating was considered for factors that decrease confidence in the results (see column 2 of **Figure 1**). Potential upgrading of the confidence rating was considered for factors that increase confidence in the results (see column 3 of **Figure 1**). Short descriptions of the factors that can decrease or increase confidence in the body of evidence for human studies are provided below (see protocol [https://ntp.niehs.nih.gov/go/785076] for additional details related to the human body of evidence, as well as considerations for experimental animal studies).

Factors to Consider for Potential Downgrading

- Risk of bias—addresses whether the body of evidence did not account for critical factors in study quality or design including confounding bias, selection bias, exposure assessment, and outcome assessment. Consideration for downgrading the confidence rating is based on the entire body of evidence, and the evidence is downgraded when there is substantial bias across most studies that could lead to decreased confidence in the results and where the studies without substantial bias could not support the confidence rating. Individual studies are evaluated for risk of bias based on a set of criteria (as discussed above); magnitude and direction of the bias are also considered.

- Unexplained inconsistency—addresses inconsistencies in results across studies of similar populations and design that can be determined by assessing similarity of point estimates and extent of overlap between confidence intervals or more formally through statistical tests of heterogeneity. Sensitivity analysis can be used to assess the impact of specific variables on the outcome. Inconsistencies that can be plausibly explained by characteristics of the studies (e.g., sex-associated differences) are typically not used to support a downgrade. A downgrade would only be applied when there is an inconsistency that cannot be explained and results in reduced confidence in the body of evidence.

- Indirectness—addresses generalizability and relevance to the objective of the assessment. All exposure levels and scenarios encountered in human studies are considered direct (i.e., applicable, generalizable, and relevant to address the objective of the assessment); therefore, a downgrade for indirectness would not be applied to bodies of evidence from human studies.

- Imprecision—addresses confidence associated with variability in quantitative measures such as effect sizes. Typically, 95% confidence intervals are used as the primary method to assess imprecision, but considerations can also be made on whether studies were adequately powered. Meta-analyses can also be used to determine if the data are imprecise. When a meta-analysis is not appropriate or feasible, imprecision can be based on variability around the effect estimate. A downgrade would occur if the body of evidence was considered to be imprecise based on a meta-analysis or if serious or very serious imprecision was consistently present in

the body of evidence. A downgrade is especially likely if imprecision raised questions as to whether an overall effect was significant.

- Publication bias—downgrade if strongly detect publication bias. Publication bias is difficult to detect but may be evident if major sections of the research community are not publishing (e.g., absence of industry, academia, or government studies) on a topic or if there are multiple instances where data from conference abstracts are never published in peer-reviewed journals. In addition, there are methods included in conducting a meta-analysis to detect if there is potential for publication bias, including the use of fit-and-trim models, which help identify how publication bias may affect the results of the meta-analysis. Although a meta-analysis is not included in this systematic review, there are two published meta-analyses (Duan *et al.* 2018, Choi *et al.* 2012) in addition to one conducted by NTP (manuscript development in progress) that can be used to address publication bias.

Factors to Consider for Potential Upgrading

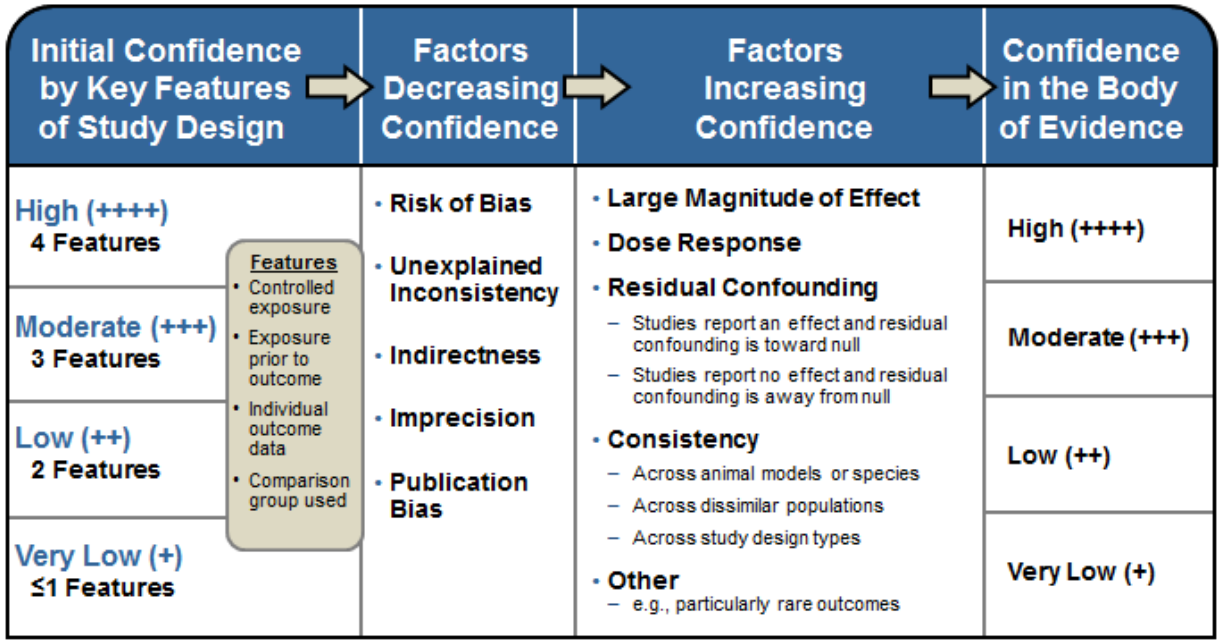
- Large magnitude of effect—factors to consider include the outcome being measured and the dose or exposure range assessed. The confidence can be upgraded if the body of evidence is suggestive of a large magnitude of effect. GRADE provides guidance on what can be considered a large magnitude of effect based on relative risk (i.e., suggests one upgrade in confidence if relative risk is greater than 2 and two upgrades in confidence if greater than 5). However, not all studies provide data as a risk estimate, and smaller changes, such as increases in blood pressure, may have greater impact on health at the population level. Consideration for an upgrade is not based on a single study, and what constitutes a large magnitude of effect will depend on the outcome and the potential public health impact.

- Dose response—patterns of dose response are evaluated within and across studies. Confidence will be increased when there is sufficient evidence of a dose-response pattern across multiple studies in the body of evidence.

- Consistency—does not apply in this evaluation. The consideration of a potential upgrade for consistency is primarily for non-human animal evidence where it would be applied to address increased confidence based on observation of consistent effects across multiple non-human animal species. For human evidence, this factor would generally not be applied. Human studies are instead evaluated for issues of consistency that could result in downgrading confidence for unexplained inconsistency (see Factors to Consider for Potential Downgrading above).

- Consideration of residual confounding—applies to observational studies and refers to consideration of unmeasured determinants that are likely to be distributed unevenly across groups. Residual confounding can push results in either direction, but confidence in the results is increased when the body of evidence is biased by factors that counter the observed effect and would cause an underestimation of the effect. Confounding that would cause an overestimation of the effect is considered under the risk-of-bias considerations for decreasing confidence.

Figure 1. Assessing Confidence in the Body of Evidence



Confidence ratings were assessed by the evaluation team for accuracy and consistency, and discrepancies were resolved by consensus and consultation with technical advisors as needed. Confidence ratings for the primary outcomes are summarized in evidence profile tables for each outcome.

RESULTS

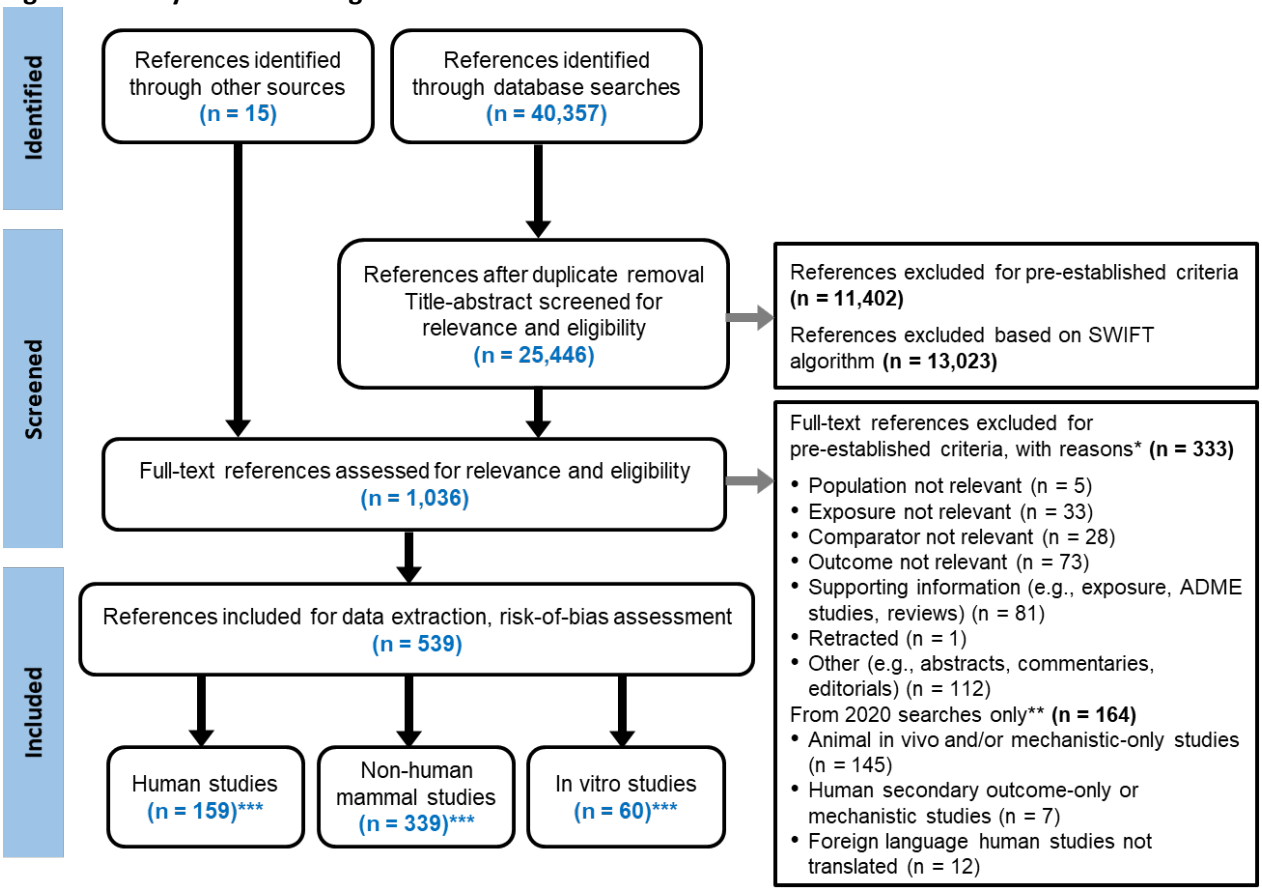
Literature Search Results

The electronic database searches retrieved 25,522 unique references with 15 additional references identified by technical advisors or from reviewing reference lists in published reviews and included studies. During title and abstract screening, 1,036 references were moved to full-text review and 24,501 were excluded (11,478 by manual screening for not satisfying the PECO criteria and 13,023 based on the SWIFT algorithm). Among the 1,036 references that underwent full-text review, 539 studies were considered PECO relevant (see **Appendix 2** for list of included studies). A few studies assessed data for more than one evidence stream (human, non-human mammal, and/or in vitro), and several studies assessed more than one type of outcome (e.g., primary and secondary outcomes). Included studies breakdown as follows:

- 159 human studies (78 primary only; 13 secondary only; 5 primary and secondary; 6 primary and thyroid; 2 secondary and thyroid; and 55 thyroid only);
- 339 non-human mammal studies (7 primary only; 186 secondary only; 67 primary and secondary; 6 primary, secondary, and thyroid; 4 secondary and thyroid; and 69 thyroid only); and,
- 60 in vitro/mechanistic studies (48 neurological and 12 thyroid).

Additional details on the screening results are provided in **Appendix 2**. These screening results are outlined in a study selection diagram that reports numbers of studies excluded at each stage and documents the reason for exclusion at the full text review stage (see **Figure 2**) [using reporting practices outlined in Moher *et al.* (2009)].

Figure 2. Study Selection Diagram[a]



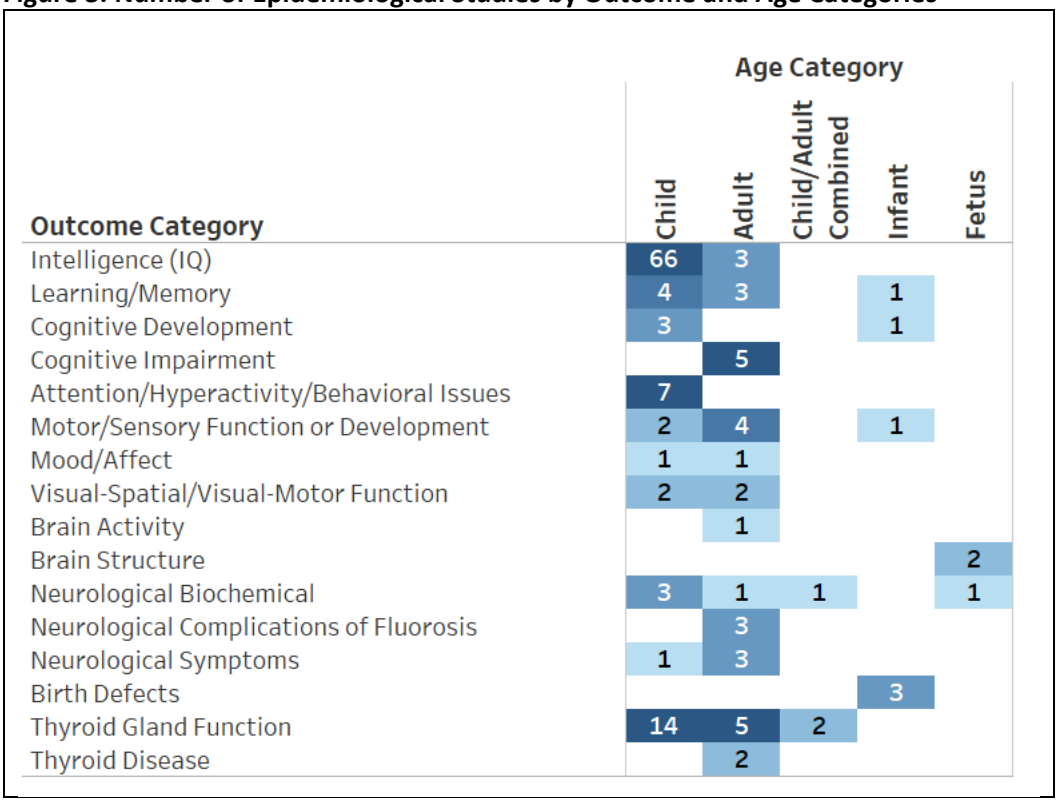
Notes:
[a]An interactive PRISMA is available here: https://hawcproject.org/summary/visual/assessment/405/Figure-2/.
* Studies may have been excluded for more than one reason; the first reason identified was recorded.
** Includes all studies from all literature searches, see **Methods** section for extraction and search update information.
*** Publications may contain more than on evidence stream, so the numbers will not total the 539 included studies.

Human Neurodevelopmental and Cognitive Data

The body of literature that evaluates the association between fluoride exposure and neurodevelopmental and cognitive effects in humans is relatively robust with a large number of studies (n = 92) that cover a wide array of endpoints (see **Figure 3**). Sixty-six human studies investigated IQ in children. Additional studies evaluated learning and memory (n = 8 studies) or other cognitive development effects (e.g., total neurobehavioral scores and total mental capacity index in children,

cognitive impairment in adults; n = 14 studies).[8] For this review, the evidence in children and adults was evaluated separately to address potential differences in the health impact of fluoride exposure during development versus adulthood.

Figure 3. Number of Epidemiological Studies by Outcome and Age Categories[a]



Notes:
[a]Interactive figure and additional study details in Tableau®.
(https://public.tableau.com/profile/ntp.visuals#!/vizhome/Fluoride_Epi_2021Update/Figure5?publish=yes)
Choi *et al.* (2015) used subtests of the omnibus IQ test reported by the authors as Wechsler Intelligence Scale for Children-Revised (WISC-IV) to evaluate visuospatial abilities (using block design) and executive function (using digit span). These endpoints are included in the intelligence (IQ) outcome category as they are subsets of the IQ tests.
Three additional publications based on subsamples (i.e., 50–60 children) of the larger Yu *et al.* (2018) cohort were identified (Zhao *et al.* 2020, Zhao *et al.* 2019, Zhou *et al.* 2019) and are not included in the counts of this figure.

Because the majority of studies evaluated intelligence, the following section focuses on IQ effects in children followed by separate discussions on other measures of cognitive function and neurobehavioral effects in children and cognitive effects in adults. Congenital neurological malformations and neurological complications of fluorosis are not considered further due to the limited number of studies and the heterogeneity of outcomes evaluated in these studies.

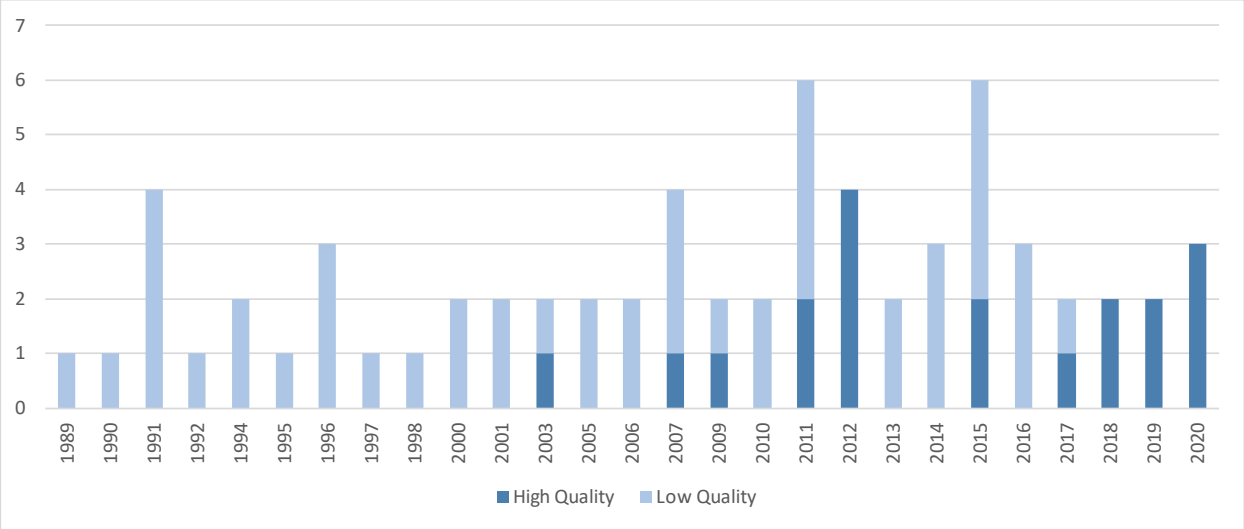
IQ in Children

Sixty-six epidemiological studies were identified that evaluated the effects of fluoride exposure on children's IQ. Nineteen of the 66 IQ studies were determined to have low potential for bias (i.e., were of

[8]Some studies are included in more than one endpoint category (e.g., IQ and other cognitive developmental effects); therefore, these counts are not mutually exclusive.

high quality). Looking across the literature, there has been a progression over the years in the quality of studies conducted to assess the association between fluoride exposure and IQ in children, with more recent studies including better study designs, larger sample sizes, and more sophisticated statistical analysis. Older studies often had limitations related to study design or methods, and most of the high risk-of-bias studies (i.e., studies of low quality) were published prior to the 2006 NRC evaluation of fluoride in drinking water. In contrast, 18 of the low risk-of-bias studies were published after the 2006 NRC evaluation of fluoride in drinking water, and over half of those 18 studies were published between 2015 and 2020 (**Figure 4**).

Figure 4. Number of High- and Low-quality Studies of Fluoride Exposure and IQ in Children by Year of Publication



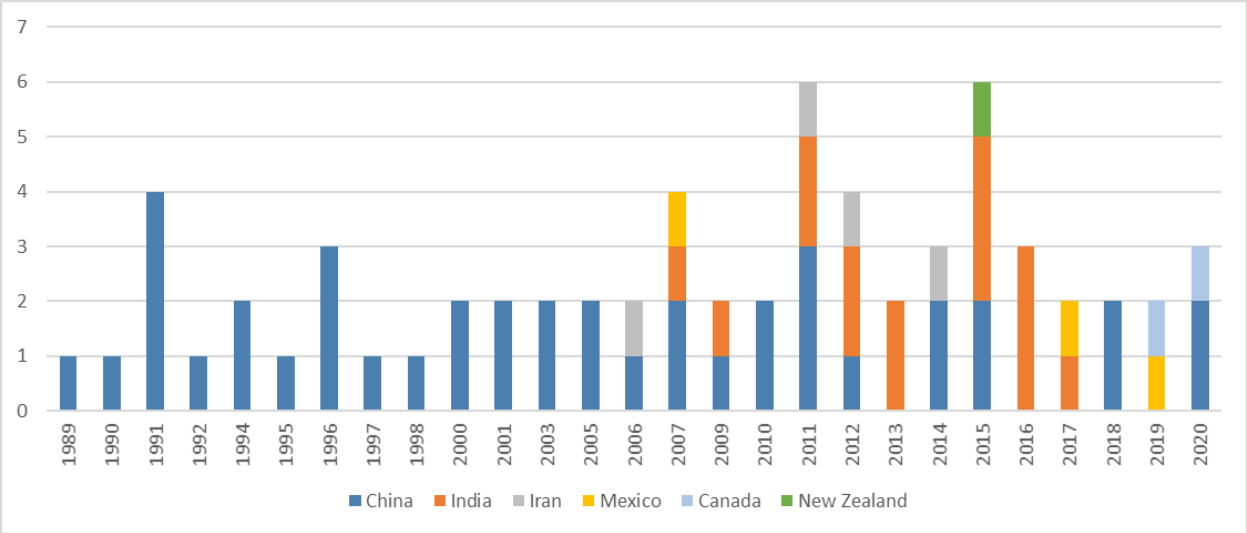
There are several characteristics of recent studies that contribute to higher study quality in the overall body of literature on children's IQ and fluoride including:

- Demonstration that exposure occurred prior to outcome assessment (an important factor when considering confidence in study results; see **Figure 1**) either by study design (e.g., for prospective cohort studies) or analysis (e.g., prevalence of dental fluorosis in children, limiting study populations to children who lived in the same area for long periods of time).
- Improved reporting of key study details that are necessary to evaluate study quality and allow for a more precise analysis of risk of bias.
- Increased consideration of potential confounders (e.g., socioeconomic status) and potential co-exposures (e.g., arsenic or lead intake).
- Increased use of individual-level exposure measures (urine or water), as well as prenatal fluoride exposure, to assess either individual-level fluoride exposure or—if still using group-level data—to confirm that regions being compared had differences in fluoride exposure.
- Utilization of more sophisticated sampling techniques for the study populations (e.g., stratified multistage random sampling).
- Application of more sophisticated regression approaches (e.g., piecewise linear regression models, multi-level regression with random effects, or generalized additive models for longitudinal measurements of fluoride).
- For studies using individual-level exposure measures, application of more sophisticated regression techniques to account for clustering at the cohort level by using cohort as a fixed or

random effect and by accounting for numerous potential confounders that capture the cohort effect.

In addition, newer studies represent more diverse study populations across several countries (**Figure 5**), whereas all identified peer-reviewed studies that were published prior to 2006 took place in a single country (China). The majority of high quality, low risk-of-bias studies exhibit these important study design and analysis characteristics, as discussed further in subsequent sections.

Figure 5. Number of Studies of Fluoride Exposure and IQ in Children by Country and Year of Publication



All available studies were considered in this evaluation; however, review of the body of evidence focused on the high quality, low risk-of-bias studies for two main reasons. First, there are fewer limitations and greater confidence in the results of the high quality studies. Second, there are a relatively large number of high quality studies (n = 19), such that the body of evidence from these studies could be used to evaluate confidence in the association between fluoride exposure and changes in children's IQ. Therefore, the remainder of the discussion on IQ in children focuses on the 19 studies with low risk of bias. The high risk-of-bias studies are discussed briefly relative to their overall support of findings from the low risk-of-bias studies.

Low Risk-of-bias IQ Studies

Overview of Studies

Nineteen studies with low potential for bias evaluated the association between fluoride exposure and IQ in children. These IQ studies were conducted in multiple study populations across 5 countries and included more than 7,000 children. Specifically, of the 19 low risk-of-bias studies of IQ in children:

- 10 were conducted in 4 areas of China on 7 study populations[9];
- 3 were conducted in 3 areas of Mexico on 3 study populations;

[9]In this document, "study population" refers to a defined population on which an original body of research was conducted. The published work drawn from that original body of research is often referred to as a "study." IQ studies that report on the same study populations are identified in **Table 6**.

- 2 were conducted in Canada using the same study population;
- 3 were conducted in 3 areas of India on 3 study populations; and
- 1 was conducted in Iran.

Most studies measured fluoride in the drinking water or urine (child or maternal) with a few that measured fluoride in serum. The IQ studies used a variety of tests to measure IQ. Because IQ tests should be culturally relevant, the IQ tests used often differed between studies reflecting adjustments for the range in populations studied (e.g., western vs. Asian populations). In some cases, different IQ tests were used to study similar populations. Overall, these studies used IQ tests appropriate for the population and were age appropriate.

Table 6 provides a summary of study characteristics and key IQ and fluoride findings for the 19 low risk-of-bias studies. Several of these studies conducted multiple analyses and reported results on multiple endpoints. The purpose of the table is to summarize key findings (independent of whether an effect is indicated) from each study and is not meant to be a comprehensive summary of all results from each study. For each study, results are summarized for each exposure measure assessed, but results from multiple analyses using the same exposure measure may not all be presented unless results from multiple analyses yielded conflicting results.

Table 6. Studies on IQ in Children[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[c,d]
China					
Xiang *et al.* (2003a)[1]	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children [512]	Drinking water Mean (SD): 0.36 (0.15) (control), 2.47 (0.79) (high fluoride) mg/L Children's urine Mean (SD): 1.11 (0.39) (control), 3.47 (1.95) (high fluoride) mg/L Village of residence (non-endemic v. endemic fluorosis)	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significant dose-related effect of fluoride on IQ score based on drinking water quintile levels with significantly lower IQ scores observed at water fluoride levels of 1.53 mg/L or higher; % of subjects with IQ < 80 was significantly increased at water levels 2.46 mg/L or higher; significant inverse correlation between IQ and urinary fluoride (Pearson correlation coefficient of −0.164); mean IQ scores for children in non-endemic region (100.41 ± 13.21) significantly higher than endemic region (92.02 ± 13.00) No statistical adjustment for confounders
Ding *et al.* (2011)	Cross-sectional Inner Mongolia (Hulunbuir City)/elementary school children [331]	Children's urine Range: 0.1–3.55 mg/L Drinking water (reported but not used in analyses) Mean (SD): 1.31 (1.05) mg/L	Children (ages 7–14 years)	IQ: Combined Raven's Test for Rural China	Significant association between urinary fluoride and IQ score (each 1-mg/L increase was associated with a lower IQ score of 0.59 points; 95% CI: −1.09, −0.08) Adjusted for child's age
Xiang *et al.* (2011)[1]	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children [512]	Children's serum Mean (SD): 0.041 (0.009) (control), 0.081 (0.019) (high fluoride) mg/L	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significant linear trend across quartiles of serum fluoride and children's IQ score < 80 (adjusted ORs for Q1 and Q2; Q1 and Q3; and Q1 and Q4, respectively: 1; 2.22 [95% CI: 1.42, 3.47]; and 2.48 [95% CI: 1.85, 3.32]); significant effects at ≥0.05 mg/L serum fluoride Adjusted for child's age and gender
Wang *et al.* (2012)[1]	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children [526]	Children's total fluoride intake Mean (SD): 0.78 (0.13) (control), 3.05 (0.99) (high fluoride) mg/day Village of residence (non-endemic v. endemic fluorosis) Drinking water (reported for villages but not used in analyses) Mean (SD): 0.36 (0.11) (control), 2.45 (0.80) (high fluoride) mg/L	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significantly lower mean IQ in the endemic versus non-endemic regions, as reported in Xiang *et al.* (2003a); when high exposure group was broken into 4 exposure groups based on fluoride intake, a dose-dependent decrease in IQ and increase in % with low IQ observed; significant correlation between total fluoride intake and IQ (r = −0.332); for IQ < 80, adjusted OR of total fluoride intake was 1.106 (95% CI: 1.052, 1.163) Adjusted for child's age and gender

Table 6. Studies on IQ in Children[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[c,d]
Choi *et al.* (2015)	Cross-sectional Mianning County/1st grade children [51]	Drinking water GM: 2.20 mg/L Children's urine GM: 1.64 mg/L Severity of fluorosis (Dean Index)	Children (ages 6–8 years)	IQ: WISC-IV (square root block design and digit span)	Compared to normal/ questionable fluorosis, moderate/severe fluorosis significantly associated with lower total (adjusted β = −4.28; 95% CI: −8.22, −0.33) and backward (adjusted β = −2.13; 95% CI: −4.24, −0.02) digit span scores; linear correlations between total digit span and log-transformed urinary fluoride (adjusted β = −1.67; 95% CI: −5.46, 2.12) and log-transformed drinking water fluoride (adjusted β = −1.39; 95% CI: −6.76, 3.98) observed but not significant; forward digit span had similar results as backward and total, but was not statistically significant; squareroot block design not significantly associated with any measure of fluoride exposure Adjusted for child's age, child's gender, parity, illness before 3 years old, household income last year, and caretaker's age and education
Zhang *et al.* (2015b)	Cross-sectional Tianjin City (Jinnan District)/school children [180]	Drinking water Mean: 0.63 (control), 1.40 (endemic fluorosis) mg/L (SD not reported) Children's urine Mean (SD): 1.1 (0.67) (control), 2.4 (1.01) (endemic fluorosis) mg/L Children's serum Mean (SD): 0.06 (0.03) (control), 0.18 (0.11) (endemic fluorosis) mg/L	Children (ages 10–12 years)	IQ: Combined Raven's Test for Rural China	Significant correlation between IQ score and children's serum fluoride (r = −0.47) and urinary fluoride (r = −0.45); significant difference in mean IQ score for high-fluoride area (defined as >1 mg/L in drinking water; 102.33 ± 13.46) compared with control area (109.42 ± 13.30); % of subjects with IQ < 90 significantly increased in high-fluoride area (28.7%) vs. low-fluoride area (8.33%); not significantly associated with water fluoride as a continuous variable Adjusted for child's age and gender, if applicable
Cui *et al.* (2018)	Cross-sectional Tianjin City (districts Jinghai and Dagang)/school children [323]	Children's urine Range (log-transformed): −1.2–2.2	Children (ages 7–12 years)	IQ: Combined Raven's Test for Rural China	Significant correlation between IQ score and log-transformed urinary fluoride (adjusted β = −2.47; 95% CI: −4.93, −0.01) Adjusted for child age, mother's education, family member smoking, stress, and anger

Table 6. Studies on IQ in Children[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[c,d]
Yu *et al.* (2018)[2]*	Cross-sectional Tianjin City (7 towns)/children [2,886]	Drinking water Mean (SD): 0.50 (0.27) (normal), 2.00 (0.75) (high) mg/L Children's urine Mean (SD): 0.41 (0.49) (normal), 1.37 (1.08) (high) mg/L	Children (ages 7–13 years)	IQ: Combined Raven's Test for Rural China	Significant difference in mean IQ scores in high water fluoride areas (>1.0 mg/L; 106.4 ± 12.3 IQ) compared to the normal water fluoride areas (≤1.0 mg/L; 107.4 ± 13.0); distribution of the IQ scores also significantly different (p = 0.003); every 0.5-mg/L increase in water fluoride was associated with a 4.29 lower IQ score (95% CI: −8.09, −0.48) between 3.40 and 3.90 mg/L; no significant association between 0.2 and 3.40 mg/L; every 0.5-mg/L increment of urinary fluoride was associated with a 2.67 lower IQ score (95% CI: −4.67, −0.68) between 1.60 and 2.50 mg/L but not at 0.01–1.60 mg/L or 2.50–5.54 mg/L. Adjusted for child's age, child's gender, maternal education, paternal education, and low birth weight
Cui *et al.* (2020)	Cross-sectional Tianjin City (all districts)/school children (potentially some overlap with Cui *et al.* (2018)) [498]	Children's urine <1.6–≥2.5 mg/L	Children (ages 7–12 years)	IQ: Combined Raven's Test	Decreasing mean (± SD) IQ score with increasing urinary fluoride levels (statistical significance not reached based on a one-way ANOVA) <1.6 mg/L: 112.16 ± 11.50 1.6-2.5 mg/L: 112.05 ± 12.01 ≥2.5 mg/L: 110 ± 14.92 No statistical adjustment for confounders
Wang *et al.* (2020b)[2]	Cross-sectional Tianjin City (villages not specified)/school children [571]	Drinking water Mean (SD): 1.39 (1.01) mg/L Children's urine Mean (SD): 1.28 (1.30) mg/L	Children (ages 7–13 years)	IQ: Combined Raven's Test for Rural China	Significant associations between IQ and water and urinary fluoride concentrations in boys and girls combined based on both quartiles and continuous measures (water: 1.587 decrease in IQ score per 1-mg/L increase; urine: 1.214 decrease in IQ score per 1-mg/L increase); no significant modification effect of gender Adjusted for child's age, child's gender, BMI, maternal education, paternal education, household income, and low birth weight

Table 6. Studies on IQ in Children[a,b]					
Study	**Study Design (Location/Subjects) [n]**	**Exposure Measures and Summary Statistics**	**Assessment Timing**	**Outcome and Methods**	**Summary of IQ Results[c,d]**
Mexico					
Rocha-Amador *et al.* (2007)	Cross-sectional Moctezuma and Salitral in San Luis Potosi State and 5 de Febrero of Durango State /elementary school children [132]	Drinking water Mean (SD): 0.8 (1.4), 5.3 (0.9), 9.4 (0.9) mg/L (3 rural areas) Children's urine Mean (SD): 1.8 (1.5), 6.0 (1.6), 5.5 (3.3) mg/L (3 rural areas)	Children (ages 6–10 years)	IQ: WISC-Revised Mexican Version	Significant associations between log-transformed fluoride and IQ scores (full IQ adjusted βs of −10.2 [water] and −16.9 [urine]; CIs not reported); arsenic also present, but the effect was smaller (full-scale IQ adjusted βs of −6.15 [water] and −5.72 [urine]; CIs not reported) Adjusted for blood lead, mother's education, SES, height-for-age z-scores, and transferrin saturation
Bashash *et al.* (2017)	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [299] IQ analysis [211]	Maternal urine during pregnancy Mean (SD): 0.90 (0.35) mg/L Children's urine Mean (SD): 0.82 (0.38) mg/L	Children (ages 6–12 years)	IQ: WASI-Spanish Version	Significantly lower child IQ score per 0.5-mg/L increase in maternal urinary fluoride (adjusted β = −2.50; 95% CI: −4.12, −0.59); no significant association with children's urine Adjusted for gestational age, weight at birth, child's gender, parity (being the first child), age at outcome measurement, and maternal characteristics including smoking history (ever smoked during the pregnancy vs. nonsmoker), marital status (married vs not married), age at delivery, education, IQ, and cohort
Soto-Barreras *et al.* (2019)	Cross-sectional Chihuahua/school children [161]	Children's urine Range: 0.11–2.10 mg/L Drinking water Range: 0.05–2.93 mg/L Fluoride exposure dose (summary statistics not reported) Fluorosis index (summary statistics not reported)	Children (ages 9–10 years)	IQ: Raven's Colored Progressive Matrices	No significant difference in urinary fluoride, drinking water fluoride, fluoride exposure dose, or fluorosis index in subjects across different IQ grades No statistical adjustment for confounders

Table 6. Studies on IQ in Children[a,b]					
Study	**Study Design (Location/Subjects) [n]**	**Exposure Measures and Summary Statistics**	**Assessment Timing**	**Outcome and Methods**	**Summary of IQ Results[c,d]**
Canada					
Green *et al.* (2019)[3]	Cohort (prospective) 10 cities/Maternal-Infant Research on Environmental Chemicals (MIREC) [512] Non-Fluoridated [238] Fluoridated [162] Boys [248] Girls [264]	Maternal urine during pregnancy Mean (SD): 0.51 (0.36) mg/L (0.40 [0.27] mg/L in non-fluoridated areas and 0.69 [0.42] mg/L in fluoridated areas) Maternal fluoride intake during pregnancy Mean (SD): 0.54 (0.44) mg/day (0.30 [0.26] and 0.93 [0.43] mg/day, respectively) Drinking water Mean (SD): 0.31 (0.23) mg/L (0.13 [0.06] and 0.59 [0.08] mg/L, respectively)	Children (ages 3−4 years)	IQ: full scale, performance, and verbal using Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	Significantly lower full-scale IQ (adjusted β = −4.49; 95% CI: −8.38, −0.60) and performance IQ (adjusted β = −4.63; 95% CI: −9.01, −0.25) per 1-mg/L increase in maternal urinary fluoride in boys, but not girls (adjusted β = 2.40; 95% CI: −2.53, 7.33 and adjusted β = 4.51; 95% CI: −1.02, 10.05, respectively) or boys and girls combined (adjusted β = −1.95; 95% CI: −5.19, 1.28 and adjusted β = −1.24; 95% CI: −4.88, 2.40, respectively); significantly lower full-scale IQ (adjusted β = −3.66; 95% CI: −7.16, −0.15) per 1-mg increase in maternal fluoride intake (no sex interaction); significantly lower full-scale IQ (adjusted β = −5.29; 95% CI: −10.39, −0.19) per 1-mg/L increase in water fluoride concentration (no sex interaction); no significant associations observed between measures of fluoride and verbal IQ Adjusted for city, HOME score, maternal education, race, child's gender, and prenatal secondhand smoke exposure

Table 6. Studies on IQ in Children[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[c,d]
Till *et al.* (2020)[3]	Cohort (prospective) 10 cities/ MIREC [398] Non-Fluoridated [247] Fluoridated [151] Breastfed as infants [200] Formula-fed as infants [198]	Drinking water Mean (SD) for breastfed infants: 0.13 (0.06) mg/L in non-fluoridated areas and 0.58 (0.08) mg/L in fluoridated areas For formula fed infants: 0.13 (0.05) mg/day in non-fluoridated areas and 0.59 (0.07) mg/L in fluoridated areas Infant fluoride intake Mean (SD) for breastfed infants: 0.02 (0.02) mg/day in non-fluoridated areas and 0.12 (0.07) mg/day in fluoridated areas For formula fed infants : 0.08 (0.04) mg/day in non-fluoridated areas and 0.34 (0.12) mg/day in fluoridated areas Maternal urine during pregnancy Mean (SD) breastfed: 0.42 (0.28) mg/L in non-fluoridated areas and 0.70 (0.39) mg/L in fluoridated areas formula-fed: 0.38 (0.27) mg/L in non-fluoridated areas and 0.64 (0.37) mg/L in fluoridated areas	Children (ages 3−4 years)	IQ: full-scale, performance, and verbal using Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	Drinking water breastfed infants: Lower (not significant) full-scale IQ (adjusted β = −1.34, 95% CI: −5.04, 2.38) per 0.5 mg/L increase in water fluoride concentration; significantly lower performance IQ (adjusted β = −6.19, 95% CI: −10.45, −1.94) formula fed infants: Significantly lower full-scale IQ (adjusted β = −4.40, 95% CI: −8.34, −0.46) per 0.5 mg/L increase in water fluoride concentration; significantly lower performance IQ (adjusted β = −9.26, 95% CI: −13.77, −4.76) Infant fluoride intake breastfed: No results reported formula fed: Lower (not significant) full-scale IQ (adjusted β = −2.69, 95% CI: −709, 3.21) per 0.5 mg/L increase in fluoride intake from formula; significantly lower performance IQ (adjusted β = −8.76, 95% CI: −14.18, −3.34) Maternal urine during pregnancy+ Lower (not significant) full-scale IQ (adjusted β = −1.08, 95% CI: −1.54, 0.47) per 0.5 mg/L increase in maternal urinary fluoride++; lower (not significant) performance IQ (adjusted β = −1.31, 95% CI: −3.63, 1.03)++ Lower (not significant) performance IQ (adjusted β = −1.50, 95% CI: −3.41, 0.43) per 0.5 mg/L increase in maternal urinary fluoride+++; significantly lower full-scale IQ (adjusted β = −2.38, 95% CI: −4.62, −0.27)+++ No association between verbal IQ scores and any measure of fluoride exposure +Maternal urinary fluoride analyzed as covariate in the drinking water and infant fluoride intake from formula models, and not in an individual model ++After additional adjustment for drinking water and breastfeeding status +++After additional adjustment for infant fluoride intake from formula All models adjusted for maternal education, maternal race, child's age at IQ testing, child's sex, HOME total score, and second-hand smoke status in the child's house (separate analysis also adjusted for mother's urinary fluoride)

Table 6. Studies on IQ in Children[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[c,d]
India					
Sudhir *et al.* (2009)	Cross-sectional Nalgonda District (Andhra Pradesh)/school children [1,000]	Drinking water Level 1: <0.7 mg/L Level 2: 0.7–1.2 mg/L Level 3: 1.3–4.0 mg/L Level 4: >4.0 mg/L	Children (ages 13–15 years)	IQ: Raven's Standard Progressive Matrices	Significant increase in mean and distributions of IQ grades (i.e., increase in proportion of children with intellectual impairment) with increasing drinking water fluoride levels No statistical adjustment for confounders
Saxena *et al.* (2012)	Cross-sectional Madhya Pradesh/school children [170]	Drinking water ≥1.5 mg/L (high fluoride group) Children's urine Range: 1.7–8.4 mg/L	Children (age 12 years)	IQ: Raven's Standard Progressive Matrices	Significant correlations between IQ score and water (r = 0.534) and urinary (r = 0.542) fluoride levels; significant increase in mean IQ grade (i.e., increase in proportion of children with intellectual impairment) with increasing water fluoride quartile; no significant differences in the levels of urinary lead or arsenic in children from the different groups Confounders included in the analysis were not reported
Trivedi *et al.* (2012)	Cross-sectional Kachchh, Gujarat/school children (6th and 7th grades) [84]	Mean (SE) Low fluoride villages: drinking water: 0.84 (0.38) mg/L Children's urine: 0.42 (0.23) mg/L High fluoride villages: drinking water: 2.3 (0.87) mg/L Children's urine: 2.69 (0.92) mg/L	Children (ages 12–13 years)	IQ: questionnaire prepared by Professor JH Shah (97% reliability rating)	Significantly lower mean IQ score in high fluoride villages (92.53 ± 3.13) compared to the low fluoride villages (97.17 ± 2.54); differences significant for boys and girls combined as well as separately No statistical adjustment for confounders

Table 6. Studies on IQ in Children[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Summary of IQ Results[c,d]
Iran					
Seraj *et al.* (2012)	Cross-sectional Makoo/school children [293]	Drinking water Mean (SD): 0.8 (0.3) (normal), 3.1 (0.9) (medium), 5.2 (1.1) (high) mg/L	Children (ages 6–11 years)	IQ: Raven's Colored Progressive Matrices	Significant correlation between water fluoride and IQ score (adjusted β = −3.865; CIs not reported); significantly higher mean IQ score in normal area (97.77 ± 18.91) compared with medium (89.03 ± 12.99) and high (88.58 ± 16.01) areas Adjusted for child's age, child's gender, child's education level, mother's education level, father's education level, and fluorosis intensity

Studies with the same number superscript are based on the same study population.

*Three additional publications based on subsample (i.e., 50–60 children) of the larger Yu *et al.* (2018) cohort were identified (Zhao *et al.* 2020, Zhou *et al.* 2019, Zhao *et al.* 2019); however, these publications focused on mechanistic considerations and are not included in the study totals for IQ because the main study by Yu *et al.* (2018) is considered a better representation of the IQ results.

[a]Includes low risk-of-bias studies.

[b]Definitions: ANOVA: analysis of variance; **GM**: geometric mean; **HOME**: Home Observation Measurement of the Environment; **IQ**: intelligence quotient; **WASI**: Wechsler Abbreviated Scale of Intelligence (Spanish version); **WISC-IV**: Wechsler Intelligence Scale for Children-Revised (as reported by Choi *et al.* 2015).

[c]Associations between IQ and fluoride levels were reported quantitatively, as possible. For studies with multiple analyses and results, the table summarizes key findings and is not a comprehensive summary of all findings. Results also indicate when a study found no association between IQ and fluoride, provided as a qualitative statement of no association.

[d]See **Figure D1** through **Figure D7** for additional study results.

Summary of Results

Overall Findings

The results from 18 of the 19 high quality (low risk-of-bias) studies (3 prospective cohort and 15 cross-sectional studies from 13 different study populations) that evaluated IQ in children provide consistent evidence that exposure to fluoride is associated with lower IQ scores (see "Summary of IQ Results" in **Table 6**)(Bashash *et al.* 2017, Choi *et al.* 2015, Ding *et al.* 2011, Rocha-Amador *et al.* 2007, Saxena *et al.* 2012, Seraj *et al.* 2012, Xiang *et al.* 2003a, Xiang *et al.* 2011, Zhang *et al.* 2015b, Yu *et al.* 2018, Green *et al.* 2019, Cui *et al.* 2018, Cui *et al.* 2020, Wang *et al.* 2020b, Wang *et al.* 2012, Sudhir *et al.* 2009, Till *et al.* 2020, Trivedi *et al.* 2012). Only one study did not observe evidence of an association between fluoride exposure and IQ; however, results were not provided in a manner that allowed for a direct comparison with other low risk-of-bias studies (see **Appendix 4** for details). A strength of the findings across 18 of 19 low risk-of-bias studies was the consistent association between increased fluoride levels (generally above the WHO Drinking Water Quality Guideline [1.5 mg/L] (WHO 2011)) and lower IQ scores among studies of varying study designs, exposure measures, and study populations. In studies that analyzed the sexes separately (n = 5 studies with 2 studies reporting on the same study population), consistent findings of lower IQ associated with fluoride exposure were generally reported for both sexes. There is some indication of differential susceptibility between sexes, but ultimately it is unclear if one gender is more susceptible to the effects of fluoride exposure than the other due to too few high quality studies that analyzed exposure and outcome by gender separately and a lack of consistent findings that one gender is more susceptible. The body of evidence from the 19 low risk-of-bias studies is described in further detail below. Prospective cohort studies are discussed first, as this study design can establish a temporal relationship between exposure and outcome which would contribute to demonstrating causality, therefore, providing the strongest evidence for an association between fluoride exposure during development and IQ in children.

Results by Study Design

Prospective cohort studies

As discussed above, all three prospective cohort studies found an association between increasing maternal or children's fluoride exposure and lower IQ in children (Green *et al.* 2019, Till *et al.* 2020, Bashash *et al.* 2017). Two of the studies (Green *et al.* 2019, Till *et al.* 2020) were based on the same study population, but the authors used different measures of fluoride exposure to evaluate IQ. Multiple analyses were conducted in each prospective study and, although not every analysis found a statistically significant association, together the three studies provided consistent evidence that maternal fluoride levels were associated with lower IQ scores in the children.

Bashash *et al.* (2017) observed a statistically significant association (p-value = 0.01) between lower IQ scores in children and prenatal fluoride exposure measured by maternal urinary fluoride (measured during all three trimesters and included if at least one measurement was available). An increase of 0.5 mg/L of maternal urinary fluoride was associated with a 2.5-point lower IQ score [95% CI: −4.12, −0.59] in boys and girls combined (see **Figure D7**). This study also reported an inverse association between IQ level and children's urinary fluoride levels (single spot urine sample); however, this specific result did not achieve statistical significance (a 0.5-mg/L increase of child urinary fluoride was associated with a 0.89-point lower IQ score [95% CI: −2.63, 0.85]) (Bashash *et al.* 2017).

Green *et al. (*2019) also reported inverse associations between IQ scores in children and multiple measures of prenatal fluoride exposure, including maternal urinary fluoride, maternal fluoride intake, and water fluoride concentrations in the Maternal-Infant Research on Environmental Chemicals cohort,

consisting of 10 cities in Canada. Green *et al. (*2019) observed a statistically significantly lower IQ for boys associated with maternal urinary fluoride averaged across trimesters (4.49-point lower IQ score [95% CI: −8.38, −0.60; p-value = 0.02] in IQ per 1-mg/L increase in maternal urinary fluoride); however, results were not significant in boys and girls combined (1.95-point lower IQ [95% CI: −5.19, 1.28]) and were positive in girls (2.40-point increase [95% CI: −2.53, 7.33] in IQ). Other measures of prenatal exposure (maternal fluoride intake or water fluoride concentrations) were associated with lower IQ scores in boys and girls combined; the authors found no significant effect measure modification between child sex and fluoride exposure in these analyses so did not report boys and girls separately (Green *et al.* 2019). Specifically, when evaluating the association between estimated maternal fluoride intake based on maternal water and beverage consumption during pregnancy and IQ in children, a 1-mg increase in daily maternal consumption of fluoride during pregnancy was associated with a significantly lower IQ score of 3.66 points in boys and girls combined (95% CI: −7.16, −0.15; p-value = 0.04). Similarly, water fluoride concentrations for pregnant women from fluoridated areas (mean water fluoride levels of 0.59 ± 0.08 mg/L) versus pregnant women from non-fluoridated areas (mean water fluoride levels of 0.13 ± 0.06 mg/L) were associated with a significant 5.29-point lower IQ score per 1-mg/L increase in fluoride in both boys and girls combined (95% CI: −10.39, −0.19; p-value < 0.05) (Green *et al.* 2019).

In a study of the same study population as Green *et al.* (2019) that used fluoride intake from formula or water concentrations in formula-fed versus breastfed infants, Till *et al.* (2020) observed significantly lower performance IQ scores with higher fluoride regardless of the comparison used (p-values <= 0.004). They did not observe any effect on verbal IQ, and full-scale IQ was only significantly lower in formula-fed infants using water fluoride concentrations as the exposure measure (p-value = 0.03). Breastfed infants and fluoride intake from formula also showed inverse associations but were not significant.

Taken together, the three prospective cohort studies (based on two study populations) indicate consistency across different types of analysis and across two study populations that fluoride exposure during development is associated with lower IQ scores.

Cross-sectional studies

As with the prospective cohort studies, the cross-sectional studies reported a consistent association between fluoride exposure and lower IQ scores in children. Fifteen of the 16 low risk-of-bias cross-sectional studies provide consistent evidence that exposure to fluoride is associated with lower IQ scores. Fourteen of these 15 studies (with the exception of Cui *et al.* 2020) reported significant associations.

Cross-sectional studies can have limitations in assessing causality, as the study design often cannot ensure that exposure preceded outcome. This uncertainty reduces confidence in study findings compared to prospective cohort studies—which, by design, establish that exposure occurred prior to outcome—and is captured in the outcome assessment. In some cases, cross-sectional studies do provide indicators of prior exposure (e.g., prevalence of dental fluorosis, limiting study populations to subjects who lived in the same area for long periods of time). Evidence that exposure occurred prior to the outcome of interest increases the confidence in results and any potential association reported in these studies. Of the 16 low risk-of-bias cross-sectional studies, 12 studies established that exposure preceded the outcome assessment (Choi *et al.* 2015, Ding *et al.* 2011, Seraj *et al.* 2012, Yu *et al.* 2018, Sudhir *et al.* 2009, Rocha-Amador *et al.* 2007, Saxena *et al.* 2012, Wang *et al.* 2020b, Wang *et al.* 2012, Xiang *et al.* 2011, Xiang *et al.* 2003a, Soto-Barreras *et al.* 2019). Five studies from different study populations indicated that a large portion of the exposed children had dental fluorosis (ranging from 43–100%) at the time of the assessment (Choi *et al.* 2015, Ding *et al.* 2011, Seraj *et al.* 2012, Yu *et al.* 2018, Sudhir *et*

al. 2009). Because dental fluorosis occurs when fluoride is consumed during enamel formation (usually during the first 6–8 years of life), the presence of dental fluorosis suggests that exposures to fluoride occurred prior to the outcome assessment. Nine studies from six study populations (including Yu *et al.* (2018), Sudhir *et al.* (2009) listed above) excluded subjects that had not lived in the study area for a specified period of time, sometimes since birth (Rocha-Amador *et al.* 2007, Saxena *et al.* 2012, Yu *et al.* 2018, Wang *et al.* 2020b, Wang *et al.* 2012, Xiang *et al.* 2011, Xiang *et al.* 2003a, Soto-Barreras *et al.* 2019, Sudhir *et al.* 2009). Because these areas were generally known to be fluoride-endemic areas for long periods of time, it can generally be assumed that in these nine studies, exposure occurred prior to the outcome. Taken together, 12 cross-sectional studies from 9 study populations provide indicators of prior exposure.

Cross-sectional study variations

Overall, the cross-sectional studies provide consistent evidence that fluoride exposure is associated with lower IQ scores in the children. Several cross-sectional studies conducted multiple analyses (e.g., reported results for multiple exposure metrics, endpoints, subpopulations). Although some of these variations are heterogeneous and are not comparable across studies, the consistent results across multiple metrics increase our confidence in the data. **Table 6** summarizes key results for each of the low risk-of-bias cross-sectional studies, and a few examples of the within-study variations in results are provided below.

Nine cross-sectional studies (from 6 study populations) assessed the association between IQ and multiple exposure measures (Choi *et al.* 2015, Saxena *et al.* 2012, Rocha-Amador *et al.* 2007, Wang *et al.* 2020b, Yu *et al.* 2018, Zhang *et al.* 2015b, Xiang *et al.* 2003a, Xiang *et al.* 2011, Wang *et al.* 2012). Lower IQ was consistently observed across exposure measures in these studies; however, Choi *et al.* (2015), a small pilot study (n = 51), reported a variation in statistical significance by exposure measure (see **Figure D7**). Choi *et al.* (2015) also observed some variation in results by outcome assessed (i.e., square root block design and digit span [forward, backward, and total]). It was the only cross-sectional study that did not provide a full IQ score, but instead provided results by specific subtests. The study authors observed a consistent inverse association between fluoride exposure and results from the digit span subtest (which specifically assesses executive function); however, results from the square root block design, a subtest of the WISC-IV omnibus IQ test that specifically assesses visuospatial function, was not associated with fluoride exposure. Note that Rocha-Amador *et al*. (2009) also assessed visuospatial function, and the authors reported a significant association (p-value < 0.001) between fluoride exposure and decreased visuospatial constructional ability using the Rey-Osterrieth Complex Figure (ROCF) Test. Ultimately, too few studies were identified that reported results by subtest of omnibus IQ tests or assessed domains other than IQ (e.g., visuospatial function) to examine or explain the variation by outcome observed in Choi *et al.* (2015). The only other studies that provided a breakdown of the full IQ score were the prospective cohort studies by Green (2019) and Till (2020), which provided results for full-scale IQ as well as results for performance and verbal IQ. In both of these studies, verbal IQ was not associated with fluoride exposure, but performance and full-scale IQ were associated with fluoride exposure. There are too few studies to evaluate if there is a specific aspect of IQ testing that is affected by exposure to fluoride, but the studies nonetheless provide consistent evidence that fluoride exposure is associated with lower IQ.

Yu *et al.* (2018) reported an overall association between lower IQ and higher fluoride exposure across multiple analyses, but observed some variation in IQ results by urinary exposure level. The authors reported inverse associations between IQ and children's medium-range urinary fluoride levels (1.60–2.50 mg/L) and children' high-range urinary fluoride levels (2.50–5.54 mg/L) although change in IQ score

was greater in the medium-range urinary fluoride group (2.67 points lower [95% CI: −4.67, −0.68]) for every 0.5-mg/L increment of urinary fluoride, than in the high-range urinary fluoride group (0.84 points lower [95% CI: −2.18, 0.50]) (see **Figure D7**). No association was reported at low-range urinary fluoride levels (0.01–1.60 mg/L). Note that Yu *et al.* (2018) also reported an inverse association between IQ and drinking water fluoride levels at 3.40–3.90 mg/L (4.29-point lower IQ score [95% CI: −8.09, −0.48]) for every 0.5-mg/L increment of water fluoride; a 0.04-point lower IQ score [95% CI: −0.33, 0.24] was observed for 0.5-mg/L increments of water fluoride at levels of 0.20–3.40 mg/L). The variation by exposure level in urine could not be verified in the analysis of drinking water exposures because there were only two water exposure groups (low and high). In a second study (Wang *et al.* 2020b), authors conducted a categorical analysis using urinary fluoride quartiles with reported betas per quartile. As observed in Yu *et al.* (2018), there were decreasing trends in IQ within each quartile; however, unlike Yu *et al.* (2018), Wang *et al.* (2020b) observed a larger decrease in IQ with each increasing urinary quartile and observed similar results using water fluoride quartiles (Wang *et al.* 2020b). Note that Wang *et al.* (2020b) cannot be compared directly to Yu *et al.* (2018) for evaluation at the higher exposure levels because the two studies do not use the same categorical exposure ranges. Although additional studies may have looked at different exposure levels, none of these studies provided results in the same manner as Yu *et al.* (2018) and Wang *et al.* (2020b) (i.e., betas by exposure category). Instead, these other studies provided an overall beta or mean IQ scores by exposure level. Despite the noted variations among these studies, the overall results still support a consistent association between fluoride exposure and lower IQ.

Two studies (Cui *et al.* 2018, Zhang *et al.* 2015b) observed associations between lower IQ in children and exposure to fluoride, with variations in results in subpopulations of children with different polymorphisms (see **Figure D7**). These were the only two studies that considered polymorphism as a subanalysis. Cui *et al.* (2018) observed a significant association between log-transformed children's single spot urinary fluoride and lower IQ scores (2.47-point lower IQ scores [95% CI: −4.93, −0.01; p-value = 0.049] per unit increase in urinary fluoride), and the association was strongest in subjects with a TT polymorphism (compared to children with a CC or CT polymorphism) in the dopamine receptor D2 (DRD2) gene (12.31-point lower IQ score [95% CI: −18.69, −5.94; p-value < 0.001] per unit increase in urinary fluoride) which, according to the authors, probably resulted in a reduced D2 receptor density (Cui *et al.* 2018). Similarly, Zhang *et al.* (2015b) observed a significant association between lower IQ scores and children's single spot urinary fluoride (2.42-point lower IQ scores [95% CI: −4.59, −0.24; p-value = 0.030] per unit increase in urinary fluoride), and the association was strongest in subjects with a val/val polymorphism (compared to children who carried the heterozygous or homozygous variant genotypes [met/val or met/met]) in the catechol-O-methyltransferase (COMT) gene (9.67-point lower IQ score [95% CI: −16.80, −2.55; p-value = 0.003] per unit increase in urinary fluoride).

Overall, the cross-sectional studies support a consistent pattern of findings that increased fluoride exposure is associated with lower IQ scores in children. Slight within-study variations occur that may be associated with study variables such as IQ domains or subsets of IQ tests in a few studies that conducted multiple analyses, but these variations are heterogenous and cannot be further explored with the available studies. Despite these few variations, the overall evidence of an effect on IQ is apparent.

Exposure Measure and Study Population Factors

Low risk-of-bias studies provide consistent evidence that increased fluoride levels are associated with lower IQ scores across studies using different exposure measures. In addition to water fluoride levels, studies measured fluoride exposure using single serum samples in children (Xiang *et al.* 2011, Zhang *et al.* 2015b), single spot urine samples in children (Xiang *et al.* 2003a, Rocha-Amador *et al.* 2007, Ding *et*

al. 2011, Saxena *et al.* 2012, Zhang *et al.* 2015b, Cui *et al.* 2018, Yu *et al.* 2018, Wang *et al.* 2020b), and prenatal maternal urinary measures (Bashash *et al.* 2017, Green *et al.* 2019), all of which were demonstrated to be consistently associated with lower IQ scores (see **Figure D6** and **Figure D7**). Urine levels encompass all sources of fluoride exposure and provide a better measure of the totality of exposure. As noted previously, even though some studies measured single spot samples, which may not be representative of peak exposure, these studies generally provided evidence that fluoride exposure had been occurring for some time. The consistency in the study results across studies that used different measures of fluoride exposure and different lifestages at which fluoride was measured helps strengthen the body of evidence.

The low risk-of-bias studies provide consistent evidence that increased fluoride levels are associated with lower IQ scores across studies of different study populations. These 19 high quality studies represent diverse populations (n = 15 study populations) across 5 countries. Eighteen of the 19 studies conducted in Canada (n = 2), China (n = 10), India (n = 3), Mexico (n = 2), and Iran (n = 1) provide evidence that exposure to fluoride is associated with lower IQ scores; 1 study conducted in Mexico did not observe an association but reported results in a manner that did not allow for a direct comparison with the other studies (see **Appendix 4** for details). The overall consistency in the study results across study populations adds strength to the body of evidence.

Exposure Levels

As described in this section, the body of evidence for studies assessing the effect of fluoride exposure on IQ in children provides consistent evidence of an association between fluoride exposure and lower IQ in children; however, there is less certainty in the data at lower fluoride exposures (e.g., <1.5 mg/L in drinking water or equivalent total fluoride exposures from all sources). In the September 6, 2019, draft of this monograph, NTP conducted a qualitative analysis of children's IQ studies that 1) evaluated lower fluoride exposures (<1.5 mg/L) in drinking water and/or urine and 2) provided information to evaluate dose response (i.e., provided three or more fluoride exposure groups or a dose-response curve in their publication) in the lower fluoride exposure range. Nine low risk-of-bias studies met these criteria. Based on a qualitative review of these studies, the evidence of an association between fluoride exposure below 1.5 mg/L and lower IQ in children appeared less consistent than results of studies at higher exposure levels.

A draft quantitative dose-response meta-analysis was prepared and included in the September 16, 2020, draft monograph (NTP 2020). This meta-analysis is undergoing further refinement in preparation for separate publication and may further inform a discussion on the association between fluoride exposure levels and IQ in children.

Gender Considerations

Gender differences were examined in five of the low risk-of-bias studies (in four study populations) (Green *et al.* 2019, Wang *et al.* 2020b, Trivedi *et al.* 2012, Xiang *et al.* 2003a, Wang *et al.* 2012). In general, gender differences were difficult to assess for trends within different study populations because few studies in the body of evidence analyzed exposure and stratified results by gender. Although these five studies reported IQ scores separately for boys and girls, only two of these studies analyzed fluoride exposure for boys and girls separately (Green *et al.* 2019, Wang *et al.* 2020b), which is essential for evaluating whether a differential change in IQ by gender may be related to higher susceptibility in one gender or higher exposure in that gender. The remaining three studies stratified results by gender (Trivedi *et al.* 2012, Xiang *et al.* 2003a, Wang *et al.* 2012), but the analyses were based on area-level exposure data (e.g., low fluoride village compared to high fluoride village) and not drinking water or

urinary fluoride concentrations. In the five studies that reported results by gender separately, consistent findings of lower IQ associated with fluoride exposure were generally reported for both sexes. There was some variation in the results between sexes across study populations and exposure measures, but there is insufficient evidence to determine if one gender is more susceptible to the effects of fluoride exposure than the other.

Green *et al.* (2019) observed a significant inverse association between maternal urinary fluoride levels and IQ scores in boys (p-values <= 0.04) but not girls in a Canadian population. Green *et al.* (2019) did not find any sex differences in the association between IQ and water fluoride concentrations. Wang *et al.* (2020b) evaluated Chinese boys and girls separately and combined and observed statistically significant decreasing trends in IQ in all groups by urinary fluoride quartiles (p-values for trend <= 0.035) (see **Figure D7**). Similarly, when evaluated as a continuous variable, spot urinary fluoride levels (per 1-mg/L increase) were significantly associated with lower IQ scores in girls (−1.379 [95%CI: −2.628, −0.129; p-value = 0.031]), boys (−1.037 [95% CI: −2.040, −0.035; p-value = 0.043]), and in the sexes combined (−1.214 [95%CI: −1.987, −0.442; p-value = 0.002]). Based on water fluoride quartiles, Wang *et al.* (2020b) found that there was a significant trend in the sexes combined although the decreasing trend in boys and girls separately did not achieve statistical significance (p-values = 0.077 and 0.055, respectively). When water fluoride levels were evaluated as a continuous variable (per 1-mg/L increase), there were significant associations between lower IQ scores in girls (−1.649 [95%CI: −3.201, −0.097]; p-value = 0.037), boys (−1.422 [95%CI: −2.792, −0.053; p-value = 0.042]), and the sexes combined (−1.587 [95%CI: −2.607, −0.568]; p-value = 0.002).

The remaining three studies that reported results by gender based comparisons of areas of high and low urinary or water fluoride did not report exposure levels separately for boys and girls, which decreases the utility of the data to evaluate differential susceptibility by gender. Trivedi *et al.* (2012) observed significantly lower IQ in children in high fluoride Indian villages compared to low fluoride villages with decreases observed in boys and girls separately or combined (p-values <= 0.05) (see **Figure D2**). Xiang *et al.* (2003a) and Wang *et al.* (2012) provide data on the same study population in China. There was a significantly lower IQ in the high fluoride area compared to the low fluoride area in boys and girls separately and in the sexes combined (p-values < 0.01), although the difference was greater in girls. Because fluoride exposure was not analyzed for boys and girls separately, it is unclear if the greater change in IQ scores in girls than boys could be attributable to higher susceptibility to fluoride exposure in girls or differences in fluoride exposure by gender.

In summary, it is unclear if one gender is more susceptible to the effects of fluoride exposure than the other due to a limited number of studies that analyzed exposure and outcome by gender and a lack of a consistent pattern of findings that one gender is more susceptible. Green *et al.* (2019) did not observe an association between maternal urinary fluoride levels and IQ scores in girls but did observe a significant association in boys. Although this is an indication of higher sensitivity in boys in this analysis, the authors did not detect this gender difference using other measures of prenatal exposure (maternal fluoride intake or water fluoride concentrations). Wang *et al.* (2020b) and Trivedi *et al.* (2012) reported statistically significant associations in both boys and girls without indication that one gender may be more susceptible. Although Xiang *et al.* (2003a) and Wang *et al.* (2012) reported a greater change in IQ in girls than boys, the studies used area-level exposure data and authors did not determine if fluoride exposure differed in boys versus girls. Therefore, it is unclear if this differential result by gender is an indication of higher susceptibility in girls or if it could be explained by a difference in exposure by gender. Overall, there are too few studies that analyzed exposure and outcome by gender separately to properly evaluate if there is differential susceptibility to fluoride exposure by gender, and results from

the five low risk-of-bias studies that do evaluate gender differences indicate that there is no consistent difference by gender across the different study populations.

Summary of Key Findings for Low Risk-of-bias Children's IQ Studies

In summation, the high quality studies (i.e., studies with low potential for bias) demonstrate consistently lower IQ scores with higher fluoride exposure (e.g., greater than the WHO Drinking Water Quality Guideline [≥1.5 mg/L] (WHO 2011) or equivalent total exposures from all sources). The consistency in association is observed among studies of varying study designs, exposure measures, and study populations. Although some studies that conducted multiple analyses observed within-study variations in results (e.g., differences between subsets of IQ tests), these variations were unique to individual studies and do not detract from the overall consistency in the findings that higher fluoride is associated with lower IQ scores.

High Risk-of-bias IQ Studies

The results from 47 studies with high potential for bias that evaluated IQ in children provide consistent supporting evidence of decrements in IQ associated with exposures to fluoride. Forty-one of the 47 studies reported an association between high fluoride exposure and lower IQ scores in children.

Risk of Bias for IQ Studies in Children

The confidence in the human body of evidence was based on studies with the lowest potential for bias. A total of 19 studies on IQ in children had little or no risk-of-bias concerns, representing a relatively large body of evidence for low risk-of-bias studies (i.e., 15 study populations across 5 countries evaluating more than 7,000 children). These 19 studies are considered low risk of bias because they were rated probably low or definitely low risk of bias for at least two of the three key risk-of-bias questions and did not have any other risk-of-bias concerns that would indicate serious issues with the studies. Thirteen of the 19 studies were rated definitely low or probably low risk of bias for all risk-of-bias questions, and the remaining 6 studies were rated probably high risk of bias for a single question that was judged to have minimal impact on overall potential for bias. None of the 19 studies had a rating of definitely high risk of bias for any question. Risk-of-bias ratings for individual studies for all questions are available in **Figure A3-1** through **Figure A3-4**, with risk-of-bias ratings for IQ studies in children available in **Figure A3-5** through **Figure A3-8** and **Appendix 4**. Although the low risk-of-bias studies had minimal or no concerns, the studies with high overall potential for bias had a number of risk-of-bias concerns, including potential confounding, poor exposure characterization, poor outcome assessment, and, in many cases, potential concern with participant selection. The key risk-of-bias questions are discussed below.

Confounding for IQ Studies in Children

Low Risk-of-bias Studies

As discussed above, there are 19 studies considered to have low risk of bias when assessed across all risk-of-bias domains. Sixteen of 19 low risk-of-bias studies were considered to have low potential for bias due to confounding because the authors addressed potential confounders through study design or analysis. All 16 of these studies addressed the three key potential confounders for all studies: child's age, child's sex, and socioeconomic status. Co-exposures as potential confounders were addressed through study design or analysis in most of the low risk-of-bias studies and were ultimately not considered a concern in 18 of 19 studies (further discussed below). Other potential confounders, including health factors, smoking, and parental characteristics, were also addressed in many of the low risk-of-bias studies (see **Figure 6**).

Co-exposures to arsenic and lead were not considered a concern in 18 of 19 low risk-of-bias studies either because studies addressed the potential co-exposures, the co-exposures were not considered an issue in the study population, or the impact of the potential bias on the results was not a concern. Potential confounding related to co-exposure to arsenic was accounted for or determined not to be of concern in 15 of 19 low risk-of-bias studies, and all 15 studies observed an association between lower IQ and fluoride exposure. Co-exposure to arsenic was not accounted for in four low risk-of-bias studies and was the main potential concern in these studies; however, three of these studies (Xiang *et al.* 2011, Wang *et al.* 2012, Xiang *et al.* 2003a) were still considered low risk of bias for confounding due to the fact that arsenic was observed in the low fluoride comparison areas (which would bias the effect toward the null), but an effect was still observed. In this case, the lack of adjustment for arsenic strengthens the evidence for an association and does not represent a potential concern. The other study did not address arsenic and as noted above was in an area that had potential for arsenic exposure to occur Soto-Barreras *et al.* (2019) and is the only low risk-of-bias study that did not observe an association between lower IQ and fluoride exposure. Fourteen studies considered co-exposure to lead and all observed an association between lower IQ and fluoride exposure. Five studies did not consider co-exposures to lead; however, for all of these studies co-exposure to lead was considered unlikely to have an impact in these study populations as there was no evidence that lead was prevalent or occurring in relation to fluoride (Cui *et al.* 2018, Cui *et al.* 2020, Till *et al.* 2020, Trivedi *et al.* 2012, Soto-Barreras *et al.* 2019).

There is considerable variation in the specific confounders considered across the 19 low risk-of-bias studies. The consistency of results across these low risk-of-bias studies suggests that confounding is not a concern in this body of evidence. Each of the 18 low risk-of-bias studies that observed an association between fluoride and IQ (see **Summary of Results** section above) considered a unique combination of potential confounders. The findings of these studies consistently provide evidence of an association between lower IQ in children and exposure to fluoride regardless of the inclusion or absence of consideration for any one or combination of potential confounders of interest. For example, maternal or family member smoking was addressed in 7 of the 19 low risk-of-bias studies, and this did not appear to affect the conclusions. All 7 studies that accounted for smoking found evidence of an association between fluoride exposure and lower IQ scores as did 11 of the 12 studies that did not account for smoking. Similarly, all 16 studies that addressed the three key confounders (age, sex, SES) (16 of 16 studies), and two of the three studies that did not fully account for them, also found evidence of an association between lower IQ scores and fluoride exposure. In summary, when considering the impact of each potential confounder (or combinations of potential confounders considered) on the consistency of results, no trends are discernable that would suggest that bias due to confounding has impacted or would explain the consistency in findings across the body of evidence that fluoride exposure is associated with lower IQ in children.

Five of the low risk-of-bias studies confirmed the robustness of the results by conducting sensitivity analyses (Bashash *et al.* 2017, Green *et al.* 2019, Yu *et al.* 2018, Wang *et al.* 2020b, Till *et al.* 2020), and none of the sensitivity analyses adjusting for additional confounders found meaningful shifts in the association between fluoride exposure and IQ or other measures of cognitive function. Bashash *et al.* (2017) found that adjusting for HOME score increased the association between maternal urinary fluoride and children's IQ. Green *et al.* (2019) reported that adjusting for lead, mercury, manganese, perfluorooctanoic acid, and arsenic concentrations did not substantially alter the associations with IQ. Sensitivity analyses by Yu *et al.* (2018) that adjusted for covariates (including age, sex, and socioeconomic status) did not find differences in the results compared to the primary analyses. Wang *et al.* (2020b) found the results of the sensitivity analysis to be the same as the results from the primary

analysis. Till *et al.* (2020) observed that adjusting for maternal urinary fluoride levels had little effect on the results.

Among the 19 low risk-of-bias studies, three studies were identified that have potential for bias due to confounding (Ding *et al.* 2011, Soto-Barreras *et al.* 2019, Cui *et al.* 2020). This was mainly due to lack of details on confounders considered key for all studies (i.e., age, sex, and SES). See **Appendix 4** for further discussion of the risk-of-bias concerns regarding confounding for individual studies. Although these three studies have some potential for bias due to confounding, these studies are considered to have low potential for bias overall, as they have low potential for bias for the other two key risk-of-bias questions (exposure characterization and outcome assessment), and no other major concerns for bias were identified. Consistent with the 16 studies that adequately addressed confounding, two of these three studies also provide evidence of an association between fluoride exposure and lower IQ scores in children.

Taken together and considering the consistency in the results despite the variability across studies in potential confounders considered, bias due to confounding is not believed to be a concern in the body of evidence. The potential for the consistency in results to be attributable to bias due to confounding in the 19 low risk-of-bias studies is considered low.

Figure 6. Potential Confounders Considered in Low Risk-of-bias IQ Studies Conducted in Children

Study (Location)[1]	Potential Confounding Factors Considered[2]																Notes	Reported Effect of Fluoride[4]
	Subject Characteristics				Other Exposures					Socioeconomic Factors		Parental Characteristics				Other[3]		
	Age	Sex	Race/Ethnicity	Health Factors[3]	Arsenic	Smoking	Iodine	Lead	Other[3]	SES	Caregiving Environment (e.g., HOME score)	Demographics[3]	Reproductive Factors[3]	Health Factors[3]	IQ			
Overall RoB Rating for Confounding: Probably Low																		
Bashash 2017 (Mexico)	√	√	–	–	√	√	–	√	√	√	√	√	√	–	√	√	Other exposures: Hg, Ca Demographics: maternal age Reproductive: parity, birth order, birth weight, gestational age at delivery Other: cohort	Yes
Choi 2015 (China)	√	√	–	√	√	–	–	√	–	√	–	√	√	√	–	√	Health: subject Fe deficiency, illnesses before age 3, medical history of subject and caretakers Demographics: parental age Reproductive: parity Other: residential history	Yes
Cui 2018 (China)	√	√	√	√	√	√	√	–	–	√	–	√	√	√	–	√	Health: subject BMI, stress/anger/anxiety/depression, psychological trauma, having a cold, in relatives: thyroid diseases, cancer, mental retardation Demographics: maternal age Reproductive: abnormal birth Other: alcohol consumption, proximity to factory, physical activity, various dietary factors, environmental noise	Yes
Green 2019 (Canada)	√	√	√	–	√	√	–	√	√	√	√	√	√	–	–	√	Other exposures: Hg, Mn, PFOA Demographics: parental age, pre-pregnancy BMI Reproductive: parity, weeks of gestation, birth weight, maternal chronic condition during pregnancy Other: alcohol consumption, birth country, voiding interval in urine sampling, breastfeeding duration	Yes
Rocha-Amador 2007 (Mexico)	√	√	–	√	√	–	–	√	–	√	–	–	–	–	–	–	Health: subject height and weight by age, transferrin saturation	Yes
Saxena 2012 (India)	√	√	–	√	√	–	√	√	–	√	–	–	–	–	–	√	Health: subject height for age ratio, weight for height ratio, medical history Other: residential history	Yes
Seraj 2012 (Iran)	√	√	–	–	√	–	√	√	–	√	–	–	–	–	–	√	Other: fluorosis intensity	Yes
Sudhir 2009 (India)	√	√	–	–	√	–	–	√	–	√	–	–	–	–	–	√	Other: staple food consumed, liquids routinely consumed, aids used for oral hygiene maintenance (fluoridated or nonfluoridated)	Yes
Till 2020 (Canada)	√	√	√	–	√	√	–	–	–	√	√	–	–	–	–	√	Other: city	Yes
Trivedi 2012 (India)	√	√	–	–	√	–	√	–	–	√	–	–	–	–	–	–		Yes
Wang 2012 (China)	√	√	–	√	–	–	√	√	–	√	–	–	–	√	–	√	Health: medical conditions Other: transportation, natural environment, and lifestyle	Yes
Wang 2020b (China)	√	√	–	√	√	√	√	√	–	√	–	–	√	–	–	√	Health: subject BMI, exclusions based on diseases affecting intellegence, history of trauma or neurological disorders, positive screening test history Reproductive: low birth weight Other: alcohol consumption	Yes
Xiang 2003 (China)	√	√	–	–	–	–	√	√	–	√	–	–	–	–	–	–		Yes
Xiang 2011 (China)	√	√	–	–	–	–	√	√	–	√	–	–	–	–	–	–		Yes
Yu 2018 (China)	√	√	–	√	√	√	√	√	√	√	–	–	√	–	–	√	Health: subject BMI Reproductive: disease history during pregnancy, delivery conditions Other: dental fluorosis prevalence, consanguineous marriage	Yes
Zhang 2015b (China)	√	√	–	√	√	–	√	√	√	√	–	–	–	–	–	√	Health: subject physical and mental health status Other: thyroid hormone levels, residential history, having knowledge of fluorosis, COMT genotype	Yes
Overall RoB Rating for Confounding: Probably High																		
Cui 2020 (China)	–	√	–	√	√	√	√	–	–	√	–	√	√	√	–	√	Health: stress/anger/anxiety, having a cold, in relatives: mental retardation Demographics: maternal age Reproductive: abnormal birth, smoking and drinking during pregnancy Other: thyroid horomone levels	Yes[5]
Ding 2011 (China)	√	–	–	–	√	–	√	√	–	–	–	–	–	–	–	–	–	Yes
Soto-Barreras 2019 (Mexico)	√	√	–	–	–	–	–	–	–	√	–	–	–	–	–	–		No

Notes:

[1]Includes all low risk-of-bias IQ studies in children. Studies are organized as those with an overall risk-of-bias rating for confounding as probably low (green) followed by those with an overall risk-of-bias rating for confounding as probably high (yellow).

[2]Potential confounding factors and/or effect modifiers represented here are those considered important for this evaluation. See study details provided in HAWC for information on additional confounders.

Factors outlined in blue are key potential confounders for all studies (subject age, subject sex, SES) and arsenic (which is of particular importance to some study populations).
A √ indicates that a factor was considered (and may or may not have been adjusted for in final model). For arsenic, a √ might also be used when arsenic was not expected to be an issue because there is no evidence to indicate that the co-exposure was prevalent or occurring in relation to fluoride. See risk-of-bias explanations in **Appendix 4** (or HAWC) for details. A hyphen (–) indicates that the factor was not considered.
[3]See the "Notes" column for additional details.
[4]Extent of reported effects varies by study. "Yes" indicates that study authors provided evidence of an association between lower IQ scores and fluoride exposure.
[5]Study reported lower IQ scores with increasing fluoride exposure, but the results did not achieve statistical significance.

High Risk-of-bias Studies

Most high risk-of-bias studies (n = 47) considered potential confounders to some degree through study design or analysis; however, when considering the full scale of potential concerns of bias due to confounding, all but three of these studies were rated probably or definitely high risk of bias for not adequately addressing potential confounders. The majority of high risk-of-bias studies accounted for one or two of the three potential confounding variables considered key for all studies (age, sex, SES) but did not address all three, while also not addressing other potential confounders considered important for the specific study population and outcome. Potential confounding related to important co-exposures (e.g., arsenic) was often not addressed in high risk-of-bias studies. In studies where there was high exposure to fluoride via drinking water with high naturally-occurring fluoride or from the use of coal-containing fluoride, most researchers did not account for potential exposures to arsenic, which is commonly found in coal and drinking water in fluoride-endemic areas of China and Mexico.

Despite the lack of adequate consideration of confounders in the vast majority of high risk-of-bias studies, the results across most of these studies (41 of 47) consistently provide evidence of an association between fluoride exposure and IQ, supporting the results observed in the low risk-of-bias studies. This finding suggests that confounding is likely less of a concern for the body of evidence as a whole than for any individual study. Although the high risk-of-bias studies may have more potential for bias due to confounding compared to the low risk-of-bias studies, the consistent IQ findings across high and low risk-of-bias studies indicate that the results cannot be explained based on bias due to confounding.

Exposure Characterization in IQ Studies

Low Risk-of-bias Studies

In general, there were few if any risk-of-bias concerns regarding exposure characterization in the low risk-of-bias studies. These studies mainly had individual exposure data based on urine or water measures with appropriate analyses. Although there are concerns related to using urine samples (see **Risk-of-bias Considerations for Human Studies** section for details), many studies provide evidence to suggest that urinary fluoride is a reasonable measure of exposure. Using three methods to account for urine dilution, Till *et al.* (2018) reported that adjusted risk estimates did not differ from unadjusted estimates. Analyzing the same study population as Till *et al.* (2018), Green *et al.* (2019) found that adjusting for time of urine collection or time of collection since last void during pregnancy did not substantially affect associations with IQ results in either boys or girls. In addition, adjusting the maternal urinary fluoride for creatinine did not substantially alter the association observed (Green *et al.* 2019). To provide a more accurate and sensitive measurement of maternal urinary fluoride than a single measurement provides, Green *et al.* (2019) only included participants with valid fluoride measurements at all trimesters in their analysis. Other studies also measured urinary fluoride multiple times throughout pregnancy (Bashash *et al.* 2017). Some studies demonstrated correlations between the urinary fluoride

and fluoride in the drinking water, fluorosis, or estimated dose based on drinking water concentrations and consumption (Green *et al.* 2019, Saxena *et al.* 2012, Zhang *et al.* 2015b, Ding *et al.* 2011, Choi *et al.* 2015, Yu *et al.* 2018). Till *et al.* (2018) demonstrated that there was a linear association between urinary fluoride concentrations in pregnant women and drinking water fluoride concentrations regardless of method to correct for urine dilution or whether or not adjustments were made for dilution. Bashash *et al.* (2017) excluded exposure outliers but found that doing so did not substantively change the results. Taken together, these studies suggest that urinary fluoride is a reasonable measure of exposure despite some of the potential issues.

All but one low risk-of-bias study was rated probably or definitely low risk of bias for exposure assessment. Seraj *et al.* (2012) had potential exposure misclassification and was rated probably high risk of bias for exposure assessment. Villages were categorized as normal (0.5–1 ppm), medium (3.1 ± 0.9 ppm), or high (5.2 ± 1.1 ppm) based on average fluoride content in drinking water in varying seasons over a 12-year period. Mild fluorosis observed in children in the normal fluoride level group indicates that there may have been higher exposure in this group at some point in the past; however, this would bias the results towards the null, and the children in the normal fluoride group had a significantly higher IQ score compared to the medium and high fluoride groups ($p = 0.001$). There were also significant associations between lower IQ scores and fluorosis intensity (p-value = 0.014) and water fluoride concentration when evaluated as a continuous variable (p-values < 0.001). Although there is potential for exposure bias, the apparent exposure misclassification and inclusion of children with higher fluoride exposure in the normal group indicate that the association may be greater than observed in this study.

High Risk-of-bias Studies

A frequent, critical limitation among the high risk-of-bias studies was lack of information regarding exposure or poor exposure characterization. Many of the high risk-of-bias studies only compared subjects living in two regions with differing levels of fluoride exposure, and although most of them did provide some differentiation in levels of fluoride between the areas, limited or no individual exposure information was reported. Among studies that provided drinking water levels of fluoride in two areas being compared, sufficient information to determine if the individual study subjects were exposed to these levels was often not reported. Some studies also lacked information on fluoride analysis methods and timing of the exposure measurements. In some cases (n = 3), study areas that were considered endemic for dental and/or skeletal fluorosis were compared to non-endemic areas, or high-fluoride areas were compared to low-fluoride areas, with no other information provided on fluoride levels in the areas (Sun *et al.* 1991, Li *et al.* 2003 [translated in Li *et al.* 2008c], Ren *et al.* 1989 [translated in Ren *et al.* 2008]). Although living in an area endemic for fluorosis could be an indicator of exposure, these studies did not specify if the study subjects themselves had fluorosis. Another study used only dental fluorosis as a measure of fluoride exposure in subjects that were all from an endemic area with similar drinking water fluoride levels (Li *et al.* 2010). In one case, multiple sources of fluoride exposure were assessed separately without properly controlling for the other sources of exposure, which could bias the results (Broadbent *et al.* 2015). Broadbent *et al.* (2015) assessed fluoride exposure in three ways: use of community water in fluoridated area versus non-fluoridated area, use of fluoride toothpaste (never, sometimes, always), or use of fluoride tablets prior to age 5 (ever, never). The same children were used for each analysis without accounting for fluoride exposure through other sources. For example, there were 99 children included in the non-fluoridated area for the community water evaluation, but there is no indication that these 99 children were not some of the 139 children that ever used supplemental fluoride tablets or the 634 children that always used fluoride toothpaste. Therefore, comparing fluoridated areas to non-fluoridated areas without accounting for other sources of exposure that might occur in these non-fluoridated areas would bias the results towards the null.

Outcome Assessment for IQ Studies

Low Risk-of-bias Studies

The low risk-of-bias studies have few concerns regarding outcome assessment. Eighteen of the 19 studies used appropriate methods for measuring IQ in the study population being assessed, and blinding of outcome assessors was not a concern. Fourteen of the studies reported blinding of the outcome assessors, or correspondence with the study authors indicated that it was not likely an issue. For the remainder of the studies, it was assumed that the outcome assessors were most likely blind because exposure was assessed via urine or drinking water obtained at the same time as the outcome assessment in general population studies. One IQ study (Sudhir *et al.* 2009) had concerns for potential bias in the outcome assessment due to lack of information to determine if blinding at the time of the outcome assessment was a concern (see **Appendix 4** for details).

High Risk-of-bias Studies

Among the studies with high risk of bias, the main limitation in the outcome assessment was the lack of reporting on blinding of the outcome assessor (i.e., whether the outcome was assessed without knowledge of exposure). Although there is little concern that the children's knowledge of their own exposure would bias the way they took the IQ tests, there is potential for bias if the tests were administered by an interviewer, or if the scoring of results could be subjective (e.g., drawing tests), and the interviewer or scorer had knowledge of the children's exposure. Most of the studies did not provide sufficient information on the person scoring or administering the tests or other information on the assessment methods to alleviate concerns for potential interviewer or reviewer bias.

High risk-of-bias studies were mainly carried out in two separate populations without information provided that the tests were conducted in a central location, and in many cases the methods indicated that the tests were conducted at the schools in the study area indicating that there was likely knowledge of the exposure. In some cases, the outcomes were not considered sensitive measures (e.g., Seguin Form Board Test to test for IQ), or the test was not considered appropriate for the study population (e.g., a test validated in a western population was used on a rural Chinese population).

Confidence Assessment of Findings on IQ in Children

There is moderate confidence in the body of evidence that fluoride exposure is associated with lower IQ in children based on the consistent evidence of an association between high total fluoride exposure (mainly greater than the WHO Drinking Water Quality Guideline [≥1.5 mg/L] (WHO 2011), but also high exposure via fluoridated salt and food) and lower IQ. Eighteen of the 19 studies reported associations between lower IQ scores and higher fluoride levels. These studies were conducted in 5 different countries on more than 7,000 children from 15 different study populations demonstrating consistency across multiple populations. There is consistency in results across study design with associations observed in the prospective cohort studies (n = 3) and cross-sectional studies with 12 of the cross-sectional studies providing evidence of prior long-term, chronic fluoride exposure. There is also consistency in results across studies using different exposure measures, including urinary and drinking water fluoride. The initial moderate confidence rating in the body of evidence is based on 15 studies that have 3 of the 4 key features shown in **Figure 1** (i.e., where exposure occurred prior to outcome, that evaluated individual-based outcomes and used a comparison group). Factors to consider for increasing or decreasing the confidence in the body of evidence are provided in **Figure 1**. Discussion of these factors for upgrading or downgrading the confidence in the evidence is presented below.

- **Risk of bias**: Only studies that were considered to have low risk of bias were included in the moderate confidence rating; therefore, there was no downgrade for risk-of-bias concerns.

- **Unexplained inconsistencies:** The data are consistent and there was no downgrade for this factor. In terms of IQ data, 18 studies observed lower IQ with higher fluoride exposure. The one study that did not observe an effect did not provide results in a comparable manner and is not considered unexplained.

- **Indirectness:** IQ in humans is a direct measure of the effect of interest and therefore no adjustment in confidence is warranted.

- **Imprecision:** There is no evidence of imprecision that would warrant a downgrade. Eighteen studies reported lower IQ with higher fluoride, and no issues with imprecision were identified to challenge the significance of the effect.

- **Publication bias:** There is no strong evidence of publication bias; therefore, no downgrade was applied for publication bias. Two published meta-analyses (Duan *et al.* 2018, Choi *et al.* 2012) did not indicate strong evidence of publication bias. The draft meta-analysis conducted by NTP in the September 16, 2020, draft monograph found no publication bias among the low risk-of-bias studies (NTP 2020). Among high risk-of-bias studies, adjusting for publication bias using the trim-and-fill analysis estimated that, in the absence of publication bias, the inverse direction of effect and statistical significance remained, thus indicating that there was no need to downgrade for publication bias.

- **Large magnitude of effect:** Although some individual studies indicate a large magnitude of effect, the magnitude of effect was not the same across all studies. Therefore, the overall data would not support an upgrade due to a large magnitude of effect.

- **Dose-response:** There is evidence of a dose-response relationship that could justify an upgrade to the confidence in the body of evidence. However, many of the studies that provide data to evaluate dose response were judged to be high risk of bias. In addition, the data appear less clear in the lower dose range. The refined NTP meta-analysis (in preparation) may further inform this issue.

- **Residual confounding:** Xiang *et al.* (2003a), Xiang *et al.* (2011), and Wang *et al.* (2012) studied the same population where arsenic occurred in the area with low fluoride, but did not occur in the area with high fluoride. This would have biased the results toward the null, but there were significantly lower IQ scores in the area with high fluoride. The remaining studies do not provide enough information to consider residual confounding as an impactful factor for the body of evidence. Therefore, the overall data would not support an upgrade due to residual confounding.

- **Consistency:** The high quality studies demonstrate a consistent pattern of findings that fluoride exposure is associated with lower IQ scores in children; however, the consideration of a potential upgrade for consistency in the methods is primarily for non-human animal evidence where it would be applied to address increased confidence for consistent effects across multiple non-human animal species. For human evidence, it is generally not applied and the data would only be considered in deciding whether to downgrade for unexplained inconsistency. Therefore, no upgrade is applied for consistency.

Although the OHAT approach for evidence integration allows for the initial confidence in the body of evidence to be increased from moderate to high confidence based on evidence of a dose response,

these procedures are not algorithms and require scientific judgments. The NTP judgement is that the magnitude of effect and the overall strength and quality of the human literature base provide a moderate confidence in the body of evidence that fluoride is associated with lower IQ in children (see **Discussion** section for strengths and limitations of the evidence base). Note that additional, well-designed prospective cohort studies with individual-level exposure data and outcome measures could provide increased confidence in the association between lower IQ in children and fluoride exposure.

Other Neurodevelopmental or Cognitive Effects in Children

Low Risk-of-bias Studies

Overview of Studies

Nine low risk-of-bias studies evaluated the association between fluoride exposure and cognitive neurodevelopmental effects other than IQ in children. These 9 studies were conducted in multiple study populations in three countries, specifically:

- 3 were conducted in 3 areas of China on 3 study populations;
- 4 were conducted in 2 areas of Mexico on 3 study populations; and
- 2 were conducted in Canada using the same study population.

There is considerable heterogeneity across studies, particularly in the different health outcomes evaluated and ages assessed. Most studies measured fluoride in the drinking water or urine (child or maternal) with one study using severity of dental fluorosis as an exposure measure in addition to drinking water and children's urine. Two of the studies were conducted on infants, with one evaluating effects within 72 hours of birth (Li *et al.* 2004 [translated in Li *et al.* 2008a]) and the other evaluating effects at 3 to 15 months of age (Valdez Jimenez *et al.* 2017). The remaining studies were conducted in children of varying ages ranging from 4 to 17 years. Other cognitive neurodevelopmental outcomes assessed include neurobehavioral effects in infants, learning and memory impairment, and learning disabilities such as attention deficit hyperactivity disorder (ADHD). Few studies measured the same health outcomes, used the same outcome assessment methods, or evaluated the same age groups.

Table 7 provides a summary of study characteristics and key findings related to other cognitive neurodevelopmental outcomes and fluoride exposure for the nine low risk-of-bias studies. The different tests conducted and the populations on which the tests were conducted are also indicated in **Table 7**. Several of these studies conducted multiple analyses and reported results on multiple endpoints. The purpose of the table is to summarize key findings (independent of whether an effect was found) from each study and is not meant to be a comprehensive summary of all results. For each study, results are summarized for each exposure measure assessed. Results from multiple analyses using the same exposure measure may not all be presented unless conflicting results were reported.

Table 7. Studies on Other Neurodevelopmental and Cognitive Function in Children[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurological Outcome Summary[c]
China					
Li *et al.* (2004) [translated in Li *et al.* 2008a]	Cross-sectional Zhaozhou County, Heilongjiang Province/neonates [91]	Drinking water Range: 0.5–1.0 mg/L (control); 1.7–6.0 mg/L (high) Maternal urine during pregnancy Mean (SD): 1.74 (0.96) mg/L (control); 3.58 (1.47) mg/L (high)	Neonates (24–72 hours after delivery)	Neurodevelopmental: Neonatal behavioral neurological assessment (NBNA)	Significant differences in neurobehavioral assessment total scores between high-fluoride (36.48 ± 1.09) and control groups (38.28 ± 1.10) (subjects divided into high fluoride group and control group based on drinking water fluoride levels in place of residence); significant differences in total score of behavioral capability that includes measures of non-biological visual orientation reaction and biological visual and auditory orientation reaction between the two groups (11.34 ± 0.56 in controls compared to 10.05 ± 0.94 in high-fluoride group) No statistical adjustment for confounders
Choi *et al.* (2015)	Cross-sectional Mianning County/1st grade children [51]	Drinking water GM: 2.20 mg/L Children's urine GM: 1.64 mg/L Severity of fluorosis (Dean Index)	Children (ages 6–8 years)	Learning and memory: Neuropsychological tests including WRAML Visual motor ability: WRAVMA Motor ability: Finger tapping task Manual dexterity: Grooved pegboard test	Outcomes unrelated to the IQ test not significantly associated with any fluoride exposure measure Adjusted for child's age, child's gender, parity, illness before 3 years old, household income last year, and caretaker's age and education
Wang *et al.* (2020a)	Cross-sectional Tongxu County/school children [325]	Children's urine Mean (SD): 1.54 (0.89) mg/L	Children (ages 7–13 years)	ADHD and behavior measures: Conners' Parent Rating Scale-Revised (Chinese version) (CPRS-48)	Significant association between psychosomatic problems and urinary fluoride level (per 1-mg/L increase; β = 4.01; 95% CI: 2.74, 5.28; OR for T-score >70 = 1.97; 95% CI: 1.19, 3.27); no associations between urinary fluoride level and ADHD index or other behavioral measures Adjusted for child's age, child's gender, child's BMI, urinary creatinine, mother migrated and father migrated

Table 7. Studies on Other Neurodevelopmental and Cognitive Function in Children[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurological Outcome Summary[c]
Mexico					
Rocha-Amador *et al.* (2009)	Cross-sectional Durango/elementary school children [80]	Children's urine GM (SD): 5.6 (1.7) mg/L	Children (ages 6–11 years)	Visuospatial organization and visual memory: Rey-Osterrieth Complex Figure Test, children's version	Significant correlation between urinary fluoride and visuospatial organization (r = −0.29) and visual memory scores (r = −0.27); no significant correlation with arsenic Adjusted for age
Valdez Jimenez *et al.* (2017)	Cohort (Prospective) Durango City and Lagos de Moreno/infants [65]	Maternal urine Range: 0.16–8.2 mg/L (all trimesters) Drinking water Range: 0.5–12.5 mg/L (all trimesters)	Infants (ages 3–15 months)	Mental development index (MDI): Bayley Scales of Infant Development II (BSDI-II) Psychomotor developmental index (PDI): Bayley Scales of Infant Development II (BSDI-II)	Significant correlation between maternal urinary fluoride and MDI score during first trimester (adjusted β = −19.05; SE = 8.9) and second trimester (adjusted β = −19.34; SE = 7.46); no significant associations between maternal urinary fluoride and PDI score; analyses of outcomes using drinking water fluoride not performed Adjusted for gestational age, child's age, marginality index, and type of drinking water
Bashash *et al.* (2017)[1]	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [299] GCI analysis [287]	Maternal urine during pregnancy Mean (SD): 0.90 (0.35) mg/L Children's urine Mean (SD): 0.82 (0.38) mg/L	Children (age 4 years)	General cognitive index (GCI): McCarthy Scales of Children's Abilities (MSCA)	Significant effect between maternal urinary fluoride and offspring GCI score (adjusted β = −3.15; 95% CI: −5.42, −0.87); associations with children's urine not significant Adjusted for gestational age, weight at birth, child's gender, parity (being the first child), age at outcome measurement, and maternal characteristics including smoking history (ever smoked during the pregnancy vs. nonsmoker), marital status (married vs not married), age at delivery, IQ, education, and cohort

Table 7. Studies on Other Neurodevelopmental and Cognitive Function in Children[a,b]					
Study	**Study Design (Location/Subjects) [n]**	**Exposure Measures and Summary Statistics**	**Assessment Timing**	**Outcome and Methods**	**Neurological Outcome Summary[c]**
Bashash *et al.* (2018)[1]	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [210]	Maternal urine during pregnancy Mean 0.85 (95% CI: 0.81, 0.90) mg/L	Children (ages 6–12 years)	ADHD: Conners' Rating Scales-Revised (CRS-R)	Significant associations between maternal urinary fluoride and CRS-R scores including Cognitive Problems + Inattention Index (adjusted β = 2.54; 95% CI: 0.44, 4.63), DSM-IV Inattention Index (adjusted β = 2.84; 95% CI: 0.84, 4.84), DSM-IV ADHD Total Index (adjusted β = 2.38; 95% CI: 0.42, 4.34), and ADHD Index (adjusted β = 2.47; 95% CI: 0.43, 4.50) Adjusted for gestational age, birth weight, child's gender, parity, age at outcome measurement, and maternal characteristics including smoking history (ever smoked vs. nonsmoker), marital status (married vs. not married), education, socioeconomic status, and cohort
Canada					
Barberio *et al.* (2017b)[2]	Cross-sectional General population/ Canadian Health Measures Survey (Cycles 2 and 3) [2,221]	Children's urine Mean Cycle 2: 32.06 (95% CI: 29.65, 34.46) µmol/L Mean Cycle 3: 26.17 (95% CI: 22.57, 29.76) µmol/L	Children (ages 3–12 years)	Learning disability, ADHD (Cycle 2 only): Parent or child self-report	Significant increase in adjusted OR for learning disability (adjusted OR = 1.02; 95% CI: 1.00, 1.03) only when Cycle 2 and 3 were combined using unadjusted urinary fluoride; no significant associations found between urinary fluoride and ADHD (only evaluated in Cycle 2); no significant associations found when using creatinine- or specific gravity-adjusted urinary fluoride Adjusted for child's age, child's gender, household income adequacy, and highest attained education in the household

Table 7. Studies on Other Neurodevelopmental and Cognitive Function in Children[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurological Outcome Summary[c]
Riddell *et al.* (2019)[2]	Cross-sectional General population/ Canadian Health Measures Survey (Cycles 2 and 3) [3,745]	Drinking water Mean (SD): 0.23 (0.24) mg/L; non-fluoridated water-0.04 (0.06) mg/L, fluoridated water-0.49 (0.22) Community water fluoridation status (yes or no) Children's urine Mean (SD): 0.61 (0.39) mg/L; non-fluoridated water-0.46 (0.32) mg/L, fluoridated water-0.82 (0.54)	Children (ages 6–17 years)	Hyperactivity/inattention: Strengths and Difficulties Questionnaire (SDQ) ADHD: parent or self-reported physician diagnosis	Significantly increased risk of ADHD with fluoride in tap water (adjusted OR = 6.10 per 1-mg/L increase; 95% CI: 1.60, 22.8) or community water fluoridation status (1.21; 95% CI: 1.03, 1.42), but not with urinary fluoride; similar results observed with attention symptoms based on the SDQ scores Adjusted for child's age, child's gender, child's BMI, ethnicity, parental education, household income, blood lead, and smoking in the home

Studies with the same number superscript are based on the same study population.

[a]Includes low risk-of-bias studies.

[b]Definitions: **ADHD**: attention-deficit/hyperactivity disorder; **BMI**: body mass index; **GCI**: General Cognitive Index; **GM**: geometric mean; **HOME**: Home Observation Measurement of the Environment; **IQ**: intelligence quotient; **MSCA**: McCarthy Scales of Children's Abilities; **WASI**: Wechsler Abbreviated Scale of Intelligence (Spanish version); **WISC-IV**: Wechsler Intelligence Scale for Children-Revised (as reported by Choi *et al.* 2015); **WRAML**: Wide Range Assessment of Memory and Learning; **WRAVMA**: Wide Range Assessment of Visual Motor Ability.

[c]Associations between other cognitive neurodevelopmental outcomes in children and fluoride levels were reported quantitatively, as possible. For studies with multiple analyses and results, the table summarizes key findings and is not a comprehensive summary of all findings. Results also indicated when a study found no association, provided as a qualitative statement of no association.

Summary of Results

Overall Findings

Although discussed together in this section, there are various health outcomes assessed in the nine low risk-of-bias studies of other neurodevelopmental outcomes, including neurobehavioral scores in infants (2 studies), cognitive tests in children other than IQ (3 studies), and ADHD or learning disabilities (4 studies) in children. Altogether, the results from eight of nine low risk-of-bias studies (three prospective cohort studies and five cross-sectional studies from seven different study populations) provide evidence of significant associations between fluoride exposure and cognitive neurodevelopmental outcomes in children other than decrements in IQ (see **Figure D8** through **Figure D10**) (Bashash *et al.* 2017, Valdez Jimenez *et al.* 2017, Bashash *et al.* 2018, Li *et al.* 2004 [translated in Li *et al.* 2008a], Rocha-Amador *et al.* 2009, Barberio *et al.* 2017b, Wang *et al.* 2020a, Riddell *et al.* 2019). Only one cross-sectional study did not find a significant association between fluoride exposure and a measure of cognitive neurodevelopment (Choi *et al.* 2015).

Although there is heterogeneity in the outcomes assessed and a limited number of directly comparable studies, the data provide additional evidence (beyond the consistent evidence of an association between fluoride exposure and IQ) of an effect of fluoride exposure on cognitive neurodevelopment. The body of evidence from the nine low risk-of-bias studies is described in further detail below and are grouped into outcome categories of studies that are most comparable.

Results in Infants

Two studies evaluated neurobehavioral effects in infants either shortly after birth or at 3 to 15 months of age (Li *et al.* 2004 [translated in Li *et al.* 2008a], Valdez Jimenez *et al.* 2017). Both studies observed a significant association between higher fluoride exposure and lower neurobehavioral scores. In neonates (1–3 days old), the high fluoride group (3.58 ± 1.47 mg/L fluoride based on spot maternal urine collected just prior to birth) had significantly lower total neurobehavioral assessment scores (36.48 ± 1.09 versus 38.28 ± 1.10 in controls; p-value < 0.05) and total behavioral capacity scores (10.05 ± 0.94 versus 11.34 ± 0.56 in controls; p-value < 0.05) compared to the control group (1.74 ± 0.96 mg/L fluoride) as measured by a standard neonatal behavioral neurological assessment (NBNA) method (Li *et al.* 2004 [translated in Li *et al.* 2008a]). In infants 3 to 15 months of age, the Mental Development Index (MDI)—which measures functions including hand-eye coordination, manipulation, understanding of object relations, imitation and early language development—was significantly inversely correlated with maternal urinary fluoride in both the first and second trimesters (adjusted βs = −19.05 with standard error of 8.9 for first trimester [p-value = 0.04] and −19.34 with standard error of 7.46 for second trimester [p-value = 0.013]) (Valdez Jimenez *et al.* 2017). Note that this study did not find an association between maternal fluoride during any trimester and Psychomotor Developmental Index (PDI), which measures gross motor development (adjusted βs = 6.28 and 5.33 for first and second trimesters, respectively; no variance provided) (Valdez Jimenez *et al.* 2017).

Results for Cognitive Tests Other Than IQ in Children

Three studies conducted tests on cognitive function in children that were not part of an IQ test. None of the studies conducted the same tests, but two of the three studies observed associations between fluoride exposure and lower test scores. The General Cognitive Index (GCI) of the McCarthy Scales of Children's Abilities (MSCA) in 4-year-old children was significantly inversely associated with maternal creatinine-adjusted urinary fluoride levels during pregnancy (collected during each trimester) (adjusted β = −3.15 [95% CI: −5.42, −0.87; p-value = 0.01] in a model adjusting for main covariates including gestational age, weight at birth, sex, maternal smoking, and indicators of socioeconomic status). The

association remained even after adjusting for maternal bone lead (adjusted β = −5.63 [95% CI: −8.53, −2.72; p-value < 0.01]) (Bashash *et al.* 2017) (see **Figure D10**). Choi *et al.* (2015), however, evaluated cognitive function endpoints in addition to IQ and found no significant associations between concurrent water or urinary fluoride levels and Wide Range Assessment of Visual Motor Ability (WRAVMA) scores, finger tapping test scores, and grooved pegboard test scores, although there were some significant associations based on degree of fluorosis (see **Figure D10**). Another study using construction and memory scores in children 6–11 years old observed significantly lower scores with increasing concurrent child single spot urinary fluoride even after adjusting for age (−0.29 and −0.27 for copy [p-value < 0.001] and immediate recall [p-value < 0.001], respectively [CIs not reported]). Although these children were also exposed to arsenic, the presence of arsenic could not explain the changes as test scores were not significantly associated with urinary arsenic levels (−0.05 and 0.02 for copy and immediate recall, respectively [CIs not reported]) (Rocha-Amador *et al.* 2009) (see **Figure D9**).

Attention-related Disorders Including ADHD and Learning Disabilities in Children

Four studies evaluated attention related disorders or learning disabilities. All four studies found an association between increased fluoride and increased ADHD or learning disability; however, studies varied in the exposure metrics and outcomes measure. Bashash *et al.* (2018) evaluated behaviors associated with ADHD in children ages 6–12 years using the Conners' Rating Scales-Revised (CRS-R) and observed significant associations between maternal urinary fluoride (measured during each trimester) and ADHD-like symptoms, particularly those related to inattention (an increase in 0.5 mg/L of maternal urinary fluoride was significantly associated with a 2.84-point increase [95% CI: 0.84, 4.84; p-value = 0.0054] in the DSM-IV Inattention Index and a 2.54-point increase [95% CI: 0.44, 4.63; p-value = 0.0178] in the Cognitive Problems and Inattention Index). These two scales contributed to the global ADHD Index and the DSM-IV ADHD Total Index, which were also significantly associated with higher levels of prenatal fluoride exposure (an increase of 0.5 mg/L in maternal urinary fluoride was associated with a 2.38-point increase [95% CI: 0.42, 4.34; p-value = 0.0176] in the DSM-IV ADHD Total Index and a 2.47-point increase [95% CI: 0.43, 4.50; p-value = 0.0175] in the ADHD Index) (see **Figure D10**). Significant associations were not observed between maternal urinary fluoride concentrations during pregnancy and child performance on measures of hyperactivity nor were there any significant results in children using the Conners' Continuous Performance Test (CPT-II, 2nd Edition), a computerized test of sustained attention and inhibitory control (Bashash *et al.* 2018). Wang *et al.* (2020a) also used a Conners' Parent Rating Scale (Chinese version) to assess behavioral outcomes in children ages 7–13 years, but only found a significant association between spot urinary fluoride concentrations in children (model adjusted for creatinine) and psychosomatic problems (adjusted OR for T-score > 70 = 1.97 [95% CI: 1.19, 3.27; p-value = 0.009] and adjusted β = 4.01 [95% CI: 2.74, 5.28; p-value < 0.001]). No associations were found between spot urinary fluoride and ADHD index or other behavioral measures.

Barberio *et al.* (2017b) evaluated learning disabilities in children 3–12 years of age, including ADHD, attention deficit disorder (ADD), and dyslexia, as part of the Canadian Health Measures Survey and found a small but significantly increased risk in self-reported (children 12 years of age) or parent- or guardian-reported (children 3–11 years of age) learning disabilities associated with higher spot urinary fluoride levels in children (adjusted OR = 1.02; 95% CI: 1.00, 1.03; p-value < 0.05) (see **Figure D11**); however, significant associations were not observed in analyses using creatinine- or specific gravity-adjusted urinary fluoride (Barberio *et al.* 2017b). Barberio *et al.* (2017b) also reported no associations between single spot urinary fluoride and ADHD in children ages 3 to 12 years. Riddell *et al.* (2019) used the same Canadian Health Measured Survey, but evaluated children 6–17 years old. Riddell *et al.* (2019) found a significantly increased risk for ADHD diagnosis with both tap water fluoride (adjusted OR per 1-mg/L increase = 6.10; 95% CI: 1.60, 22.8; p-value < 0.05) and community water fluoridation status

(adjusted OR = 1.21; 95% CI: 1.03, 1.42; p-value < 0.05). A similar increase in hyperactivity-inattention symptoms score based on the Strengths and Difficulties Questionnaire was observed with both tap water fluoride (adjusted β per 1-mg/L increase = 0.31; 95% CI: 0.04, 0.58; p-value < 0.05) and community fluoridation status (adjusted β = 0.11; 95% CI: 0.02, 0.20; p-value < 0.05). As was observed with Barberio *et al.* (2017b), Riddell *et al.* (2019) did not observe associations between specific-gravity adjusted spot urinary fluoride concentrations and either ADHD diagnosis (adjusted OR per 1-mg/L increase = 0.96; 95% CI: 0.63, 1.46) or hyperactivity-inattention symptoms (adjusted β = 0.31; 95% CI: −0.04, 0.66).

Summary of Key Findings for Low Risk-of-bias Studies of Other Neurodevelopmental and Cognitive Effects in Children

In summation, the high quality studies (i.e., studies with low potential for bias) provide evidence of an association between fluoride exposure and neurodevelopmental and cognitive effects in children other than IQ; however, the body of evidence is limited by heterogeneity in the outcomes evaluated and few comparable studies. Across these outcomes, eight of nine studies reported a significant association between fluoride exposure and a measure neurodevelopment or cognition other than IQ, which provides support to the consistency in evidence based on children's IQ studies of an association between fluoride exposure and adverse effects on cognitive neurodevelopment.

High Risk-of-bias Studies

High risk-of-bias studies (n = 5) also provide some evidence of associations between fluoride exposure and neurodevelopmental or cognitive effects in children other than effects on IQ, but the results are inconsistent and address different outcomes (Li *et al.* 1994 [translated in Li *et al.* 2008b], Shannon *et al.* 1986, Malin and Till 2015, Morgan *et al.* 1998, Mustafa *et al.* 2018).

Risk of Bias for Neurodevelopmental or Cognitive Effect Studies in Children

The confidence in the human body of evidence was based on studies with the lowest potential for bias (i.e., studies that rated probably low or definitely low risk of bias for at least two of the three key risk-of-bias questions and did not have any other risk-of-bias concerns that would indicate serious issues with the studies). Each of the nine low risk-of-bias studies on other neurodevelopmental effects in children had little or no risk-of-bias concerns. Four of the nine studies were rated definitely low or probably low risk of bias for all risk-of-bias questions, and the remaining five studies were rated probably high risk of bias for a single question that was judged to have minimal impact on overall potential bias. None of the nine studies had a rating of definitely high risk of bias for any question. Although the nine low risk-of-bias studies had minimal or no concerns, the five studies with high overall potential for bias (n = 5) had several risk-of-bias concerns related to one or more of the three key risk-of-bias questions (confounding, exposure characterization, and outcome assessment). The key risk-of-bias questions are discussed below. Risk-of-bias ratings for other neurodevelopmental effect studies in children are available in **Figure A3-9** through **Figure A3-12** and **Appendix 4** for the low and high risk-of-bias studies.

Confounding for Other Neurodevelopmental Studies in Children

Low Risk-of-bias Studies

As discussed above, there are nine studies considered to have low risk of bias when assessed across all risk-of-bias domains. Seven of nine low risk-of-bias studies were considered to have low potential for bias due to confounding because the authors addressed potential confounders through study design or analysis. All seven of these studies addressed the three potential confounders considered key for all studies (child's age, child's sex, and socioeconomic status) and also addressed arsenic as a potential co-exposure of concern through study design or analysis. Other potential confounders, including health factors, smoking, and parental characteristics, were also addressed in many of the low risk-of-bias

studies. One of the studies (Bashash *et al.* 2018) examined several potential confounders in sensitivity analyses involving subsets of participants, including HOME scores, child contemporaneous fluoride exposure measured by child urinary fluoride adjusted for specific gravity, and maternal lead and mercury exposures. The authors reported that none of the sensitivity analyses indicated appreciable changes in the fluoride-related association with behaviors related to ADHD, nor did they find evidence of effect modification between sex and maternal urinary fluoride.

Among the nine low risk-of-bias studies, two studies were identified that have potential for bias due to confounding (Rocha-Amador *et al.* 2009, Valdez Jimenez *et al.* 2017). Although both of these studies adjusted for several confounders through analysis or study design, Valdez Jimenez *et al.* (2017) did not address a potential concern for co-exposure to arsenic, and Rocha-Amador *et al.* (2009) does not appear to adjust for SES or address why it would not be a concern in the study population (see **Appendix 4** for further details). Although these two studies have some potential for bias due to confounding, these studies are considered to have low potential for bias overall, as they have low potential for bias for the other two key risk-of-bias questions (exposure characterization and outcome assessment), and no other major concerns for bias were identified. Consistent with the IQ studies, bias due to confounding is not likely a concern for the low risk-of-bias studies.

High Risk-of-bias Studies

The five high risk-of-bias studies in the human body of evidence did not adequately address potential confounders through study design or analysis. The same concerns due to potential confounding noted previously for the high risk-of-bias children's IQ studies were also present in the other neurodevelopmental high risk-of-bias studies including not addressing the three potential confounding variables considered key for all studies (age, sex, SES) and/or not addressing potential co-exposures (e.g., arsenic) in areas of potential concern.

Exposure Characterization in Other Neurodevelopmental Studies in Children

Low Risk-of-bias Studies

There were no risk-of-bias concerns regarding exposure assessment in the low risk-of-bias studies. All of the low risk-of-bias studies had individual exposure data based on urine or water measures with appropriate analyses, and most of the urinary fluoride studies accounted for urinary dilution when appropriate. Although there are concerns related to using urine samples (see **Risk-of-bias Considerations for Human Studies** section for details), the studies that used maternal urine measured urinary fluoride multiple times throughout pregnancy (Bashash *et al.* 2017, Bashash *et al.* 2018, Valdez Jimenez *et al.* 2017). Another study demonstrated correlations between urinary fluoride and fluoride in the drinking water, fluorosis, or estimated dose based on water (Choi *et al.* 2015). Bashash *et al.* (2017) excluded exposure measurement outliers but found that doing so did not change the results in a meaningful way.

High Risk-of-bias Studies

A frequent critical limitation among the high risk-of-bias studies was lack of information regarding exposure or poor exposure characterization. In the high risk-of-bias studies that assessed the association between fluoride exposure and other neurodevelopmental and cognitive effects in children, fluoride exposure assessment was based on dental fluorosis, municipality-level water fluoridation prevalence data, number of years living in an area with fluorinated water, or group-level water samples. See the **Exposure Characterization in IQ Studies** section for further discussion on the limitations of exposure assessments in high risk-of-bias studies.

Outcome Assessment in Other Neurodevelopmental Studies in Children

Low Risk-of-bias Studies

The low risk-of-bias studies have few concerns regarding outcome assessment. Six of the nine studies used appropriate methods for measuring other neurodevelopmental effects in the study population being assessed, and blinding of outcome assessors was either reported or not a concern.

Among the nine low risk-of-bias studies, three studies were identified that have a potential for bias due to outcome assessment. One of the studies (Wang *et al.* 2020a) had potential concern for bias due to lack of information regarding the blinding of outcome assessors. Two of the studies (Riddell *et al.* 2019, Barberio *et al.* 2017b) were based on the same study population in Canada where different questions were asked in Cycles 2 (2009–2011) and 3 (2012–2013) of the Canadian Health Measures Survey (CHMS) to ascertain learning disabilities including ADHD. In Cycle 2, subjects were asked if they had a learning disability diagnosed by a health professional and, if yes, were asked what kind. In Cycle 3, CHMS did not ask what kind of learning disability was diagnosed nor was a reason for the question omission provided. Because no reason was provided for the removal of the question, and because a question on learning disability without the specific diagnosis may be more prone to bias, this change in questioning from Cycles 2 to 3 is a potential concern. Blinding was not considered an issue in these two studies, but the methods for obtaining the information are considered to be less than ideal methods for measuring learning disabilities including ADHD. Although the questionnaire asked about doctor diagnosis of a learning disability, there was no confirmation with medical records. Moreover, these questionnaires were not validated like the Conners' Rating Scales, which would have been a better method for assessing ADHD. Although the outcome assessment methods are less than ideal, there was no direct evidence that they were conducted incorrectly or that the methods would bias the results in any specific direction. Because this was the only concern in these studies, they were considered to have low risk of bias overall.

High Risk-of-bias Studies

Among the studies on other neurodevelopmental effects with high potential for bias, there were several reasons for studies to be considered probably or definitely high risk of bias for outcome assessment. One study (Shannon *et al.* 1986) was considered to have probably high risk of bias based on lack of information regarding blinding of outcome assessors. One study was considered definitely high risk of bias because outcome was assessed based on a parent-completed questionnaire, and the study authors noted that the parents were informed of the study's intent and were requested to provide information on fluoride history. Other studies used outcome assessment methods that were not validated or utilized group-level measurements (i.e., school performance).

Confidence Assessment of Findings on Other Neurodevelopmental Effects in Children

The high quality studies (i.e., studies with low potential for bias) provide evidence of an association between fluoride exposure and other cognitive neurodevelopmental effects, including lower neurobehavioral scores in infants, cognitive effects other than IQ in children, and increased attention-related disorders including ADHD in children; however, there is low confidence in this body of evidence due to limitations in the data set, including the heterogeneity in the outcomes assessed, a limited number of directly comparable studies, and differences in outcome assessment methods even when studies evaluated similar outcomes. Although there are limitations in the body of evidence, the low risk-of-bias studies demonstrate a relationship between higher fluoride exposure and neurodevelopmental effects, even in very young children, which supports the consistency in evidence shown in children's IQ studies of an association between fluoride exposure and adverse effects on cognitive neurodevelopment.

Cognitive Effects in Adults

Low Risk-of-bias Studies

Overview of Studies

Two low risk-of-bias studies evaluated the association between fluoride exposure and cognitive effect in adults (Jacqmin *et al.* 1994, Li *et al.* 2016). These two studies used the same test for cognitive function (i.e., Mini-Mental State or MMS Examination) and used drinking water fluoride levels to assess fluoride exposure. Li *et al.* (2016) also measured urinary fluoride. Both studies were cross-sectional in design. One study was conducted in France (Jacqmin *et al.* 1994) and the other in China (Li *et al.* 2016). Both studies were conducted in older populations (i.e., over 60 or 65 years of age).

Table 8 provides a summary of study characteristics and key findings related fluoride exposure and to cognitive effects in adults for the two low risk-of-bias studies. The purpose of the table is to summarize key findings (independent of whether an effect was found) from each study and is not meant to be a comprehensive summary of all results. For each study, results are summarized for each exposure measure assessed. Results from multiple analyses using the same exposure measure may not all be presented unless conflicting results were reported.

Table 8. Studies on Cognitive Function in Adults[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurological Outcome Summary[c]
Jacqmin *et al.* (1994)	Cross-sectional France (Gironde and Dordogne)/elderly adults [3,490]	Drinking water Range: 0.03–2.03 mg	Adults (ages ≥ 65 years)	Cognitive function: MMS Examination	No significant increase in the prevalence of cognitive impairment with increasing fluoride quartiles No statistical adjustment for confounders
Li *et al.* (2016)	Cross-sectional China (Inner Mongolia)/adults [511]	Drinking water daily fluoride intake Mean (SD): 2.23 (2.23) (normal group), 3.62 (6.71) (cognitive impairment group) mg Urine Mean (SD): 1.46 (1.04) (normal group), 2.47 (2.88) (cognitive impairment group) mg/L Fluorosis score Mean (SD): 0.74 (0.98) (normal group), 1.29 (1.01) (cognitive impairment group)	Adults (ages ≥ 60 years)	Cognitive function: MMS Examination	Subjects with cognitive impairment had a significantly higher skeletal fluorosis score and urinary fluoride concentrations; odds of increasing severity of cognitive impairment increased with urinary fluoride concentrations, but were not statistically significant; no significant association with total daily water fluoride intake Adjusted for sex, age, education, marital status (married vs. not married), alcohol consumption (non-drinkers, light drinkers, moderate to heavy drinkers), smoking history (never smoker, ex-smoker, light smoker, heavy smoker), and serum homocysteine levels

[a]Includes low risk-of-bias studies.

[b]Definitions: **GM**: geometric mean; **MMS**: Mini-Mental State.

[c]Associations between cognitive effects in adults and fluoride levels were reported quantitatively, as possible. For studies with multiple analyses and results, the table summarizes key findings and is not a comprehensive summary of all findings. Results also indicate when a study found no association, provided as a qualitative statement of no association.

Summary of Results

Results from two low risk-of-bias studies in adults were not consistent when assessing evidence for a potential association between fluoride exposure and cognitive impairment (based on the MMS Examination) (Jacqmin *et al.* 1994, Li *et al.* 2016). Jacqmin *et al.* (1994) did not find an association between drinking water fluoride and cognitive impairment in populations in France (n = 3,490) and found prevalence rates of cognitive impairment to be the same regardless of fluoride exposure (see **Figure D12**). In contrast, Li *et al.* (2016) did find significantly higher urinary fluoride levels and skeletal fluorosis scores in the cognitively-impaired group compared with the control group in an analysis of 38 cognitively-impaired cases and 38 controls matched for several confounders including age, gender, education, alcohol consumption, and smoking (p-value < 0.05). However, the authors found no significant correlation between cognitive impairment and total daily water fluoride intake (adjusted ORs = 0.94 [95% CI: 0.85, 1.04] and 0.86 [95% CI: 0.69, 1.06] in the moderate and severe cognitive impairment groups, respectively) or urinary fluoride levels (adjusted ORs = 1.12 [95% CI: 0.89, 1.42] and 1.25 [95% CI: 0.87, 1.81] in the moderate and severe cognitive impairment groups, respectively) in subjects from fluorosis-endemic areas of China (n = 511).

High Risk-of-bias Studies

The results from five out of seven high risk-of-bias studies provide evidence of cognitive impairment in adults associated with exposure to fluoride; however, there was heterogeneity in the outcomes assessed, a limited number of directly comparable studies, and some variability in results (e.g., variation in IQ results across studies). Due to the limited number of low risk-of-bias studies identified that assess cognitive impairment in adults, the results from the high risk-of-bias studies are summarized in greater detail below than has been done in this document for the IQ in children and other neurodevelopmental and cognitive effects in children bodies of evidence.

In aluminum factory workers (exposed to gaseous and particular fluoride emissions during the production of aluminum metal), significant decreases in IQ (Duan *et al.* 1995), diminished performance on several neurobehavioral core battery tests (NCTBs) (Guo *et al.* 2001 [translated in Guo *et al.* 2008b]), and impaired psychomotor performance and memory were observed (Yazdi *et al.* 2011). One study conducted on adult subjects with fluorosis (dental and skeletal) from a fluorosis-endemic area compared with healthy subjects from a non-endemic area observed significant differences for some cognitive function tests (i.e., tests of speech fluency, recognition, and working memory) but not others and generally did not observe a significant change in IQ except in the operation scores (Shao 2003). One prospective cohort study evaluated exposure to fluoride in children at age of 5 years, based on whether or not the children resided in areas with community water fluoridation or used fluoride toothpaste or fluoride tablets, and found no clear differences in IQ scores of the subjects at age 38 years (Broadbent *et al.* 2015). One additional study suggested that populations living in areas with higher drinking water fluoride had lower levels of dementia (Still and Kelley 1980); however, the study was not focused on effects of fluoride, but rather if fluoride was able to reduce the risk associated with aluminum by competing with aluminum and reducing the aluminum bioavailability. Therefore, the study was considered inadequate to evaluate the effects of fluoride on dementia (Still and Kelley 1980). A more recent study in Scotland evaluated dementia rates associated with aluminum and fluoride drinking water concentrations and observed a significant increased risk of dementia per standard deviation increase in fluoride (p-value < 0.001) with the risk of dementia more than doubled in the highest quartile of fluoride exposure (56.3 µg/L) compared to the lowest quartile (<44.4 µg/L). The authors also found a significantly increased risk of dementia associated with increased aluminum levels at all quartiles compared with the reference group (p-values < 0.05) but found no statistical interaction between aluminum and fluoride levels in relation with dementia (Russ *et al.* 2019). In addition to studies that

reported on cognitive impairment and exposure to fluoride, two high risk-of-bias studies were identified that reported impaired motor and sensory function (Rotton *et al.* 1982) and a higher prevalence of self-reported headaches, insomnia, and lethargy (Sharma *et al.* 2009) associated with fluoride exposure.

Risk of Bias for Cognitive Effect Studies in Adults

Due to the small number of studies with a low potential for bias and a lack of risk-of-bias issues (see **Figure A3-13** and **Figure A3-14**), the key risk-of-bias domains (confounding, exposure characterization, outcome assessment) are not discussed separately in respective subsections, as was done for the IQ in Children and Other Neurodevelopmental and Cognitive Effects in Children bodies of evidence. The high risk-of-bias studies had concerns across several domains (see **Figure A3-15** and **Figure A3-16**), but there were still relatively few studies. Therefore, the discussion for high risk-of-bias studies is also not separated into subsections by key domain.

Low Risk-of-bias Studies

Both low risk-of-bias studies on cognitive effects in adults had little or no risk-of-bias concerns. One study was rated definitely low or probably low risk of bias for all risk-of-bias questions (Li *et al.* 2016), and the other study was rated probably high risk of bias for a single question that was judged to have minimal impact on overall potential bias (Jacqmin *et al.* 1994). Jacqmin *et al.* (1994) had potential concern for bias due to confounding because smoking was not addressed as a potential confounder, which has potential to impact risk for Alzheimer's disease and rates could vary by parish (the target population consisted of men and women from 75 civil parishes in Southwestern France).

High Risk-of-bias Studies

There were several issues in the seven studies in adults considered to have high potential for bias. Four of the seven studies had potential concern for bias due to lack of information on the comparison groups, or the comparison groups were considered not to be appropriate. All seven studies had potential concern for bias regarding potential confounders not being addressed including possible co-exposures in occupational studies (e.g., aluminum) and smoking. Five of the seven studies had potential concern for bias due to lack of information regarding exposure characterization or poor exposure characterization with the most utilized exposure measure in these studies being a comparison between exposed and unexposed areas. In one case (Broadbent *et al.* 2015), multiple sources of fluoride exposure were assessed separately without properly controlling for the other sources of exposure, which could bias the results (see **Exposure Characterization in IQ Studies** for further details). Four studies also had potential for bias based on limitations in the outcome assessment, which was mainly due to lack of blinding of outcome assessors or lack of sufficient details on how the outcomes were assessed.

Confidence Assessment of Findings on Cognitive Effects in Adults

The body of evidence available to examine the association between exposure to fluoride and cognitive effects in adults is limited to two low risk-of-bias cross-sectional studies. The human body of evidence in adults is considered inadequate to evaluate whether fluoride exposure is associated with cognitive effects due to low confidence in the human data in adults, a limited number of studies, and a lack of evidence of an effect.

Mechanistic Data in Humans

Eight low risk-of-bias studies that evaluated fluoride exposure and mechanistic data in humans were considered potentially relevant to neurological effects. Effects on the thyroid were specifically evaluated because the NRC 2006 report identified this as a possible effect of fluoride (NRC 2006), and changes in thyroid hormones have been identified as a mechanism for neurodevelopmental effects (Haschek and Rousseaux 1991). These included effects on thyroid hormones in children (Kheradpisheh *et al.* 2018b,

Kheradpisheh *et al.* 2018a, Malin *et al.* 2018), adults (Kheradpisheh *et al.* 2018b, Kheradpisheh *et al.* 2018a, Malin *et al.* 2018), or children and adults combined (Barberio *et al.* 2017a). In addition, some studies evaluated self-reported thyroid conditions in children and adults combined (Barberio *et al.* 2017a) and thyroid diseases in adults (Kheradpisheh *et al.* 2018b, Peckham *et al.* 2015) (see **Figure A3-17** and **Figure A3-18**). Although the low risk-of-bias studies provide some evidence of mechanistic effects (primarily changes in thyroid stimulating hormone [TSH] levels in children), the studies were too heterogeneous or limited in number to make any determination on mechanism (see **Figure 7**).

Among the seven low risk-of-bias studies that reported on changes in thyroid hormones, three studies were conducted in children (Zhang *et al.* 2015b, Singh *et al.* 2014, Kumar *et al.* 2018) and reported increases in TSH levels. Zhang *et al.* (2015b) reported significant increases in TSH in children from a fluorosis-endemic area (median fluoride drinking water concentration = 1.40 mg/L; interquartile range = 1.23–1.57 mg/L) compared with a non-fluorosis-endemic area (median fluoride drinking water concentration = 0.63 mg/L; interquartile range = 0.58–0.68 mg/L), while 3,5,3'-triiodothyronine (T_3) or thyroxine (T_4) were not significantly different between the two groups. Similarly, Singh *et al.* (2014) observed significantly higher TSH levels in children without dental fluorosis who lived in a fluorosis-endemic area (fluoride drinking water concentrations of 1.6–5.5 mg/L) compared with children without dental fluorosis who lived in a non-fluorosis-endemic area (fluoride drinking water concentrations of 0.98–1.00 mg/L). When all children (with and without dental fluorosis) in the endemic area were compared to children from the non-endemic area, the TSH levels were higher in children from the fluorosis-endemic area although results did not reach statistical significance ($p = 0.057$). Significant differences in T_4 or T_3 were not observed between groups (Singh *et al.* 2014). Kumar *et al.* (2018) also observed a significant increase in TSH levels in children from a fluorosis endemic area (1.5–5.8 mg/L fluoride) compared with a control area (0.94–1.08 mg/L fluoride). There were also decreases in T_3 and T_4, but results were not statistically significant.

Barberio *et al.* (2017a) evaluated fluoride effects on TSH levels in children and adults combined and found no relationship between fluoride exposure (measures in urine and tap water) and TSH levels. In the one study that evaluated thyroid hormone levels in adults but not children, Kheradpisheh *et al.* (2018b) found a significant increase in TSH associated with higher fluoride concentrations in drinking water in both adults with and without thyroid diseases such as hypothyroidism, hyperthyroidism, thyroid nodules, or thyroid cancer. Significant increases in T_3 were associated with higher fluoride in drinking water in adults without thyroid diseases, but increases in T_3 were not significant in adults with thyroid diseases. A significant association between T_4 and higher fluoride in drinking water was not observed in adults with or without thyroid diseases (Kheradpisheh *et al.* 2018b).

Other than changes in hormone levels, there is limited evidence of fluoride-related mechanistic effects in the three low risk-of-bias studies that evaluated thyroid-related effects. Barberio *et al.* (2017a) found no relationship between fluoride exposure and self-reported thyroid conditions in children and adults (children were older than 12). Kheradpisheh *et al.* (2018b) also found no association between fluoride exposure and hypothyroidism in an adult population in Iran. One study found a significantly higher prevalence of hypothyroidism in areas with higher fluoride concentrations in drinking water (>0.7 mg/L) compared with areas with lower fluoride drinking water concentrations (≤0.7 mg/L) (Peckham *et al.* 2015).

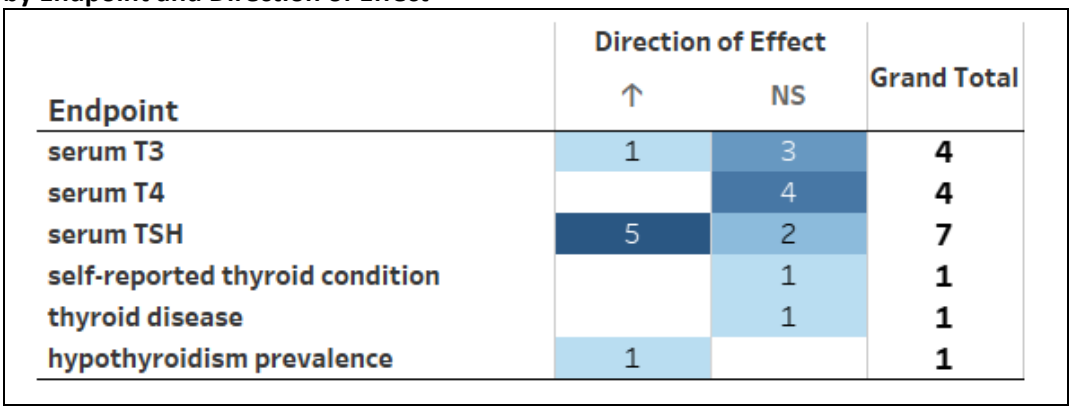
Sixteen high risk-of-bias studies were available that evaluated mechanistic data in humans associated with fluoride exposure, including effects on thyroid hormones in children (n = 9 studies); thyroid hormones in adults (Michael 1996, Yasmin 2013); catecholamines in adults (Michael *et al.* 1996) or in

subjects of unknown ages (Chinoy and Narayana 1992); acetylcholinesterase (AChE) or serotonin levels in children (Singh *et al.* 2013, Lu *et al.* 2019); brain histopathology or biochemistry in aborted fetuses (Du *et al.* 1992 [translated in Du *et al.* 2008], Yu *et al.* 1996 [translated in Yu *et al.* 2008]); and mitochondrial fission/fusion molecules in children (Zhao *et al.* 2019). Similar to the low risk-of-bias studies, the high risk-of-bias studies provide some evidence of mechanistic effects (primarily changes in TSH levels in children); however, the data are insufficient to identify a clear mechanism by which fluoride causes neurodevelopmental or cognitive effects in humans.

Among high risk-of-bias studies (see **Figure A3-19** and **Figure A3-20**), varying results were reported in 11 studies that evaluated fluoride exposure and effects on thyroid hormones, and a few of these studies (Lin *et al.* 1991, Yang *et al.* 1994 [translated in Yang *et al.* 2008], Wang *et al.* 2001) were complicated by high or low iodine in the high fluoride area. When considering fluoride effects on each of the hormones individually, similar to results from low risk-of-bias studies, the most consistent evidence of fluoride-associated effects on a thyroid hormone was reported as changes in TSH levels in children, although there was some variation in the direction of effect. Six of the nine high risk-of-bias studies that evaluated changes in TSH levels in children reported increases in TSH levels with higher fluoride (Lin *et al.* 1991, Susheela *et al.* 2005, Wang *et al.* 2001, Yang *et al.* 1994 [translated in Yang *et al.* 2008], Yao *et al.* 1996, Yasmin *et al.* 2013). Two of the nine high risk-of-bias studies reported decreases in TSH levels in children with higher fluoride (Khandare *et al.* 2017, Khandare *et al.* 2018). One of the nine studies found no significant alterations in TSH levels in children from fluorosis-endemic areas (Hosur *et al.* 2012) (see **Figure 8**).

When considering fluoride-associated effects on TSH, T_3, and T_4 levels together, studies that evaluated changes in all three thyroid hormones reported varying combinations of increases, decreases, or no changes in levels across the three hormones, although among the eight low and high risk-of-bias studies that evaluated the effects of fluoride exposure on TSH, T_3, and T_4 levels and reported increases in TSH levels in children, seven of the eight studies found no alterations in T_3 levels (one study found an increase in T_3), and six of the eight studies found no alterations in T_4 levels (two studies found an increase in T_4). Studies also displayed variation by age in fluoride-associated effects on TSH, T_3, and T_4. Due to the dynamic relationship between the thyroid gland, the pituitary gland, and the production and clearance of TSH, T_3, and T_4, the variations in results are not unexpected and do not eliminate the possibility of a mechanistic link between thyroid effects and neurodevelopmental or cognitive effects; however, the data do not support a clear indication that thyroid effects are a mechanism by which fluoride causes these effects in humans.

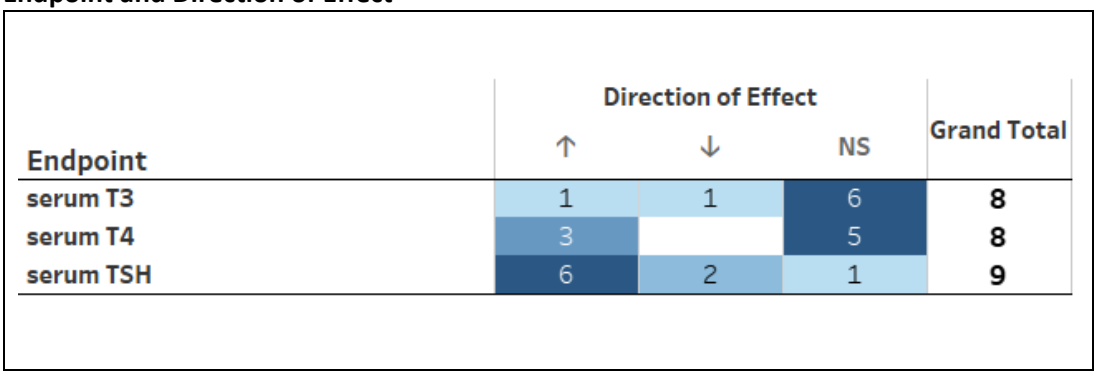
Figure 7. Number of Low Risk-of-bias Studies that Evaluated Thyroid Hormones in Children and Adults by Endpoint and Direction of Effect*



Endpoint	Direction of Effect ↑	Direction of Effect NS	Grand Total
serum T3	1	3	4
serum T4		4	4
serum TSH	5	2	7
self-reported thyroid condition		1	1
thyroid disease		1	1
hypothyroidism prevalence	1		1

*Interactive figure and additional study details in Tableau® (https://public.tableau.com/profile/ntp.visuals#!/vizhome/Fluoride_EpiThyroid_UPDATE/Figures6and7). This figure displays study counts for low risk-of-bias studies in both children and adults, as these counts are most relevant to the summary of fluoride-related mechanistic effects in low risk-of-bias studies. Counts for high risk-of bias studies and studies by age (i.e., children, adults, or children/adults combined) can also be accessed in the interactive figure in Tableau®. Study counts are tabulated by significance (unless if study footnotes in Tableau indicate that statistical significance was not tested)—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with significantly increased results.

Figure 8. Number of High Risk-of-bias Studies that Evaluated Thyroid Hormones in Children by Endpoint and Direction of Effect*



Endpoint	Direction of Effect ↑	Direction of Effect ↓	Direction of Effect NS	Grand Total
serum T3	1	1	6	8
serum T4	3		5	8
serum TSH	6	2	1	9

*Interactive figure and additional study details in Tableau® (https://public.tableau.com/profile/ntp.visuals#!/vizhome/Fluoride_EpiThyroid_UPDATE/Figures6and7). This figure displays study counts for high risk-of-bias studies in children, as these counts are most relevant to the summary of fluoride-related effects on thyroid hormones in high risk-of-bias studies. Counts for low risk-of bias studies, studies in adults, or all studies combined, can also be accessed in the interactive figure in Tableau®. Study counts are tabulated by significance (unless if study footnotes in Tableau indicate that statistical significance was not tested)—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with significantly increased results.

In addition to evaluating thyroid hormone levels, a few high risk-of-bias studies evaluated other mechanistic data associated with fluoride exposure; however, the data are insufficient to identify a clear mechanism by which fluoride might cause neurodevelopmental or cognitive effects in humans. Serum epinephrine and norepinephrine were significantly increased in a fluoride-endemic region (not reported whether subjects were children or adults) compared to a non-endemic region (Chinoy and Narayana 1992). Serum adrenaline and noradrenaline were significantly increased in adults in a fluoride-endemic

area (fluoride in the drinking water ranged from 1.0–6.53 ppm) compared to a control area (fluoride in the drinking water ranged from 0.56–0.72 ppm) (Michael *et al.* 1996). Serum AChE was significantly reduced in children from a high fluoride region compared to a lower fluoride region (Singh *et al.* 2013). Serum serotonin was significantly increased in children from Turkey who were drinking water containing 2.5 mg/L of fluoride compared to children drinking bottled water or water containing <0.5 mg/L of fluoride (Lu *et al.* 2019). Aborted fetuses from high fluoride areas in China were found to have histological changes in the brain and significant changes in neurotransmitter levels compared to a control area (Du *et al.* 1992 [translated in Du *et al.* 2008], Yu *et al.* 1996 [translated in Yu *et al.* 2008]).

There are also two more recent low risk-of-bias studies that evaluated polymorphisms in dopamine-related genes; however, a determination on mechanism cannot be made at this time due to the limited number of studies. For children (10–12 years old) with a Val158Met polymorphism in the *COMT* gene (i.e., catechol-O-methyltransferase), which results in slower degradation and greater availability of dopamine within the brain, a stronger association between increasing urinary fluoride levels and decreasing IQ was reported (Zhang *et al.* 2015b). For children (7–12 years old) with a dopamine receptor-2 (DRD2) Taq 1A polymorphism (which is involved in reduced D2 receptor density and availability) and the TT (variant) genotype, a significant inverse relationship between log urine fluoride and IQ was observed; however, this significant relationship was not observed in children with the CC (wild-type) or CT (hybrid) genotypes (Cui *et al.* 2018).

Animal Learning and Memory Data

The NTP provided a review of the experimental animal evidence in the earlier draft monographs (NTP 2020) and agrees with the NASEM review committee comments (NASEM 2020, 2021)(placeholder to cite NTP 2021 Response to NASEM comments) that the experimental animal database is of poor quality, with many studies suffering from major reporting deficiencies. The NTP acknowledges that further efforts to disentangle the potential for motor activity deficits to influence tests of learning and memory in the fluoride literature are warranted. Overall, these general issues and deficiencies with the experimental animal database led to NTP's conclusion that the animal studies are currently *inadequate* to inform the question of an association between fluoride exposures and neurodevelopmental and cognitive effects in humans. Therefore, this systematic review does not include an experimental animal section.

Mechanistic Data in Animals

There are a wide variety of studies in animals that evaluate mechanistic effects potentially related to neurological changes following oral fluoride exposure (see **Appendix 5**); however, the mechanisms underlying fluoride-associated cognitive neurodevelopmental effects are not well characterized and review of the data did not identify a mode of action for fluoride effects on IQ in children. Categories of mechanistic endpoints with the largest amount of available data include changes in biochemical components of the brain or neurons, neurotransmitters, oxidative stress, histopathology, and thyroid function. Limiting the data to studies with at least one exposure at or below 20 ppm fluoride drinking water equivalents (gavage and dietary exposures were back calculated into equivalent drinking water concentrations for comparison) still provided a sufficient number of studies for evaluation of these mechanistic endpoints. This evaluation is provided in **Appendix 5**. Neurotransmitter and biochemical changes in the brain and neurons were considered to be the mechanistic areas with the greatest potential to demonstrate effects of fluoride on the brain of animals in the lower dose range and provide evidence of changes in the brain that may relate to lower IQ in children (see **Appendix 5**). Histological data can be useful in determining whether effects are occurring in the brain at lower fluoride concentrations; however, author descriptions of these effects may be limited thereby making it difficult

to directly link histological changes in the brain to learning and memory effects. Oxidative stress is considered a general mechanistic endpoint that cannot be specifically linked to neurodevelopmental or cognitive effects in humans; however, like histopathology, it may help in identifying changes in the brain occurring at lower concentrations of fluoride. Although any effects in the brain or neurological tissue at lower concentrations of fluoride may support reduced IQ in humans, it may be difficult to distinguish the potential effects of fluoride on learning and memory functions from other neurological or general health outcomes.

In Vitro Data on Neurodevelopmental or Cognitive Effects

Although in vitro data were collected as part of the systematic review process, NTP determined that the information on neurological effects obtained from these studies is too general, and results cannot necessarily be attributed to effects on learning and memory or other cognitive functions at this time. The in vitro data may help support specific mechanisms identified from in vivo mechanistic data; however, as described above, no specific mechanism has been determined for fluoride effects on learning and memory or other neurodevelopmental or cognitive outcomes.

DISCUSSION

This systematic review evaluated the available animal and human literature concerning the association between fluoride exposure and cognitive neurodevelopment. The available data on potential mechanisms to evaluate biological plausibility were also assessed. This review addresses whether exposure to fluoride could present a potential hazard (i.e., has the potential to cause harm at any exposure level). Benefits of fluoride with respect to oral health are not addressed in this review.

This review extended the NTP's previous evaluation of the experimental animal data (NTP 2016). Although the animal data provide some evidence of effects of fluoride on neurodevelopment, they give little insight into the question of whether fluoride influences IQ. This is due to the deficiencies identified in the animal body of evidence. Mechanistic studies in humans provide some evidence of adverse neurological effects of fluoride. However, these studies were too heterogenous and limited in number to make any determination on biological plausibility.

The literature on adults is also limited; therefore, it was determined that there is low confidence in the body of evidence from studies that evaluate fluoride exposure and adult cognition. Compared to the literature in adults, there is a much more extensive literature in children.

The literature in children was separated into studies assessing IQ and studies assessing other cognitive neurodevelopmental outcomes. There is low confidence in the body of evidence from studies that evaluate fluoride exposure and other cognitive neurodevelopmental outcomes in children. Altogether, the results from eight of nine high quality studies (three prospective cohort and five cross-sectional studies from seven different study populations) provide some evidence that fluoride is associated with other cognitive neurodevelopmental outcomes in children. The data also suggest that neurodevelopmental effects occur in very young children. However, the number of studies is limited and there is too much heterogeneity in the outcomes measured and methods used to directly compare studies of any one outcome. Additional studies on outcomes such as attention deficit hyperactivity disorder (ADHD) and other attention-related disorders, where there is some evidence of an effect of fluoride exposure, would be necessary to critically assess the data.

Most of the epidemiological studies (n = 66) assessed the association between fluoride exposure and IQ in children. Although all studies, both high and low quality, were considered, this evaluation focuses on the high quality, low risk-of-bias studies in children for two reasons. First, there are fewer limitations and greater confidence in the results of the high quality studies. Second, there is a relatively large number of high quality studies (n = 19), such that the body of evidence from these studies could be used to evaluate confidence in the association between fluoride exposure and changes in children's IQ.

This review finds, with moderate confidence, that fluoride exposure is consistently associated with lower IQ in children. The inverse association between fluoride exposure and children's IQ was consistent across different study populations, study locations, study quality/risk of bias determinations, study designs, exposure measures, and types of exposure data (group-level and individual-level). There were 19 low risk-of-bias studies that were conducted in 15 study populations, across 5 countries, and evaluating more than 7,000 children. Of these 19 studies, 18 reported that higher fluoride exposure is associated with lower IQ. These include 3 prospective cohort studies and 15 cross-sectional studies (12 of which indicated that exposure likely preceded the outcome). Forty-one of 47 low quality studies in children also found evidence of an inverse relationship between fluoride exposure and IQ.

Many studies in this assessment relied on drinking water fluoride levels (both group-level measures and individual-level measures), rather than measures of total fluoride exposures, to establish exposed versus "unexposed" or reference groups. Although fluoride in water is a major source of exposure [comprising 40 to 70% of total exposure (US EPA 2010)], other sources provide variable amounts that depend on personal preferences and habits. The use of dental products containing fluoride and consuming foods and beverages prepared with fluoridated water can also result in measurable exposures (US EPA 2010). Green *et al.* (2019) suggested that significant exposures occur from black tea consumption. Thus, drinking water fluoride levels may, but usually do not, reflect total fluoride exposures. This could be a potential limitation in studies that rely on water fluoride data to assess fluoride exposure (in particular, earlier studies). However, because water is only part of a person's total exposure to fluoride, this limitation would likely result in an underestimate of exposure to fluoride. Also, this limitation is less of a concern in areas where fluoride in the drinking water is high because drinking water likely contributes a large proportion of the total fluoride intake in those areas as compared to areas where fluoride in the drinking water is lower.

This review found that the quality of exposure assessment has improved over the years. More recent studies by Valdez Jimenez *et al.* (2017), Bashash *et al.* (2017), and Green *et al.* (2019) used individual measures of urinary fluoride, either maternal urine collected prenatally or children's urine, which confirmed the inverse association between total fluoride exposure and children's IQ and other cognitive neurodevelopmental effects. Studies using different types of exposure measures reported similar findings of an association, which strengthens confidence in earlier studies that reported IQ deficits with increasing group-level fluoride exposures. However, there is less certainty in the quantitative estimates of the magnitude of IQ deficits from earlier studies that used group-level exposure measures than the estimates from more recent studies that used individual-level exposure measures.

It is worth noting that there are circumstances where typical children's water consumption considered with water fluoride levels may substantially underestimate total fluoride exposures. One example is bottle fed infants where nutrition is provided by powdered formula that is rehydrated with fluoridated water (Till *et al.* 2020). To decrease an exclusively formula fed infant's exposure to fluoride, for the purpose of reducing risk of dental fluorosis, the Centers for Disease Control and Prevention (CDC) recommends using low fluoride bottled water to mix with infant formula (CDC 2015). A few studies also support the possibility of heightened sensitivities to the detrimental cognitive effects of fluoride exposures in individuals with certain genetic polymorphisms in dopamine receptor D2, or catechol-O-methyltransferase (Cui *et al.* 2018, Zhang *et al.* 2015b), potentially impacting dopamine catabolism and receptor sensitivity. Differential exposures to fluoride and genetic susceptibilities of children to fluoride may represent special situations that would appear to warrant further research.

The following section briefly recaps the strength of the epidemiological evidence for an inverse association between fluoride exposures and cognitive neurodevelopmental deficits. This is followed by a more detailed listing of limitations of the evidence base and limitations of the systematic review, with some suggestions of areas where further research may be most beneficial.

Strengths of the Evidence Base

Strengths in the epidemiological evidence base include:

- Sixty-six studies directly addressing the relationship between high fluoride exposure and deficits in children's IQ or other measures of neurological function.

- Twelve high quality cross-sectional studies with low risk of bias providing evidence that exposure occurred prior to outcome assessment.
- Studies from diverse geographic locations that included data for more than 7,000 children.
- Nineteen high quality studies evaluating the same outcome (i.e., IQ) and nine evaluating other neurodevelopmental outcomes.
- Consistency in the reported responses to fluoride exposure in studies of both low and high quality.
- Consistency in the reported responses to fluoride exposure across different study populations, study designs, and exposure measures.
- Similar findings of studies with group- and individual-level information on exposure and outcomes.
- Wide variety of confounders either addressed by study design or captured across the evidence base, with no consistent patterns that would suggest an alternative explanation.

Limitations of the Evidence Base

Limitations in the epidemiological studies with low risk of bias include:

- Few studies available that assessed the association between fluoride exposure and cognitive function (particularly IQ) in adults and attention-related disorders including ADHD in children and adults.
- Heterogeneity in outcomes assessed for other neurobehavioral outcomes, limiting the assessment of other possible effects in children.
- Studies rarely separated the results by gender or provided information to indicate that gender was not a modifying factor.
- Effects of lower total fluoride exposures on children's IQ remain unclear. More studies at lower exposure levels are needed to fully understand potential effects in ranges typically found in the United States (i.e., <1.5 mg/L in water).
- No studies investigating the association between fluoride exposure and neurodevelopmental or cognitive effects in adults or children have been conducted in the United States.

Limitations in the epidemiological studies with high risk of bias include:

- Many of the original publications were in a non-English language and provided limited details on methodology.
- Studies lacked information regarding exposure and/or had serious limitations in the exposure assessment. Exposure assessment concerns include limited individual exposure information, a lack of information on fluoride sampling methods and timing of the exposure measurements, a lack of quantitation of levels of fluoride in drinking water in a few studies, and a lack of individual-level information on fluorosis in areas reported to be endemic for fluorosis.
- The comparison groups in studies conducted in areas endemic for fluorosis still may have been exposed to high levels of fluoride or levels similar to those used in water fluoridation in the United States. This factor may have limited the ability to detect true effects.

- Studies did not provide sufficient direct information (e.g., participation rates or methods for selection) to evaluate selection bias.
- Failure to address potential confounders was an issue for many studies. Some studies conducted simple statistical analyses without accounting for any potential confounders in the analysis, although many noted similarities between the study populations. In cases where adjustments in analyses were made, often these studies did not account for potential confounders considered critical for that study population and outcome.
- Studies conducted in areas with high, naturally-occurring fluoride levels in drinking water often did not account for potential exposures to arsenic or iodine deficiencies in study subjects in areas where these were likely to occur.
- Studies lacked information on whether the outcome assessors were blind to the exposure group, including studies that examined children in their schools and subjects from high-fluoride communities.

Limitations in the animal and mechanistic evidence base include:

- The overall poor quality of experimental animal studies and the relatively few well-designed and well-performed studies at lower exposure levels (i.e., <20 ppm, which is roughly equivalent to human exposure of <4 ppm).
- Poor understanding of the specific molecular events responsible for fluoride's adverse effects on neurological function.

A key data gap in the human and animal bodies of evidence includes the need for mechanistic insight into fluoride-related neurodevelopmental or cognitive changes.

Limitations of the Systematic Review

There are few limitations of this systematic review. The human body of evidence included a large database of observational studies. Most of the observational studies were cross-sectional; however, 12 of these were considered to provide sufficient evidence that exposure occurred prior to the outcome. In addition, the systematic review covered a wide range of study designs, populations, and measures of fluoride exposure. The systematic review was designed to cover reports on all potential mechanistic data including effects on the thyroid. After review of the studies evaluating thyroid effects, studies that only evaluated goiters and other effects on thyroid size were not considered in this review. This is not considered a limitation because changes in thyroid size are not functional changes to the thyroid that could specifically indicate a mechanism for thyroid involvement in neurodevelopment. In addition, review of the mechanistic data was limited to in vivo studies with at least one concentration below 20 ppm. This is not considered a limitation for the systematic review because the mechanistic body of evidence was used to evaluate biological plausibility for the effects observed in humans; therefore, data were limited to concentrations that would be more reflective of human exposures. The decision to not more closely evaluate the in vitro data is not considered a limitation because there were sufficient in vivo data, and no key events were identified where in vitro data would provide additional insight.

Following the recommendation of the NASEM committee in its review of the September 16, 2020, draft monograph, the experimental animal section has been removed and is not included in this monograph. Although the deficiencies identified in the animal body of evidence support this removal (see **Animal Learning and Memory Data** for further explanation), the NTP acknowledges that the absence of the

experimental animal data is a limitation of this systematic review. For the purpose of this review, the NTP considers the experimental animal data to be *inadequate* to inform whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans.

REFERENCES

Adedara IA, Abolaji AO, Idris UF, Olabiyi BF, Onibiyo EM, Ojuade TD, Farombi EO. 2017a. Neuroprotective influence of taurine on fluoride-induced biochemical and behavioral deficits in rats. *Chem Biol Interact* 261: 1-10.

Adedara IA, Olabiyi BF, Ojuade TD, Idris UF, Onibiyo EM, Farombi EO. 2017b. Taurine reverses sodium fluoride-mediated increase in inflammation, caspase-3 activity, and oxidative damage along the brain-pituitary-gonadal axis in male rats. *Can J Phys Pharmacol* 95: 1019-1029.

Akinrinade ID, Memudu AE, Ogundele OM. 2015a. Fluoride and aluminium disturb neuronal morphology, transport functions, cholinesterase, lysosomal and cell cycle activities. *Pathophysiology* 22: 105-115.

Akinrinade ID, Memudu AE, Ogundele OM, Ajetunmobi OI. 2015b. Interplay of glia activation and oxidative stress formation in fluoride and aluminium exposure. *Pathophysiology* 22: 39-48.

An J, Mei S, Liu A, Fu Y, Wang C. 1992. [Effect of high level of fluoride on children's intelligence]. *Chin J Control Endem Dis* 7(2): 93-94.

Arbuckle TE, Fraser WD, Fisher M, Davis K, Liang CL, Lupien N, Bastien S, Velez MP, von Dadelszen P, Hemmings DG, Wang J, Helewa M, Taback S, Sermer M, Foster W, Ross G, Fredette P, Smith G, Walker M, Shear R, Dodds L, Ettinger AS, Weber JP, D'Amour M, Legrand M, Kumarathasan P, Vincent R, Luo ZC, Platt RW, Mitchell G, Hidiroglou N, Cockell K, Villeneuve M, Rawn DF, Dabeka R, Cao XL, Becalski A, Ratnayake N, Bondy G, Jin X, Wang Z, Tittlemier S, Julien P, Avard D, Weiler H, Leblanc A, Muckle G, Boivin M, Dionne G, Ayotte P, Lanphear B, Séguin JR, Saint-Amour D, Dewailly E, Monnier P, Koren G, Ouellet E. 2013. Cohort profile: The maternal-infant research on environmental chemicals research platform. *Paediatr Perinat Epidemiol* 27(4): 415-425.

Baba NA, Raina R, Verma PK, Sultana M. 2014. Alterations in plasma and tissue acetylcholinesterase activity following repeated oral exposure of chlorpyrifos alone and in conjunction with fluoride in Wistar rats. *Proc Indian Natl Sci Acad B Biol Sci* 84: 969-972.

Banala R, Nagapuri K, Mohd K, Reddy M, Karnati P. 2018. Carica Papaya leaf extract as a neuroprotective agent against behavioral and neurotransmitter changes in brain of the rat treated with sodium fluoride in pre- and post-natal periods. *Pharmacogn Mag* 14(55): 123-131.

Barberio AM, Hosein FS, Quinonez C, McLaren L. 2017a. Fluoride exposure and indicators of thyroid functioning in the Canadian population: Implications for community water fluoridation. *J Epidemiol Community Health* 71: 1019-1025.

Barberio AM, Quinonez C, Hosein FS, McLaren L. 2017b. Fluoride exposure and reported learning disability diagnosis among Canadian children: Implications for community water fluoridation. *Can J Public Health* 108: 229-239.

Bartos M, Gumilar F, Gallegos CE, Bras C, Dominguez S, Monaco N, Esandi MDC, Bouzat C, Cancela LM, Minetti A. 2018. Alterations in the memory of rat offspring exposed to low levels of fluoride during gestation and lactation: Involvement of the alpha7 nicotinic receptor and oxidative stress. *Reprod Toxicol* 81: 108-114.

Bashash M, Marchand M, Hu H, Till C, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Green R, Schnaas L, Mercado-Garcia A, Hernandez-Avila M, Tellez-Rojo MM. 2018. Prenatal fluoride exposure and attention deficit hyperactivity disorder (ADHD) symptoms in children at 6-12 years of age in Mexico City. *Environ Int* 121(Pt 1): 658-666.

Bashash M, Thomas D, Hu H, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Ettinger AS, Wright R, Zhang Z, Liu Y, Schnaas L, Mercado-Garcia A, Tellez-Rojo MM, Hernandez-Avila M. 2017. Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6-12 years of age in Mexico. *Environ Health Perspect* 125(9): 1-12.

Bhatnagar M, Rao P, Sushma J, Bhatnagar R. 2002. Neurotoxicity of fluoride: Neurodegeneration in hippocampus of female mice. *Indian J Exp Biol* 40: 546-554.

Bhatnagar M, Sukhwal P, Suhalka P, Jain A, Joshi C, Sharma D. 2011. Effects of fluoride in drinking water on NADPH-diaphorase neurons in the forebrain of mice: A possible mechanism of fluoride neurotoxicity. *Fluoride* 44: 195-209.

Broadbent JM, Thomson WM, Moffitt TE, Poulton R. 2015. Community water fluoridation and intelligence response. *Am J Public Health* 105: 3-4.

CDC (Centers for Disease Control and Prevention). 2013. *Community water fluoridation: Fluoridation statistics*. Atlanta, GA. Available: https://www.cdc.gov/fluoridation/statistics/2012stats.htm [accessed 19 August 2019].

CDC (Centers for Disease Control and Prevention). 2015. *Community water fluoridation FAQs: Infant formula* Atlanta, GA. Available: https://www.cdc.gov/fluoridation/faqs/infant-formula.html [accessed 22 September 2021].

Chen Y. 2012. Organophosphate-induced brain damage: Mechanisms, neuropsychiatric and neurological consequences, and potential therapeutic strategies. *Neurotox* 33: 391-400.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 1991. [Research on the intellectual development of children in high fluoride areas]. *Chin J Control Endem Dis* 6(Suppl): 99-100.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 2008. Research on the intellectual development of children in high fluoride areas. *Fluoride* 41: 120-124.

Chinoy NJ, Narayana MV. 1992. Studies on fluorosis in Mehsana District of North Gujarat. *Proc Zool Soc* 45: 157-161.

Choi AL, Sun G, Zhang Y, Grandjean P. 2012. Developmental fluoride neurotoxicity: A systematic review and meta-analysis. *Environ Health Perspect* 120: 1362-1368.

Choi AL, Zhang Y, Sun G, Bellinger DC, Wang K, Yang XJ, Li JS, Zheng Q, Fu Y, Grandjean P. 2015. Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study. *Neurotoxicol Teratol* 47: 96-101.

Chouhan S, Flora SJS. 2008. Effects of fluoride on the tissue oxidative stress and apoptosis in rats: Biochemical assays supported by IR spectroscopy data. *Toxicology* 254: 61-67.

Chouhan S, Lomash V, Flora SJ. 2010. Fluoride-induced changes in haem biosynthesis pathway, neurological variables and tissue histopathology of rats. *J Appl Toxicol* 30: 63-73.

Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. 2020. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. *Neurosci Lett* 729: 134981.

Cui Y, Zhang B, Ma J, Wang Y, Zhao L, Hou C, Yu J, Zhao Y, Zhang Z, Nie J, Gao T, Zhou G, Liu H. 2018. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. *Ecotoxicol Environ Saf* 165: 270-277.

Ding Y, Sun H, Han H, Wang W, Ji X, Liu X, Sun D. 2011. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. *J Hazard Mater* 186: 1942-1946.

Du L, Wan C, Cao X, Liu J. 1992. [The effect of fluorine on the developing human brain]. *Chin J Pathol* 21(4): 218-220.

Du L, Wan C, Cao X, Liu J. 2008. The effect of fluorine on the developing human brain. *Fluoride* 41: 327-330.

Duan J, Zhao M, Wang L, Fang D, Wang Y, Wang W. 1995. A comparative analysis of the results of multiple tests in patients with chronic industrial fluorosis. *Guizhou Med J* 18(3): 179-180.

Duan Q, Jiao J, Chen X, Wang X. 2018. Association between water fluoride and the level of children's intelligence: A dose-response meta-analysis. *Public Health* 154: 87-97.

Gais S, Schonauer M. 2017. Untangling a cholinergic pathway from wakefulness to memory. *Neuron* 94(4): 696-698.

Gao Q, Liu YJ, Guan ZZ. 2009. Decreased learning and memory ability in rats with fluorosis: Increased oxidative stress and reduced cholinesterase activity in the brain. *Fluoride* 42: 277-285.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008a. [Effects of fluoride on learning and memory and cholinesterase activity in rat brains]. *Chin J Endemiol* 27: 128-130.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008b. [Level of oxidative stress in rat brains and learning and memory function of rats with chronic fluorosis]. *Chin J Endemiol* 27: 371-373.

Garman RH, Li AA, Kaufmann W, Auer RN, Bolon B. 2016. Recommended methods for brain processing and quantitative analysis in rodent developmental neurotoxicity studies. *Toxicol Pathol* 44(1): 14-42.

Green R, Lanphear B, Hornung R, Flora D, Martinez-Mier EA, Neufeld R, Ayotte P, Muckle G, Till C. 2019. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. *JAMA Pediatr*: E1-E9.

Guner S, Uyar-Bozkurt S, Haznedaroglu E, Mentes A. 2016. Dental fluorosis and catalase immunoreactivity of the brain tissues in rats exposed to high fluoride pre- and postnatally. *Biol Trace Elem Res* 174: 150-157.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 1991. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. *Chin J Epidemiol* 10(2): 98-100.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 2008a. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. *Fluoride* 41: 125-128.

Guo ZY, He YH, Zhu QX. 2001. [Research on the neurobehavioral function of workers occupationally exposed to fluoride]. *Ind Health & Occup Dis* 27(6): 346-348.

Guo ZY, He YH, Zhu QX. 2008b. Research on the neurobehavioral function of workers occupationally exposed to fluoride. *Fluoride* 41: 152-155.

Guyatt GH, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, Debeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schunemann HJ. 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 64(4): 383-394.

Haschek W, Rousseaux C, eds. 1991. *Handbook of toxicologic pathology*. 1st ed.: Academic Press.

Health Canada. 2015. *Third report on human biomonitoring of environmental chemicals in Canada - Results of the Canadian Health Measures Survey Cycle 3 (2012–2013)*. Ottawa, Ontario: Canadian Minstry of Health. Available: https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/ewh-semt/alt_formats/pdf/pubs/contaminants/chms-ecms-cycle3/chms-ecms-cycle3-eng.pdf.

Higgins JP, Green S. 2011. *Cochrane handbook for systematic reviews of interventions,* In: The Cochrane Collaboration. Vol 4, New York, NY: John Wiley & Sons.

Hosur MB, Puranik RS, Vanaki S, Puranik SR. 2012. Study of thyroid hormones free triiodothyronine (FT3), free thyroxine (FT4) and thyroid stimulating hormone (TSH) in subjects with dental fluorosis. *Eur J Dent* 6: 184-190.

Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, Sedykh A, Thayer K, Merrick BA, Walker V, Rooney A, Shah RR. 2020. SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation. *Environ Int* 138: 105623.

Jacqmin H, Commenges D, Letenneur L, Barberger-Gateau P, Dartigues JF. 1994. Components of drinking water and risk of cognitive impairment in the elderly. *Am J Epidemiol* 139: 48-57.

Jia B, Zong L, Lee JY, Lei J, Zhu Y, Xie H, Clemens JL, Feller MC, Na Q, Dong J, McLane MW, Jones-Beatty K, Burd I. 2019. Maternal supplementation of low dose fluoride alleviates adverse perinatal outcomes following exposure to intrauterine inflammation. *Scientific Reports* 9(1): 2575.

Jiang C, Zhang S, Liu H, Guan Z, Zeng Q, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Zhang X, Cui Y, Yu L, Wang A. 2014. Low glucose utilization and neurodegenerative changes caused by sodium fluoride exposure in rat's developmental brain. *Neuromolecular Med* 16: 94-105.

Jones S, Burt BA, Petersen PE, Lennon MA. 2005. The effective use of fluorides in public health. *Bull World Health Organ* 83: 670-676.

Kang J, Cheng Y, Wu K, Lin S, He G, Jin Y. 2011. [Effect of exposure to fluoride and arsenic in drinking water of Hangjinhouqi on children's intelligence]. *Chin School Health*: 679-681.

Khan AM, Raina R, Dubey N, Verma PK. 2017. Effect of deltamethrin and fluoride co-exposure on the brain antioxidant status and cholinesterase activity in Wistar rats. *Drug Chem Toxicol* 41(2): 1-5.

Khandare AL, Gourineni SR, Validandi V. 2017. Dental fluorosis, nutritional status, kidney damage, and thyroid function along with bone metabolic indicators in school-going children living in fluoride-affected hilly areas of Doda District, Jammu and Kashmir, India. *Environ Monit Assess* 189: 579.

Khandare AL, Validandi V, Gourineni SR, Gopalan V, Nagalla B. 2018. Dose-dependent effect of fluoride on clinical and subclinical indices of fluorosis in school going children and its mitigation by supply of safe drinking water for 5 years: An Indian study. *Environ Monit Assess* 190: 110.

Kheradpisheh Z, Mahvi AH, Mirzaei M, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018a. Correlation between drinking water fluoride and TSH hormone by ANNs and ANFIS. *J Environ Health Sci Eng* 16(1): 11-18.

Kheradpisheh Z, Mirzaei M, Mahvi AH, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018b. Impact of drinking water fluoride on human thyroid hormones: A case-control study. *Sci Rep* 8: 2674.

Kumar V, Chahar P, Kajjari S, Rahman F, Bansal DK, Kapadia JM. 2018. Fluoride, thyroid hormone derangements and its correlation with tooth eruption pattern among the pediatric population from endemic and non-endemic fluorosis areas. *J Contemp Dent Pract* 19(12): 1512-1516.

Li F, Chen X, Huang R, Xie Y. 2009. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. *J Environ Health* 26(4): 838-840.

Li J, Yao L, Q.L. S, Wu CY. 2004. [Effects of high fluoride level on neonatal neurobehavioral development]. *Chin J Endemiol* 23(5): 463-465.

Li J, Yao L, Shao QL, Wu CY. 2008a. Effects of high fluoride level on neonatal neurobehavioral development. *Fluoride* 41: 165-170.

Li M, Gao Y, Cui J, Li Y, Li B, Liu Y, Sun J, Liu X, Liu H, Zhao L, Sun D. 2016. Cognitive impairment and risk factors in elderly people living in fluorosis areas in China. *Biol Trace Elem Res* 172: 53-60.

Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. 2010. [Investigation and analysis of children's IQ and dental fluorosis in high fluoride area]. *Chin J Pest Control* 26(3): 230-231.

Li Y, Li X, Wei S. 1994. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. *J West Chin University Med Sci* 25(2): 188-191.

Li Y, Li X, Wei S. 2008b. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. *Fluoride* 41: 331-335.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2003. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. *Chin J Public Health Manag* 19(4): 337-338.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2008c. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. *Fluoride* 41: 161-164.

Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. 1991. [High fluoride and low iodine environment and subclinical cretinism in Xinjiang]. *Endem Dis Bull* 6(2): 62-67.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. Alterations of nAChRs and ERK1/2 in the brains of rats with chronic fluorosis and their connections with the decreased capacity of learning and memory. *Toxicol Lett* 192: 324-329.

Lou DD, Guan ZZ, Liu YJ, Liu YF, Zhang KL, Pan JG, Pei JJ. 2013. The influence of chronic fluorosis on mitochondrial dynamics morphology and distribution in cortical neurons of the rat brain. *Arch Toxicol* 87: 449-457.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S, Mentes A, DenBesten PK. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. *Physiol Behav* 206: 76-83.

Ma Q, Huang H, Sun L, Zhou T, Zhu J, Cheng X, Duan L, Li Z, Cui L, Ba Y. 2017. Gene-environment interaction: Does fluoride influence the reproductive hormones in male farmers modified by ERα gene polymorphisms? *Chemosphere* 188: 525-531.

Malin AJ, Riddell J, McCague H, Till C. 2018. Fluoride exposure and thyroid function among adults living in Canada: Effect modification by iodine status. *Environ Int* 121(Pt 1): 667-674.

Malin AJ, Till C. 2015. Exposure to fluoridated water and attention deficit hyperactivity disorder prevalence among children and adolescents in the United States: An ecological association. *Environ Health* 14: 17.

McPherson CA, Zhang G, Gilliam R, Brar SS, Wilson R, Brix A, Picut C, Harry GJ. 2018. An evaluation of neurotoxicity following fluoride exposure from gestational through adult ages in Long-Evans hooded rats. *Neurotoxicol Res*: 1-18.

Mesram N, Nagapuri K, Banala RR, Nalagoni CR, Karnati PR. 2016. Quercetin treatment against NaF induced oxidative stress related neuronal and learning changes in developing rats. *J King Saud Univ Sci* 29: 221-229.

Michael M, Barot VV, Chinoy NJ. 1996. Investigations of soft tissue functions in fluorotic individuals of North Gujarat. *Fluoride* 29: 63-71.

Miller K, Howard B, Phillips J, Shah M, Mav D, Thayer K, Shah R. 2016. SWIFT-Active screener: Reducing literature screening effort through machine learning for systematic reviews, Cochrane Colloquium Seoul, Seoul, Korea.

Moher D, Liberati A, Tetzlaff J, Altman D. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med* 6(7): e1000097.

Morgan L, Allred E, Tavares M, Bellinger D, Needleman H. 1998. Investigation of the possible associations between fluorosis, fluoride exposure, and childhood behavior problems. *Pediatr Dent* 20: 244-252.

Mustafa DE, Younis UM, Elhag SAA. 2018. The relationship between the fluoride levels in drinking water and the schooling performance of children in rural areas of Khartoum State, Sudan. *Fluoride* 51(2): 102-113.

Nageshwar M, Sudhakar K, Reddy KP. 2018. Quercetin ameliorates oxidative stress, neural damage of brain and behavioral impairment of rat with fluoride exposure. *Int J Pharm Sci Res* 9(8): 3247-3256.

NASEM (National Academies of Sciences, Engineering and Medicine). 2020. *Review of the draft NTP monograph: Systematic review of fluoride exposure and neurodevelopmental and cognitive health effects*. Washington, DC: The National Academies Press. Available: https://doi.org/10.17226/25715.

NASEM (National Academies of Sciences, Engineering and Medicine). 2021. *Review of the revised NTP Monograph on the systematic review of fluoride exposure and neurodevelopmental and cognitive health effects: A letter report*. Washington, DC. Available: https://doi.org/10.17226/26030.

NIOSH (National Institute for Occupational Safety and Health). 1984. *Fluoride in urine*. In: Manual of Analytical Methods Vol 11. Method 8308. Washington, DC: US Department of Health and Human Services: 1-3.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. *Environ Pollut* 233: 889-899.

Nkpaa KW, Onyeso GI. 2018. Rutin attenuates neurobehavioral deficits, oxidative stress, neuro-inflammation and apoptosis in fluoride treated rats. *Neurosci Lett* 682: 92-99.

NRC (National Research Council). 2006. *Committee on fluoride in drinking water, board on environmental studies and toxicology. Fluoride in drinking water: A scientific review of EPA's standards.* Available: http://www.nap.edu/catalog/11571/fluoride-in-drinking-water-a-scientific-review-of-epas-standards [accessed 19 August 2019].

NTP (National Toxicology Program). 2016. *Systematic literature review on the effects of fluoride on learning and memory in animal studies*. NTP Research Report 1. Research Triangle Park, NC. Available: https://ntp.niehs.nih.gov/ntp/results/pubs/rr/reports/01fluoride_508.pdf [accessed 19 August 2019].

NTP (National Toxicology Program). 2020. *Revised draft NTP Monograph on the systematic review of fluoride exposure and neurodevelopmental and cognitive health effects*. Research Triangle Park, NC. Available: https://www.nationalacademies.org/event/10-19-2020/docs/DDA97C9170D1A255D69C004CEB77B698E8D005011EFB.

OEHHA (California Office of Environmental Health Hazard Assessment). 2011. *Meeting synopsis and slide presentations: carcinogen identification committee meeting held on October 12, 2011*. Available: http://oehha.ca.gov/prop65/public_meetings/cic101211synop.html [accessed 19 August 2019].

Peckham S, Lowery D, Spencer S. 2015. Are fluoride levels in drinking water associated with hypothyroidism prevalence in England? A large observational study of GP practice data and fluoride levels in drinking water. *J Epidemiol Community Health* 69: 619-624.

Podgorski J, Berg M. 2020. Global threat of arsenic in groundwater. *Science* 368(6493): 845-850.

Pulungan ZS, Sofro ZM, Partadiredja G. 2016. Sodium fluoride does not affect the working memory and number of pyramidal cells in rat medial prefrontal cortex. *Anat Sci Int* 93(1): 128-138.

Reddy YP, Tiwari SK, Shaik AP, Alsaeed A, Sultana A, Reddy PK. 2014. Effect of sodium fluoride on neuroimmunological parameters, oxidative stress and antioxidative defenses. *Toxicol Mech Methods* 24: 31-36.

Ren D, Li K, Liu D. 1989. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. *Chin J Control Endem Dis* 4(4): 251.

Ren D, Li K, Liu D. 2008. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. *Fluoride* 41: 319-320.

Riddell JK, Malin AJ, Flora D, McCague H, Till C. 2019. Association of water fluoride and urinary fluoride concentrations with attention deficit hyperactivity disorder in Canadian youth. *Environ Int* 133: 105190.

Rocha-Amador D, Navarro M, Trejo-Acevedo A, Carrizales L, Perez-Maldonado I, Diaz-Barriga F, Calderon J. 2009. Use of the Rey-Osterrieth Complex Figure Test for neurotoxicity evaluation of mixtures in children. *Neurotox* 30: 1149-1154.

Rocha-Amador D, Navarro ME, Carrizales L, Morales R, Calderon J. 2007. Decreased intelligence in children and exposure to fluoride and arsenic in drinking water. *Cad Saude Publica* 23(Suppl 4): S579-587.

Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. *Environ Health Perspect* 122(7): 711-718.

Rooney AA, Cooper GS, Jahnke GD, Lam J, Morgan RL, Ratcliffe JM, Kraft AD, Schünemann HJ, Schwingl P, Walker TD, Thayer KA, Lunn RM. 2016. How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. *Environ Int* 92-93: 617-629.

Rotton J, Tikofsky RS, Feldman HT. 1982. Behavioral effects of chemicals in drinking water. *J Appl Psychol* 67: 230-238.

Russ TC, Killin LOJ, Hannah J, Batty GD, Deary IJ, Starr JM. 2019. Aluminium and fluoride in drinking water in relation to later dementia risk. *Brit J Psychol* 14: 1-6.

Santa-Marina L, Jimenez-Zabala A, Molinuevo A, Lopez-Espinosa M, Villanueva C, Riano I, Ballester F, Sunyer J, Tardon A, Ibarluzea J. 2019. Fluorinated water consumption in pregnancy and neuropsychological development of children at 14 months and 4 years of age. *Environ Epidemiol* 3: 386-387.

Saxena S, Sahay A, Goel P. 2012. Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India. *J Neurosci Rural Pract* 3: 144-149.

SCHER (Scientific Committee on Health and Environmental Risks). 2011. *Critical review of any new evidence on the hazard profile, health effects, and human exposure to fluoride and the fluoridating agents of drinking water*. European Commission Directorate-General for Health and Consumers Scientific Committees. Available: http://ec.europa.eu/health/scientific_committees/environmental_risks/docs/scher_o_139.pdf [accessed 19 August 2019].

Seraj B, Shahrabi M, Shadfar M, Ahmadi R, Fallahzadeh M, Eslamlu HF, Kharazifard MJ. 2012. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. *J Dent* 9: 221-229.

Shan KR, Qi XL, Long YG, Nordberg A, Guan ZZ. 2004. Decreased nicotinic receptors in PC12 cells and rat brains influenced by fluoride toxicity: A mechanism relating to a damage at the level in post-transcription of the receptor genes. *Toxicology* 200: 169-177.

Shannon FT, Fergusson DM, Horwood LJ. 1986. Exposure to fluoridated public water supplies and child health and behaviour. *New Zeal Med J* 99: 416-418.

Shao Q. 2003. [Study of cognitive function impairment caused by chronic fluorosis]. *Chin J Endemiol* 22(4): 336-338.

Sharma JD, Sohu D, Jain P. 2009. Prevalence of neurological manifestations in a human population exposed to fluoride in drinking water. *Fluoride* 42: 127-132.

Singh N, Verma KG, Verma P, Sidhu GK, Sachdeva S. 2014. A comparative study of fluoride ingestion levels, serum thyroid hormone & TSH level derangements, dental fluorosis status among school children from endemic and non-endemic fluorosis areas. *Springerplus* 3: 7.

Singh V, Singh C, Sandeep T, Sandeep K, Vikas G, Mukesh T, Anurag T. 2013. A correlation between serum vitamin, acetylcholinesterase activity and IQ in children with excessive endemic fluoride exposure in Rajasthan, India. *Int Res J Medical Sci* 1(3): 12-16.

Soto-Barreras U, Escalante-Villalobos KY, Holguin-Loya B, Perez-Aguirre B, Nevarez-Rascon A, Martinez-Martinez RE, Loyola-Rodriguez JP. 2019. Effect of fluoride in drinking water on dental caries and IQ in children. *Fluoride* 52: 474-482.

Still CN, Kelley P. 1980. The incidence of primary degenerative dementia vs. water fluoride content in South Carolina, USA. *Neurotox* 1: 125-132.

Sudhakar K, Reddy KP. 2018. Protective effects of Abelmoschus moschatus seed extract on neurotransmitter system of developing brain of Wistar rats with gestational and post-natal exposure of sodium fluoride. *Int J Green Pharm* 12: S131-S139.

Sudhir KM, Chandu GN, Prashant GM, Subba Reddy VV. 2009. Effect of fluoride exposure on intelligence quotient (IQ) among 13-15 year old school children of known endemic area of fluorosis, Nalgonda District, Andhra Pradesh. *J Indian Assoc Public Health Dent* 2009(13): 88-94.

Sun M, Li S, Wang Y, Li F. 1991. [Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis]. *J Guiyang Med Coll* 16(3): 204-206.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2000. [Effects of high fluoride drinking water on the cerebral function of mice]. *Chin J Epidemiol* 19: 262-263.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2008. Effects of high fluoride drinking water on the cerebral function of mice. *Fluoride* 41: 148-151.

Susheela AK, Bhatnagar M, Vig K, Mondal NK. 2005. Excess fluoride ingestion and thyroid hormone derangements in children living in Delhi, India. *Fluoride* 38: 98-108.

Till C, Green R, Flora D, Hornung R, Martinez-Mier EA, Blazer M, Farmus L, Ayotte P, Muckle G, Lanphear B. 2020. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. *Environ Int* 134: 105315.

Till C, Green R, Grundy JG, Hornung R, Neufeld R, Martinez-Mier EA, Ayotte P, Muckle G, Lanphear B. 2018. Community water fluoridation and urinary fluoride concentrations in a national sample of pregnant women in Canada. *Environ Health Perspect* 126(10): 107001.

Trivedi M, Sangai N, Patel R, Payak M, Vyas S. 2012. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. *Fluoride* 45(4): 377-383.

Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. 2007. Effect of high fluoride water on intelligence of school children in India. *Fluoride* 40: 178-183.

Tsunoda M, Aizawa Y, Nakano K, Liu Y, Horiuchi T, Itai K, Tsunoda H. 2005. Changes in fluoride levels in the liver, kidney, and brain and in neurotransmitters of mice after subacute administration of fluoride. *Fluoride* 38: 284-292.

US DHHS (U.S. Department of Health and Human Services). 2015. *U.S. Public Health Service recommendation for fluoride concentration in drinking water for the prevention of dental caries*. 318-331. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4547570/ [accessed 19 August 2019].

US EPA (U.S. Environmental Protection Agency). 2010. *Fluoride: Exposure and relative source contribution analysis*. 820-R-10-015. Washington, DC. Available: http://www.epa.gov/dwstandardsregulations/fluoride-risk-assessment-and-relative-source-contribution [accessed 19 August 2019].

Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, Costilla-Salazar R, Calderon Hernandez J, Alcaraz Contreras Y, Rocha-Amador DO. 2017. In utero exposure to fluoride and cognitive development delay in infants. *Neurotox* 59: 65-70.

Villa A, Anabalon M, Zohouri V, Maguire A, Franco AM, Rugg-Gunn A. 2010. Relationships between fluoride intake, urinary fluoride excretion and fluoride retention in children and adults: An analysis of available data. *Caries Res* 44(1): 60-68.

Wang A, Duan L, Huang H, Ma J, Zhang Y, Ma Q, Guo Y, Li Z, Cheng X, Zhu J, Zhou G, Ba Y. 2020a. Association between fluoride exposure and behavioural outcomes of school-age children: A pilot study in China. *Int J Environ Health Res*: 1-10.

Wang G, Gao M, Zhang M, Yang M, Xiang Q. 2012. Correlation between total fluoride intake and children's IQ. *J Southeast Univ Med Ed*: 743-746.

Wang M, Liu L, Li H, Li Y, Liu H, Hou C, Zeng Q, Li P, Zhao Q, Dong L, Zhou G, Yu X, Liu L, Guan Q, Zhang S, Wang A. 2020b. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. *Environ Int* 134: 105229.

Wang X, Wang L, Hu P, Guo X, Luo X. 2001. [Effects of high iodine and high fluorine on children's intelligence and thyroid function]. *Chin J Endemiol* 20(4): 288-290.

Watanabe M, Kono K, Orita Y, Dote T, Usuda K, Takahashi Y, Yoshida Y. 1995. Influence of dietary fluoride intake on urinary fluoride concentration and evaluation of corrected levels in spot urine. *Fluoride* 28(2): 61-70.

WHO (World Health Organization). 2008. *Guidelines for drinking-water quality [electronic resource]: Incorporating 1st and 2nd addenda*. Third Edition. Vol. 1. Geneva, Switzerland. Available: https://apps.who.int/iris/bitstream/handle/10665/204411/9789241547611_eng.pdf?sequence=1&isAllowed=y.

WHO (World Health Organization). 2011. *Guidelines for drinking-water quality*. Fourth edition. Available: https://apps.who.int/iris/bitstream/handle/10665/44584/9789241548151_eng.pdf?sequence=1&isAllowed=y&ua=1.

Xiang Q, Liang Y, Chen B, Chen L. 2011. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. *Fluoride* 44: 191-194.

Xiang Q, Liang Y, Chen L, Wang C, Chen B, Chen X, Zhou M. 2003a. Effect of fluoride in drinking water on children's intelligence. *Fluoride* 36: 84-94.

Xiang Q, Wang Y, Yang M, Zhang M, Xu Y. 2013. Level of fluoride and arsenic in household shallow well water in Wamiao and Xinhuai villages in Jiangsu province, China. *Fluoride* 46: 192-197.

Xiang QY, Liang YX, Zhou MS, Zang HB. 2003b. Blood lead of children in Wamiao-Xinhuai intelligence study. *Fluoride* 36: 198-199.

Yang Y, Wang X, Guo X, Hu P. 2008. The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine. *Fluoride* 41: 336-339.

Yang Y, Wang X, X. G, Hu P. 1994. [The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine]. *Chin J Epidemiol* 15(4): 296-298.

Yao L, Zhou J, Wang S, Cui K, Lin F. 1996. Analysis on TSH and intelligence level of children with dental Fluorosis in a high fluoride area. *Lit Inf Prev Med* 2(1): 26-27.

Yasmin S, Ranjan S, D'Souza D. 2013. Effect of excess fluoride ingestion on human thyroid function in Gaya Region, Bihar, India. *Toxicol Environ Chem* 95: 1235-1243.

Yazdi SM, Sharifian A, Dehghani-Beshne M, Momeni VR, Aminian O. 2011. Effects of fluoride on psychomotor performance and memory of aluminum potroom workers. *Fluoride* 44: 158-162.

Yu X, Chen J, Li Y, Liu H, Hou C, Zeng Q, Cui Y, Zhao L, Li P, Zhou Z, Pang S, Tang S, Tian K, Zhao Q, Dong L, Xu C, Zhang X, Zhang S, Liu L, Wang A. 2018. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. *Environ Int* 118: 116-124.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 1996. [Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis]. *Chin J Endemiol* 15(5): 257-259.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 2008. Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis. *Fluoride* 41: 134-138.

Zhang KL, Lou DD, Guan ZZ. 2015a. Activation of the AGE/RAGE system in the brains of rats and in SH-SY5Y cells exposed to high level of fluoride might connect to oxidative stress. *Neurotoxicol Teratol* 48: 49-55.

Zhang S, Zhang X, Liu H, Qu W, Guan Z, Zeng Q, Jiang C, Gao H, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Cui Y, Yu L, Wang A. 2015b. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. *Toxicol Sci* 144: 238-245.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao Q, Tian Z, Zhou G, Niu Q, Chen J, Li P, Dong L, Xia T, Zhang S, Wang A. 2020. SIRT1-dependent mitochondrial biogenesis supports therapeutic effects of resveratrol against neurodevelopment damage by fluoride. *Theranostics* 10: 4822-4838.

Zhou G, Tang S, Yang L, Niu Q, Chen J, Xia T, Wang S, Wang M, Zhao Q, Liu L, Li P, Dong L, Yang K, Zhang S, Wang A. 2019. Effects of long-term fluoride exposure on cognitive ability and the underlying mechanisms: Role of autophagy and its association with apoptosis. *Toxicol Appl Pharmacol* 378: 114608.

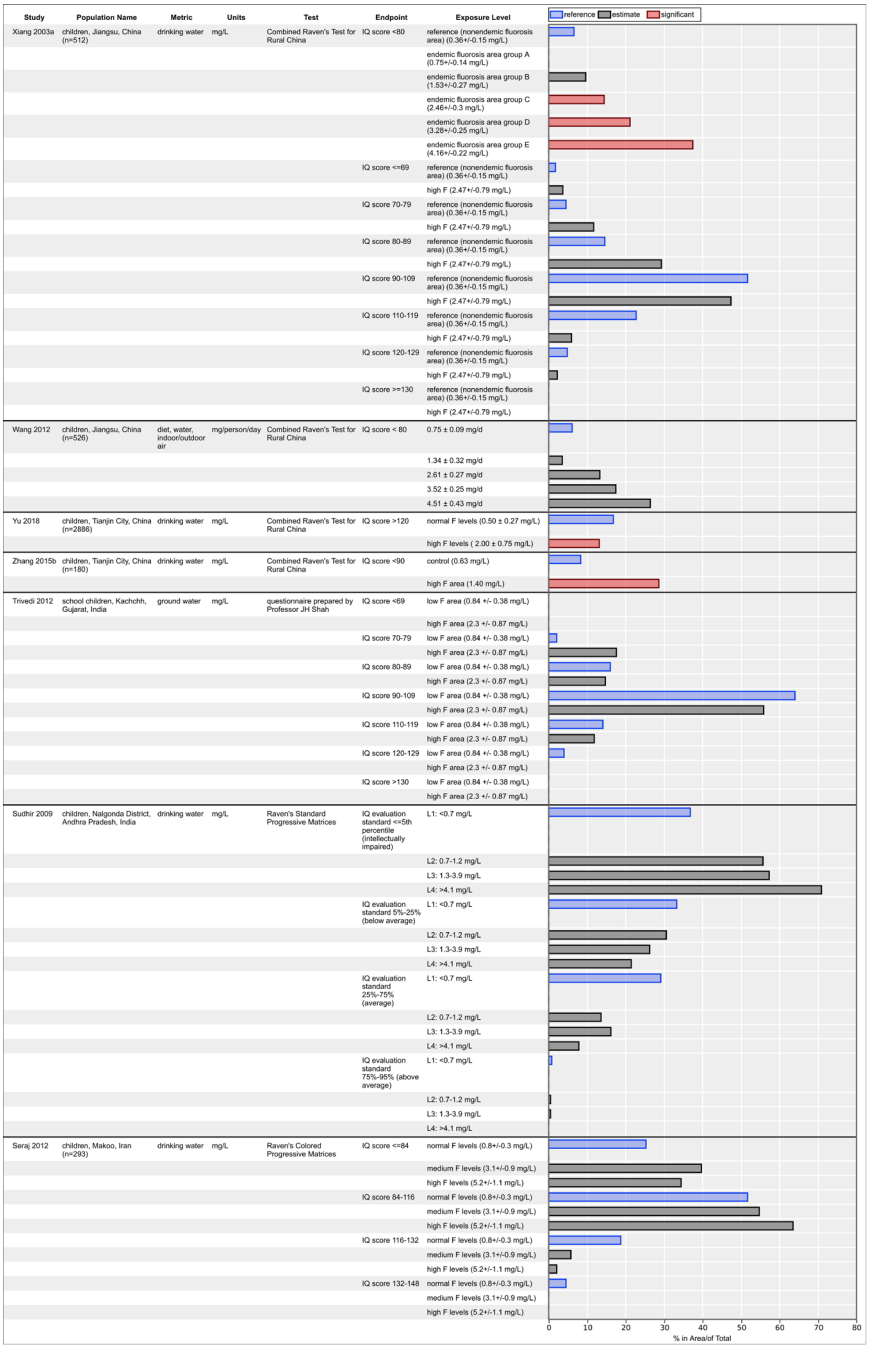
Zhou T, Duan L-J, Ding Z, Yang R-P, Li S-H, Xi Y, Cheng X-M, Hou J-X, Wen S-B, Chen J, Cui L-X, Ba Y. 2012. Environmental fluoride exposure and reproductive hormones in male living in endemic fluorosis villages in China. *Life Sci J* 9(4): 1-7.

Zohouri FV, Swinbank CM, Maguire A, Moynihan PJ. 2006. Is the fluoride/creatinine ratio of a spot urine sample indicative of 24-h urinary fluoride? *Community Dent Oral Epidemiol* 34(2): 130-138.

DATA FIGURES

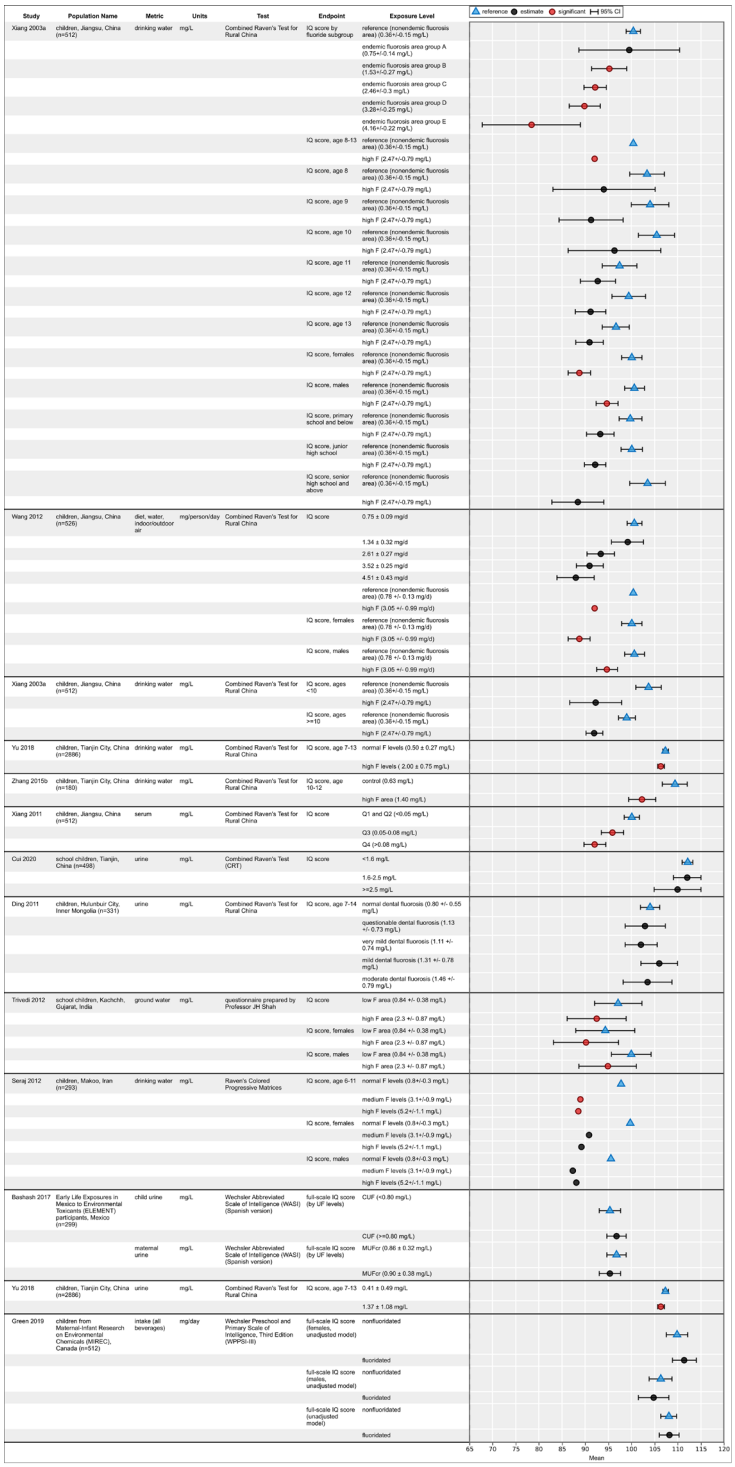
Neurodevelopmental or Cognitive Effects and Outcomes

Figure D1. IQ Distribution in Children by Fluoride Exposure (low risk-of-bias studies; presented as % in area or % of total group)



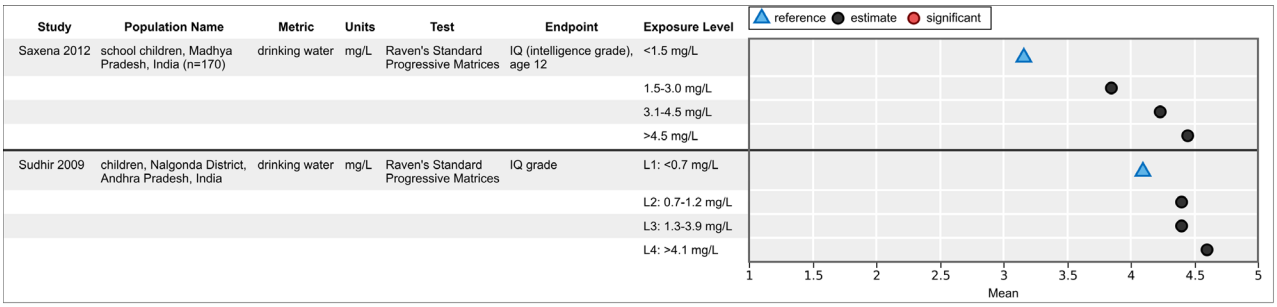
Interactive figure and additional study details in HAWC here. "F" represents fluoride. Differences in intelligence between the reference group and treatment groups were statistically significant although significance was not reported separately for each score level.

Figure D2. Mean IQ in Children by Fluoride Exposure (low risk-of-bias studies)



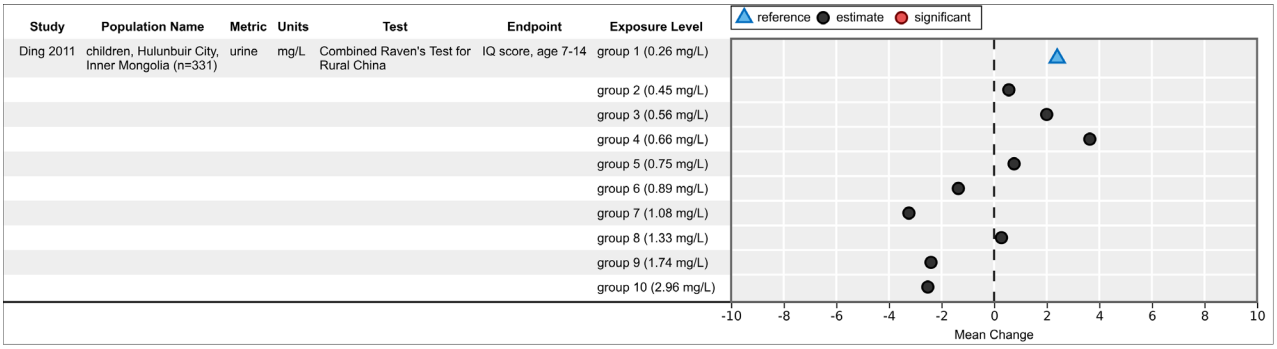
Interactive figure and additional study details in HAWC here. "F" represents fluoride. Three additional publications based on subsample of the larger Yu *et al.* (2018) cohort were identified (Zhao *et al.* 2019, Zhou *et al.* 2019, Zhao *et al.* 2020); however, results from these studies are not presented here. The main study by Yu *et al.* (2018) is considered a better representation of the IQ results. For all studies, SDs are available and can be viewed in HAWC by clicking the data points within the plot area; however, 95% CIs could not be calculated for Seraj *et al.* (2012) because Ns are not available for exposure groups.

Figure D3. Intelligence Grade in Children by Fluoride Exposure (low risk-of-bias studies; presented as mean)



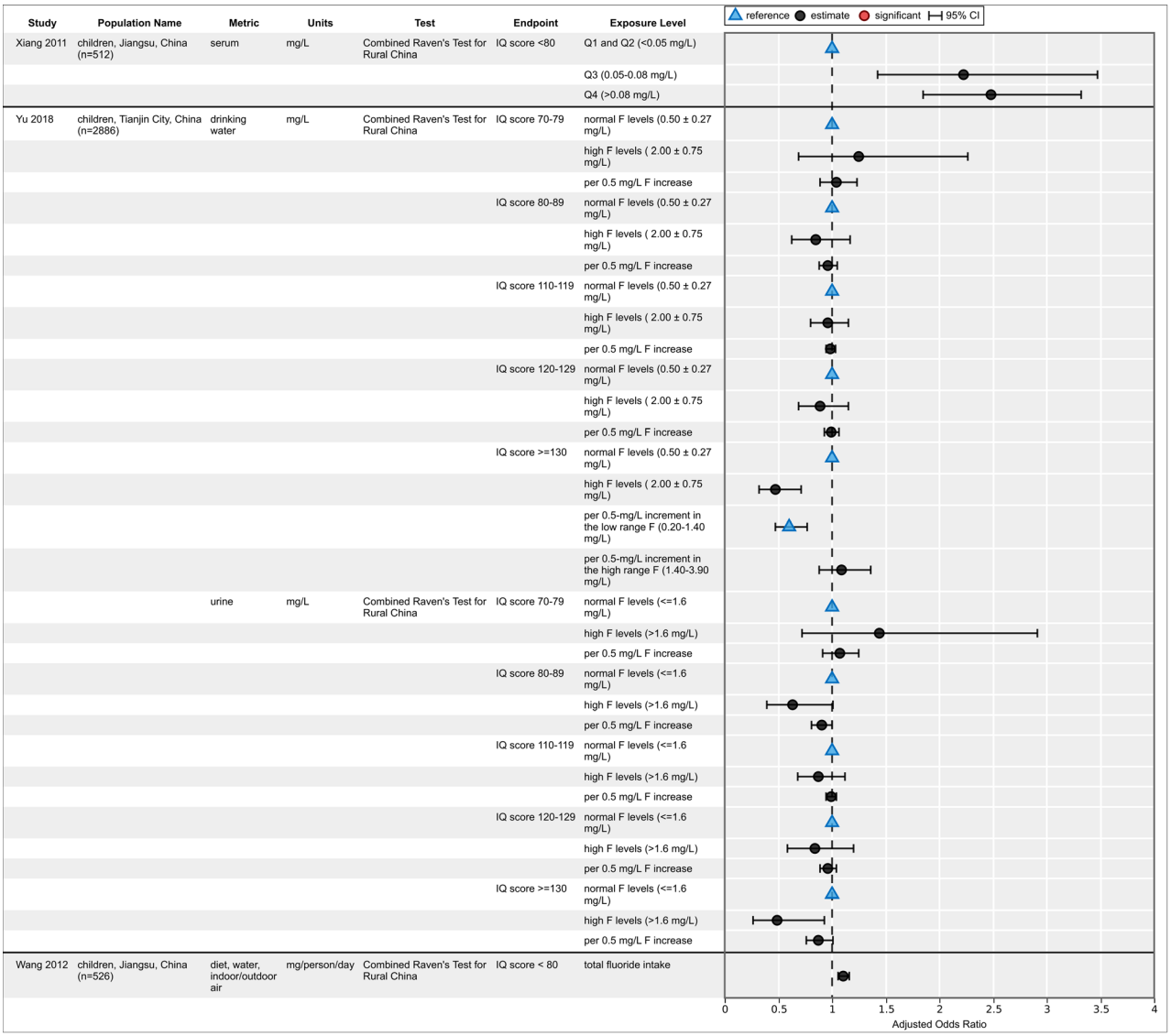
Interactive figure and additional study details in HAWC here. For Saxena *et al.* (2012), children's intelligence was measured using the Raven's Standard Progressive Matrices. Children's scores were converted to percentile and specific grades were allotted based on the percentiles. Grades ranged from intellectually superior (Grade I) to intellectually impaired (Grade V). Results for Soto-Barreras *et al.* (2019) are not presented here. Outcomes in the study were presented as levels of fluoride exposure associated with each intelligence grade. Results reported were not significant.

Figure D4. Mean Change in IQ in Children by Fluoride Exposure (low risk-of-bias studies)



Interactive figure and additional study details in HAWC here. For Ding *et al.* (2011), SDs are available and can be viewed in HAWC by clicking the data points within the plot area; however, 95% CIs could not be calculated because Ns for each exposure group are not available.

Figure D5. IQ Score in Children by Fluoride Exposure (low risk-of-bias studies; presented as adjusted OR)

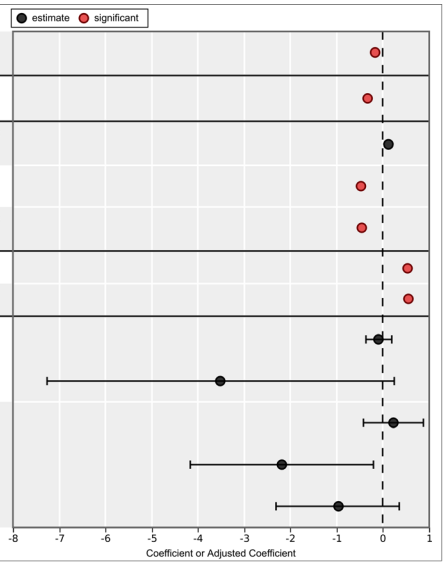


Study	Population Name	Metric	Units	Test	Endpoint	Exposure Level
Xiang 2011	children, Jiangsu, China (n=512)	serum	mg/L	Combined Raven's Test for Rural China	IQ score <80	Q1 and Q2 (<0.05 mg/L)
						Q3 (0.05-0.08 mg/L)
						Q4 (>0.08 mg/L)
Yu 2018	children, Tianjin City, China (n=2886)	drinking water	mg/L	Combined Raven's Test for Rural China	IQ score 70-79	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score 80-89	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score 110-119	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score 120-129	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5 mg/L F increase
					IQ score >=130	normal F levels (0.50 ± 0.27 mg/L)
						high F levels (2.00 ± 0.75 mg/L)
						per 0.5-mg/L increment in the low range F (0.20-1.40 mg/L)
						per 0.5-mg/L increment in the high range F (1.40-3.90 mg/L)
		urine	mg/L	Combined Raven's Test for Rural China	IQ score 70-79	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score 80-89	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score 110-119	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score 120-129	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
					IQ score >=130	normal F levels (<=1.6 mg/L)
						high F levels (>1.6 mg/L)
						per 0.5 mg/L F increase
Wang 2012	children, Jiangsu, China (n=526)	diet, water, indoor/outdoor air	mg/person/day	Combined Raven's Test for Rural China	IQ score < 80	total fluoride intake

Interactive figure and additional study details in HAWC here. For Xiang *et al.* (2011), there was a significant linear trend across different levels of serum fluoride for IQ score < 80 ($p < 0.001$). For Yu *et al.* (2018), significance levels by IQ score were not reported.

Figure D6. Correlations between IQ Score and Fluoride Exposure in Children (low risk-of-bias studies; presented as coefficient)

Study	Population Name	Metric	Units	Test	Endpoint	Exposure Level	Stat Description
Xiang 2003a	children, Jiangsu, China (n=512)	urine	mg/L	Combined Raven's Test for Rural China	IQ score	continuous F measure	Pearson's correlation coefficient adjusted for creatinine
Wang 2012	children, Jiangsu, China (n=526)	diet, water, indoor/outdoor air	mg/person/day	Combined Raven's Test for Rural China	IQ score	total fluoride intake	correlation coefficient
Zhang 2015b	children, Tianjin City, China (n=180)	drinking water	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
		serum	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
		urine	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
Saxena 2012	school children, Madhya Pradesh, India (n=170)	drinking water	mg/L	Raven's Standard Progressive Matrices	IQ (intelligence grade), age 12	continuous F measure	regression coefficient (R)
		urine	ppm	Raven's Standard Progressive Matrices	IQ (intelligence grade), age 12	continuous F measure	regression coefficient (R)
Yu 2018	children, Tianjin City, China (n=2886)	drinking water	mg/L	Combined Raven's Test for Rural China	change in IQ score (unadjusted model)	per 0.5-mg/L increment in the low range F (0.20-3.40 mg/L)	unadjusted beta
						per 0.5-mg/L increment in the high range F (3.40–3.90 mg/L)	unadjusted beta
		urine	mg/L	Combined Raven's Test for Rural China	change in IQ score (unadjusted model)	per 0.5-mg/L increment in the low range F (0.01-1.60 mg/L)	unadjusted beta
						per 0.5-mg/L increment in the medium range F (1.60-2.50 mg/L)	unadjusted beta
						per 0.5-mg/L increment in the high range F (2.50-5.54 mg/L)	unadjusted beta



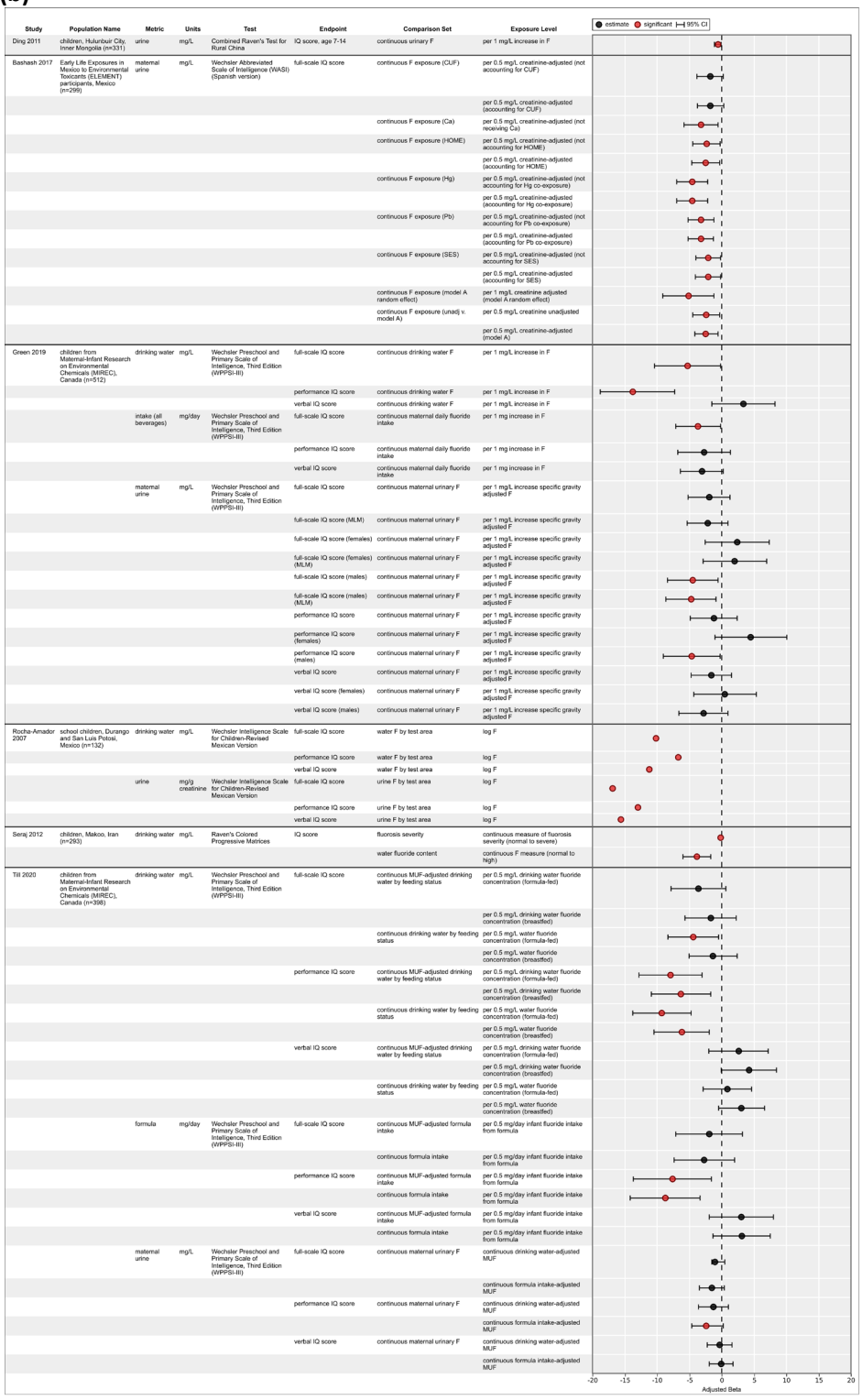
Interactive figure and additional study details in HAWC here. "F" represents fluoride. For Saxena *et al.* (2012), a significant relationship between water fluoride level and intelligence grade was observed. Increasing intelligence grades reflected increasing levels of impairment (reduced intelligence) in children. Zhao *et al.* (2020) and Zhou *et al.* (2019) also had correlations, but these were based on a subsample of the Yu *et al.* (2018) study (which presented betas and provided a better representation of the IQ data).

Figure D7. Correlations between IQ Score and Fluoride Exposure in Children (low risk-of-bias studies; presented as adjusted beta)—(a) China; (b) all other areas

(a)

Study	Population Name	Metric	Units	Test	Endpoint	Comparison Set	Exposure Level
Choi 2015	first-grade children, China (n=51)	drinking water	mg/L	Wechsler Intelligence Scale for Children-IV	digit span backward test score	continuous exposure_water	log10 F
						fluorosis (geometric mean water fluoride)	normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
					digit span forward test score	continuous exposure_water	log10 F
						fluorosis (geometric mean water fluoride)	normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
					digit span total test score	continuous exposure_water	log10 F
						fluorosis (geometric mean water fluoride)	normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
					squareroot block design test score	continuous exposure_water	log10 F
						fluorosis (geometric mean water fluoride)	normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
		urine	mg/L	Wechsler Intelligence Scale for Children-IV	digit span backward test score	continuous exposure_urine	log10 F
					digit span forward test score	continuous exposure_urine	log10 F
					digit span total test score	continuous exposure_urine	log10 F
					squareroot block design test score	continuous exposure_urine	log10 F
Cui 2018	school children, Tianjin, China (n=323)	urine	mg/L	Combined Raven's Test for Rural China	IQ score	urinary fluoride (TT subgroup)	continuous F measure
						urinary fluoride (overall)	continuous F measure
Wang 2020b	school children 7-13 years old, Tianjin, China (n=571)	drinking water	mg/L	Combined Raven's Test for Rural China (CRT-RC2)	IQ score	continuous water	per 1 mg/L water F
						water quartiles	Q1 (≤0.70 mg/L)
							Q2 (0.70–1.00 mg/L)
							Q3 (1.00–1.90 mg/L)
							Q4 (>1.90 mg/L)
					IQ score (boys)	continuous water	per 1 mg/L water F
						water quartiles	Q1 (≤0.70 mg/L)
							Q2 (0.70–1.00 mg/L)
							Q3 (1.00–1.90 mg/L)
							Q4 (>1.90 mg/L)
					IQ score (girls)	continuous water	per 1 mg/L water F
						water quartiles	Q1 (≤0.70 mg/L)
							Q2 (0.70–1.00 mg/L)
							Q3 (1.00–1.90 mg/L)
							Q4 (>1.90 mg/L)
		urine	mg/L	Combined Raven's Test for Rural China (CRT-RC2)	IQ score	continuous urine	per 1 mg/L urinary F
						urine quartiles	Q1 (≤0.15 mg/L)
							Q2 (0.15–0.41 mg/L)
							Q3 (0.41–2.28 mg/L)
							Q4 (>2.28 mg/L)
					IQ score (boys)	continuous urine	per 1 mg/L urinary F
						urine quartiles	Q1 (≤0.15 mg/L)
							Q2 (0.15–0.41 mg/L)
							Q3 (0.41–2.28 mg/L)
							Q4 (>2.28 mg/L)
					IQ score (girls)	continuous urine	per 1 mg/L urinary F
						urine quartiles	Q1 (≤0.15 mg/L)
							Q2 (0.15–0.41 mg/L)
							Q3 (0.41–2.28 mg/L)
							Q4 (>2.28 mg/L)
Yu 2018	children, Tianjin City, China (n=2886)	drinking water	mg/L	Combined Raven's Test for Rural China	change in IQ score	water F incremental w/ range	per 0.5-mg/L increment in the low range F (0.20-3.40 mg/L)
							per 0.5-mg/L increment in the high range F (3.40–3.90 mg/L)
		urine	mg/L	Combined Raven's Test for Rural China	change in IQ score	urine F incremental w/ range	per 0.5-mg/L increment in the low range F (0.01-1.60 mg/L)
							per 0.5-mg/L increment in the medium range F (1.60-2.50 mg/L)
							per 0.5-mg/L increment in the high range F (2.50-5.54 mg/L)
Zhang 2015b	children, Tianjin City, China (n=180)	urine	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	urinary fluoride	continuous F measure
						val/val stratum	continuous F measure

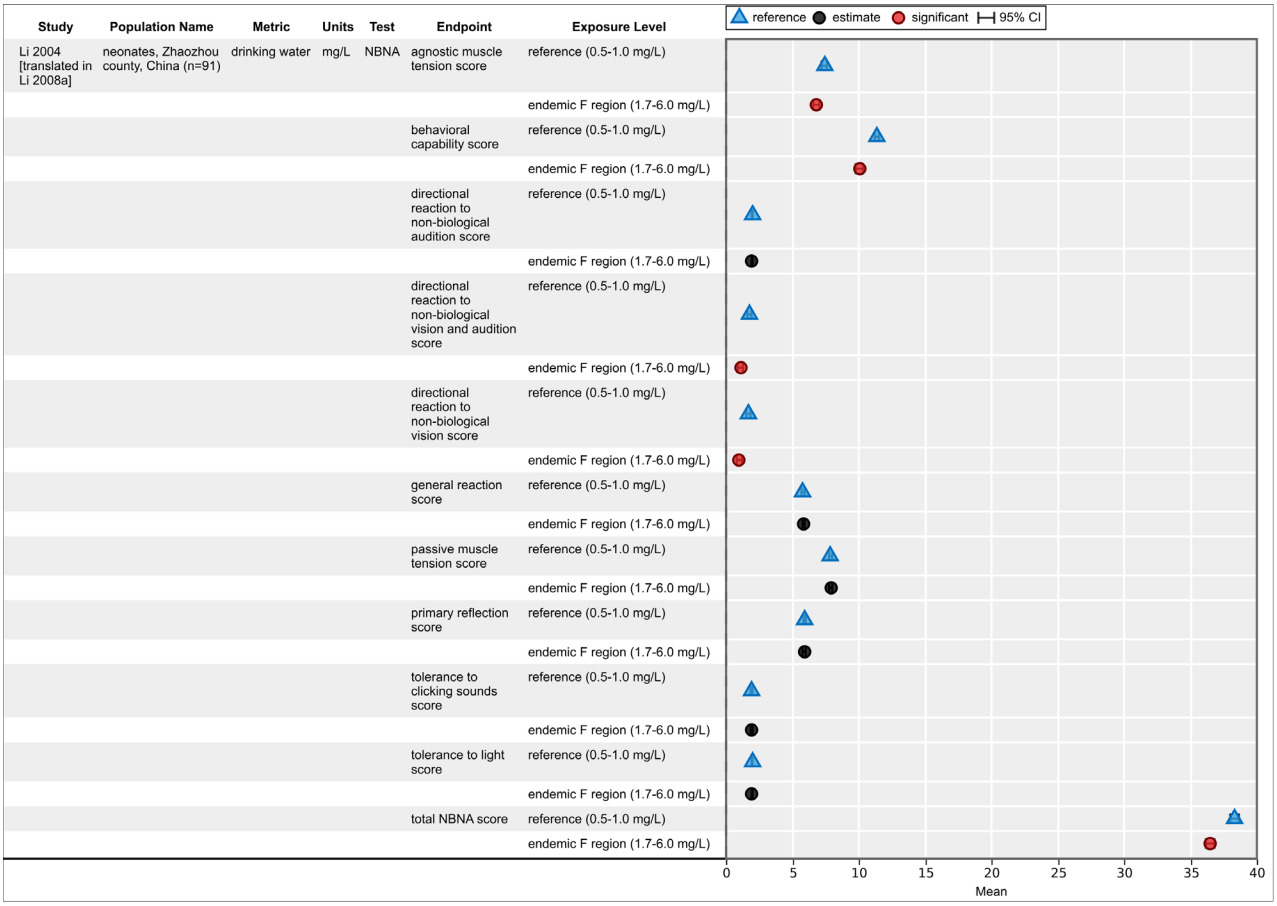
(b)



Study	Population Name	Metric	Units	Test	Endpoint	Comparison Set	Exposure Level
Ding 2011	children, Hulunbuir City, Inner Mongolia (n=331)	urine	mg/L	Combined Raven's Test for Rural China	IQ score, age 7-14	continuous urinary F	per 1 mg/L increase in F
Bashash 2017	Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants, Mexico (n=299)	maternal urine	mg/L	Wechsler Abbreviated Scale of Intelligence (WASI) (Spanish version)	full-scale IQ score	continuous F exposure (CUF)	per 0.5 mg/L creatinine-adjusted (not accounting for CUF)
							per 0.5 mg/L creatinine-adjusted (accounting for CUF)
						continuous F exposure (Ca)	per 0.5 mg/L creatinine-adjusted (not receiving Ca)
						continuous F exposure (HOME)	per 0.5 mg/L creatinine-adjusted (not accounting for HOME)
							per 0.5 mg/L creatinine-adjusted (accounting for HOME)
						continuous F exposure (Hg)	per 0.5 mg/L creatinine-adjusted (not accounting for Hg co-exposure)
							per 0.5 mg/L creatinine-adjusted (accounting for Hg co-exposure)
						continuous F exposure (Pb)	per 0.5 mg/L creatinine-adjusted (not accounting for Pb co-exposure)
							per 0.5 mg/L creatinine-adjusted (accounting for Pb co-exposure)
						continuous F exposure (SES)	per 0.5 mg/L creatinine-adjusted (not accounting for SES)
							per 0.5 mg/L creatinine-adjusted (accounting for SES)
						continuous F exposure (model A random effect)	per 1 mg/L creatinine adjusted (model A random effect)
						continuous F exposure (unadj v. model A)	per 0.5 mg/L creatinine unadjusted
							per 0.5 mg/L creatinine-adjusted (model A)
Green 2019	children from Maternal-Infant Research on Environmental Chemicals (MIREC), Canada (n=512)	drinking water	mg/L	Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	full-scale IQ score	continuous drinking water F	per 1 mg/L increase in F
					performance IQ score	continuous drinking water F	per 1 mg/L increase in F
					verbal IQ score	continuous drinking water F	per 1 mg/L increase in F
		intake (all beverages)	mg/day	Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	full-scale IQ score	continuous maternal daily fluoride intake	per 1 mg increase in F
					performance IQ score	continuous maternal daily fluoride intake	per 1 mg increase in F
					verbal IQ score	continuous maternal daily fluoride intake	per 1 mg increase in F
		maternal urine	mg/L	Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	full-scale IQ score	continuous maternal urinary F	per 1 mg/L increase specific gravity adjusted F
					full-scale IQ score (MLM)	continuous maternal urinary F	per 1 mg/L increase specific gravity adjusted F
					full-scale IQ score (females)	continuous maternal urinary F	per 1 mg/L increase specific gravity adjusted F
					full-scale IQ score (females) (MLM)	continuous maternal urinary F	per 1 mg/L increase specific gravity adjusted F
					full-scale IQ score (males)	continuous maternal urinary F	per 1 mg/L increase specific gravity adjusted F
					full-scale IQ score (males) (MLM)	continuous maternal urinary F	per 1 mg/L increase specific gravity adjusted F
					performance IQ score	continuous maternal urinary F	per 1 mg/L increase specific gravity adjusted F
					performance IQ score (females)	continuous maternal urinary F	per 1 mg/L increase specific gravity adjusted F
					performance IQ score (males)	continuous maternal urinary F	per 1 mg/L increase specific gravity adjusted F
					verbal IQ score	continuous maternal urinary F	per 1 mg/L increase specific gravity adjusted F
					verbal IQ score (females)	continuous maternal urinary F	per 1 mg/L increase specific gravity adjusted F
					verbal IQ score (males)	continuous maternal urinary F	per 1 mg/L increase specific gravity adjusted F
Rocha-Amador 2007	school children, Durango and San Luis Potosi, Mexico (n=132)	drinking water	mg/L	Wechsler Intelligence Scale for Children-Revised Mexican Version	full-scale IQ score	water F by test area	log F
					performance IQ score	water F by test area	log F
					verbal IQ score	water F by test area	log F
		urine	mg/g creatinine	Wechsler Intelligence Scale for Children-Revised Mexican Version	full-scale IQ score	urine F by test area	log F
					performance IQ score	urine F by test area	log F
					verbal IQ score	urine F by test area	log F
Seraj 2012	children, Makoo, Iran (n=293)	drinking water	mg/L	Raven's Colored Progressive Matrices	IQ score	fluorosis severity	continuous measure of fluorosis severity (normal to severe)
						water fluoride content	continuous F measure (normal to high)
Till 2020	children from Maternal-Infant Research on Environmental Chemicals (MIREC), Canada (n=396)	drinking water	mg/L	Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	full-scale IQ score	continuous MUF-adjusted drinking water by feeding status	per 0.5 mg/L drinking water fluoride concentration (formula-fed)
							per 0.5 mg/L drinking water fluoride concentration (breastfed)
						continuous drinking water by feeding status	per 0.5 mg/L water fluoride concentration (formula-fed)
							per 0.5 mg/L water fluoride concentration (breastfed)
					performance IQ score	continuous MUF-adjusted drinking water by feeding status	per 0.5 mg/L drinking water fluoride concentration (formula-fed)
							per 0.5 mg/L drinking water fluoride concentration (breastfed)
						continuous drinking water by feeding status	per 0.5 mg/L water fluoride concentration (formula-fed)
							per 0.5 mg/L water fluoride concentration (breastfed)
					verbal IQ score	continuous MUF-adjusted drinking water by feeding status	per 0.5 mg/L drinking water fluoride concentration (formula-fed)
							per 0.5 mg/L drinking water fluoride concentration (breastfed)
						continuous drinking water by feeding status	per 0.5 mg/L water fluoride concentration (formula-fed)
							per 0.5 mg/L water fluoride concentration (breastfed)
		formula	mg/day	Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	full-scale IQ score	continuous MUF-adjusted formula intake	per 0.5 mg/day infant fluoride intake from formula
						continuous formula intake	per 0.5 mg/day infant fluoride intake from formula
					performance IQ score	continuous MUF-adjusted formula intake	per 0.5 mg/day infant fluoride intake from formula
						continuous formula intake	per 0.5 mg/day infant fluoride intake from formula
					verbal IQ score	continuous MUF-adjusted formula intake	per 0.5 mg/day infant fluoride intake from formula
						continuous formula intake	per 0.5 mg/day infant fluoride intake from formula
		maternal urine	mg/L	Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	full-scale IQ score	continuous maternal urinary F	continuous drinking water-adjusted MUF
							continuous formula intake-adjusted MUF
					performance IQ score	continuous maternal urinary F	continuous drinking water-adjusted MUF
							continuous formula intake-adjusted MUF
					verbal IQ score	continuous maternal urinary F	continuous drinking water-adjusted MUF
							continuous formula intake-adjusted MUF

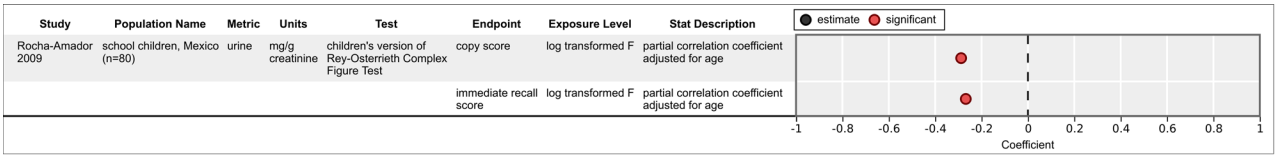
Interactive figure and additional study details in HAWC here for part (a) and here for part (b). "F" represents fluoride. For Yu *et al.* (2018), authors note an obvious decrease in the IQ score at water fluoride exposure levels between 3.40 mg/L and 3.90 mg/L and a similar adverse effect on IQ scores at urinary fluoride exposure levels from 1.60 mg/L to 2.50 mg/L, and so the changes in IQ score are indicated as significant; however, significance levels by change in IQ score were not reported.

Figure D8. Mean Motor/Sensory Scores in Children by Fluoride Exposure (low risk-of-bias studies)



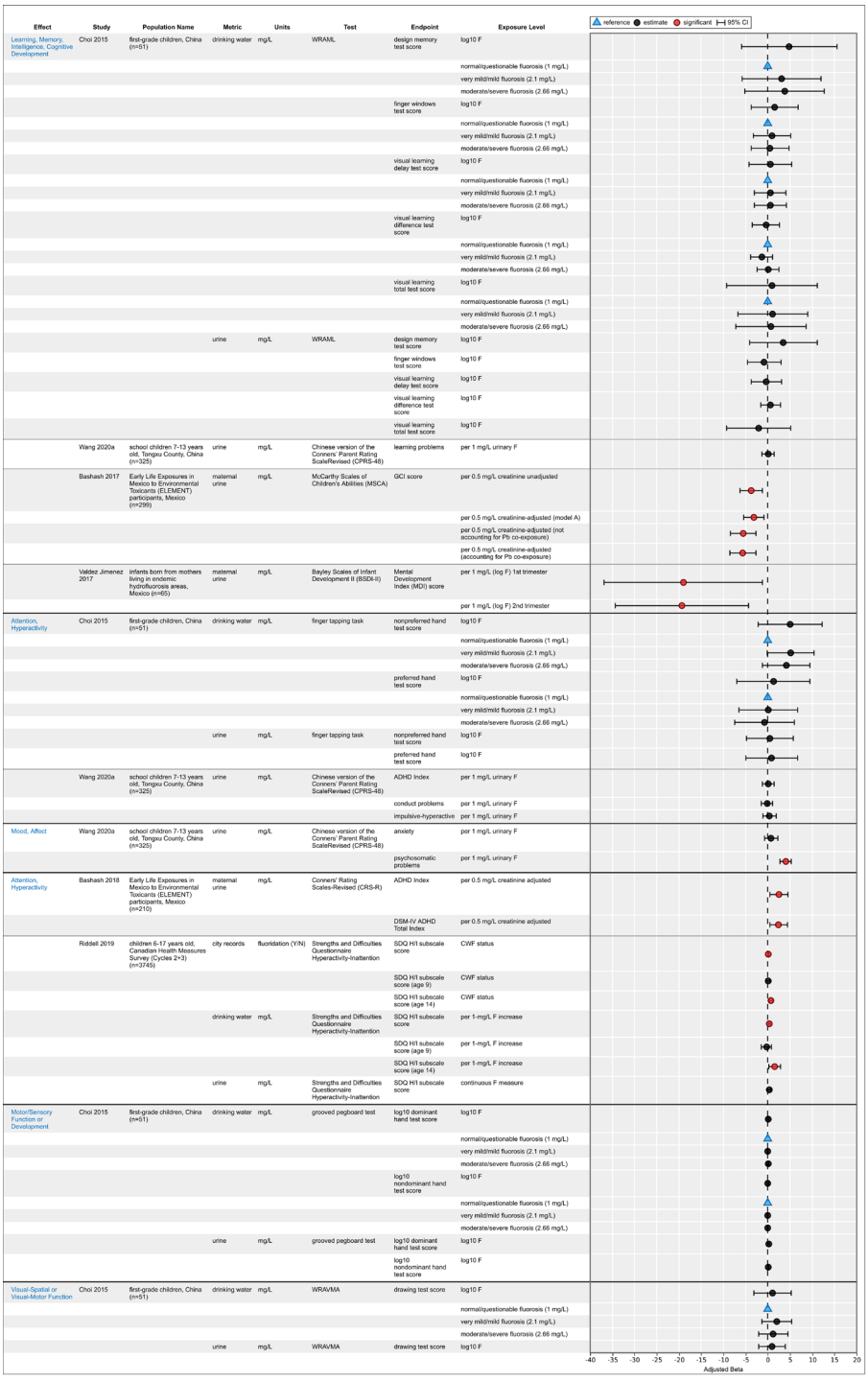
Interactive figure and additional study details in HAWC here. "F" represents fluoride. 95% CIs are small and are within figure symbols and may be difficult to see. Values for SDs and 95% CIs can be viewed in HAWC by clicking the data points within the plot area.

Figure D9. Correlations between Other Neurological Effects and Fluoride Exposure in Children (low risk-of-bias studies; presented as coefficient)



Interactive figure and additional study details in HAWC here. "F" represents fluoride.

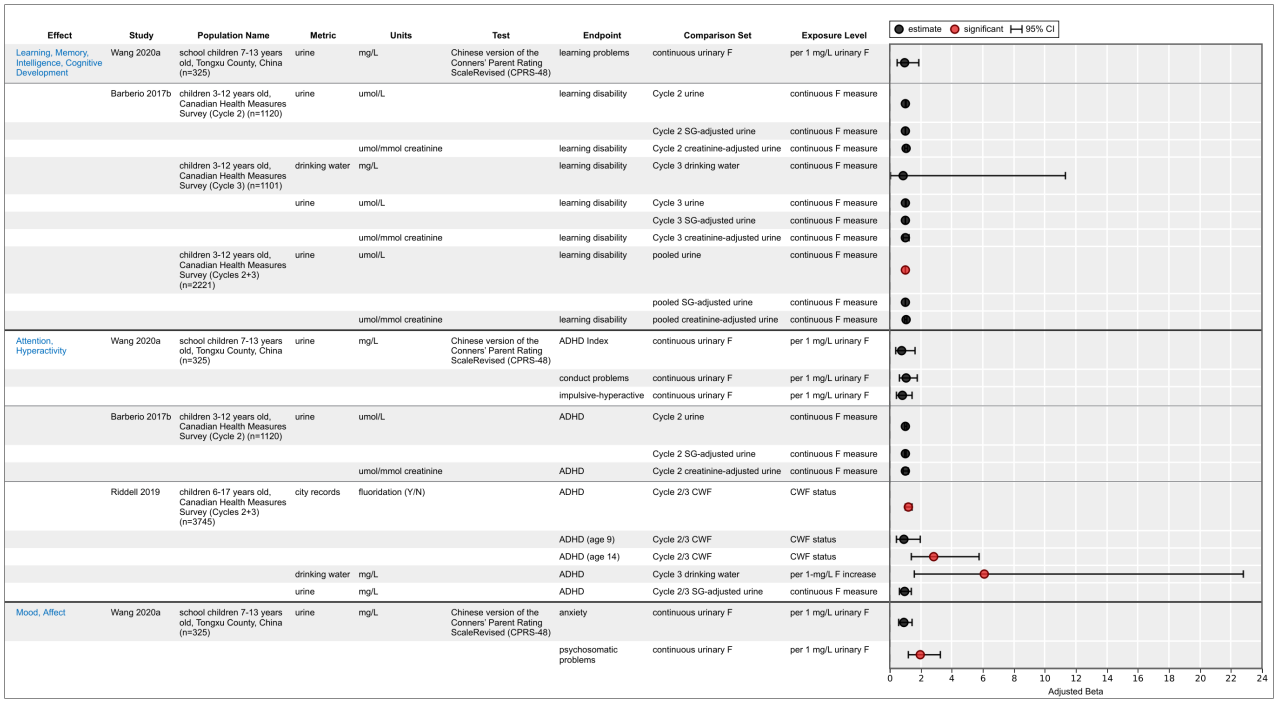
Figure D10. Correlations between Other Neurological Effects and Fluoride Exposure in Children (low risk-of-bias studies; presented as adjusted beta)



Effect	Study	Population Name	Metric	Units	Test	Endpoint	Exposure Level
Learning, Memory, Intelligence, Cognitive Development	Choi 2015	first-grade children, China (n=51)	drinking water	mg/L	WRAML	design memory test score	log10 F
							normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
						finger windows test score	log10 F
							normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
						visual learning delay test score	log10 F
							normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
						visual learning difference test score	log10 F
							normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
						visual learning total test score	log10 F
							normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
			urine	mg/L	WRAML	design memory test score	log10 F
						finger windows test score	log10 F
						visual learning delay test score	log10 F
						visual learning difference test score	log10 F
						visual learning total test score	log10 F
	Wang 2020a	school children 7-13 years old, Tongxu County, China (n=325)	urine	mg/L	Chinese version of the Conners' Parent Rating Scale Revised (CPRS-48)	learning problems	per 1 mg/L urinary F
	Bashash 2017	Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants, Mexico (n=299)	maternal urine	mg/L	McCarthy Scales of Children's Abilities (MSCA)	GCI score	per 0.5 mg/L creatinine unadjusted
							per 0.5 mg/L creatinine-adjusted (model A)
							per 0.5 mg/L creatinine-adjusted (not accounting for Pb co-exposure)
							per 0.5 mg/L creatinine-adjusted (accounting for Pb co-exposure)
	Valdez Jimenez 2017	infants born from mothers living in endemic hydrofluorosis areas, Mexico (n=65)	maternal urine	mg/L	Bayley Scales of Infant Development II (BSID-II)	Mental Development Index (MDI) score	per 1 mg/L (log F) 1st trimester
							per 1 mg/L (log F) 2nd trimester
Attention, Hyperactivity	Choi 2015	first-grade children, China (n=51)	drinking water	mg/L	finger tapping task	nonpreferred hand test score	log10 F
							normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
						preferred hand test score	log10 F
							normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
			urine	mg/L	finger tapping task	nonpreferred hand test score	log10 F
						preferred hand test score	log10 F
	Wang 2020a	school children 7-13 years old, Tongxu County, China (n=325)	urine	mg/L	Chinese version of the Conners' Parent Rating Scale Revised (CPRS-48)	ADHD Index	per 1 mg/L urinary F
						conduct problems	per 1 mg/L urinary F
						impulsive-hyperactive	per 1 mg/L urinary F
Mood, Affect	Wang 2020a	school children 7-13 years old, Tongxu County, China (n=325)	urine	mg/L	Chinese version of the Conners' Parent Rating Scale Revised (CPRS-48)	anxiety	per 1 mg/L urinary F
						psychosomatic problems	per 1 mg/L urinary F
Attention, Hyperactivity	Bashash 2018	Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants, Mexico (n=210)	maternal urine	mg/L	Conners' Rating Scales-Revised (CRS-R)	ADHD Index	per 0.5 mg/L creatinine adjusted
						DSM-IV ADHD Total Index	per 0.5 mg/L creatinine adjusted
	Riddell 2019	children 6-17 years old, Canadian Health Measures Survey (Cycles 2+3) (n=3740)	city records	fluoridation (Y/N)	Strengths and Difficulties Questionnaire Hyperactivity-Inattention	SDQ H/I subscale score	CWF status
						SDQ H/I subscale score (age 9)	CWF status
						SDQ H/I subscale score (age 14)	CWF status
			drinking water	mg/L	Strengths and Difficulties Questionnaire Hyperactivity-Inattention	SDQ H/I subscale score	per 1-mg/L F increase
						SDQ H/I subscale score (age 9)	per 1-mg/L F increase
						SDQ H/I subscale score (age 14)	per 1-mg/L F increase
			urine	mg/L	Strengths and Difficulties Questionnaire Hyperactivity-Inattention	SDQ H/I subscale score	continuous F measure
Motor/Sensory Function or Development	Choi 2015	first-grade children, China (n=51)	drinking water	mg/L	grooved pegboard test	log10 dominant hand test score	log10 F
							normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
						log10 nondominant hand test score	log10 F
							normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
			urine	mg/L	grooved pegboard test	log10 dominant hand test score	log10 F
						log10 nondominant hand test score	log10 F
Visual-Spatial or Visual-Motor Function	Choi 2015	first-grade children, China (n=51)	drinking water	mg/L	WRAVMA	drawing test score	log10 F
							normal/questionable fluorosis (1 mg/L)
							very mild/mild fluorosis (2.1 mg/L)
							moderate/severe fluorosis (2.66 mg/L)
			urine	mg/L	WRAVMA	drawing test score	log10 F

Interactive figure and additional study details in HAWC here. "F" represents fluoride. Bashash *et al.* (2018) observed significant associations between maternal urinary fluoride and ADHD-like symptoms related to inattention (an increase in 0.5 mg/L of maternal urinary fluoride was associated with a 2.84-point increase in the DSM-IV Inattention Index and a 2.54-point increase in Cognitive Problems and Inattention Index). These two scales contributed to the global ADHD Index and the DSM-IV ADHD Total Index shown here.

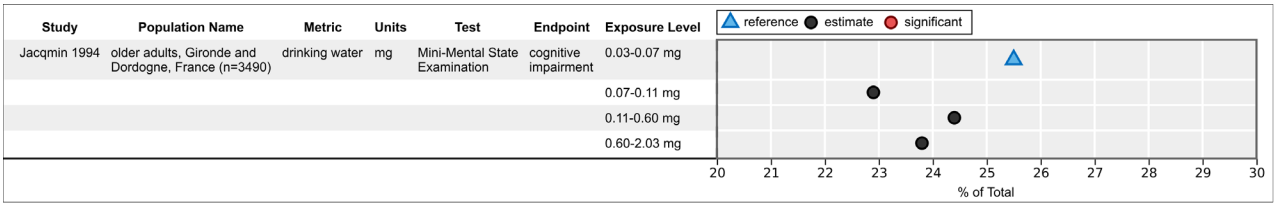
Figure D11. Correlations between Other Neurological Effects and Fluoride Exposure in Children (low risk-of-bias studies; presented as adjusted OR)



Effect	Study	Population Name	Metric	Units	Test	Endpoint	Comparison Set	Exposure Level
Learning, Memory, Intelligence, Cognitive Development	Wang 2020a	school children 7-13 years old, Tongxu County, China (n=325)	urine	mg/L	Chinese version of the Conners' Parent Rating ScaleRevised (CPRS-48)	learning problems	continuous urinary F	per 1 mg/L urinary F
	Barberio 2017b	children 3-12 years old, Canadian Health Measures Survey (Cycle 2) (n=1120)	urine	umol/L		learning disability	Cycle 2 urine	continuous F measure
							Cycle 2 SG-adjusted urine	continuous F measure
				umol/mmol creatinine		learning disability	Cycle 2 creatinine-adjusted urine	continuous F measure
		children 3-12 years old, Canadian Health Measures Survey (Cycle 3) (n=1101)	drinking water	mg/L		learning disability	Cycle 3 drinking water	continuous F measure
			urine	umol/L		learning disability	Cycle 3 urine	continuous F measure
							Cycle 3 SG-adjusted urine	continuous F measure
				umol/mmol creatinine		learning disability	Cycle 3 creatinine-adjusted urine	continuous F measure
		children 3-12 years old, Canadian Health Measures Survey (Cycles 2+3) (n=2221)	urine	umol/L		learning disability	pooled urine	continuous F measure
							pooled SG-adjusted urine	continuous F measure
				umol/mmol creatinine		learning disability	pooled creatinine-adjusted urine	continuous F measure
Attention, Hyperactivity	Wang 2020a	school children 7-13 years old, Tongxu County, China (n=325)	urine	mg/L	Chinese version of the Conners' Parent Rating ScaleRevised (CPRS-48)	ADHD Index	continuous urinary F	per 1 mg/L urinary F
						conduct problems	continuous urinary F	per 1 mg/L urinary F
						impulsive-hyperactive	continuous urinary F	per 1 mg/L urinary F
	Barberio 2017b	children 3-12 years old, Canadian Health Measures Survey (Cycle 2) (n=1120)	urine	umol/L		ADHD	Cycle 2 urine	continuous F measure
							Cycle 2 SG-adjusted urine	continuous F measure
				umol/mmol creatinine		ADHD	Cycle 2 creatinine-adjusted urine	continuous F measure
	Riddell 2019	children 6-17 years old, Canadian Health Measures Survey (Cycles 2+3) (n=3745)	city records	fluoridation (Y/N)		ADHD	Cycle 2/3 CWF	CWF status
						ADHD (age 9)	Cycle 2/3 CWF	CWF status
						ADHD (age 14)	Cycle 2/3 CWF	CWF status
			drinking water	mg/L		ADHD	Cycle 3 drinking water	per 1-mg/L F increase
			urine	mg/L		ADHD	Cycle 2/3 SG-adjusted urine	continuous F measure
Mood, Affect	Wang 2020a	school children 7-13 years old, Tongxu County, China (n=325)	urine	mg/L	Chinese version of the Conners' Parent Rating ScaleRevised (CPRS-48)	anxiety	continuous urinary F	per 1 mg/L urinary F
						psychosomatic problems	continuous urinary F	per 1 mg/L urinary F

Interactive figure and additional study details in HAWC here. "F" represents fluoride. Drinking water results for Barberio *et al.* (2017b) have a large confidence interval and are not completely visible in the figure. 95% CIs are 0.068–11.33 and can be viewed in HAWC by clicking the OR within the plot area.

Figure D12. Cognitive Impairment in Adults by Fluoride Exposure (low risk-of-bias studies; presented as % of total group)



Study	Population Name	Metric	Units	Test	Endpoint	Exposure Level
Jacqmin 1994	older adults, Gironde and Dordogne, France (n=3490)	drinking water	mg	Mini-Mental State Examination	cognitive impairment	0.03-0.07 mg
						0.07-0.11 mg
						0.11-0.60 mg
						0.60-2.03 mg

Interactive figure and additional study details in HAWC here. Results from Li *et al.* (2016) suggested that fluoride exposure may be a risk factor for cognitive impairment in elderly subjects; however, results from the study were not conducive to presentation in this visualization.

ABOUT THIS REVIEW

Sources of Support

National Institute of Environmental Health Sciences/Division of the National Toxicology Program

Contributors

Evaluation Team

The evaluation team is composed of federal staff and contractor staff support.

Name	Affiliation
Kyla Taylor, PhD	NIEHS/DNTP, Project Lead
John Bucher, PhD	NIEHS/DNTP
Andrew Rooney, PhD	NIEHS/DNTP
Vickie Walker	NIEHS/DNTP
Cynthia J. Willson, PhD, DVM, DACVP	ILS
Louise Assem, PhD	ICF [f]
Carlye A. Austin, PhD	ICF [d]
Robyn Blain, PhD	ICF [abcdefg]
Natalie Blanton, MPH	ICF [a]
Kristin Bornstein, PhD	ICF [f]
Canden Byrd	ICF [a]
Michelle Cawley	ICF [c]
Jonathan Cohen, PhD	ICF [g]
Ryan Cronk, PhD	ICF [d]
Sorina E. Eftim, PhD	ICF [fg]
Anna Engstrom, PhD	ICF [abcdf]
Jeremy S. Frye, MLS	ICF [a]
Kaitlin Geary	ICF [c]
Susan Goldhaber, MPH	ICF [e]
Ali Goldstone, MPH	ICF [fg]
Pamela Hartman, MEM	ICF [abceg]
Cara Henning, PhD	ICF [a]
Melinda Hoang, MPH	ICF [d]
Tao Hong	ICF [ef]
Jennifer Hsieh, MSPH	ICF [f]
Penelope Kellar	ICF [a]
Courtney Lemeris	ICF [a]
Cynthia J. Lin, PhD MSPH	ICF [g]
Alex Lindahl, MPH	ICF [f]
Yi Lu, PhD	ICF [f]
Kristen Magnuson, MESM	ICF [abcde]
Rachel McGill	ICF [c]
Shannon McGinnis, PhD	ICF [d]

Name	Affiliation
William Mendez, PhD	ICF [b]
Whitney Mitchell	ICF [a]
Revathi Muralidharan	ICF [de]
Johanna Rochester, PhD	ICF [def]
Amanda Ross, Med	ICF [d]
Pam Ross, MSPH	ICF [df]
Jennifer Seed, PhD	ICF [de]
Karen E. Setty, MESM, PhD	ICF [d]
Codi Sharp	ICF [de]
Robert Shin, MHS	ICF [e]
Kelly Shipkowski, PhD	ICF [df]
Christopher A. Sibrizzi, MPH	ICF [abcdefg]
Raquel A. Silva, PhD	ICF [d]
Samantha J. Snow, PhD, DABT	ICF [ef]
Parnian Soleymani, MS	ICF [d]
Anna Stamatogiannakis	ICF [a]
Nicole Vetter, MLS	ICF [a]
River Williams, BS	ICF [c]

Note: the roles of individual contractors differed: [a] indicates monograph development; [b] indicates review of data, results, and analyses; [c] indicates database and HAWC support; [d] indicates literature screening, [e] indicates data extraction, [f] indicates risk-of-bias assessment, [g] indicates meta-analysis

Peer Reviewers

The peer reviewers were outside experts selected for their experience with fluoride, developmental neurobehavioral toxicity, and systematic review procedures. Peer reviewers were screened for conflict of interest prior to their service and did not report any conflicts of interest. Service as a peer reviewer does not necessarily indicate that the reviewer endorses the final document.

Protocol Reviewers

Name	Affiliation
Joseph Braun, PhD [a]	Brown University
Marie Sutton, PhD [a]	Health Research Board
Thomas Zoeller, PhD [a]	University of Massachusetts, Amherst
Thomas Webster, PhD [a]	Boston University
Gail Wasserman, PhD [a]	Columbia University
Suril Mehta, MPH [b]	NIEHS
Tianjing Li, PhD [b]	University of Colorado Denver

[a] Reviewer of the systematic review protocol; [b] Reviewer of meta-analysis protocol
no conflicts of interest declared

Technical Review of Draft Monograph

Name	Affiliation
Freya Kamal, PhD	NIEHS (retired)

no conflicts of interest declared

Protocol History and Revisions

Date	Activity or revision
December 14, 2016	**Draft evaluation protocol reviewed:** sent to technical advisors for peer review
April 10, 2017	**Draft human risk-of-bias protocol reviewed;** sent to technical advisors for peer review
May 2, 2017	**Draft animal risk-of-bias protocol reviewed;** sent to technical advisors for peer review
June 2017	**Evaluation protocol finalized:** Review protocol finalized for use and posting
May 2019	**Revised protocol:** Revised review protocol posted
September 2020	**Revised protocol:** Revised review protocol posted

APPENDICES

Appendix 1. Literature Search Strategy

The strategy for this search is broad for the consideration of neurodevelopmental or cognitive endpoints and comprehensive for fluoride as an exposure or treatment in order to ensure inclusion of relevant papers. The search terms for PubMed are provided below. The specific search strategies for other databases are available in the protocol (https://ntp.niehs.nih.gov/go/785076).

Database	Search Terms
PUBMED	((Fluorides[mh:noexp] OR fluorides, topical[mh] OR sodium fluoride[mh] OR Fluorosis, Dental[mh] OR fluorosis[tiab] OR fluorid*[tiab] OR flurid*[tiab] OR fluorin*[tiab] OR florin*[tiab]) NOT (18F[tiab] OR f-18[tiab] OR 19F[tiab] OR f-19[tiab] OR f-labeled[tiab] OR "fluorine-18"[tiab] OR "fluorine-19"[tiab] OR pet-scan[tiab] OR radioligand*[tiab])) AND ((Aryl Hydrocarbon Hydroxylases[mh] OR Aryl Hydrocarbon Receptor Nuclear Translocator[mh] OR Behavior and Behavior Mechanisms[mh] OR Gene Expression Regulation[mh] OR Glucuronosyltransferase[mh] OR Intelligence tests[mh] OR Malate Dehydrogenase[mh] OR Mediator Complex Subunit 1[mh] OR Mental disorders[mh] OR Mental processes[mh] OR Monocarboxylic Acid Transporters[mh] OR Myelin Basic Protein[mh] OR nervous system[mh] OR nervous system diseases[mh] OR nervous system physiological phenomena[mh] OR Neurogranin[mh] OR Oligodendroglia[mh] OR Peroxisome Proliferator-Activated Receptors[mh] OR Psychological Phenomena and Processes[mh] OR Receptors, thyroid hormone[mh] OR Receptors, thyrotropin[mh] OR Retinoid X Receptors[mh] OR thyroid diseases[mh] OR thyroid hormones[mh] OR Thyrotropin-releasing hormone[mh] OR Thyroxine-Binding Proteins[mh] OR Pregnane X Receptor[supplementary concept] OR thyroid-hormone-receptor interacting protein[supplementary concept] OR Constitutive androstane receptor[supplementary concept] OR Academic performance[tiab] OR auditory[tiab] OR cortical[tiab] OR delayed development[tiab] OR developmental impairment[tiab] OR developmental-delay*[tiab] OR developmental-disorder*[tiab] OR euthyroid[tiab] OR gait[tiab] OR glia*[tiab] OR gliogenesis[tiab] OR hyperactiv*[tiab] OR impulse-control[tiab] OR iodide peroxidase[tiab] OR IQ[tiab] OR ischemi*[tiab] OR locomotor[tiab] OR mental deficiency[tiab] OR mental development[tiab] OR mental illness[tiab] OR mental-deficit[tiab] OR mobility[tiab] OR mood[tiab] OR morris-maze[tiab] OR morris-water[tiab] OR motor abilit*[tiab] OR Motor activities[tiab] OR motor performance[tiab] OR nerve[tiab] OR neural[tiab] OR neurobehav*[tiab] OR Neurocognitive impairment[tiab] OR neurodegenerat*[tiab] OR Neurodevelopment*[tiab] OR neurodisease*[tiab] OR neurologic*[tiab] OR neuromuscular[tiab] OR neuron*[tiab] OR neuropath*[tiab] OR obsessive compulsive[tiab] OR OCD[tiab] OR olfaction[tiab] OR olfactory[tiab] OR open-field-test[tiab] OR passive avoidance[tiab] OR plasticity[tiab] OR senil*[tiab] OR sociab*[tiab] OR speech*[tiab] OR spelling[tiab] OR stereotypic-movement*[tiab] OR synap*[tiab] OR tauopath*[tiab] OR Thyroglobulin[tiab] OR Thyroid disease*[tiab] OR thyroid gland[tiab] OR thyroid hormone*[tiab] OR thyronine*[tiab] OR visual motor[tiab] OR Visuospatial processing[tiab] OR water maze[tiab]) OR ((active-avoidance[tiab] OR ADHD[tiab] OR alzheimer*[tiab] OR amygdala[tiab] OR antisocial[tiab] OR anxiety[tiab] OR anxious[tiab] OR asperger*[tiab] OR attention deficit[tiab] OR autism[tiab] OR autistic[tiab] OR behavioral[tiab] OR behaviors[tiab] OR behavioural[tiab] OR behaviours[tiab] OR bipolar[tiab] OR cerebellum[tiab] OR cognition[tiab] OR cognitive[tiab] OR communication-disorder*[tiab] OR comprehension[tiab] OR cranial[tiab] OR dementia[tiab] OR dendrit*[tiab] OR dentate-gyrus[tiab] OR depression[tiab] OR dextrothyroxine[tiab] OR diiodothyronine*[tiab] OR diiodotyrosine[tiab] OR down syndrome[tiab] OR dyslexia[tiab] OR entorhinal cortex[tiab] OR epilep*[tiab] OR gangli*[tiab] OR goiter[tiab] OR graves-disease[tiab] OR hearing[tiab] OR hippocamp*[tiab] OR human development[tiab] OR hyperthyroid*[tiab] OR hypothalam*[tiab] OR hypothyroid*[tiab] OR impulsiv*[tiab] OR Intellectual disability[tiab] OR intelligence[tiab] OR language[tiab] OR learning[tiab] OR lewy bod*[tiab] OR

Database	Search Terms
	long-term potentiation[tiab] OR long-term synaptic depression[tiab] OR memory[tiab] OR mental disorder*[tiab] OR mental recall[tiab] OR monoiodotyrosine[tiab] OR Motor activity[tiab] OR motor skill*[tiab] OR multiple sclerosis[tiab] OR myxedema[tiab] OR Nervous system[tiab] OR nervous-system[tiab] OR neurit*[tiab] OR optic[tiab] OR palsy[tiab] OR panic[tiab] OR parahippocamp*[tiab] OR paranoia[tiab] OR paranoid[tiab] OR parkinson*[tiab] OR perception[tiab] OR perforant*[tiab] OR personality[tiab] OR phobia[tiab] OR problem solving[tiab] OR proprioception[tiab] OR psychomotor[tiab] OR reflex[tiab] OR risk taking[tiab] OR schizophrenia[tiab] OR seizure*[tiab] OR sensation*[tiab] OR sleep[tiab] OR smell[tiab] OR spatial behavior[tiab] OR stroke[tiab] OR substantia-nigra[tiab] OR taste[tiab] OR thyroiditis[tiab] OR thyrotoxicosis[tiab] OR Thyrotropin[tiab] OR thyroxine[tiab] OR triiodothyronine[tiab] OR vision[tiab]) NOT medline[sb]))

Appendix 2. Detailed Literature Search Results and List of Included Studies

Detailed Literature Search Results

Literature Search Results Counts and Title and Abstract Screening

The electronic database searches retrieved 25,522 unique references in total (20,883 references during the initial search conducted in December 2016, 3,733 references during the literature search updates [including the final updated search conducted for the primary epidemiology studies on May 1, 2020], and 906 references from the supplemental Chinese database searches); 15 additional references were identified by technical advisors or from reviewing reference lists in published reviews and included studies. As a result of title and abstract screening, 1,036 references were moved to full-text review, and 24,501 references were excluded (11,478 by manual screening for not satisfying the PECO criteria and 13,023 based on the SWIFT algorithm).

Full-text Review

Among the 1,036 references that underwent full-text review, 497 references were excluded during the full-text review with reasons for exclusion documented at this stage; 337 references were excluded for not satisfying the PECO criteria; and 160 references from the May 2020 searches (main literature search update and supplemental Chinese database searches) were excluded for not including information that would materially advance the human, animal in vivo, or mechanistic findings (see the **Literature Search** section for a description of the methodology). These screening results are outlined in a study selection diagram that reports numbers of studies excluded for each reason at the full text review stage (see **Figure 2**) [using reporting practices outlined in Moher *et al.* (2009)]. After full-text review, 539 studies were considered relevant with primary neurological outcomes, secondary neurological outcomes, and/or outcomes related to thyroid function. A few studies assessed data for more than one evidence stream (human, non-human mammal, and/or in vitro), and several human and animal studies assessed more than one type of outcome (e.g., primary and secondary outcomes). The number of included studies is summarized below. There are:

- 159 human studies (78 primary only; 13 secondary only; 5 primary and secondary; 6 primary and thyroid; 2 secondary and thyroid; and 55 thyroid only);
- 339 non-human mammal studies (7 primary only; 186 secondary only; 67 primary and secondary; 6 primary, secondary, and thyroid; 4 secondary and thyroid; and 69 thyroid only); and,
- 60 in vitro/mechanistic studies (48 neurological and 12 thyroid).

One publication contained human, experimental non-human mammal, and in vitro data. Three publications contained both human and experimental non-human mammal data. Fourteen publications contained data relevant to both experimental non-human mammal studies and in vitro studies.

List of Included Studies

Studies in Humans

As described in **Figure 2**, 159 human studies were included; however, full data extraction was only conducted on studies with neurological outcomes or thyroid hormone data. Data extraction was completed using HAWC. Data were extracted from a subset of included studies in humans (n = 116) and are available in HAWC based on outcome. The following lists of references are organized as studies that are available in HAWC followed by studies that are not available in HAWC. Specifically, data for primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence) and secondary neurobehavioral outcomes (anxiety, aggression, motor activity, or biochemical changes), as well as thyroid hormone level data, were extracted from included human studies and are available in HAWC. Data for included studies identified through the 2020 literature search update were only extracted for primary neurodevelopmental or cognitive outcomes; a subset of these studies (n = 5) also included secondary neurobehavioral outcomes and/or thyroid hormone level data that were not extracted because those data would not materially advance the human or mechanistic findings. Included human studies that only evaluated other thyroid-related effects such as goiters or thyroid size (n = 43) were not extracted and are not available in HAWC. The list below presents the 159 human studies that were included in the review. An overview of the screening results is outlined in the study selection diagram (**Figure 2**) that reports numbers of included studies as well as numbers of studies excluded for each reason at the full text review stage.

Studies Available in HAWC

An J, Mei S, Liu A, Fu Y, Wang C. 1992. [Effect of high level of fluoride on children's intelligence]. *Chin J Control Endem Dis* 7(2): 93-94.

Aravind A, Dhanya RS, Narayan A, Sam G, Adarsh VJ, Kiran M. 2016. Effect of fluoridated water on intelligence in 10-12-year-old school children. *J Int Soc Prev Community Dent* 6(Suppl 3): S237-S242.

Bai A, Li Y, Fan Z, Li X, Li P. 2014. [Intelligence and growth development of children in coal-burning-borne arsenism and fluorosis areas: An investigation study]. *Chin J Endemiol* 33(2): 160-163.

Barberio AM, Hosein FS, Quinonez C, McLaren L. 2017. Fluoride exposure and indicators of thyroid functioning in the Canadian population: Implications for community water fluoridation. *J Epidemiol Community Health* 71: 1019-1025.

Barberio AM, Quinonez C, Hosein FS, McLaren L. 2017. Fluoride exposure and reported learning disability diagnosis among Canadian children: Implications for community water fluoridation. *Can J Public Health* 108: 229-239.

Bashash M, Thomas D, Hu H, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Ettinger AS, Wright R, Zhang Z, Liu Y, Schnaas L, Mercado-Garcia A, Tellez-Rojo MM, Hernandez-Avila M. 2017. Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6-12 years of age in Mexico. *Environ Health Perspect* 125(9): 1-12.

Bashash M, Marchand M, Hu H, Till C, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Green R, Schnaas L, Mercado-Garcia A, Hernandez-Avila M, Tellez-Rojo MM. 2018. Prenatal fluoride exposure and attention deficit hyperactivity disorder (ADHD) symptoms in children at 6-12 years of age in Mexico City. *Environ Int* 121(Pt 1): 658-666.

Broadbent JM, Thomson WM, Moffitt TE, Poulton R. 2015. Community water fluoridation and intelligence response. *Am J Public Health* 105: 3-4.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 1991. [Research on the intellectual development of children in high fluoride areas]. *Chin J Control Endem Dis* 6(Suppl): 99-100.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 2008. Research on the intellectual development of children in high fluoride areas. *Fluoride* 41: 120-124.

Chinoy NJ, Narayana MV. 1992. Studies on fluorosis in Mehsana District of North Gujarat. *Proc Zool Soc* 45: 157-161.

Choi AL, Zhang Y, Sun G, Bellinger DC, Wang K, Yang XJ, Li JS, Zheng Q, Fu Y, Grandjean P. 2015. Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study. *Neurotoxicol Teratol* 47: 96-101.

Cui Y, Zhang B, Ma J, Wang Y, Zhao L, Hou C, Yu J, Zhao Y, Zhang Z, Nie J, Gao T, Zhou G, Liu H. 2018. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. *Ecotoxicol Environ Saf* 165: 270-277.

Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. 2020. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. *Neurosci Lett* 729: 134981.

Das K, Mondal NK. 2016. Dental fluorosis and urinary fluoride concentration as a reflection of fluoride exposure and its impact on IQ level and BMI of children of Laxmisagar, Simlapal Block of Bankura District, W.B., India. *Environ Monit Assess* 188: 218.

Ding Y, Sun H, Han H, Wang W, Ji X, Liu X, Sun D. 2011. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. *J Hazard Mater* 186: 1942-1946.

Du L, Wan C, Cao X, Liu J. 1992. [The effect of fluorine on the developing human brain]. *Chin J Pathol* 21(4): 218-220.

Du L, Wan C, Cao X, Liu J. 2008. The effect of fluorine on the developing human brain. *Fluoride* 41: 327-330.

Duan J, Zhao M, Wang L, Fang D, Wang Y, Wang W. 1995. A comparative analysis of the results of multiple tests in patients with chronic industrial fluorosis. Guizhou Med J 18(3): 179-180.
Erickson JD, Hay S. 1976. Water fluoridation and congenital malformations: No association. *J Am Dent Assoc* 93: 981-984.

Erickson JD, Hay S. 1976. Water fluoridation and congenital malformations: No association. *J Am Dent Assoc* 93: 981-984.

Erickson JD. 1980. Down syndrome, water fluoridation, and maternal age. *Teratology* 21: 177-180.

Eswar P, Nagesh L, Devaraj CG. 2011. Intelligent quotients of 12-14 year old school children in a high and low fluoride village in India. *Fluoride* 44: 168-172.

Fan Z, Dai H, Bai A, Li P, Li T, Li G. 2007. Effect of high fluoride exposure in children's intelligence. *J Environ Health* 24(10): 802-803.

Green R, Lanphear B, Hornung R, Flora D, Martinez-Mier EA, Neufeld R, Ayotte P, Muckle G, Till C. 2019. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. *JAMA Pediatr*: E1-E9.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 1991. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. *Chin J Epidemiol* 10(2): 98-100.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 2008. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. *Fluoride* 41: 125-128.

Guo ZY, He YH, Zhu QX. 2001. [Research on the neurobehavioral function of workers occupationally exposed to fluoride]. *Ind Hlth & Occup Dis* 27(6): 346-348.

Guo ZY, He YH, Zhu QX. 2008. Research on the neurobehavioral function of workers occupationally exposed to fluoride. *Fluoride* 41: 152-155.

He H, Cheng ZS, Liu WQ. 1989. [Effects of fluorine on the human fetus]. *J Control Endem Dis* 4(3): 136-138.

He H, Cheng ZS, Liu WQ. 2008. Effects of fluorine on the human fetus. *Fluoride* 41: 321-326.

He MX, Zhang CN. 2010. [Investigation of children's intelligence quotient and dental fluorosis in drinking water-type of endemic fluorosis area in Pucheng County, Shaanxi Province before and after drinking water change]. *Chin J Endemiol* 29: 547-548.

Hong FG, Cao YX, Yang D, Wang H. 2001. [Research on the effects of fluoride on child intellectual development under different environmental conditions]. *Chin Prim Health Care* 15(3): 56-57.

Hong FG, Cao YX, Yang D, Wang H. 2008. Research on the effects of fluoride on child intellectual development under different environmental conditions. *Fluoride* 41: 156-160.

Hosur MB, Puranik RS, Vanaki S, Puranik SR. 2012. Study of thyroid hormones free triiodothyronine (FT3), free thyroxine (FT4) and thyroid stimulating hormone (TSH) in subjects with dental fluorosis. *Eur J Dent* 6: 184-190.

Jacqmin H, Commenges D, Letenneur L, Barberger-Gateau P, Dartigues JF. 1994. Components of drinking water and risk of cognitive impairment in the elderly. *Am J Epidemiol* 139: 48-57.

Kang J, Cheng Y, Wu K, Lin S, He G, Jin Y. 2011. Effect of exposure to fluoride and arsenic in drinking water of Hangjinhouqi on children's intelligence. *Chinese School Health*: 679-681.

Karimzade S, Aghaei M, Mahvi AH. 2014. Investigation of intelligence quotient in 9-12 year-old children exposed to high- and low-drinking water fluoride in West Azerbaijan Province, Iran. *Fluoride* 47: 9-14.

Khan SA, Singh RK, Navit S, Chadha D, Johri N, Navit P, Sharma A, Bahuguna R. 2015. Relationship between dental fluorosis and intelligence quotient of school going children in and around Lucknow District: A cross-sectional study. *J Clin Diagn Res* 9(11): 10-15.

Khandare AL, Gourineni SR, Validandi V. 2017. Dental fluorosis, nutritional status, kidney damage, and thyroid function along with bone metabolic indicators in school-going children living in fluoride-affected hilly areas of Doda District, Jammu and Kashmir, India. *Environ Monit Assess* 189: 579.

Khandare AL, Validandi V, Gourineni SR, Gopalan V, Nagalla B. 2018. Dose-dependent effect of fluoride on clinical and subclinical indices of fluorosis in school going children and its mitigation by supply of safe drinking water for 5 years: An Indian study. *Environ Monit Assess* 190: 110.

Kheradpisheh Z, Mahvi AH, Mirzaei M, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018. Correlation between drinking water fluoride and TSH hormone by ANNs and ANFIS. *J Environ Health Sci Eng* 16(1): 11-18.

Kheradpisheh Z, Mirzaei M, Mahvi AH, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018. Impact of drinking water fluoride on human thyroid hormones: A case-control study. *Sci Rep* 8: 2674.

Kumar V, Chahar P, Kajjari S, Rahman F, Bansal DK, Kapadia JM. 2018. Fluoride, thyroid hormone derangements and its correlation with tooth eruption pattern among the pediatric population from endemic and non-endemic fluorosis areas. *J Contemp Dent Pract* 19(12): 1512-1516.

Kundu H, Basavaraj P, Singla A, Gupta R, Singh K, Jain S. 2015. Effect of fluoride in drinking water on children's intelligence in high and low fluoride areas of Delhi. *J Indian Assoc Public Health Dent* 13(2): 116-121.

Lamberg M, Hausen H, Vartiainen T. 1997. Symptoms experienced during periods of actual and supposed water fluoridation. *Community Dent Oral Epidemiol* 25: 291-295.

Li F, Chen X, Huang R, Xie Y. 2009. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. *J Environ Health* 26(4): 838-840.

Li J, Yao L, Shao QL, Wu CY. 2004. [Effects of high fluoride level on neonatal neurobehavioral development]. *Chin J Endemiol* 23(5): 463-465.

Li J, Yao L, Shao QL, Wu CY. 2008. Effects of high fluoride level on neonatal neurobehavioral development. *Fluoride* 41: 165-170.

Li M, Gao Y, Cui J, Li Y, Li B, Liu Y, Sun J, Liu X, Liu H, Zhao L, Sun D. 2016. Cognitive impairment and risk factors in elderly people living in fluorosis areas in China. *Biol Trace Elem Res* 172: 53-60.

Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. 2010. Investigation and analysis of children’s intelligence and dental fluorosis in high fluoride area. *J Med Pest Control* 26(3): 230-231.

Li XS, Zhi JL, Gao RO. 1995. Effect of fluoride exposure on the intelligence of children. *Fluoride* 28: 189-192.

Li Y, Li X, Wei S. 1994. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. *J West Chin University Med Sci* 25(2): 188-191.

Li Y, Li X, Wei S. 2008. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. *Fluoride* 41: 331-335.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2003. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. *Chin J Public Health Manag* 19(4): 337-338.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2008. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. *Fluoride* 41: 161-164.

Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. 1991. High fluoride and low iodine environment and subclinical cretinism in Xinjiang. *Endem Dis Bull* 6(2): 62-67.

Liu S, Lu Y, Sun Z, Wu L, Wang X, Yan S. 2000. [Report on the intellectual ability of children living in high-fluoride water areas]. *Chin J Control Endem Dis* 15(4): 231-232.

Liu SL, Lu Y, Sun ZR, Wu L, Lu WL, Wang XW, Yan S. 2008. Report on the intellectual ability of children living in high-fluoride water areas. *Fluoride* 41: 144-147.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S, Mentes A, DenBesten PK. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. *Physiol Behav* 206: 76-83.

Lu Y, Sun ZR, Wu LN, Wang X, Lu W, Liu SS. 2000. Effect of high-fluoride water on intelligence in children. *Fluoride* 33: 74-78.

Malin AJ, Till C. 2015. Exposure to fluoridated water and attention deficit hyperactivity disorder prevalence among children and adolescents in the United States: An ecological association. *Environ Health* 14: 17.

Malin AJ, Riddell J, McCague H, Till C. 2018. Fluoride exposure and thyroid function among adults living in Canada: Effect modification by iodine status. *Environ Int* 121(Pt 1): 667-674.

Michael M, Barot VV, Chinoy NJ. 1996. Investigations of soft tissue functions in fluorotic individuals of North Gujarat. *Fluoride* 29: 63-71.

Mondal D, Dutta G, Gupta S. 2016. Inferring the fluoride hydrogeochemistry and effect of consuming fluoride-contaminated drinking water on human health in some endemic areas of Birbhum District, West Bengal. *Environ Geochem Health* 38: 557-576.

Morgan L, Allred E, Tavares M, Bellinger D, Needleman H. 1998. Investigation of the possible associations between fluorosis, fluoride exposure, and childhood behavior problems. *Pediatr Dent* 20: 244-252.

Mustafa DE, Younis UM, Elhag SAA. 2018. The relationship between the fluoride levels in drinking water and the schooling performance of children in rural areas of Khartoum State, Sudan. *Fluoride* 51(2): 102-113.

Nagarajappa R, Pujara P, Sharda AJ, Asawa K, Tak M, Aapaliya P, Bhanushali N. 2013. Comparative assessment of intelligence quotient among children living in high and low fluoride areas of Kutch, India: A pilot study. *Iran J Public Health* 42: 813-818.

Peckham S, Lowery D, Spencer S. 2015. Are fluoride levels in drinking water associated with hypothyroidism prevalence in England? A large observational study of GP practice data and fluoride levels in drinking water. *J Epidemiol Community Health* 69: 619-624.

Poureslami HR, Horri A, Garrusi B. 2011. A comparative study of the IQ of children age 7-9 in a high and a low fluoride water city in Iran. *Fluoride* 44: 163-167.

Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. 1990. [Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children]. *J Control Endem Dis* 5(4): 203-204.

Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. 2008. Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children. *Fluoride* 41: 115-119.

Ratan SK, Rattan KN, Pandey RM, Singhal S, Kharab S, Bala M, Singh V, Jhanwar A. 2008. Evaluation of the levels of folate, vitamin B12, homocysteine and fluoride in the parents and the affected neonates with neural tube defect and their matched controls. *Pediatr Surg Int* 24: 803-808.

Razdan P, Patthi B, Kumar JK, Agnihotri N, Chaudhari P, Prasad M. 2017. Effect of fluoride concentration in drinking water on intelligence quotient of 12-14-year-old children in Mathura District: A cross-sectional study. *J Int Soc Prev Community Dent* 7: 252-258.

Ren D, Li K, Liu D. 1989. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. *Chin J Control Endem Dis* 4(4): 251.

Ren D, Li K, Liu D. 2008. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. *Fluoride* 41: 319-320.

Riddell JK, Malin AJ, Flora D, McCague H, Till C. 2019. Association of water fluoride and urinary fluoride concentrations with attention deficit hyperactivity disorder in Canadian youth. *Environ Int* 133: 105190.

Rocha-Amador D, Navarro ME, Carrizales L, Morales R, Calderon J. 2007. Decreased intelligence in children and exposure to fluoride and arsenic in drinking water. *Cad Saude Publica* 23 Suppl 4: S579-587.

Rocha-Amador D, Navarro M, Trejo-Acevedo A, Carrizales L, Perez-Maldonado I, Diaz-Barriga F, Calderon J. 2009. Use of the Rey-Osterrieth Complex Figure Test for neurotoxicity evaluation of mixtures in children. *Neurotox* 30: 1149-1154.

Rotton J, Tikofsky RS, Feldman HT. 1982. Behavioral effects of chemicals in drinking water. *J Appl Psychol* 67: 230-238.

Russ TC, Killin LOJ, Hannah J, Batty GD, Deary IJ, Starr JM. 2019. Aluminium and fluoride in drinking water in relation to later dementia risk. *Brit J Psychol* 14: 1-6.

Saxena S, Sahay A, Goel P. 2012. Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India. *J Neurosci Rural Pract* 3: 144-149.

Sebastian ST, Sunitha S. 2015. A cross-sectional study to assess the intelligence quotient (IQ) of school going children aged 10-12 years in villages of Mysore District, India with different fluoride levels. *J Indian Soc Pedod Prev Dent* 33: 307-311.

Seraj B, Shahrabi M, Falahzade M, Falahzade F, Akhondi N. 2006. Effect of high fluoride concentration in drinking water on children's intelligence. *J Dent Med* 19(2): 80-86.

Seraj B, Shahrabi M, Shadfar M, Ahmadi R, Fallahzadeh M, Eslamlu HF, Kharazifard MJ. 2012. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. *J Dent* 9: 221-229.

Shannon FT, Fergusson DM, Horwood LJ. 1986. Exposure to fluoridated public water supplies and child health and behaviour. *New Zeal Med J* 99: 416-418.

Shao Q. 2003. Study of cognitive function impairment caused by chronic fluorosis. *Chin J Endemiol* 22(4): 336-338.

Sharma JD, Sohu D, Jain P. 2009. Prevalence of neurological manifestations in a human population exposed to fluoride in drinking water. *Fluoride* 42: 127-132.

Shivaprakash PK, Ohri K, Noorani H. 2011. Relation between dental fluorosis and intelligence quotient in school children of Bagalkot District. *J Indian Soc Pedod Prev Dent* 29: 117-120.

Singh A, Jolly SS, Devi P, Bansal BC, Singh SS. 1962. Endemic fluorosis: An epidemiological, biochemical and clinical study in the Bhatinda District of Panjab. *Indian J Med Res* 50: 387-398.

Singh N, Verma KG, Verma P, Sidhu GK, Sachdeva S. 2014. A comparative study of fluoride ingestion levels, serum thyroid hormone & TSH level derangements, dental fluorosis status among school children from endemic and non-endemic fluorosis areas. *Springerplus* 3: 7.

Singh V, Singh C, Sandeep T, Sandeep K, Vikas G, Mukesh T, Anurag T. 2013. A correlation between serum vitamin, acetylcholinesterase activity and IQ in children with excessive endemic fluoride exposure in Rajasthan, India. *Int Res J Medical Sci* 1(3): 12-16.

Soto-Barreras U, Escalante-Villalobos KY, Holguin-Loya B, Perez-Aguirre B, Nevarez-Rascon A, Martinez-Martinez RE, Loyola-Rodriguez JP. 2019. Effect of fluoride in drinking water on dental caries and IQ in children. *Fluoride* 52: 474-482.

Still CN, Kelley P. 1980. The incidence of primary degenerative dementia vs. water fluoride content in South Carolina, USA. *Neurotox* 1: 125-132.

Sudhir KM, Chandu GN, Prashant GM, Subba Reddy VV. 2009. Effect of fluoride exposure on intelligence quotient (IQ) among 13-15 year old school children of known endemic area of fluorosis, Nalgonda District, Andhra Pradesh. *J Indian Assoc Public Health Dent* 2009(13): 88-94.

Sun M, Li S, Wang Y, Li F. 1991. Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis. *J Guiyang Med Coll* 16(3): 204-206.

Susheela AK, Bhatnagar M, Vig K, Mondal NK. 2005. Excess fluoride ingestion and thyroid hormone derangements in children living in Delhi, India. *Fluoride* 38: 98-108.

Tamboli BL, Mathur GM, Mathur AP, Lalla SK, Goyal OP. 1980. Prevalence of fluorosis in Pratabpura and Surajpura villages, District Ajmer (Rajasthan). *Indian J Med Res* 71: 57-67.

Till C, Green R, Flora D, Hornung R, Martinez-Mier EA, Blazer M, Farmus L, Ayotte P, Muckle G, Lanphear B. 2020. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. *Environ Int* 134: 105315.

Tripathi P, Sultana N. 2007. Fluoride content of groundwater and prevalence of dental, skeletal and neurological stage of fluorosis in Tehsil Purwa of Unnao. *Indian J Environ Prot* 27: 737-739.

Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. 2007. Effect of high fluoride water on intelligence of school children in India. *Fluoride* 40: 178-183.

Trivedi M, Sangai N, Patel R, Payak M, Vyas S. 2012. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. *Fluoride* 45(4): 377-383.

Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, Costilla-Salazar R, Calderon Hernandez J, Alcaraz Contreras Y, Rocha-Amador DO. 2017. In utero exposure to fluoride and cognitive development delay in infants. *Neurotox* 59: 65-70.

Wang A, Duan L, Huang H, Ma J, Zhang Y, Ma Q, Guo Y, Li Z, Cheng X, Zhu J, Zhou G, Ba Y. 2020. Association between fluoride exposure and behavioural outcomes of school-age children: A pilot study in China. *Int J Environ Health Res*: 1-10.

Wang G, Yang D, Jia F, Wang H. 1996. [A study of the IQ levels of four- to seven-year-old children in high fluoride areas]. *Endem Dis Bull* 11(1): 60-62.

Wang G, Yang D, Jia F, Wang H. 2008. A study of the IQ levels of four- to seven-year-old children in high fluoride areas. *Fluoride* 41: 340-343.

Wang G, Gao M, Zhang M, Yang M, Xiang Q. 2012. Correlation between total fluoride intake and children's IQ. *J Southeast Univ Med Ed*: 743-746.

Wang M, Liu L, Li H, Li Y, Liu H, Hou C, Zeng Q, Li P, Zhao Q, Dong L, Zhou G, Yu X, Liu L, Guan Q, Zhang S, Wang A. 2020. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. *Environ Int* 134: 105229.

Wang S, Wang L, Hu P, Guo S, Law S. 2001. [Effects of high iodine and high fluorine on children's intelligence and thyroid function]. *Chin J Endemiol* 20(4): 288-290.

Wang SX, Wang ZH, Cheng XT, Li J, Sang ZP, Zhang XD, Han LL, Qiao XY, Wu ZM, Wang ZQ. 2005. [Investigation and evaluation on intelligence and growth of children in endemic fluorosis and arsenism areas]. *Chin J Endemiol* 24: 179-182.

Wang SX, Wang ZH, Cheng XT, Li J, Sang ZP, Zhang XD, Han LL, Qiao XY, Wu ZM, Wang ZQ. 2007. Arsenic and fluoride exposure in drinking water: Children's IQ and growth in Shanyin County, Shanxi Province, China. *Environ Health Perspect* 115: 643-647.

Wang SY, Zhang HX, Fan W, Fang SJ, Kang PP, Chen XG, Yu MJ. 2005. [The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children]. *J Appl Clin Ped* 20(9): 897-899.

Wang SY, Zhang HX, Fan W, Fang SJ, Kang PP, Chen XG, Yu MJ. 2008. The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children. *Fluoride* 41: 344-348.

Wang Z, Wang S, Zhang X, Li J, Zheng X, Hu C. 2006. Investigation of children's growth and development under long-term fluoride exposure. *Chin J Control Endem Dis* 21(4): 239-241.

Wei N, Li Y, Deng J, Xu S, Guan Z. 2014. [The effects of comprehensive control measures on intelligence of school-age children in coal-burning-borne endemic fluorosis areas]. *Chin J Endemiol* 33(2): 320-322.

Xiang Q, Liang Y, Chen L, Wang C, Chen B, Chen X, Zhou M. 2003. Effect of fluoride in drinking water on children's intelligence. *Fluoride* 36: 84-94.

Xiang Q, Liang Y, Chen B, Chen L. 2011. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. *Fluoride* 44: 191-194.

Xu Y, Lu C, Zhang X. 1994. Effect of fluoride on children's intelligence. *Endem Dis Bull* 2: 83-84.

Yang Y, Wang X, Guo X, Hu P. 1994. [The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine]. *Chin J Epidemiol* 15(4): 296-298.

Yang Y, Wang X, Guo X, Hu P. 2008. The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine. *Fluoride* 41: 336-339.

Yao L, Zhou J, Wang S, Cui K, Lin F. 1996. Analysis of TSH levels and intelligence of children residing in high fluorosis areas. *Lit Inf Prev Med* 2(1): 26-27.

Yao Y. 1997. Comparative assessment of the physical and mental development of children in endemic fluorosis area with water improvement and without water improvement. *Lit Inf Prev Med* 3(1): 42-43.

Yasmin S, Ranjan S, D'Souza D. 2013. Effect of excess fluoride ingestion on human thyroid function in Gaya Region, Bihar, India. *Toxicol Environ Chem* 95: 1235-1243.

Yazdi SM, Sharifian A, Dehghani-Beshne M, Momeni VR, Aminian O. 2011. Effects of fluoride on psychomotor performance and memory of aluminum potroom workers. *Fluoride* 44: 158-162.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 1996. [Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis]. *Chin J Endemiol* 15(5): 257-259.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 2008. Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis. *Fluoride* 41: 134-138.

Yu X, Chen J, Li Y, Liu H, Hou C, Zeng Q, Cui Y, Zhao L, Li P, Zhou Z, Pang S, Tang S, Tian K, Zhao Q, Dong L, Xu C, Zhang X, Zhang S, Liu L, Wang A. 2018. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. *Environ Int* 118: 116-124.

Zhang J, Yao H, Chen Y. 1998. [Effect of high level of fluoride and arsenic on children's intelligence]. *Chin J Public Health* 17(2): 57.

Zhang S, Zhang X, Liu H, Qu W, Guan Z, Zeng Q, Jiang C, Gao H, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Cui Y, Yu L, Wang A. 2015. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. *Toxicol Sci* 144: 238-245.

Zhao LB, Liang GH, Zhang DN, Wu XR. 1996. Effect of a high fluoride water supply on children's intelligence. *Fluoride* 29: 190-192.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao Q, Tian Z, Zhou G, Niu Q, Chen J, Li P, Dong L, Xia T, Zhang S, Wang A. 2020. SIRT1-dependent mitochondrial biogenesis supports therapeutic effects of resveratrol against neurodevelopment damage by fluoride. *Theranostics* 10: 4822-4838.

Zhou G, Tang S, Yang L, Niu Q, Chen J, Xia T, Wang S, Wang M, Zhao Q, Liu L, Li P, Dong L, Yang K, Zhang S, Wang A. 2019. Effects of long-term fluoride exposure on cognitive ability and the underlying mechanisms: Role of autophagy and its association with apoptosis. *Toxicol Appl Pharmacol 378*: 114608.

Studies Not Available in HAWC

Bachinskii PP, Gutsalenko OA, Naryzhniuk ND, Sidora VD, Shliakhta AI. 1985. [Action of the body fluorine of healthy persons and thyroidopathy patients on the function of hypophyseal-thyroid the system]. *Probl Endokrinol* 31: 25-29.

Balabolkin MI, Mikhailets ND, Lobovskaia RN, Chernousova NV. 1995. [The interrelationship of the thyroid and immune statuses of workers with long-term fluorine exposure]. *Ter Arkh* 67: 41-42.

Baum K, Boerner W, Reiners C, Moll E. 1981. [Bone density and thyroid function in adolescents in relation to fluoride content of drinking water]. *Fortschr Med* 99: 1470-1472.

Berry WTC, Whittles JH. 1963. Absence of effect of fluoride upon the incidence of thyroid enlargements in Wiltshire schoolgirls. *Mon Bull Minist Health Public Health Lab Serv* 22: 50-52.

Cherkinskii SN, Zaslavskaia RM. 1956. [Significance of fluorides in potable water in the development of endemic goiter]. *Probl Endokrinol Gormonoter* 2: 70-75.

Choubisa SL. 2001. Endemic fluorosis in southern Rajasthan, India. *Fluoride* 34: 61-70.

Chuka A, Zhukovskil V, Mirku I, Postel'Niku D. 1964. Prezhdevremennoe starenie naseleniya v zone rasprostraneniya endemicheskogo zoba. *Vestnik Akad Med Nauk Sssr* 19: 23-27.

Dai HX, Zeng P, Wang KY, Zhang XG, Ma ZJ, Zhou YG, Fan ZX, Guo SH. 2013. [Analysis of a survey results of patients with suspected high iodine goiter in Liuji Town Fuping County of Shaanxi Province]. *Chin J Endemiol* 32: 408-411.

Day T, Powell-Jackson P. 1972. Fluoride, water hardness, and endemic goitre. *Lancet* 299(7761): 1135-1138.

Desai VK, Solanki DM, Bansal RK. 1993. Epidemiological study of goitre in endemic fluorosis district of Gujarat. *Fluoride* 26: 187-190.

Díaz-Cadórniga FJ, Delgado E, Tartón T, Valdés MM, Méndez A, Fernández MT, Rojo C. 2003. Endemic goiter associated with high iodine intake in primary school children in the Saharawi Arab Democratic Republic. *Endocrinol Nutr* 50: 357-362.

Eichner R, Borner W, Henschler D, Kohler W, Moll E. 1981. [Osteoporosis therapy and thyroid function. Influence of 6 months of sodium fluoride treatment on thyroid function and bone density]. *Fortschr Med* 99: 342-348.

Fiorentini S, Galeazzi M, Visintin B. 1947. II fluoro in natura come agente morbigeno II. La fluorosi die Campagnano di Roma. III. Un focolaio di fluorosi umana a Campagnano di Roma. IV. Osservazioni radiologiche sui processi alveolari, sulle ossa mascellari, e sul paradenzio degli abitanti die Campagnano. V. Zona fluorotica intorno a Campagnano di Roma. VI. Frequenza e caratteri clinici della carie dentale in soggetti fluorotici. *Rend Ist Superiore Sanita* 10: 721-804.

Gabovich RD, Verzhikovskaia NV. 1958. [Effect of fluoride compounds on absorption of radioactive iodine by the thyroid gland in humans and in experimental conditions]. *Probl Endokrinol Gormonoter* 4: 49-54.

Galletti P, Held HR, Korrodi H, Wegmann T. 1956. [Investigations on the possible damaging effect of fluorine on the thyroid gland]. *Helv Med Acta* 23: 601-605.

Galletti PM, Joyet G, Jallut O. 1957. [Effect of sodium fluoride on thyroid function in Basedow's Disease]. *Helv Med Acta* 24: 209-215.

Galletti PM, Joyet G. 1958. Effect of fluorine on thyroidal iodine metabolism in hyperthyroidism. *J Clin Endocrinol Metab* 18: 1102-1110.

Gas'kov AI, Savchenkov MF, Iushkov NN. 2005. [The specific features of the development of iodine deficiencies in children living under environmental pollution with fluorine compounds]. *Gig Sanit*: 53-55.

Gedalia I, Brand N. 1963. The relationship of fluoride and iodine in drinking water in the occurrence of goiter. *Arch Int Pharmacodyn Ther* 142: 312-315.

Grimm H. 1973. [The physical development of schoolchildren under the influence of drinking water fluoridation in Karl Marx Stadt]. *Dtsch Gesundheitsw* 28: 2363-2369.

Hasling C, Nielsen HE, Melsen F, Mosekilde L. 1987. Safety of osteoporosis treatment with sodium fluoride, calcium phosphate and vitamin D. *Miner Electrolyte Metab* 13: 96-103.

Hidehiko T. 1958. On the relation between the distribution of endemic goiter and the fluorine content of natural water in Hidaka Province, Hokkaido. *Folia Pharmacol Jpn* 54: 225-229.

Hoffmann-Axthelm W. 1953. [Observations on the influence of fluorine on dental enamel and thyroid gland]. *Dtsch Zahnarztl Z* 8: 757-765.

Jentzer A. 1956. [Effect of fluorine on the iodine content of the human thyroid gland]. *Bull Schweiz Akad Med Wiss* 12: 539-543.

Jooste PL, Weight MJ, Kriek JA, Louw AJ. 1999. Endemic goitre in the absence of iodine deficiency in schoolchildren of the Northern Cape Province of South Africa. *Eur J Clin Nutr* 53: 8-12.

Kolomiitseva MG. 1961. [The content of fluorine in the external environment of the Upper Altai autonomous region and its role in the etiology of endemic goiter]. *Gig Sanit* 26: 101-103.

Korrodi H, Wegmann T, Galleti P, Held HR. 1955. [Caries prophylaxis and the untoward effects of fluor on the thyroid gland]. *Schweiz Med Wochenschr* 85: 1016-1019.

Kutlucan A, Kale Koroglu B, Numan Tamer M, Aydin Y, Baltaci D, Akdogan M, Ozturk M, Vural H, Ermis F. 2013. The investigation of effects of fluorosis on thyroid volume in school-age children. *Med Glas* 10: 93-98.

Latham MC, Grech P. 1967. The effects of excessive fluoride intake. *Am J Public Health* 57: 651-660.

Leone NC, Leatherwood EC, Petrie IM, Lieberman L. 1964. Effect of fluoride on thyroid gland: Clinical study. *J Am Dent Assoc* 69: 179-180.

Levi JE, Silberstein HE. 1955. Lack of effect of fluorine ingestion on uptake of iodine 131 by the thyroid gland. *J Lab Clin Med* 45: 348-351.

McGlashan N, Chelkowska E, Sasananan S. 2010. A survey of goiter morbidity in Ban Mae Toen, northwest Thailand. *Southeast Asian J Trop Med Public Health* 41: 1200-1208.

Rathore S, Meena C, Gonmei Z, Dwivedi S, Toteja GS, Bala K. 2018. Study of excess fluoride ingestion and thyroid hormone derangement in relation with different fluoride levels in drinking water among children of Jodhpur District, Rajasthan, India. *Asian J Microbiol Biotechnol Environ Sci* 20: 327-331.

Reisenauer R, Rezler D, Křemenová J, Preininger Q. 1961. [Fluorization of the waters in Czechoslovakia. IV. Endocrinological control of results of two years' fluorization of drinking-water in school children]. *Cesk Stomatol* 61: 91-97.

Romer TEZ, Kowalczyk B, Kacprzak M, Wiktorowski M. 1976. [Incidence of goiter in pubertal girls of the Piotrkow Region and iodide content in drinking water]. *Endokrynol Pol* 27: 373-380.

Savchenkov MF, Efimova NV, Manueva RS, Nikolaeva LA, Shin NS. 2016. [Thyroid gland pathology in children population exposed to the combination of iodine deficiency and fluoride pollution of environment]. *Gig Sanit* 95: 1201-1205.

Shtifanova AK. 1962. [The fluorine content in water, soil and vegetal products of the Alma-Atinsk District areas and its role in the etiology of dental caries and endemic goiter]. *Zdravookhranenie Kazakhstana*: 60-63.

Siddiqui AH. 1969. Incidence of simple goiter in areas of endemic fluorosis in Nalgonda District, Andhra Pradesh, India. *Fluoride* 2: 200-205.

Sidora VD, Shliakhta AI, Iugov VK, Kas'ianenko AS, Piatenko VG. 1983. [Indices of the pituitary-thyroid system in residents of cities with various fluorine concentrations in drinking water]. *Probl Endokrinol* 29: 32-35.

Sung FC, Chen KP, Chen CY, Tai PW, Yang CF. 1973. Studies of the effect of salt iodization on endemic goiter in Taiwan. IV. A survey of drinking water in relation to endemic goiter. *J Fomosan Med Assoc* 72: 96-103.

Tokar VI, Voroshnin VV, Sherbakov SV. 1989. [Chronic effects of fluorides on the pituitary-thyroid system in industrial workers]. *Gig Tr Prof Zabol*: 19-22.

Wespi HJ. 1954. [Iodized-fluoridized salt for the prevention of goiter and caries]. *Schweiz Med Wochenschr* 84: 885-890.

Yu YN. 1985. [Effects of chronic fluorosis on the thyroid gland]. *Chin Med J* 65: 747-7479.

Studies in Non-human Animals

As described in **Figure 2**, 339 non-human mammal studies were included; however, full data extraction was only conducted on studies with primary neurological outcomes and/or secondary functional neurological outcomes (e.g., motor activity). Data extraction was completed using HAWC. Data were extracted from a subset of included studies in animals (n = 123) and are available in HAWC based on outcome. The following lists of references are organized as studies that are available in HAWC followed by studies that are not available in HAWC. Specifically, all primary outcomes and functional neurological secondary outcomes (e.g., motor activity) were extracted from animal studies and are available in HAWC, including studies from the NTP (2016) assessment. Studies are also available in HAWC that evaluated mechanistic effects related to oral fluoride exposure at or below 20 ppm fluoride drinking water equivalents for categories of mechanistic endpoints with the largest amount of available data (i.e., biochemistry of the brain or neurons, neurotransmission, oxidative stress, and histopathology [n = 70]); however, these mechanistic data were generally not extracted. Several animal studies assessed primary neurological outcomes and/or functional neurological secondary outcomes and mechanistic effects in the four mechanistic categories listed above (n = 56). In total, 140 animal studies are available in HAWC (70 with primary neurological outcomes and/or secondary functional neurological outcomes without relevant mechanistic data; 15 with relevant mechanistic data only; and 55 with primary/or secondary functional neurological outcomes with relevant mechanistic data). Studies that evaluated other mechanistic endpoints, as well as studies that only assessed mechanistic effects at fluoride levels above 20 ppm fluoride drinking water equivalents, are not available in HAWC (n = 199). The list below presents the 339 non-human animal studies that were included in the review. An overview of the screening results is outlined in the study selection diagram (**Figure 2**) that reports numbers of included studies as well as numbers of studies excluded for each reason at the full text review stage.

Studies Available in HAWC

Adedara IA, Abolaji AO, Idris UF, Olabiyi BF, Onibiyo EM, Ojuade TD, Farombi EO. 2017. Neuroprotective influence of taurine on fluoride-induced biochemical and behavioral deficits in rats. *Chem Biol Interact* 261: 1-10.

Adedara IA, Olabiyi BF, Ojuade TD, Idris UF, Onibiyo EM, Farombi EO. 2017. Taurine reverses sodium fluoride-mediated increase in inflammation, caspase-3 activity, and oxidative damage along the brain-pituitary-gonadal axis in male rats. *Can J Phys Pharmacol* 95: 1019-1029.

Agustina F, Sofro ZM, Partadiredja G. 2018. Subchronic administration of high-dose sodium fluoride causes deficits in cerebellar purkinje cells but not motor coordination of rats. *Biol Trace Elem Res* 188(2): 424-433.

Ahmad KR, Noor S, Jabeen S, Nauroze T, Kanwal MA, Raees K, Abbas T. 2017. Amelioration by jambul fruit extract of fluoride-induced hepato-nephronal histopathologies and impaired neuromotor capacity in mice. *Fluoride* 50: 2-14.

Akinrinade ID, Memudu AE, Ogundele OM. 2015. Fluoride and aluminium disturb neuronal morphology, transport functions, cholinesterase, lysosomal and cell cycle activities. *Pathophysiology* 22: 105-115.

Akinrinade ID, Memudu AE, Ogundele OM, Ajetunmobi OI. 2015. Interplay of glia activation and oxidative stress formation in fluoride and aluminium exposure. *Pathophysiology* 22: 39-48.

Baba NA, Raina R, Verma PK, Sultana M. 2014. Alterations in plasma and tissue acetylcholinesterase activity following repeated oral exposure of chlorpyrifos alone and in conjunction with fluoride in Wistar rats. *Proc Indian Natl Sci Acad B Biol Sci* 84: 969-972.

Bagmut I, Kolisnyk I, Titkova A, Babiy L, Filipchenko S. 2018. The antioxidant system enzymes' activity in rats' brain, intoxicated with sodium fluoride in subtoxic doses. *Arch Balkan Med Union* 53(4): 506-511.

Balaji B, Kumar EP, Kumar A. 2015. Evaluation of standardized bacopa monniera extract in sodium fluoride-induced behavioural, biochemical, and histopathological alterations in mice. *Toxicol Ind Health* 31: 18-30.

Balayssac D, Richard D, Authier N, Nicolay A, Jourdan D, Eschalier A, Coudore F. 2002. Absence of painful neuropathy after chronic oral fluoride intake in Sprague-Dawley and Lou/C rats. *Neurosci Lett* 327: 169-172.

Banala RR, Karnati PR. 2015. Vitamin A deficiency: An oxidative stress marker in sodium fluoride (NaF) induced oxidative damage in developing rat brain. *Int J Dev Neurosci* 47: 298-303.

Banala R, Nagapuri K, Mohd K, Reddy M, Karnati P. 2018. Carica Papaya leaf extract as a neuroprotective agent against behavioral and neurotransmitter changes in brain of the rat treated with sodium fluoride in pre- and post-natal periods. *Pharmacogn Mag* 14(55): 123-131.

Banji D, Banji OJ, Pratusha NG, Annamalai AR. 2013. Investigation on the role of spirulina platensis in ameliorating behavioural changes, thyroid dysfunction and oxidative stress in offspring of pregnant rats exposed to fluoride. *Food Chem* 140: 321-331.

Baran-Poesina V, Negres S, Dobrescu D, Dimcevici-Poesina N, Dimcevici-Poesina A, Feghiu A, Soare T, Militaru M. 2013. Experimental pharmacological researches regarding the influence of sodium fluoride in allopathic and homeopathic doses on central nervous system's performances: A correlation between behavioral response in classic maze test and morphological aspects of cerebral cortex. *Farmacia* 61: 781-799.

Bartos M, Gumilar F, Bras C, Gallegos CE, Giannuzzi L, Cancela LM, Minetti A. 2015. Neurobehavioural effects of exposure to fluoride in the earliest stages of rat development. *Physiol Behav* 147: 205-212.

Bartos M, Gumilar F, Gallegos CE, Bras C, Dominguez S, Monaco N, Esandi MDC, Bouzat C, Cancela LM, Minetti A. 2018. Alterations in the memory of rat offspring exposed to low levels of fluoride during gestation and lactation: Involvement of the alpha7 nicotinic receptor and oxidative stress. *Reprod Toxicol* 81: 108-114.

Basha PM, Rai P, Begum S. 2011. Fluoride toxicity and status of serum thyroid hormones, brain histopathology, and learning memory in rats: A multigenerational assessment. *Biol Trace Elem Res* 144: 1083-1094.

Basha PM, Sujitha NS. 2012. Combined impact of exercise and temperature in learning and memory performance of fluoride toxicated rats. *Biol Trace Elem Res* 150: 306-313.

Bataineh HN, Nusier MK. 2006. Impact of 12-week ingestion of sodium fluoride on aggression, sexual behavior, and fertility in adult male rats. *Fluoride* 39: 293-301.

Bera I, Sabatini R, Auteri P, Flace P, Sisto G, Montagnani M, Potenza MA, Marasciulo FL, Carratu MR, Coluccia A, Borracci P, Tarullo A, Cagiano R. 2007. Neurofunctional effects of developmental sodium fluoride exposure in rats. *Eur Rev Med Pharmacol Sci* 11: 211-224.

Bhatnagar M, Rao P, Sushma J, Bhatnagar R. 2002. Neurotoxicity of fluoride: Neurodegeneration in hippocampus of female mice. *Indian J Exp Biol* 40: 546-554.

Bhatnagar M, Sukhwal P, Suhalka P, Jain A, Joshi C, Sharma D. 2011. Effects of fluoride in drinking water on NADPH-diaphorase neurons in the forebrain of mice: A possible mechanism of fluoride neurotoxicity. *Fluoride* 44: 195-209.

Chen H, Geng D. 2011. [The change of cognition induced by chronic fluoride in rats]. *Acta Academiae Medicinae Xuzhou* 31(5): 319-322.

Chen J, Niu Q, Xia T, Zhou G, Li P, Zhao Q, Xu C, Dong L, Zhang S, Wang A. 2018. ERK1/2-mediated disruption of BDNF-TrkB signaling causes synaptic impairment contributing to fluoride-induced developmental neurotoxicity. *Toxicology* 410: 222-230.

Chinoy NJ, Shah SD. 2004. Biochemical effects of sodium fluoride and arsenic trioxide toxicity and their reversal in the brain of mice. *Fluoride* 37: 80-87.

Chioca LR, Raupp IM, Da Cunha C, Losso EM, Andreatini R. 2008. Subchronic fluoride intake induces impairment in habituation and active avoidance tasks in rats. *Eur J Pharmacol* 579: 196-201.

Chouhan S, Flora SJS. 2008. Effects of fluoride on the tissue oxidative stress and apoptosis in rats: Biochemical assays supported by IR spectroscopy data. *Toxicology* 254: 61-67.

Chouhan S, Lomash V, Flora SJ. 2010. Fluoride-induced changes in haem biosynthesis pathway, neurological variables and tissue histopathology of rats. *J Appl Toxicol* 30: 63-73.

Cui YS, Zhong Q, Li WF, Liu ZH, Wang Y, Hou CC. 2017. [Effects of fluoride exposure on thyroid hormone level and intelligence in rats]. *Chin J Ind Hyg Occup Dis* 35: 888-892.

Dabrowska E. 1997. Effect of different fluorine doses on the supraoptic nucleus of the rat. *Folia Histochem Cytobiol* 35: 115-116.

Dong Y, Wang Y, Wei N, Guan Z. 2015. [Expression levels of brain muscarinic acetylcholine receptor in offspring rats of drinking-water borne fluorosis]. *Chin J Endemiol* 34: 326-330.

Dong YT, Wang Y, Wei N, Zhang QF, Guan ZZ. 2015. Deficit in learning and memory of rats with chronic fluorosis correlates with the decreased expressions of M1 and M3 muscarinic acetylcholine receptors. *Arch Toxicol* 89: 1981-1991.

Dong YT, Wei N, Qi XL, Liu XH, Chen D, Zeng XX, Guan ZZ. 2017. Attenuating effect of vitamin E on the deficit of learning and memory of rats with chronic fluorosis: The mechanism may involve muscarinic acetylcholine receptors. *Fluoride* 50: 354-364.

Dong YW, Y. Wei, N. Guan, Z. 2015. [Expression of muscarinic acetylcholine receptors in the brain of rats with chronic fluorosis]. *Chin J Endemiol* 34(2): 84-88.

Ekambaram P, Paul V. 2001. Calcium preventing locomotor behavioral and dental toxicities of fluoride by decreasing serum fluoride level in rats. *Environ Toxicol Pharmacol* 9: 141-146.

Ekambaram P, Paul V. 2002. Modulation of fluoride toxicity in rats by calcium carbonate and by withdrawal of fluoride exposure. *Pharmacol Toxicol* 90: 53-58.

Ekambaram P, Paul V. 2003. Effect of vitamin D on chronic behavioral and dental toxicities of sodium fluoride in rats. *Fluoride* 36: 189-197.

El-lethey HS, Kamel MM, Shaheed IB. 2010. Neurobehavioral toxicity produced by sodium fluoride in drinking water of laboratory rats. *J Am Sci* 6(5): 54-63.

El-lethey HS, Kamel MM. 2011. Effects of black tea in mitigation of sodium fluoride potency to suppress motor activity and coordination in laboratory rats. *J Am Sci* 7(4): 243-254.

El-lethey HS, Shaheed IB. 2011. Potential health impact of black tea against Na-F-induced alterations in territorial aggression, sexual behaviour and fertility of male rats. *Life Sci J* 8: 828-839.

Elliott L. 1967. Lack of effect of administration of fluoride on the central nervous system of rats. *Acta Pharmacol Toxicol (Copenh)* 25: 323-328.

Flace P, Benagiano V, Vermesan D, Sabatini R, Inchingolo AM, Auteri P, Ambrosi G, Tarullo A, Cagiano R. 2010. Effects of developmental fluoride exposure on rat ultrasonic vocalization, acoustic startle reflex and pre-pulse inhibition. *Eur Rev Med Pharmacol Sci* 14: 507-512.

Gabovich RD. 1962. [On the problem of the effect of fluorine in drinking water on the functional state of the central nervous system]. *Gig Sanit* 27: 10-12.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008. [Level of oxidative stress in rat brains and learning and memory function of rats with chronic fluorosis]. *Chin J Endemiol* 27: 371-373.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008. [Effects of fluoride on learning and memory and cholinesterase activity in rat brains]. *Chin J Endemiol* 27: 128-130.

Gao Q, Liu YJ, Guan ZZ. 2009. Decreased learning and memory ability in rats with fluorosis: Increased oxidative stress and reduced cholinesterase activity in the brain. *Fluoride* 42: 277-285.

Gao Y, Liu L, Young L, Huan L, Jin H. 2009. Effects of learning and memory of fluoride and the antagonism of selenium in rats. *Studies of Trace Elements and Health* 26(2): 1-3.

Ge QD, Tan Y, Luo Y, Wang WJ, Zhang H, Xie C. 2018. MiR-132, miR-204 and BDNF-TrkB signaling pathway may be involved in spatial learning and memory impairment of the offspring rats caused by fluorine and aluminum exposure during the embryonic stage and into adulthood. *Environ Toxicol Pharmacol* 63: 60-68.

Ge Y, Chen L, Yin Z, Song X, Ruan T, Hua L, Liu J, Wang J, Ning H. 2018. Fluoride-induced alterations of synapse-related proteins in the cerebral cortex of ICR offspring mouse brain. *Chemosphere* 201: 874-883.

Gopal K, Saxena R, Gupta GSD, Rana MD, Agrawal D. 2006. Fluoride induced alterations in neurobehavioural and cardiovascular responses in rats. *J Adv Zool* 27: 1-7.

Gui CZ, Ran LY, Wu CX, Long YG, He J, Zhang H, Guan ZZ. 2009. [Changes in learning and memory ability and brain cholinesterase activity in the rats with coal burning fluorosis]. *Chin J Endemiol* 28: 497-500.

Gui CZ, Ran LY, Li JP, Guan ZZ. 2010. Changes of learning and memory ability and brain nicotinic receptors of rat offspring with coal burning fluorosis. *Neurotoxicol Teratol* 32: 536-541.

Gui CZ, Ran LY, Guan ZZ. 2011. [Expression levels of brain nicotinic acetylcholine receptor mRNA and protein in coal-burning type of fluorosis rats]. *Chin J Endemiol* 30: 239-242.

Guner S, Uyar-Bozkurt S, Haznedaroglu E, Mentes A. 2016. Dental fluorosis and catalase immunoreactivity of the brain tissues in rats exposed to high fluoride pre- and postnatally. *Biol Trace Elem Res* 174: 150-157.

Han H, Du W, Zhou B, Zhang W, Xu G, Niu R, Sun Z. 2014. Effects of chronic fluoride exposure on object recognition memory and mRNA expression of SNARE complex in hippocampus of male mice. *Biol Trace Elem Res* 158: 58-64.

Hong JH, Ge YM, Ning HM, Wang JD. 2005. [Effects of High Fluoride and Low Iodine on Learning-Memory and TchE of Brain in Offspring Rats]. *Chin Prev Med* 6: 489-491.

Inkielewicz I, Krechniak J. 2004. Fluoride effects on glutathione peroxidase and lipid peroxidation in rats. *Fluoride* 37: 7-12.

Jain A, Mehta VK, Chittora RA, Mahdi A, Bhatnagar M. 2015. Melatonin ameliorates fluoride induced neurotoxicity in young rats: An in vivo evidence. *Asian J Pharm Clin Res* 8: 164-167.

Jetti R, Raghuveer CV, Mallikarjuna RC. 2016. Protective effect of ascorbic acid and Ginkgo biloba against learning and memory deficits caused by fluoride. *Toxicol Ind Health* 32: 183-187.

Jia B, Zong L, Lee JY, Lei J, Zhu Y, Xie H, Clemens JL, Feller MC, Na Q, Dong J, McLane MW, Jones-Beatty K, Burd I. 2019. Maternal supplementation of low dose fluoride alleviates adverse perinatal outcomes following exposure to intrauterine inflammation. *Sci Rep* 9(1): 2575.

Jiang C, Zhang S, Liu H, Guan Z, Zeng Q, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Zhang X, Cui Y, Yu L, Wang A. 2014. Low glucose utilization and neurodegenerative changes caused by sodium fluoride exposure in rat's developmental brain. *Neuromolecular Med* 16: 94-105.

Jiang S, Su J, Yao S, Zhang Y, Cao F, Wang F, Wang H, Li J, Xi S. 2014. Fluoride and arsenic exposure impairs learning and memory and decreases mGluR5 expression in the hippocampus and cortex in rats. *PLoS One* 9(4): e96041.

Khan AM, Raina R, Dubey N, Verma PK. 2017. Effect of deltamethrin and fluoride co-exposure on the brain antioxidant status and cholinesterase activity in Wistar rats. *Drug Chem Toxicol* 41(2): 1-5.

Kinawy AA, Al-Eidan AA. 2018. Impact of prenatal and postnatal treatment of sodium fluoride and aluminum chloride on some hormonal and sensorimotor aspects in rats. *Biol Trace Elem Res*: 1-8.

Kivrak Y. 2012. Effects of fluoride on anxiety and depression in mice. *Fluoride* 45: 302-306.

Li M, Cui J, Gao YH, Zhang W, Sun LY, Liu XN, Liu Y, Sun DJ. 2015. Pathological changes and effect on the learning and memory ability in rats exposed to fluoride and aluminum. *Toxicol Res* 4: 1366-1373.

Li X, Zhang J, Niu R, Manthari RK, Yang K, Wang J. 2019. Effect of fluoride exposure on anxiety- and depression-like behavior in mouse. *Chemosphere* 215: 454-460.

Liu F, Ma J, Zhang H, Liu P, Liu YP, Xing B, Dang YH. 2014. Fluoride exposure during development affects both cognition and emotion in mice. *Physiol Behav* 124: 1-7.

Liu WX. 1989. [Experimental study of behavior and cerebral morphology of rat pups generated by fluorotic female rat]. *Chin J Pathol* 18: 290-292.

Liu YJ, Gao Q, Wu CX, Long YG, Guan ZZ. 2009. [Modified expression of extracellular signal-regulated protein kinase signal transduction in rat brains and changed capacity of learning and memory of rats with chronic fluorosis]. *Chin J Endemiol* 28: 32-35.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. Alterations of nAChRs and ERK1/2 in the brains of rats with chronic fluorosis and their connections with the decreased capacity of learning and memory. *Toxicol Lett* 192: 324-329.

Liu YJ, Gao Q, Long YG, Yu YN, Guan ZZ. 2011. [Influence of chronic fluorosis on expression of phospho-Elk-1 in rat brains]. *Chin J Endemiol* 30: 251-255.

Lou DD, Guan ZZ, Liu YJ, Liu YF, Zhang KL, Pan JG, Pei JJ. 2013. The influence of chronic fluorosis on mitochondrial dynamics morphology and distribution in cortical neurons of the rat brain. *Arch Toxicol* 87: 449-457.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S, Mentes A, DenBesten PK. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. *Physiol Behav* 206: 76-83.

Ma J, Liu F, Liu P, Dong YY, Chu Z, Hou TZ, Dang YH. 2015. Impact of early developmental fluoride exposure on the peripheral pain sensitivity in mice. *Int J Dev Neurosci* 47: 165-171.

Manusha S, Sudhakar K, Reddy KP. 2019. Protective effects of allium sativum extract against sodium fluoride induced neurotoxicity. *Int J Pharm Sci Res* 10(2): 625-633.

McPherson CA, Zhang G, Gilliam R, Brar SS, Wilson R, Brix A, Picut C, Harry GJ. 2018. An evaluation of neurotoxicity following fluoride exposure from gestational through adult ages in Long-Evans hooded rats. *Neurotoxicol Res*: 1-18.

Mesram N, Nagapuri K, Banala RR, Nalagoni CR, Karnati PR. 2016. Quercetin treatment against NaF induced oxidative stress related neuronal and learning changes in developing rats. *J King Saud Univ Sci* 29: 221-229.

Mullenix PJ, Denbesten PK, Schunior A, Kernan WJ. 1995. Neurotoxicity of sodium fluoride in rats. *Neurotoxicol Teratol* 17: 169-177.

Nageshwar M, Sudhakar K, Reddy NCC, Reddy KP. 2017. Neuroprotective effects of curcumin on sodium fluoride induced behavioural and enzymatic changes in brain and muscles of rat. *J Environ Biol* 38: 675-681.

Nageshwar M, Sudhakar K, Reddy KP. 2018. Quercetin ameliorates oxidative stress, neural damage of brain and behavioral impairment of rat with fluoride exposure. *Int J Pharm Sci Res* 9(8): 3247-3256.

Nian W, Wang X, Shao D, Yu Q, Ouyang W, Zhang Z, Ruan Q. 2018. Effects of subchronic exposure to fluorine on hippocampal injury in mice and its molecular mechanism. *Acta Sci Circumst* 38(11): 4512-4519.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. *Environ Pollut* 233: 889-899.

Niu R, Sun Z, Wang J, Cheng Z. 2008. Effects of fluoride and lead on locomotor behavior and expression of nissl body in brain of adult rats. *Fluoride* 41: 276-282.

Niu R, Sun Z, Cheng Z, Li Z, Wang J. 2009. Decreased learning ability and low hippocampus glutamate in offspring rats exposed to fluoride and lead. *Environ Toxicol Pharmacol* 28: 254-258.

Niu R, Liu S, Wang J, Zhang J, Sun Z. 2014. Proteomic analysis of hippocampus in offspring male mice exposed to fluoride and lead. *Biol Trace Elem Res* 162: 227-233.

Niu R, Xue X, Zhao Y, Sun Z, Yan X, Li X, Feng C, Wang J. 2015. Effects of fluoride on microtubule ultrastructure and expression of Tubalpha1a and Tubbeta2a in mouse hippocampus. *Chemosphere* 139: 422-427.

Niu R, Chen H, Manthari RK, Sun Z, Wang J, Zhang J, Wang J. 2018. Effects of fluoride on synapse morphology and myelin damage in mouse hippocampus. *Chemosphere* 194: 628-633.

Nkpaa KW, Onyeso GI. 2018. Rutin attenuates neurobehavioral deficits, oxidative stress, neuro-inflammation and apoptosis in fluoride treated rats. *Neurosci Lett* 682: 92-99.

Paul V, Ekambaram P, Jayakumar AR. 1998. Effects of sodium fluoride on locomotor behavior and a few biochemical parameters in rats. *Environ Toxicol Pharmacol* 6: 187-191.

Pereira M, Dombrowski PA, Losso EM, Chioca LR, Da Cunha C, Andreatini R. 2011. Memory impairment induced by sodium fluoride is associated with changes in brain monoamine levels. *Neurotoxicol Res* 19: 55-62.

Pulungan ZS, Sofro ZM, Partadiredja G. 2016. Sodium fluoride does not affect the working memory and number of pyramidal cells in rat medial prefrontal cortex. *Anat Sci Int* 93(1): 128-138.

Raghu J, Raghuveer VC, Rao MC, Somayaji NS, Babu PB. 2013. The ameliorative effect of ascorbic acid and Ginkgo biloba on learning and memory deficits associated with fluoride exposure. *Interdiscip Toxicol* 6: 217-221.

Raju S, Sivanesan SK, Gudemalla K. 2019. Cognitive enhancement effect of Ginkgo biloba extract on memory and learning impairments induced by fluoride neurotoxicity. *Int J Res Pharm Sci* 10(1): 129-134.

Reddy MM, Karnati PR. 2015. Protective effects of aqueous extract of fruit pulp of tamarindus indica on motor activity and metabolism of the gastrocnemius muscle of rats treated with fluoride. *Int J Toxicol Pharmacol Res* 7: 241-246.

Reddy YP, Tiwari SK, Shaik AP, Alsaeed A, Sultana A, Reddy PK. 2014. Effect of sodium fluoride on neuroimmunological parameters, oxidative stress and antioxidative defenses. *Toxicol Mech Methods* 24: 31-36.

Rumiantsev GI, Novikov SM, Mel'nikova NN, Levchenko NI, Kozeeva EE. 1988. [Experimental study of the biological effect of salts of hydrofluosilicic acid]. *Gig Sanit*: 80-82.

Sarkozi K, Horvath E, Vezer T, Papp A, Paulik E. 2015. Behavioral and general effects of subacute oral arsenic exposure in rats with and without fluoride. *Int J Environ Health Res* 25: 418-431.

Shah SD, Chinoy NJ. 2004. Adverse effects of fluoride and/or arsenic on the cerebral hemisphere of mice and recovery by some antidotes. *Fluoride* 37: 162-171.

Shalini B, Sharma JD. 2015. Beneficial effects of emblica officinalis on fluoride-induced toxicity on brain biochemical indexes and learning-memory in rats. *Toxicol Int* 22: 35-39.

Shan KR, Qi XL, Long YG, Nordberg A, Guan ZZ. 2004. Decreased nicotinic receptors in PC12 cells and rat brains influenced by fluoride toxicity: A mechanism relating to a damage at the level in post-transcription of the receptor genes. *Toxicology* 200: 169-177.

Sharma C, Suhalka P, Bhatnagar M. 2018. Curcumin and resveratrol rescue cortical-hippocampal system from chronic fluoride-induced neurodegeneration and enhance memory retrieval. *Int J Neurosci*: 1-15.

Shen X, Zhang Z, Xu X. 2004. [Effect of iodine and selenium on learning memory impairment induced by fluorosis and blood biochemical criterion of rats]. *Occupation and Health* 20(1): 6-8.

Sudhakar K, Nageshwar M, Pratap Reddy K. 2017. Seed extract of Abelmoschus moschatus medik reverses NAF-induced behavioral changes through neurodegeneration and oxidative stress in brain of rat. *Asian J Pharm Clin Res* 10: 165-171.

Sudhakar K, Nageshwar M, Reddy KP. 2018. Protective effect of okra, Abelmoschus moschatus seed extract on developing brain of rats during pre- and post-natal fluoride exposure. *Int J Pharm Sci Res* 9: 1519-1528.

Sudhakar K, Nageshwar M, Reddy KP. 2018. Abelmoschus moschatus extract reverses altered pain and neurohistology of a rat with developmental exposure of fluoride. *J Appl Pharm Sci* 8(6): 94-104.

Sudhakar K, Reddy KP. 2018. Protective effects of Abelmoschus moschatus seed extract on neurotransmitter system of developing brain of Wistar rats with gestational and post-natal exposure of sodium fluoride. *Int J Green Pharm* 12: S131-S139.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2000. [Effects of high fluoride drinking water on the cerebral function of mice]. *Chin J Epidemiol* 19: 262-263.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2008. Effects of high fluoride drinking water on the cerebral function of mice. *Fluoride* 41: 148-151.

Sun Z, Zhang Y, Xue X, Niu R, Wang J. 2018. Maternal fluoride exposure during gestation and lactation decreased learning and memory ability, and glutamate receptor mRNA expressions of mouse pups. *Hum Exp Toxicol* 37: 87-93.

Trivedi MH, Verma RJ, Chinoy NJ. 2007. Amelioration by black tea of sodium fluoride-induced changes in protein content of cerebral hemisphere, cerebellum and medulla oblongata in brain region of mice. *Acta Poloniae Pharm* 64: 221-225.

Trivedi MH, Verma RJ, Chinoy NJ. 2009. Mitigation of sodium fluoride induced toxicity in mice brain by black tea infusion. *Fluoride* 42: 29-33.

Trivedi MH, Verma RJ, Sangai NP, Chinoy NJ. 2011. Black tea extract mitigation of NaF-induced lipid peroxidation in different regions of mice brains. *Fluoride* 44: 243-254.

Trivedi MH, Verma RJ, Sangai NP, Chinoy NJ. 2012. Mitigation by black tea extract of sodium fluoride induced histopathological changes in brain of mice. *Fluoride* 45: 13-26.

Tsunoda M, Aizawa Y, Nakano K, Liu Y, Horiuchi T, Itai K, Tsunoda H. 2005. Changes in fluoride levels in the liver, kidney, and brain and in neurotransmitters of mice after subacute administration of fluoride. *Fluoride* 38: 284-292.

Varner JA, Jensen KF, Horvath W, Isaacson RL. 1998. Chronic administration of aluminum-fluoride or sodium-fluoride to rats in drinking water: Alterations in neuronal and cerebrovascular integrity. *Brain Res* 784(1-2): 284-298.

Verma RJ, Trivedi MH, Chinoy NJ. 2007. Black tea amelioration of sodium fluoride-induced alterations of DNA, RNA, and protein contents in the cerebral hemisphere, cerebellum, and medulla oblongata regions of mouse brain. *Fluoride* 40: 7-12.

Wang G, Li J, Zhu H, Zhu J. 2006. Effect of different doses of chronic exposure of fluoride on rat learning and memory behavior. *Studies of Trace Elements and Health* 23(2): 1-2.

Wang JD, Ge YM, Ning HM, Wang SL. 2004. Effects of high fluoride and low iodine on biochemical indexes of the brain and learning-memory of offspring rats. *Fluoride* 37: 201-208.

Wang J, Zhang Y, Guo Z, Li R, Xue X, Sun Z, Niu R. 2018. Effects of perinatal fluoride exposure on the expressions of miR-124 and miR-132 in hippocampus of mouse pups. *Chemosphere* 197: 117-122.

Wei N, Dong Y, Wang Y, Guan Z. 2014. [Effects of chronic fluorosis on neurobehavioral development in offspring of rats and antagonistic effect of vitamin E]. *Chin J Endemiol* 33: 125-128.

Whitford GM, Whitford JL, Hobbs SH. 2009. Appetitive-based learning in rats: Lack of effect of chronic exposure to fluoride. *Neurotoxicol Teratol* 31: 210-215.

Wu CX, Gu XL, Ge YM, Zhang JH, Wang JD. 2006. Effects of high fluoride and arsenic on brain biochemical indexes and learning-memory in rats. *Fluoride* 39: 274-279.

Wu NP, Zhao ZL, Gao WH, Li XQ. 1995. [Behavioral teratology in rats exposed to fluoride.] *Chin J Endemiol* 12(5): 271-273.

Wu NP, Zhao ZL, Gao WH, Li XQ. 2008. Behavioral teratology in rats exposed to fluoride. *Fluoride* 41: 129-133.

Xu X, Shen X, Zhang Z. 2001. Effect of fluorosis on mice learning and memory behaviors and brain SOD activity and MDA content *China Public Health* 17(1): 8-10.

Yang L, Jin P, Wang X, Zhou Q, Lin X, Xi S. 2018. Fluoride activates microglia, secretes inflammatory factors and influences synaptic neuron plasticity in the hippocampus of rats. *Neurotox* 69: 108-120.

Yu Q, Shao D, Zhang R, Ouyang W, Zhang Z. 2019. Effects of drinking water fluorosis on L-type calcium channel of hippocampal neurons in mice. *Chemosphere* 220: 169-175.

Yuan J, Li Q, Niu R, Wang J. 2019. Fluoride exposure decreased learning ability and the expressions of the insulin receptor in male mouse hippocampus and olfactory bulb. *Chemosphere* 224: 71-76.

Zhang C, Ren C, Chen H, Geng R, Fan H, Zhao H, Guo K, Geng D. 2013. The analog of Ginkgo biloba extract 761 is a protective factor of cognitive impairment induced by chronic fluorosis. *Biol Trace Elem Res* 153: 229-236.

Zhang KL, Lou DD, Guan ZZ. 2015. Activation of the AGE/RAGE system in the brains of rats and in SH-SY5Y cells exposed to high level of fluoride might connect to oxidative stress. *Neurotoxicol Teratol* 48: 49-55.

Zhang J, Zhu W, Zhang Z. 2009. [The effect of fluorine exposure of pregnant rats on the learning and memory capabilities of baby rats]. *Chinese Journal of Public Health* 25(11): 1347-1348.

Zhang J, Zhu WJ, Xu XH, Zhang ZG. 2011. Effect of fluoride on calcium ion concentration and expression of nuclear transcription factor kappa-B rho65 in rat hippocampus. *Exp Toxicol Pathol* 63: 407-411.

Zhang J, Zhang Z. 2013. Effects of chronic fluorosis on camkiiα, c-FOS, BAX, and BCL-2 channel signaling in the hippocampus of rats. *Fluoride* 46: 135-141.

Zhang Z, Shen X, Xu X. 2001. [Effects of selenium on the damage of learning-memory ability of mice induced by fluoride]. *J Hyg Res* 30: 144-146.

Zhang Z, Xu X, Shen X, Xua XH. 1999. [Effect of fluoride exposure on synaptic structure of brain areas related to learning-memory in mice]. *J Hyg Res* 28(4): 210-212.

Zhang Z, Xu X, Shen X, Xua XH. 2008. Effect of fluoride exposure on synaptic structure of brain areas related to learning-memory in mice. *Fluoride* 41: 139-143.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao XL, Wu JH. 1998. Actions of sodium fluoride on acetylcholinesterase activities in rats. *Biomed Environ Sci* 11: 1-6.

Zheng X, Sun Y, Ke L, Ouyang W, Zhang Z. 2016. Molecular mechanism of brain impairment caused by drinking-acquired fluorosis and selenium intervention. *Environ Toxicol Pharmacol* 43: 134-139.

Zhu W, Zhang J, Zhang Z. 2011. Effects of fluoride on synaptic membrane fluidity and PSD-95 expression level in rat hippocampus. *Biol Trace Elem Res* 139: 197-203.

Zhu YL, Zheng YJ, LV XM, Ma Y, Zhang J. 2012. Effects of fluoride exposure on performance in water labyrinth and monoamine neurotransmitters of rats. *Journal of Xinjiang Medical University* 3: 014.

Zhu YP, Xi SH, Li MY, Ding TT, Liu N, Cao FY, Zeng Y, Liu XJ, Tong JW, Jiang SF. 2017. Fluoride and arsenic exposure affects spatial memory and activates the ERK/CREB signaling pathway in offspring rats. *Neurotox* 59: 56-64.

Studies Not Available in HAWC

Abdelaleem MM, El-Tahawy NFG, Abozaid SMM, Abdel-Hakim SA. 2018. Possible protective effect of curcumin on the thyroid gland changes induced by sodium fluoride in albino rats: Light and electron microscopic study. *Endocr Regul* 52: 59-68.

Abd-Elhakim YM, Mohammed AT, Ali HA. 2018. Impact of subchronic exposure to triclosan and/or fluoride on estrogenic activity in immature female rats: The expression pattern of calbindin-D9k and estrogen receptor alpha genes. *J Biochem Mol Toxicol* 32(2): 22027.

Abdumajidov OR. 2004. [Sex differences in lipid peroxidation and antioxidant defense of the brain tissue in intoxication with low doses of inorganic compounds]. *Uzbekiston Tibbiet Zhurnali*: 58-60.

Adebayo OL, Shallie PD, Salau BA, Ajani EO, Adenuga GA. 2013. Comparative study on the influence of fluoride on lipid peroxidation and antioxidants levels in the different brain regions of well-fed and protein undernourished rats. *J Trace Elem Med Biol* 27: 370-374.

Adedara IA, Ojuade TJD, Olabiyi BF, Idris UF, Onibiyo EM, Ajeigbe OF, Farombi EO. 2016. Taurine ameliorates renal oxidative damage and thyroid dysfunction in rats chronically exposed to fluoride. *Biol Trace Elem Res*: 1-8.

Ahmed SK, Kalleny NK, Attia AAEM, Elkateb LA. 2015. The possible protective role of chromium chloride against sodium fluoride-induced changes in the structure of the cerebellar cortex of the adult male albino rat. *Egypt J Histol* 38: 402-414.

Al Badawi MH, Mahmoud OM, Salem NA. 2016. Therapeutic potential of omega-3 against sodium fluoride toxicity on the cerebellar cortex of adult male albino rats: Histological and immunohistochemical study. *Egypt J Histol* 39: 170-178.

Alhayani A, Elshal EB, Aal IHA, Al-Shammeri E, Kabra H. 2013. Does vitamin E protect against sodium fluoride toxicity on the cerebellar cortex of albino rats? *Middle East J Sci Res* 16: 1019-1026.

Ameeramja J, Raghunath A, Perumal E. 2018. Tamarind seed coat extract restores fluoride-induced hematological and biochemical alterations in rats. *Environ Sci Pollut Res Int* 25(26): 26157-26166.

Antonyan OA. 1980. [Lipid per oxidation in fluorosis and the protective role of dietary factors]. *Zh Eksp Klin Med* 20: 381-388.

Ardelean I, Racoveanu N, Manescu S, Lupulescu A, Diaconescu M, Ghelerter L. 1964. Experimental investigations concerning the effect of fluorine on the thyroid gland. *Rum Med Rev* 8: 17-20.

Atmaca N, Atmaca HT, Kanici A, Anteplioglu T. 2014. Protective effect of resveratrol on sodium fluoride-induced oxidative stress, hepatotoxicity and neurotoxicity in rats. *Food Chem Toxicol* 70: 191-197.

Auskaps AM, Shaw JH. 1955. Hemoglobin concentration, thyroid weight and growth rate in rats during minimum fluoride ingestion. *J Nutr* 55: 611-621.

Bagmut I, Kolisnyk I, Titkova A, Petrenko T, Filipchenko S. 2018. Content of catecholamines in blood serum of rats under fluoride intoxication. *Georgian Med News* (280-281): 125-129.

Bakalyan PH, Antonyan OA. 1981. [Effect of fluorosis on glutathione peroxidase and glutathione reductase activities and sulfhydryl groups]. *Zh Eksp Klin Med* 21: 10-14.

Basha PM, Madhusudhan N. 2010. Pre and post natal exposure of fluoride induced oxidative macromolecular alterations in developing central nervous system of rat and amelioration by antioxidants. *Neurochem Res* 35: 1017-1028.

Basha PM, Madhusudhan N. 2011. Effect of maternal exposure of fluoride on oxidative stress markers and amelioration by selected antioxidants in developing central nervous system of rats. *Biologia* 66: 187-193.

Basha PM, Rai P, Begum S. 2011. Evaluation of fluoride-induced oxidative stress in rat brain: A multigeneration study. *Biol Trace Elem Res* 142: 623-637.

Basha PM, Sujitha NS. 2012. Combined influence of intermittent exercise and temperature stress on the modulation of fluoride toxicity. *Biol Trace Elem Res* 148: 69-75.

Basha PM, Saumya SM. 2013. Suppression of mitochondrial oxidative phosphorylation and TCA enzymes in discrete brain regions of mice exposed to high fluoride: Amelioration by panax ginseng (ginseng) and lagerstroemia speciosa (banaba) extracts. *Cell Mol Neurobiol* 33: 453-464.

Basha MP, Begum S, Madhusudhan N. 2014. Antioxidants in the management of fluoride induced neural oxidative stress in developing rats. *Int J Pharm Sci Res* 5: 201-206.

Benetato G, Giuran AM, Cirmaciu R, Cirje M, Petrescu A, Vacariu A. 1970. [Effect of fluorine in drinking water on the metabolism of Ca and Mg and on neuromuscular excitability: Experimental studies and clinical observations]. *Rev Roum Physiol* 7: 335-352.

Bharti VK, Srivastava RS. 2009. Fluoride-induced oxidative stress in rat's brain and its amelioration by buffalo (Bubalus bubalis) pineal proteins and melatonin. *Biol Trace Elem Res* 130: 131-140.

Bhatnagar M, Rao P, Saxena A, Bhatnagar R, Meena P, Barbar S, Chouhan A, Vimal S. 2006. Biochemical changes in brain and other tissues of young adult female mice from fluoride in their drinking water. *Fluoride* 39: 280-284.

Bilgili A, Akdogan M, Yildiz M, Eraslan G, Cetin N. 2004. The effects of fluoride on thyroid hormones in rabbits. *Indian Vet J* 81: 986-988.

Bobek S, Kahl S, Ewy Z. 1976. Effect of long-term fluoride administration on thyroid hormones level blood in rats. *Endocrinol Exp* 10: 289-295.

Bouaziz H, Ammar E, Ghorbel H, Ketata S, Jamoussi K, Ayadi F, Guermazi F, Zeghal N. 2004. Effect of fluoride ingested by lactating mice on the thyroid function and bone maturation of their suckling pups. *Fluoride* 37: 133-142.

Bouaziz H, Soussia L, Guermazi F, Zeghal N. 2005. Fluoride-induced thyroid proliferative changes and their reversal in female mice and their pups. *Fluoride* 38: 185-192.

Bouaziz HB, Amara I, Essefi M, Croute F, Zeghal N. 2010. Fluoride-induced brain damages in suckling mice. *Pestic Biochem Physiol* 96: 24-29.

Chauhan SS, Ojha S, Mahmood A. 2013. Effects of fluoride and ethanol administration on lipid peroxidation systems in rat brain. *Indian J Exp Biol* 51: 249-255.

Chen J, Chen X, Yang K, Xia T, Xie H. 2002. [Studies on DNA damage and apoptosis in rat brain induced by fluoride]. *Chin J Prev Med* 36: 222-224.

Chirumari K, Reddy PK. 2007. Dose-dependent effects of fluoride on neurochemical milieu in the hippocampus and neocortex of rat brain. *Fluoride* 40: 101-110.

Chouhan S, Yadav A, Kushwah P, Kaul RK, Flora SJS. 2011. Silymarin and quercetin abrogates fluoride induced oxidative stress and toxic effects in rats. *Mol Cell Toxicol* 7: 25-32.

Clay AB, Suttie JW. 1987. Effect of dietary fluoride on dairy cattle: Growth of young heifers. *J Dairy Sci* 70: 1241-1251.

Czechowicz K, Osada A, Slesak B. 1974. Histochemical studies on the effect of sodium fluoride on metabolism in Purkinje's cells. *Folia Histochem Cytochem* 12: 37-44.

Demole V, Lerch P. 1956. [Normality of fixation of radioactive iodine in the thyroid of rats during experimental fluorosis]. *Helv Physiol Pharmacol Acta* 14(4): 62-63.

Dhurvey V, Patil V, Thakare M. 2017. Effect of sodium fluoride on the structure and function of the thyroid and ovary in albino rats (rattus norvegicus). *Fluoride* 50: 235-246.

Domzalska E. 1966. [Influence of sodium fluoride on hypophysis, thyroid gland, parathyroid, and adrenal gland in the white rat]. *Czas Stomatol* 19: 839-844.

El-lethey HS, Kamel MM, Shaheed IB. 2011. Perinatal exposure to sodium fluoride with emphasis on territorial aggression, sexual behaviour and fertility in male rats. *Life Sci J* 8: 686-694.

Flora SJS, Mittal M, Mishra D. 2009. Co-exposure to arsenic and fluoride on oxidative stress, glutathione linked enzymes, biogenic amines and DNA damage in mouse brain. *J Neurol Sci* 285: 198-205.

Flora SJS, Mittal M, Pachauri V, Dwivedi N. 2012. A possible mechanism for combined arsenic and fluoride induced cellular and DNA damage in mice. *Metallomics* 4: 78-90.

Gabovich RD, Verzhikovskaia NV. 1958. [Effect of fluoride compounds on absorption of radioactive iodine by the thyroid gland in humans and in experimental conditions]. *Probl Endokrinol Gormonoter* 4: 49-54.

Galamini-Ligori M, Di Blasi F. 1961. [Action of sodium fluoride on the thyroid of hypophysectomized rats]. *Boll Soc Ital Biol Sper* 37: 1503-1506.

Galletti P, Held HR, Korrodi H, Wegmann T. 1956. [Investigations on the possible damaging effect of fluorine on the thyroid gland]. *Helv Med Acta* 23: 601-605.

Ge Y, Ning H, Feng C, Wang H, Yan X, Wang S, Wang J. 2006. Apoptosis in brain cells of offspring rats exposed to high fluoride and low iodine. *Fluoride* 39: 173-178.

Ge Y, Niu R, Zhang J, Wang J. 2011. Proteomic analysis of brain proteins of rats exposed to high fluoride and low iodine. *Arch Toxicol* 85: 27-33.

Ge YM, Ning HM, Wang SL, Wang JD. 2005. DNA damage in thyroid gland cells of rats exposed to long-term intake of high fluoride and low iodine. *Fluoride* 38: 318-323.

Ge YM, Ning HM, Wang SL, Wang JD. 2005. Comet assay of DNA damage in brain cells of adult rats exposed to high fluoride and low iodine. *Fluoride* 38: 209-214.

Ge YM, Ning HM, Wang SL, Wang JD. 2005. Effects of high fluoride and low iodine on brain histopathology in offspring rats. *Fluoride* 38: 127-132.

Ge YM, Ning HM, Gu XL, Yin M, Yang XF, Qi YH, Wang JD. 2013. Effects of high fluoride and low iodine on thyroid function in offspring rats. *J Integr Agric* 12: 502-508.

Guan ZZ. 1986. [Morphology of the brain of the offspring of rats with chronic fluorosis]. *Chin J Pathol* 15: 297-299.

Guan Z, Wang Y, Xiao K. 1997. [Influence of experimental fluorosis on phospholipid content and fatty acid composition in rat brain]. *Chin Med J* 77: 592-596.

Guan Z-Z, Wang Y-N, Xiao K-Q, Dai D-Y, Chen Y-H, Liu J-L, Sindelar P, Dallner G. 1998. Influence of chronic fluorosis on membrane lipids in rat brain. *Neurotoxicol Teratol* 20: 537-542.

Guan ZZ, Shan KR, Wang YN, Dallner G. 2006. [Changes of lipids and nicotinic receptors in rat brains and pheochromocytoma with fluorosis]. *Chin J Endemiol* 25: 121-124.

Gushchin SK. 1951. [Effect of sodium fluoride on iodine metabolism in rabbit tissue organs; on the etiology of endemic goiter]. *Gig Sanit* 2: 45-48.

Hamza RZ, Al-Harbi MS. 2014. Sodium fluoride induced neurotoxicity and possible antioxidant role of selenium and curcumin in male mice. *Biosci Biotechnol Res Asia* 11: 81-87.

Hamza RZ, El-Shenawy NS, Ismail HAA. 2015. Protective effects of blackberry and quercetin on sodium fluoride-induced oxidative stress and histological changes in the hepatic, renal, testis and brain tissue of male rat. *J Basic Clin Physiol Pharmacol* 26: 237-251.

Haojun Z, Yaoling W, Ke Z, Jin L, Junling W. 2012. Effects of NaF on the expression of intracellular Ca2+ fluxes and apoptosis and the antagonism of taurine in murine neuron. *Toxicol Mech Methods* 22: 305-308.

Hara K. 1980. Studies on fluorosis especially effects of fluoride on thyroid metabolism. *J Dent Health* 30: 42-57.

Harris NO, Hayes RL. 1955. A tracer study of the effect of acute and chronic exposure to sodium fluoride on the thyroid iodine metabolism of rats. *J Dent Res* 34: 470-477.

Hassan HA, Abdel-Aziz AF. 2010. Evaluation of free radical-scavenging and anti-oxidant properties of black berry against fluoride toxicity in rats. *Food Chem Toxicol* 48: 1999-2004.

Hoogstratten B, Leone NCLG, Shupe J, Greenwood DA, Lieberman J. 1965. Effect of fluorides on hematopoietic system, liver, and thyroid gland in cattle. *J Amer Med Assoc* 192: 26-32.

Inkielewicz I, Rogowska M, Krechniak J. 2006. Lipid peroxidation and antioxidant enzyme activity in rats exposed to fluoride and ethanol. *Fluoride* 39: 53-59.

Inkielewicz I, Czarnowski W. 2008. Oxidative stress parameters in rats exposed to fluoride and aspirin. *Fluoride* 41: 76-82.

Inkielewicz-Stepniak I, Czarnowski W. 2010. Oxidative stress parameters in rats exposed to fluoride and caffeine. *Food Chem Toxicol* 48: 1607-1611.

Jiang P, Li G, Zhou X, Wang C, Qiao Y, Liao D, Shi D. 2018. Chronic fluoride exposure induces neuronal apoptosis and impairs neurogenesis and synaptic plasticity: Role of GSK-3beta/beta-catenin pathway. *Chemosphere* 214: 430-435.

Jiang SF, Xi SH, Yao SQ, Tong JW, Zhang YS, Wang Q, Su J, Li MY. 2013. [Effects of fluoride, arsenic and co-exposure on expression of Bcl-2 and Bax in hippocampus and cerebral cortex of rats]. *Chin J Endemiol* 32: 365-369.

Jiang Y, Guo X, Sun Q, Shan Z, Teng W. 2016. Effects of excess fluoride and iodide on thyroid function and morphology. *Biol Trace Elem Res* 170: 382-389.

Jin TX, Guan ZZ, Zhang H. 2011. [The effect of fluoride on α subunit of calcium/calmodulin-dependent protein kinase-II mRNA and protein expression in central nervous system]. *Chin J Endemiol* 30: 247-250.

Jonderko G, Kita K, Pietrzak J, Primus-Slowinska B, Ruranska B, Zylka-Wloszczyk M, Straszecka J. 1983. [Effect of subchronic sodium fluoride poisoning on the thyroid gland of rabbits with normal and increased supply of iodine]. *Endokrynol Pol* 34: 195-203.

Kahl S, Bobek S. 1975. [Effect of fluoride administration on radiothyroxine turnower in rats]. *Endokrynol Pol* 26: 391-396.

Kahl S, Ewy Z. 1975. Effect of single and long term sodium fluoride administration on biosynthesis of the thyroid hormone in rats. *Fluoride* 8: 191-198.

Kapoor V, Prasad T, Paliwal VK. 2001. Blood biochemical constituents in calves following subclinical levels of fluoride toxicosis. *Fluoride* 34: 126-131.

Karawya FS, Zahran NM, Azzam EZ. 2015. Is water fluoridation a hidden cause of obesity? Histological study on thyroid follicular cells of albino rats. *Egypt J Histol* 38: 547-557.

Kaur T, Bijarnia RK, Nehru B. 2009. Effect of concurrent chronic exposure of fluoride and aluminum on rat brain. *Drug Chem Toxicol* 32: 215-221.

Kelimu A, Liu KT, Lian J, Hu HH, Zheng YJ, Wang TM. 2008. [Effects of vitamin C and E on the ultrastructure in liver, kidney and brain of fluorosis rats]. *Chin J Endemiol* 27: 378-381.

Kinawy AA. 2019. Synergistic oxidative impact of aluminum chloride and sodium fluoride exposure during early stages of brain development in the rat. *Environ Sci Pollut Res Int* 26(11): 10951-10960.

Knizhnikov VA. 1959. [Effect of potable water with high fluoride concentration on thyroid function]. *Gig Sanit* 24: 20-25.

Knizhnikov VA, Tsypin AB, Shcherbova EN, Bugryshev PF. 1963. [The effect of drinking water with an increased fluorine content on the bioelectrical activity of the brain and heart under experimental conditions]. *Gig Sanit* 28: 16-19.

Kondo T, Yoshida M, Kasahara K. 1976. [Acute fluorosis in female rats: Time of inhibition and recovery of cholinesterase in serum and salivary glands]. *Jpn J Dent Health* 26: 187-192.

Kowalewska M. 1974. [Biopotentials of the organ of hearing in chronic sodium fluoride poisoning]. *J Pol Otolaryngol* 28: 417-424.

Krechniak J, Inkielewicz I. 2005. Correlations between fluoride concentrations and free radical parameters in soft tissues of rats. *Fluoride* 38: 293-296.

Leonard BE. 1972. Effect of phentolamine on the increase in brain glycolysis following the intraventricular administration of dibutyryl-3,5-cyclic adenosine monophosphate and sodium fluoride to mice. *Biochem Pharmacol* 21: 115-117.

Liu G, Zhang W, Jiang P, Li X, Liu C, Chai C. 2012. Role of nitric oxide and vascular endothelial growth factor in fluoride-induced goitrogenesis in rats. *Environ Toxicol Pharmacol* 34: 209-217.

Li H, Cai Q, Wang D. 2012. [Effect of fluoride on the expression of rat thyroid peroxidase mRNA]. *Chin J Endemiol* 31: 515-517.

Li H, Cai Q, Wang D. 2012. [Effects of fluoride on rat thyroid morphology, thyroid peroxidase activity and the expression of thyroid peroxidase protein]. *Chin J Endemiol* 31: 271-274.

Liu H, Hou C, Zeng Q, Zhao L, Cui Y, Yu L, Wang L, Zhao Y, Nie J, Zhang B, Wang A. 2016. Role of endoplasmic reticulum stress-induced apoptosis in rat thyroid toxicity caused by excess fluoride and/or iodide. *Environ Toxicol Pharmacol* 46: 277-285.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. [Changes of the c-Jun N-terminal kinase in the brains of rats with chronic fluorosis]. *Chin J Endemiol* 29: 608-612.

Liu YJ, Guan ZZ, Gao Q, Pei JJ. 2011. Increased level of apoptosis in rat brains and SH-SY5Y cells exposed to excessive fluoride-A mechanism connected with activating JNK phosphorylation. *Toxicol Lett* 204: 183-189.

Lohakare J, Pattanaik AK. 2013. Effects of addition of fluorine in diets differing in protein content on urinary fluoride excretion, clinical chemistry and thyroid hormones in calves. *Brazilian J Anim Sci* 42: 751-758.

Long YG, Wang YN, Chen J, Jiang SF, Nordberg A, Guan ZZ. 2002. Chronic fluoride toxicity decreases the number of nicotinic acetylcholine receptors in rat brain. *Neurotoxicol Teratol* 24: 751-757.

Lou DD, Liu YF, Zhang KL, Yu YN, Guan ZZ. 2011. [Changes of reactive oxygen species level and mitochondria fission-fusion in cortical neurons of rats with chronic fluorosis]. *Chin J Endemiol* 30: 256-260.

Lou DD, Liu YF, Qin SL, Zhang KL, Yu YN, Guan ZZ. 2012. [Changed transcription level of mitochondrial fission and fusion gene loci in cortical neurons of rats with chronic fluorosis]. *Chin J Endemiol* 31: 125-129.

Lou DD, Zhang KL, Qin SL, Liu YF, Yu YN, Guan ZZ. 2012. [Alteration of mitochondrial distribution and gene expression of fission 1 protein in cortical neurons of rats with chronic fluorosis]. *Chin J Pathol* 41: 243-247.

Lou DD, Pan JG, Zhang KL, Qin SL, Liu YF, Yu YN, Guan ZZ. 2013. [Changed expression of mito-fusion 1 and mitochondrial fragmentation in the cortical neurons of rats with chronic fluorosis]. *Chin J Prev Med* 47: 170-174.

Lou DD, Zhang KL, Pan JG, Qin SL, Liu YF, Yu YN, Guan ZZ. 2013. [Influence of chronic fluorosis on the expression of mitochondrial fission protein dynamin-related 1 in the cortical neurons of rats]. *Chin J Prev Med* 47: 561-564.

Lou DD, Zhang KL, Qin SL, Liu YF, Liu YJ, Guan ZZ. 2013. [Effects of chronic fluorosis on 4.8 kb mitochondrial DNA in liver, kidney and brain of rats]. *Chin J Endemiol* 32: 121-124.

Lou DD, Guan ZZ, Pei JJ. 2014. Alterations of apoptosis and expressions of Bax and Bcl-2 in the cerebral cortices of rats with chronic fluorosis. *Fluoride* 47: 199-207.

Luo GY, Niu RY, Sun ZL, Zhang JH, Wang JM, Wang C, Wang JD. 2011. Reduction of CaMKII expression in the hippocampus of rats from ingestion of fluoride and/or lead. *Fluoride* 44: 63-69.

Ma T, Liu D, Song K. 1999. Cytochemical study of neuron enzyme at anterior horn of spinal cord in rats with experimental fluorosis. *J Chin Med Univ* 28: 81-82.

Ma TX, Yu HT, Song KQ. 2008. [Expression of c-fos and Caspase 8 in cerebral cortex of rats with experimental fluorosis]. *Chin J Endemiol* 27: 131-133.

Mach Z, Zygulska-Machowa H. 1959. O wplywie fluoru na przemiane J131 [Russian and English summ.]. *Endokrynol Pol* 10: 157-162.

Machida H. 1989. [A study on the rabbit thermoregulatory system effects of high dose sodium fluoride]. *Dent Sci Rep* 89: 607-626.

Madan J, Puri JP, Singh JK. 2009. Growth, feed efficiency and blood profile of buffalo calves consuming high levels of fluoride. *Trop Anim Health Prod* 41: 295-298.

Madhusudhan N, Basha PM, Begum S, Ahmed F. 2009. Fluoride-induced neuronal oxidative stress and its amelioration by antioxidants in developing rats. *Fluoride* 42: 179-187.

Madhusudhan N, Basha PM, Rai P, Ahmed F, Prasad GR. 2010. Effect of maternal fluoride exposure on developing CNS of rats: Protective role of Aloe vera, Curcuma longa and Ocimum sanctum. *Indian J Exp Biol* 48: 830-836.

Manocha SL, Warner H, Olkowski ZL. 1975. Cytochemical response of kidney, liver and nervous system of fluoride ions in drinking water. *Histochem J* 7: 343-355.

Mansour HH, Tawfik SS. 2012. Efficacy of lycopene against fluoride toxicity in rats. *Pharm Biol* 50: 707-711.

Mietkiewski K, Walczak M, Trojanowicz R. 1966. [Effect of sodium fluoride on the neurosecretory system in guinea pigs]. *Endokrynol Pol* 17: 121-131.

Mohamed NE. 2016. The role of calcium in ameliorating the oxidative stress of fluoride in rats. *Biol Trace Elem Res* 170: 128-144.

Muhlemann HR, Schneider R. 1956. [Mitotic activity of rat thyroid epithelium after administration of fluoridated drinking water]. *Schweiz Med Wochenschr* 86: 625-627.

Nabavi SF, Eslami S, Moghaddam AH, Nabavi SM. 2011. Protective effects of curcumin against fluoride-induced oxidative stress in the rat brain. *Neurophysiology* 43: 287-291.

Nabavi SF, Moghaddam AH, Nabavi SM, Eslami S. 2011. Protective effect of curcumin and quercetin on thyroid function in sodium fluoride intoxicated rats. *Fluoride* 44: 147-152.

Nabavi SF, Habtemariam S, Jafari M, Sureda A, Nabavi SM. 2012. Protective role of gallic acid on sodium fluoride induced oxidative stress in rat brain. *Bull Environ Contam Toxicol* 89: 73-77.

Nabavi SF, Nabavi SM, Latifi AM, Mirzaei M, Habtemariam S, Moghaddam AH. 2012. Mitigating role of quercetin against sodium fluoride-induced oxidative stress in the rat brain. *Pharm Biol* 50: 1380-1383.

Nabavi SF, Nabavi SM, Habtemariam S, Moghaddam AH, Sureda A, Mirzaei M. 2013. Neuroprotective effects of methyl-3-O-methyl gallate against sodium fluoride-induced oxidative stress in the brain of rats. *Cell Mol Neurobiol* 33: 261-267.

Nabavi SM, Sureda A, Nabavi SF, Latifi AM, Moghaddam AH, Hellio C. 2012. Neuroprotective effects of silymarin on sodium fluoride-induced oxidative stress. *J Fluor Chem* 142: 79-82.

Narayanaswamy M, Piler MB. 2010. Effect of maternal exposure of fluoride on biometals and oxidative stress parameters in developing CNS of rat. *Biol Trace Elem Res* 133: 71-82.

Narbutt B, Romer TE, Grabski J, Szymik N. 1971. [Influence of natrium fluoride on the structure of the rat thyroid]. *Endokrynol Pol* 22: 445-451.

Narbutt B, Romer TE, Grabski J, Szymik N. 1971. [The influence of sodium fluoride on the morphology of the thyroid gland in rats]. *Endokrynol Pol* 22: 361-365.

Niu RY, Sun ZL, Cheng ZT, Liu HT, Chen HC, Wang JD. 2008. Effects of fluoride and lead on N-methyl-D-aspartate receptor 1 expression in the hippocampus of offspring rat pups. *Fluoride* 41: 101-110.

Niu R, Wang J, Sun Z, Xue X, Yan X, Zhang J. 2015. Transcriptional regulatory dynamics of the hypothalamic-pituitary-testicular axis in male mice exposed to fluoride. *Environ Toxicol Pharmacol* 40: 557-562.

Niu R, Zhang Y, Liu S, Liu F, Sun Z, Wang J. 2015. Proteome alterations in cortex of mice exposed to fluoride and lead. *Biol Trace Elem Res* 164: 99-105.

Ogilvie AL. 1952. Histological findings in the kidney, liver, pancreas, adrenal and thyroid gland of the rat following sodium fluoride administration. *J Dent Res* 31: 598-598.

Okayasu I, Tsuchida M, Yanagisawa F. 1985. Hyperplastic nodules of thyroid parafollicular cells (C cells) in rats induced by prolonged low dose ingestion of NaF. *Fluoride* 18: 111-117.

Pal S, Sarkar C. 2014. Protective effect of resveratrol on fluoride induced alteration in protein and nucleic acid metabolism, DNA damage and biogenic amines in rat brain. *Environ Toxicol Pharmacol* 38: 684-699.

Pan Y, Lu P, Yin L, Chen K, He Y. 2015. Effect of fluoride on the proteomic profile of the hippocampus in rats. *Z Naturforsch C* 70: 151-157.

Phillips PH, Lamb AR. 1934. Histology of certain organs and teeth in chronic toxicosis due to fluorin. *Arch Path* 17: 169-176.

Portela ML. 1972. [Biochemical effects in the prolonged ingestion of fluorides in rats]. *Arch Latinoam Nutr* 22: 291-308.

Prestes DS, Filappi A, Schossler DR, Duarte FA, Dressler VL, Flores EMM, Cecim M. 2009. Functional and histological evaluations of thyroid of sheep submitted to sodium fluoride administration. *Arq Bras Med Vet Zootec* 61: 293-298.

Puentes F, Cremer HD. 1966. Experiments on fluoride-iodine antagonism in the thyroid gland. *Adv Fluorine Res* 4: 213-220.

Qian W, Miao K, Li T, Zhang Z. 2013. Effect of selenium on fluoride-induced changes in synaptic plasticity in rat hippocampus. *Biol Trace Elem Res* 155: 253-260.

Qing-Feng S, Ying-Peng X, Tian-Tong X. 2019. Matrix metalloproteinase-9 and p53 involved in chronic fluorosis induced blood-brain barrier damage and neurocyte changes. *Arch Med Sci* 15(2): 457-466.

Qiu YH, Kong DM, Yang Q, Zhao N. 2010. [Influence of high-fluoride on thyroid function and brain damage in rats]. *Chin J Endemiol* 29: 146-149.

Raghavendra M, Ravindra RK, Raghuveer YP, Narasimha JK, Uma MRV, Navakishor P. 2016. Alleviatory effects of hydroalcoholic extract of cauliflower (brassica oleracea var. botrytis) on thyroid function in fluoride intoxicated rats. *Fluoride* 49: 84-90.

Rakhov GM. 1964. [Effect of calcium and fluorine in drinking water on the iodine metabolism and status of the thyroid gland in iodine insufficiency in food]. *Gig Sanit* 29: 12-17.

Ranpariya VL, Parmar SK, Sheth NR, Chandrashekhar VM. 2011. Neuroprotective activity of matricaria recutita against fluoride-induced stress in rats. *Pharm Biol* 49: 696-701.

Reddy KP, Sailaja G, Krishnaiah C. 2009. Protective effects of selenium on fluoride induced alterations in certain enzymes in brain of mice. *J Environ Biol* 30: 859-864.

Rogalska A, Kuter K, Zelazko A, Glogowska-Gruszka A, Swietochowska E, Nowak P. 2017. Fluoride alteration of [3H]glucose uptake in Wistar rat brain and peripheral tissues. *Neurotoxicol Res* 31: 436-443.

Saka O, Hallac P, Urgancioğlu I. 1965. The effect of fluoride on the thyroid of the rat. *New Istanbul Contrib Clin Sci* 8: 87-90.

Samanta A, Chanda S, Bandyopadhyay B, Das N. 2016. Establishment of drug delivery system nanocapsulated with an antioxidant (+)-catechin hydrate and sodium meta borate chelator against sodium fluoride induced oxidative stress in rats. *J Trace Elem Med Biol* 33: 54-67.

Sarkar C, Das N, Pal S, Dinda B. 2014. Oxidative stress induced alteration of protein and nucleic acid metabolism in fluoride-intoxicated rat brain: Protection by 3α-hydroxy olean-12-en-27-oic acid isolated from neanotis wightiana. *Int J Pharm Sci Res* 5: 3047-3066.

Sarkar C, Pal S. 2014. Ameliorative effect of resveratrol against fluoride-induced alteration of thyroid function in male Wistar rats. *Biol Trace Elem Res* 162: 278-287.

Sarkar C, Pal S, Das N, Dinda B. 2014. Ameliorative effects of oleanolic acid on fluoride induced metabolic and oxidative dysfunctions in rat brain: Experimental and biochemical studies. *Food Chem Toxicol* 66: 224-236.

Seffner W, Teubener W, Runde H, Wiedner H, Vogt J, Otto G, Zschau E, Geinitz D, Franke J. 1990. Boron as an antidote to fluorosis? II. Studies on various organs of pigs. *Fluoride* 23: 68-79.

Selim AOA, El-Haleem MR, Ibrahim IH. 2012. Effect of sodium fluoride on the thyroid gland of growing male albino rats: Histological and biochemical study. *Egypt J Histol* 35: 470-482.

Shao Q, Wang Yn, Guan Z. 2000. [Influence of free radical inducer on the level of oxidative stress in brain of rats with fluorosis]. *Chin J Prev Med* 34: 330-332.

Sharma C, Suhalka P, Sukhwal P, Jaiswal N, Bhatnagar M. 2014. Curcumin attenuates neurotoxicity induced by fluoride: An in vivo evidence. *Pharmacogn Mag* 10: 61-65.

Shashi A. 1992. Studies on alterations in brain lipid metabolism following experimental fluorosis. *Fluoride* 25: 77-84.

Shashi A. 1993. Nucleic acid levels in thyroid gland in acute and chronic fluoride intoxication. *Fluoride* 26: 191-196.

Shashi A, Singh JP, Thapar SP. 1994. Effect of long-term administration of fluoride on levels of protein, free amino acids and RNA in rabbit brain. *Fluoride* 27: 155-159.

Shashi A. 2003. Histopathological investigation of fluoride-induced neurotoxicity in rabbits. *Fluoride* 36: 95-105.

Shashi A, Neetika S, Bhardwaj M. 2009. Neuronal DNA damage and apoptosis in brain of rat exposed to fluoride. *Asian J Microbiol Biotechnol Environ Sci* 11: 629-632.

Shayiq RM, Raza H, Kidwai AM. 1986. Fluoride and lipid peroxidation a comparative study in different rat tissues. *Bull Environ Contam Toxicol* 37: 70-76.

Shen QF, Li HN, Xu TT, Xia YP. 2012. [Damage of blood brain barrier of spinal cord in rats with chronic fluorosis]. *Chin Med J* 92: 2357-2361.

Shen Q, Tian R, Li H, Xu T, Xia Y. 2014. [White matter injury of spinal cord in rats with chronic fluorosis and recovery after defluoridation]. *Chin Med J* 94: 1189-1192.

Shen X, Zhang Z, Xu X. 2004. [Influence of combined iodine and fluoride on phospholipid and fatty acid composition in brain cells of rats]. *J Hyg Res* 33: 158-161.

Shivarajashankara YM, Shivashankara AR, Bhat PG, Rao SH. 2001. Effect of fluoride intoxication on lipid peroxidation and antioxidant systems in rats. *Fluoride* 34: 108-113.

Shivarajashankara YM, Shivashankara AR, Bhat PG, Rao SH. 2002. Brain lipid peroxidation and antioxidant systems of young rats in chronic fluoride intoxication. *Fluoride* 35: 197-203.

Shivarajashankara YM, Shivashankara AR, Bhat PG, Rao SM, Rao SH. 2002. Histological changes in the brain of young fluoride-intoxicated rats. *Fluoride* 35: 12-21.

Siebenhuner L, Miloni E, Burgi H. 1984. [Effects of fluoride on thyroid hormone biosynthesis: Studies in a highly sensitive test system]. *Klin Wochenschr* 62: 859-861.

Singh R, Srivastava AK, Gangwar NK. 2017. Clinico-pathological studies on the co-exposure of cypermethrin and fluoride in experimental rats with ameliorative action of Vitamin E. *Vet Pract* 18(2): 207-210.

Soni KK, Shrivastava VK. 2007. Sodium fluoride induced histopathological changes in thyroid gland of male mus musculus. *Biochem Cell Arch* 7: 317-320.

Stee EW. 1968. *Effect of sodium fluoride and AMOX (NF3O) on growth and thyroid function in the rat*. No. AMRL-TR-67-189. Wright-Patterson Air Force Base, OH: pp. 67.

Štolc V, Podoba J. 1960. Effect of fluoride on the biogenesis of thyroid hormones. *Nature* 188: 855-856.

Sugivama Y. 1967. [The effect of sodium fluoride administration on the parathyroid glands]. *Hirosaki Med J* 19: 520-529.

Sun Y, Ke L, Zheng X, Li T, Ouyang W, Zhang Z. 2016. Effects of different levels of calcium intake on brain cell apoptosis in fluorosis rat offspring and its molecular mechanism. *Biol Trace Elem Res*: 1-12.

Takata H. 1958. The effect of fluorine upon the uptake of I131 by the thyroid glands. *Folia Pharmacol Jpn* 54: 230-236.

Teng Y, Zhang J, Zhang Z, Feng J. 2017. The effect of chronic fluorosis on calcium ions and CaMKIIα, and c-fos expression in the rat hippocampus. *Biol Trace Elem Res*: 295-302.

Trabelsi M, Guermazi F, Zeghal N. 2001. Effect of fluoride on thyroid function and cerebellar development in mice. *Fluoride* 34: 165-173.

Tsuchida M, Okayasu I, Kohyama Y, Kurihara H, Tanaka H, Yanagisawa F, Date C, Hayashi M, Mui K, Asada M. 1986. Effects of long term, low dose ingestion of fluoride on the thyroid gland in rats. *Stud Environ Sci* 27: 307-312.

Vani ML, Reddy KP. 2000. Effects of fluoride accumulation on some enzymes of brain and gastrocnemius muscle of mice. *Fluoride* 33: 17-26.

Wang C, Liang C, Ma J, Manthari RK, Niu R, Wang J, Wang J, Zhang J. 2018. Co-exposure to fluoride and sulfur dioxide on histological alteration and DNA damage in rat brain. *J Biochem Mol Toxicol* 32.

Wang H, Yang Z, Zhou B, Gao H, Yan X, Wang J. 2009. Fluoride-induced thyroid dysfunction in rats: Roles of dietary protein and calcium level. *Toxicol Ind Health* 25: 49-57.

Wang J, Niu R, Sun Z, Lv L, Smith GW. 2008. Effects of protein and calcium supplementation on bone metabolism and thyroid function in protein and calcium deficient rabbits exposed to fluoride. *Fluoride* 41: 283-291.

Wang JD, Ge YM, Ning HM, Wang SL. 2004. Effects of high fluoride and low iodine on oxidative stress and antioxidant defense of the brain in offspring rats. *Fluoride* 37: 264-270.

Wang JL. 2007. [Effect of fluoride on the intracellular Ca2+ in neurons of mice]. *Chin J Endemiol* 26: 505-507.

Wang Y, Guan Z, Xiao K. 1997. [Changes of coenzyme Q content in brain tissues of rats with fluorosis]. *Chin J Prev Med* 31: 330-333.

Wang Y, Dong Y, Wei N, Guan Z. 2015. [Influence of chronic fluorosis on expression of quinone oxidoreductase-1 and heme oxygenase-1 in rat brains]. *Chin J Endemiol* 34: 250-253.

Wedzisz A, Cieciura J. 1988. Effect of small sodium fluoride feed supplements on the serum thyroid hormone content of rats. *Bromatol Chem Toksykol* 21: 174-175.

Wei N, Dong YT, Deng J, Wang Y, Qi XL, Yu WF, Xiao Y, Zhou JJ, Guan ZZ. 2018. Changed expressions of N-methyl-d-aspartate receptors in the brains of rats and primary neurons exposed to high level of fluoride. *J Trace Elem Med Biol* 45: 31-40.

Yan N, Liu Y, Liu S, Cao S, Wang F, Wang Z, Xi S. 2016. Fluoride-induced neuron apoptosis and expressions of inflammatory factors by activating microglia in rat brain. *Mol Neurobiol* 53: 4449-4460.

Yang H, Xing R, Liu S, Yu H, Li P. 2016. Gamma-Aminobutyric acid ameliorates fluoride-induced hypothyroidism in male Kunming mice. *Life Sci* 146: 1-7.

Yang H, Xing R, Liu S, Yu H, Li P. 2019. Analysis of the protective effects of gamma-aminobutyric acid during fluoride-induced hypothyroidism in male Kunming mice. *Pharm Biol* 57(1): 29-37.

Yang M, Ren Z, Zhou B, Guan Z, Yu W. 2017. [Expression of endonuclease G in the brain tissue of rats with chronic fluorosis]. *Chin J Endemiol* 36: 327-332.

Yuan SD, Xie QW, Lu FY. 1993. Changes of serotonin content and turnover rate in hypothalamus of female rat during fluorosis. *Fluoride* 26: 57-60.

Zhai JX, Guo ZY, Hu CL, Wang QN, Zhu QX. 2003. [Studies on fluoride concentration and cholinesterase activity in rat hippocampus]. *Chin J Ind Hyg Occup Dis* 21: 102-104.

Zhan CW, Huo DJ. 1988. Ultrastructural findings in liver, kidneys, thyroid-gland and cardiac-muscle of rabbits following sodium-fluoride administration. *Fluoride* 21: 32-38.

Zhan XA, Xu ZR, Li JX, Wang M. 2005. Effects of fluorosis on lipid peroxidation and antioxidant systems in young pigs. *Fluoride* 38: 157-161.

Zhan XA, Li JX, Wang M, Xu ZR. 2006. Effects of fluoride on growth and thyroid function in young pigs. *Fluoride* 39: 95-100.

Zhang KL, Lou DD, Liu YF, Qin SL, Guan ZZ. 2012. [Changes of P-glycoprotein and nuclear factor κB in the cerebral cortex of rat with chronic fluorosis]. *Chin J Endemiol* 31: 613-616.

Zhang KL, Lou DD, Guan ZZ. 2013. [Expression of receptor for advanced glycation endproducts and nuclear factor κB in brain hippocampus of rat with chronic fluorosis]. *Chin J Endemiol* 32: 625-628.

Zhang WD, Zhang Y, Liu GY, Jiang P, Chai CY. 2008. [Effects of fluoride on ultrastructure of thyroids in rats]. *Chin J Endemiol* 27: 622-624.

Zhang ZG, Wang XY, Nian WW, Liao QX, Zhang R, Ouyang W. 2017. Effects of calcium on drinking fluorosis-induced hippocampal syntaptic plasticity impairment in the offspring of rats. *Transl Neurosci* 8: 191-200.

Zhao W, Zhu H, Yu Z, Aoki K, Misumi J, Zhang X. 1998. Long-term effects of various iodine and fluorine doses on the thyroid and fluorosis in mice. *Endocr Regul* 32: 63-70.

Zhao WY. 1988. [A preliminary study of the interaction of iodine and fluoride in experimental iodine goiter and fluorosis]. *Chin J Prev Med* 22: 146-148.

Zhao XL, Gao WH, Zhao ZL. 1994. [Effects of sodium fluoride on the activity of Ca2+Mg(2+)-ATPase in synaptic membrane in rat brain]. *Chin J Prev Med* 28: 264-266.

Zhavoronkov AA, Polyakova GA. 1973. Morphological and functional state of the hypothalamo-hypophyseal neurosecretory system in experimental fluorosis. *Bull Exp Biol Med* 75: 194-196.

Zhou B, Luo G, Wang C, Niu R, Wang J. 2014. Effects of fluoride on expression of cytokines in the hippocampus of adult rats. *Fluoride* 47: 191-198.

In Vitro Experimental Studies

As described in **Figure 2**, 60 in vitro experimental studies were included; however, data extraction was not conducted on in vitro studies. Therefore, in vitro experimental studies are not available in HAWC with the exception of in vitro studies that also reported in vivo non-human animal data that meet the relevant criteria for being made available in HAWC. The following lists of references are organized as studies that are available in HAWC (n = 6) followed by studies that are not available in HAWC (n = 54).

Studies Available in HAWC

Chen J, Niu Q, Xia T, Zhou G, Li P, Zhao Q, Xu C, Dong L, Zhang S, Wang A. 2018. ERK1/2-mediated disruption of BDNF-TrkB signaling causes synaptic impairment contributing to fluoride-induced developmental neurotoxicity. *Toxicology* 410: 222-230.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. *Environ Pollut* 233: 889-899.

Shan KR, Qi XL, Long YG, Nordberg A, Guan ZZ. 2004. Decreased nicotinic receptors in PC12 cells and rat brains influenced by fluoride toxicity: A mechanism relating to a damage at the level in post-transcription of the receptor genes. *Toxicology* 200: 169-177.

Zhang KL, Lou DD, Guan ZZ. 2015. Activation of the AGE/RAGE system in the brains of rats and in SH-SY5Y cells exposed to high level of fluoride might connect to oxidative stress. *Neurotoxicol Teratol* 48: 49-55.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao XL, Wu JH. 1998. Actions of sodium fluoride on acetylcholinesterase activities in rats. *Biomed Environ Sci* 11: 1-6.

Studies Not Available in HAWC

Ardelean I, Racoveanu N, Manescu S, Lupulescu A, Diaconescu M, Ghelerter L. 1964. Experimental investigations concerning the effect of fluorine on the thyroid gland. *Rum Med Rev* 8: 17-20.

Chen J, Chen X, Yang K. 2000. [Effects of selenium and zinc on the DNA damage caused by fluoride in pallium neural cells of rats]. *J Hyg Res* 29: 216-217.

Chen L, Ning H, Yin Z, Song X, Feng Y, Qin H, Li Y, Wang J, Ge Y, Wang W. 2017. The effects of fluoride on neuronal function occurs via cytoskeleton damage and decreased signal transmission. *Chemosphere* 185: 589-594.

Chen R, Zhao LD, Liu H, Li HH, Ren C, Zhang P, Guo KT, Zhang HX, Geng DQ, Zhang CY. 2017. Fluoride induces neuroinflammation and alters Wnt signaling pathway in BV2 microglial cells. *Inflammation* 40: 1123-1130.

Cheng TJ, Chen TM, Chen CH, Lai YK. 1998. Induction of stress response and differential expression of 70 kDa stress proteins by sodium fluoride in HeLa and rat brain tumor 9L cells. *J Cell Biochem* 69: 221-231.

Deng MF, Zhu D, Liu YP, He WW, Gui CZ, Guan ZZ. 2018. Attenuation by 7-nitroindazole of fluoride-induced toxicity in SH-SY5Y cells exposed to high fluoride: Effects on nitric oxide, nitric oxide synthetase activity, nNOS, and apoptosis. *Fluoride* 51(4): 328-339.

Flores-Mendez M, Ramirez D, Alamillo N, Hernandez-Kelly LC, Del Razo LM, Ortega A. 2014. Fluoride exposure regulates the elongation phase of protein synthesis in cultured Bergmann glia cells. *Toxicol Lett* 229: 126-133.

Gao Q, Liu YH, Guan ZZ. 2008. Oxidative stress might be a mechanism connected with the decreased alpha 7 nicotinic receptor influenced by high-concentration of fluoride in SH-SY5Y neuroblastoma cells. *Toxicol In Vitro* 22: 837-843.

Goschorska M, Gutowska I, Baranowska-Bosiacka I, Piotrowska K, Metryka E, Safranow K, Chlubek D. 2018. Influence of acetylcholinesterase inhibitors used in Alzheimer's Disease treatment on the activity of antioxidant enzymes and the concentration of glutathione in THP-1 macrophages under fluoride-induced oxidative stress. *Int J Environ Res Pub Health* 16(1).

Guan ZZ, Shan KR, Xiu J, Long YG. 2005. [Fluorosis on expression of nicotinic acetylcholine receptors in protein and gene levels in human SH-SY5Y neuroblastoma cells]. *Chin J Prev Med* 39: 26-29.

Guan ZZ, Shan KR, Wang YN, Dallner G. 2006. [Changes of lipids and nicotinic receptors in rat brains and pheochromocytoma with fluorosis]. *Chin J Endemiol* 25: 121-124.

Haojun Z, Yaoling W, Ke Z, Jin L, Junling W. 2012. Effects of NaF on the expression of intracellular Ca2+ fluxes and apoptosis and the antagonism of taurine in murine neuron. *Toxicol Mech Methods* 22: 305-308.

Hong-Liang L, Qiang Z, Yu-Shan C, Lei Z, Gang F, Chang-Chun H, Liang Z, Aiguo W. 2014. Fluoride-induced thyroid cell apoptosis. *Fluoride* 47: 161-169.

Inkielewicz-Stepniak I, Radomski MW, Wozniak M. 2012. Fisetin prevents fluoride- and dexamethasone-induced oxidative damage in osteoblast and hippocampal cells. *Food Chem Toxicol* 50: 583-589.

Jin TX, Guan ZZ, Zhang H. 2011. [The effect of fluoride on α subunit of calcium/calmodulin-dependent protein kinase-II mRNA and protein expression in central nervous system]. *Chin J Endemiol* 30: 247-250.

Kariya T, Kotani M, Field JB. 1974. Effects of sodium fluoride and other metabolic inhibitors on basal and TSH stimulated cyclic AMP and thyroid metabolism. *Metab Clin Exper* 23: 967-973.

Ke L, Zheng X, Sun Y, Ouyang W, Zhang Z. 2016. Effects of sodium fluoride on lipid peroxidation and PARP, XBP-1 expression in PC12 cell. *Biol Trace Elem Res* 173: 161-167.

Lee J, Han YE, Favorov O, Tommerdahl M, Whitsel B, Lee CJ. 2016. Fluoride induces a volume reduction in CA1 hippocampal slices via MAP kinase pathway through volume regulated anion channels. *Exp Neurobiol* 25: 72-78.

Levesque L, Mizzen CA, McLachlan DR, Fraser PE. 2000. Ligand specific effects on aluminum incorporation and toxicity in neurons and astrocytes. *Brain Res* 877: 191-202.

Li H, Gao MT, Xu KY, Wang CY. 2007. Effect of sodium fluoride on the primary porcine thyroid cells and thyroid peroxidase activity. *J Clin Rehabil Tissue Eng Res* 11: 7425-7428.

Li H, Gao MT, Xu KY, Cui MY, Dai X. 2008. [Effect of fluoride on thyroid functioning in primary porcine thyrocyte]. *Chin J Endemiol* 27: 38-40.

Li H, Huang H, Xu Y, Gao Y, Liu Z. 2010. [Toxic effects of fluoride on rat cerebral cortex astrocytes in vitro]. *J Hyg Res* 39: 86-88.

Liu H, Zeng Q, Cui Y, Yu L, Zhao L, Hou C, Zhang S, Zhang L, Fu G, Liu Y, Jiang C, Chen X, Wang A. 2014. The effects and underlying mechanism of excessive iodide on excessive fluoride-induced thyroid cytotoxicity. *Environ Toxicol Pharmacol* 38: 332-340.

Liu HL, Zeng Q, Cui YS, Zhao L, Zhang L, Fu G, Hou CC, Zhang S, Yu LY, Jiang CY, Wang ZL, Chen XM, Wang AG. 2014. The role of the IRE1 pathway in excessive iodide- and/or fluoride-induced apoptosis in Nthy-ori 3-1 cells in vitro. *Toxicol Lett* 224: 341-348.

Liu YJ, Guan ZZ, Gao Q, Pei JJ. 2011. Increased level of apoptosis in rat brains and SH-SY5Y cells exposed to excessive fluoride-A mechanism connected with activating JNK phosphorylation. *Toxicol Lett* 204: 183-189.

Liu Y, Gao Q, Tang Z, Zhang X, Guan Z. 2015. [The expression and correlation between neural nicotinic acetylcholine receptor subunit α3 and mitogen-activated protein kinase cell signaling transduction pathway in human neuroblastoma cell line SH-SY5Y overexposed to fluoride]. *Chin J Endemiol* 34: 553-558.

Madaoui S, Rappaport L, Nunez J. 1974. Prostaglandins and in vitro TSH-dependent iodide binding by rat thyroid glands. *Biochimie* 56: 109-113.

Nakagawa-Yagi Y, Saito Y, Kitoh N, Ogane N, Fujisawa E, Nakamura H. 1993. Fluoride causes suppression of neurite outgrowth in human neuroblastoma via an influx of extracellular calcium. *Biochem Biophys Res Commun* 191: 727-736.

Ong J, Kerr DIB. 1995. Interactions of N-ethylmaleimide and aluminium fluoride with GABA(B) receptor function in rat neocortical slices. *Eur J Pharmacol* 287: 197-200.

Pastan I, Macchia V, Katzen R. 1968. Effect of fluoride on the metabolic activity of thyroid slices. *Endocrinology* 83: 157-160.

Rubakhova VM. 1977. [Effect of serotonin and sodium fluoride on visceral nerve conductors]. *Vyestsi Akademii Navuk BSSR Syeryya Biyalahichnykh Navuk* 1: 117-119.

Shayiq RM, Raza H, Kidwai AM. 1986. Fluoride and lipid peroxidation a comparative study in different rat tissues. *Bull Environ Contam Toxicol* 37: 70-76.

Shuhua X, Ziyou L, Ling Y, Fei W, Sun G. 2012. A role of fluoride on free radical generation and oxidative stress in BV-2 microglia cells. *Mediators Inflamm* 2012: 1-8.

Singh P, Das TK. 2019. Ultrastructural localization of 4-hydroxynonenal adducts in fluoride-exposed cells: Protective role of dietary antioxidants. *Fluoride* 52(1): 49-58.

Taylor P. 1972. Comparison of the effects of various agents on thyroidal adenyl cyclase activity with their effects on thyroid hormone release. *J Endocrinol* 54: 137-145.

Tu W, Zhang Q, Liu Y, Han LY, Wang Q, Chen PP, Zhang S, Wang AG, Zhou X. 2018. Fluoride induces apoptosis via inhibiting SIRT1 activity to activate mitochondrial p53 pathway in human neuroblastoma SH-SY5Y cells. *Toxicol Appl Pharmacol* 347: 60-69.

van der Voet GB, Schijns O, de Wolff FA. 1999. Fluoride enhances the effect of aluminium chloride on interconnections between aggregates of hippocampal neurons. *Arch Physiol Biochem* 107: 15-21.

Wang JL. 2007. [Effect of fluoride on the intracellular Ca2+ in neurons of mice]. *Chin J Endemiol* 26: 505-507.

Wang J, Gao Y, Cheng X, Yang J, Zhao Y, Xu H, Zhu Y, Yan Z, Manthari RK, Mehdi OM, Wang J. 2019. GSTO1 acts as a mediator in sodium fluoride-induced alterations of learning and memory related factors expressions in the hippocampus cell line. *Chemosphere* 226: 201-209.

Wei N, Dong YT, Deng J, Wang Y, Qi XL, Yu WF, Xiao Y, Zhou JJ, Guan ZZ. 2018. Changed expressions of N-methyl-d-aspartate receptors in the brains of rats and primary neurons exposed to high level of fluoride. *J Trace Elem Med Biol* 45: 31-40.

Willems CB-V, Sande J, Dumont JE. 1972. Inhibition of thyroid secretion by sodium fluoride in vitro. *Biochim Biophys Acta* 264: 197-204.

Woodward JJ, Harms J. 1992. Potentiation of N-methyl-D-aspartate-stimulated dopamine release from rat brain slices by aluminum fluoride and carbachol. *J Neurochem* 58: 1547-1554.

Wu J, Cheng M, Liu Q, Yang J, Wu S, Lu X, Jin C, Ma H, Cai Y. 2015. Protective role of tert-butylhydroquinone against sodium fluoride-induced oxidative stress and apoptosis in PC12 cells. *Cell Mol Neurobiol* 35: 1017-1025.

Xia T, Zhang M, He WH, He P, Wang AG. 2007. [Effects of fluoride on neural cell adhesion molecules mRNA and protein expression levels in primary rat hippocampal neurons]. *Chin J Prev Med* 41: 475-478.

Xu B, Xu Z, Xia T, He P, Gao P, He W, Zhang M, Guo L, Niu Q, Wang A. 2011. Effects of the Fas/Fas-L pathway on fluoride-induced apoptosis in SH-SY5Y cells. *Environ Toxicol* 26: 86-92.

Xu Z, Xu B, Xia T, He W, Gao P, Guo L, Wang Z, Niu Q, Wang A. 2013. Relationship between intracellular Ca2+ and ROS during fluoride-induced injury in SH-SY5Y cells. *Environ Toxicol* 28: 307-312.

Yamashita K, Field JB. 1972. Elevation of cyclic guanosine 3,5; monophosphate levels in dog thyroid slices caused by acetylcholine and sodium fluoride. *J Biol Chem* 247: 7062-7066.

Yan L, Liu S, Wang C, Wang F, Song Y, Yan N, Xi S, Liu Z, Sun G. 2013. JNK and NADPH oxidase involved in fluoride-induced oxidative stress in BV-2 microglia cells. *Mediators Inflamm* 2013: 895-975.

Zhang CY, Chen R, Wang F, Ren C, Zhang P, Li Q, Li HH, Guo KT, Geng DQ, Liu CF. 2016. EGb-761 attenuates the anti-proliferative activity of fluoride via DDK1 in PC-12 cells. *Neurochem Res* 42(2): 606-614.

Zhang M, Wang A, He W, He P, Xu B, Xia T, Chen X, Yang K. 2007. Effects of fluoride on the expression of NCAM, oxidative stress, and apoptosis in primary cultured hippocampal neurons. *Toxicology* 236: 208-216.

Zhang M, Wang A, Xia T, He P. 2008. Effects of fluoride on DNA damage, S-phase cell-cycle arrest and the expression of NF-kappaB in primary cultured rat hippocampal neurons. *Toxicol Lett* 179: 1-5.

Zhang S, Zheng X, Sun Y, Wang Y, Zhang Z. 2015. Alterations in oxidative stress and apoptosis in cultured PC12 cells exposed to fluoride. *Fluoride* 48: 213-222.

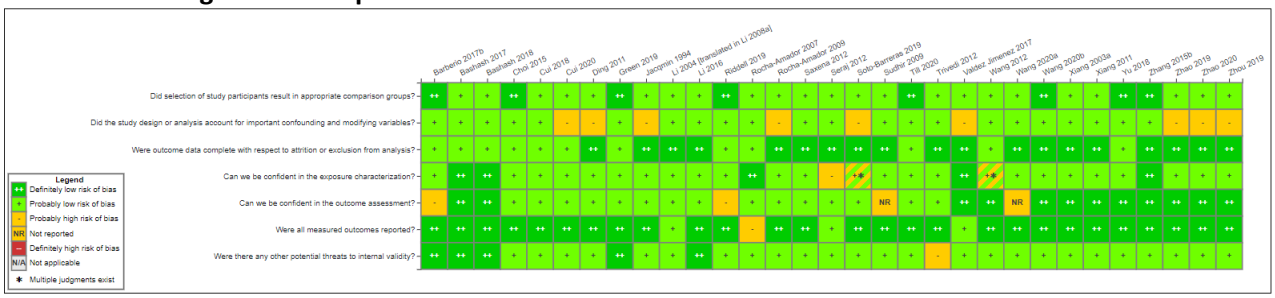
Zhao L, Xiao Y, Deng CM, Tan LC, Guan ZZ. 2016. Protective effect of lovastatin on neurotoxicity of excessive fluoride in primary hippocampal neurons. *Fluoride* 49: 36-46.

Zhao XL, Gao WH, Zhao ZL. 1994. [Effects of sodium fluoride on the activity of Ca2+Mg(2+)-ATPase in synaptic membrane in rat brain]. *Chin J Prev Med* 28: 264-266.

Appendix 3. Risk-of-bias Figures

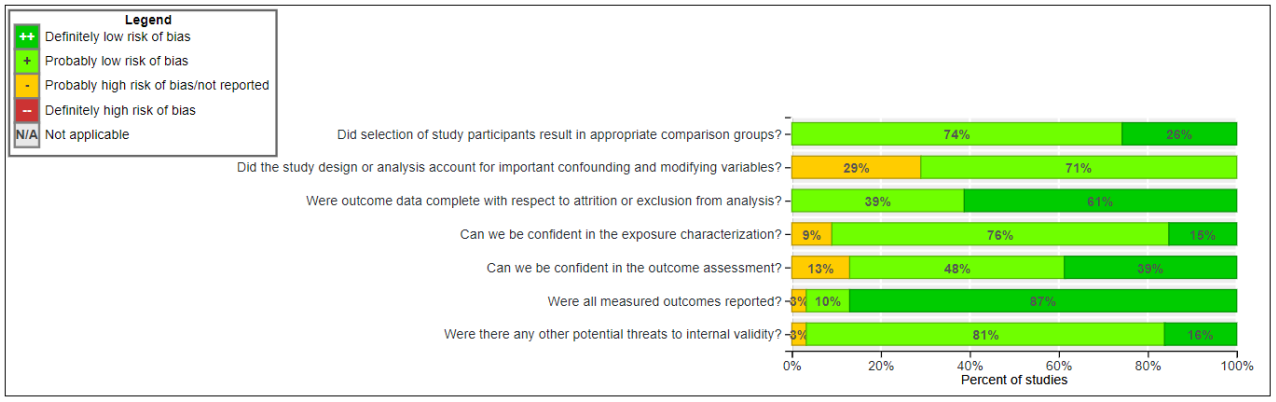
Studies in Humans

Figure A3-1. Risk-of-bias Heatmap for Low Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure



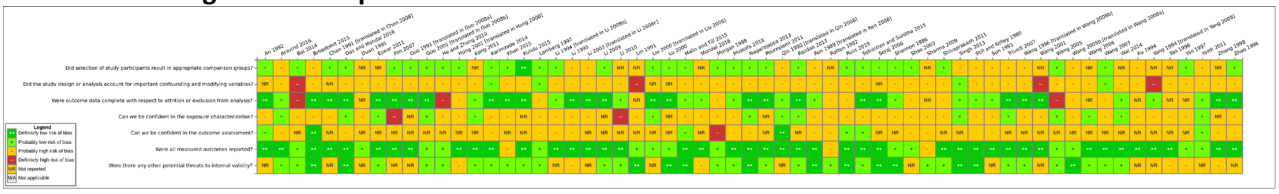
Interactive figure and additional study details in HAWC here.

Figure A3-2. Risk-of-bias Bar Chart for Low Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure



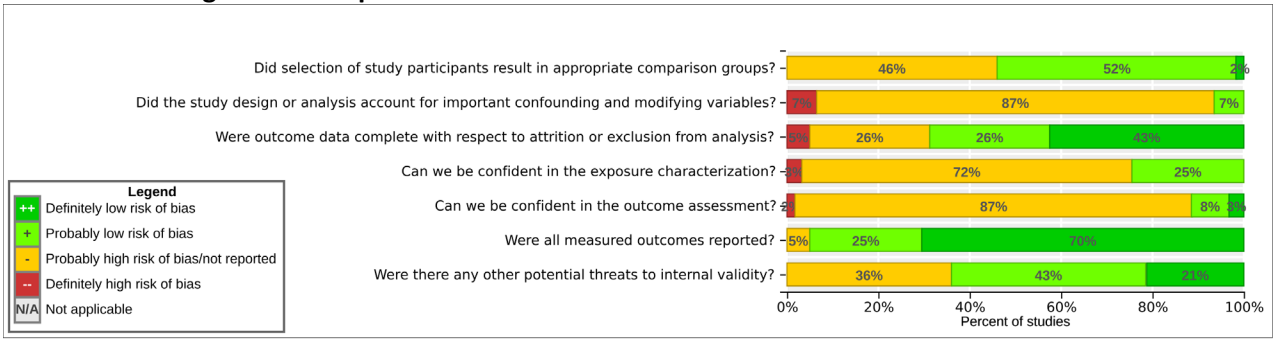
Interactive figure and additional study details in HAWC here.

Figure A3-3. Risk-of-bias Heatmap for High Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure



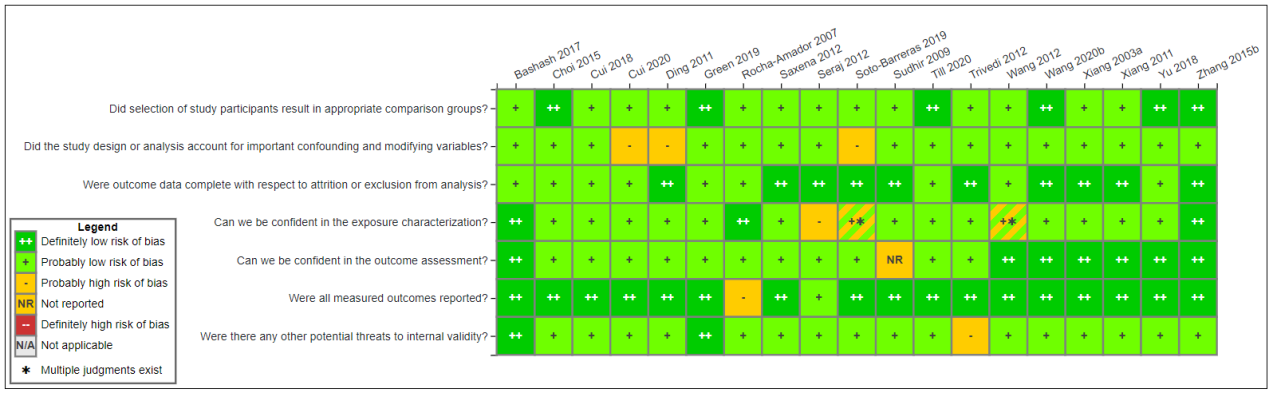
Interactive figure and additional study details in HAWC here.

Figure A3-4. Risk-of-bias Bar Chart for High Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure



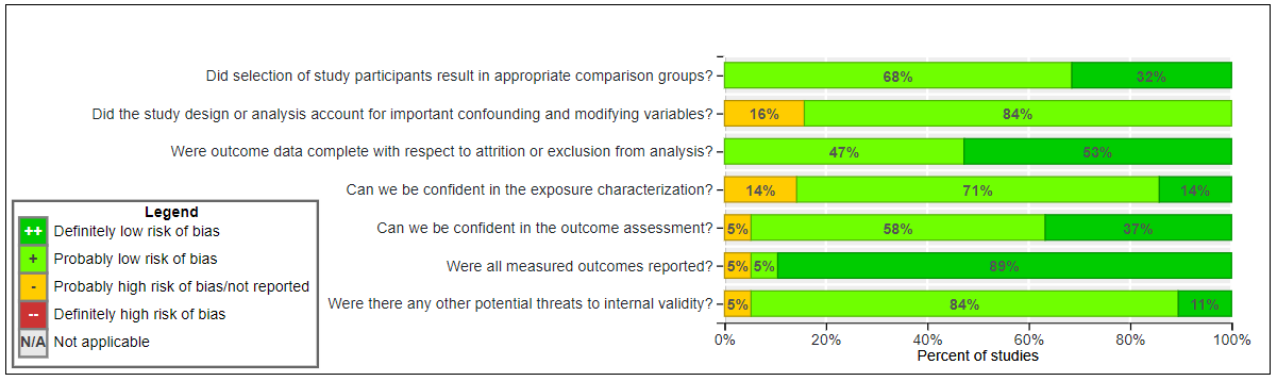
Interactive figure and additional study details in HAWC here.

Figure A3-5. Risk-of-bias Heatmap for Low Risk-of-bias Children's IQ Studies Following Fluoride Exposure



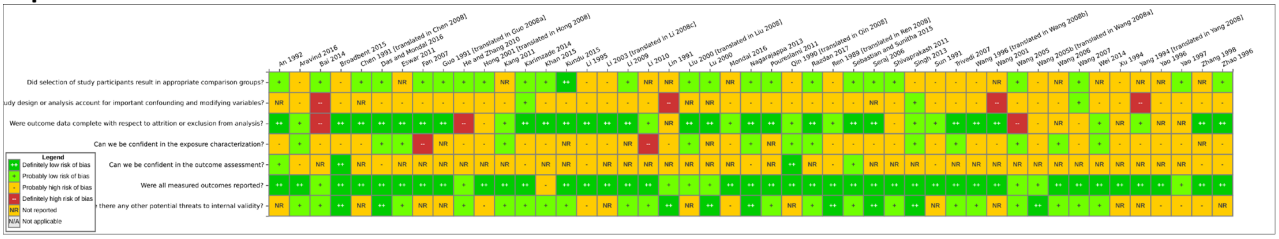
Interactive figure and additional study details in HAWC here.

Figure A3-6. Risk-of-bias Bar Chart for Low Risk-of-bias Children's IQ Studies Following Fluoride Exposure



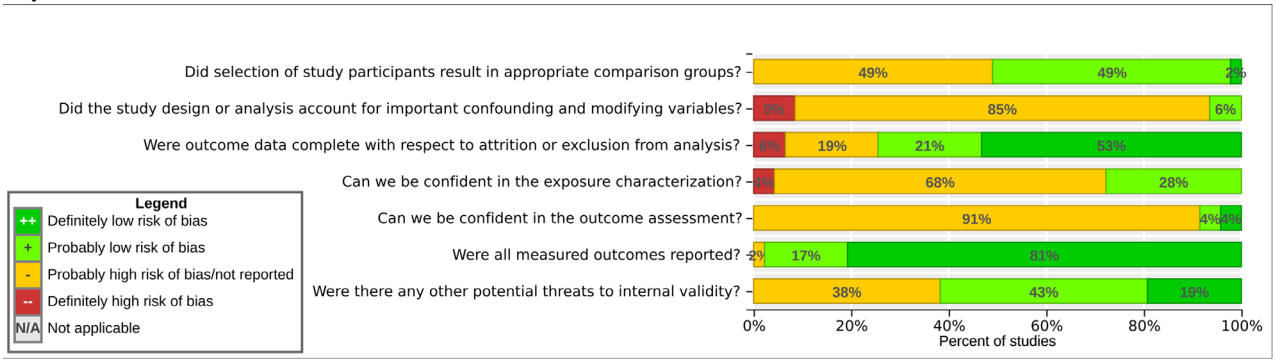
Interactive figure and additional study details in HAWC here.

Figure A3-7. Risk-of-bias Heatmap for High Risk-of-bias Children's IQ Studies Following Fluoride Exposure



Interactive figure and additional study details in HAWC here.

Figure A3-8. Risk-of-bias Bar Chart for High Risk-of-bias Children's IQ Studies Following Fluoride Exposure

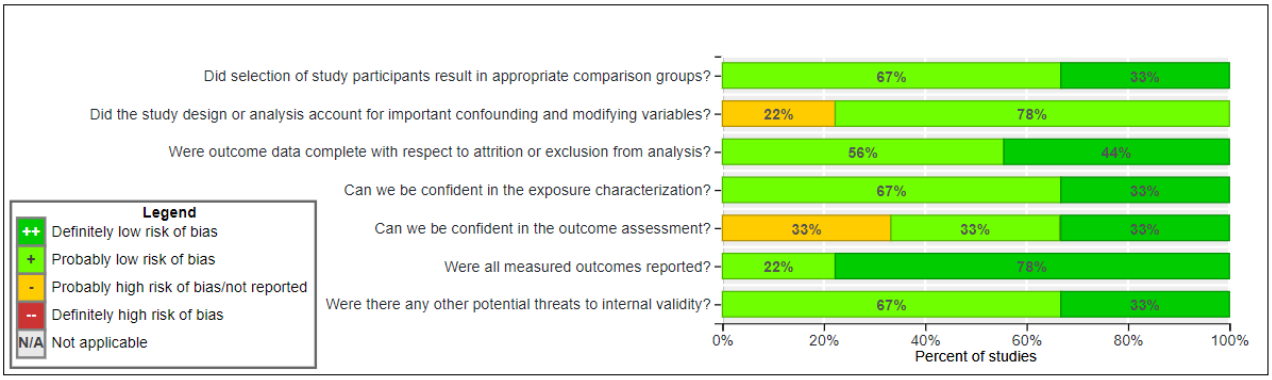


Interactive figure and additional study details in HAWC here.

Figure A3-9. Risk-of-bias Heatmap for Low Risk-of-bias Children's Other Neurodevelopmental Effect Studies Following Fluoride Exposure

Interactive figure and additional study details in HAWC here.

Figure A3-10. Risk-of-bias Bar Chart for Low Risk-of-bias Children's Other Neurodevelopmental Effect Studies Following Fluoride Exposure

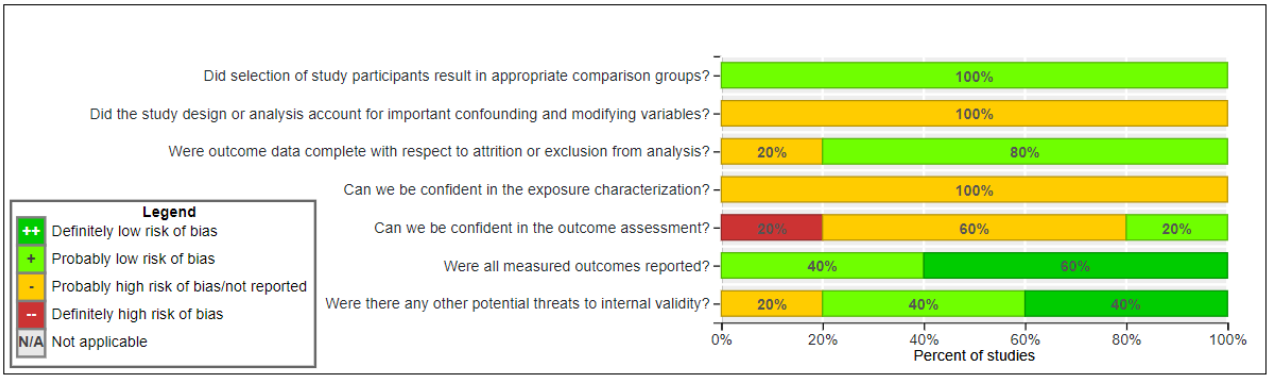


Interactive figure and additional study details in HAWC here.

Figure A3-11. Risk-of-bias Heatmap for High Risk-of-bias Children's Other Neurodevelopmental Effect Studies Following Fluoride Exposure

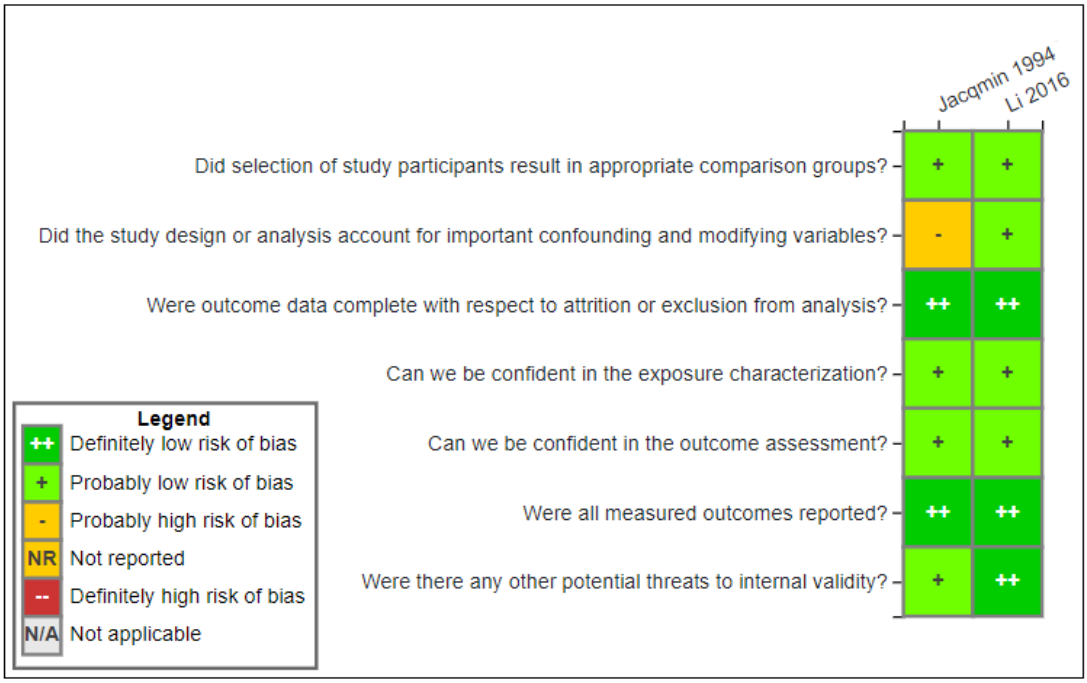
Interactive figure and additional study details in HAWC here.

Figure A3-12. Risk-of-bias Bar Chart for High Risk-of-bias Children's Other Neurodevelopmental Effect Studies Following Fluoride Exposure



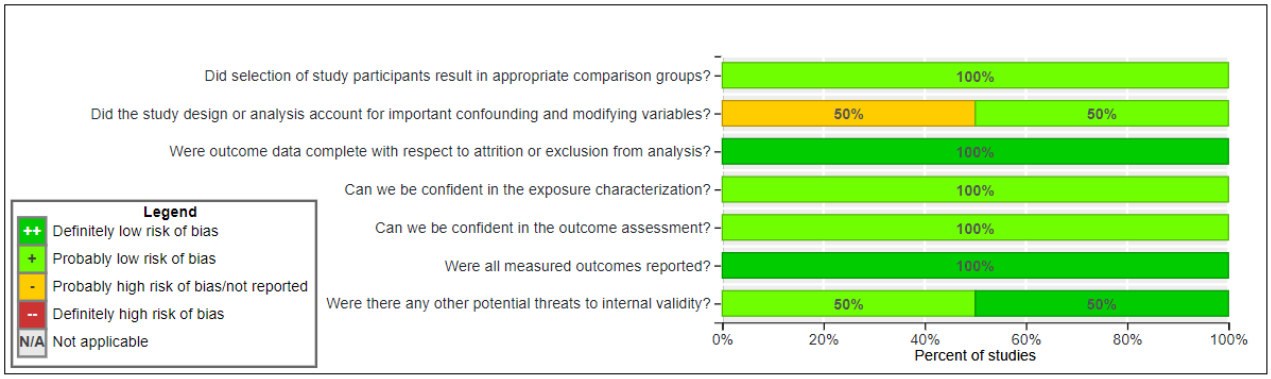
Interactive figure and additional study details in HAWC here.

Figure A3-13. Risk-of-bias Heatmap for Low Risk-of-bias Adult Cognitive Studies Following Fluoride Exposure



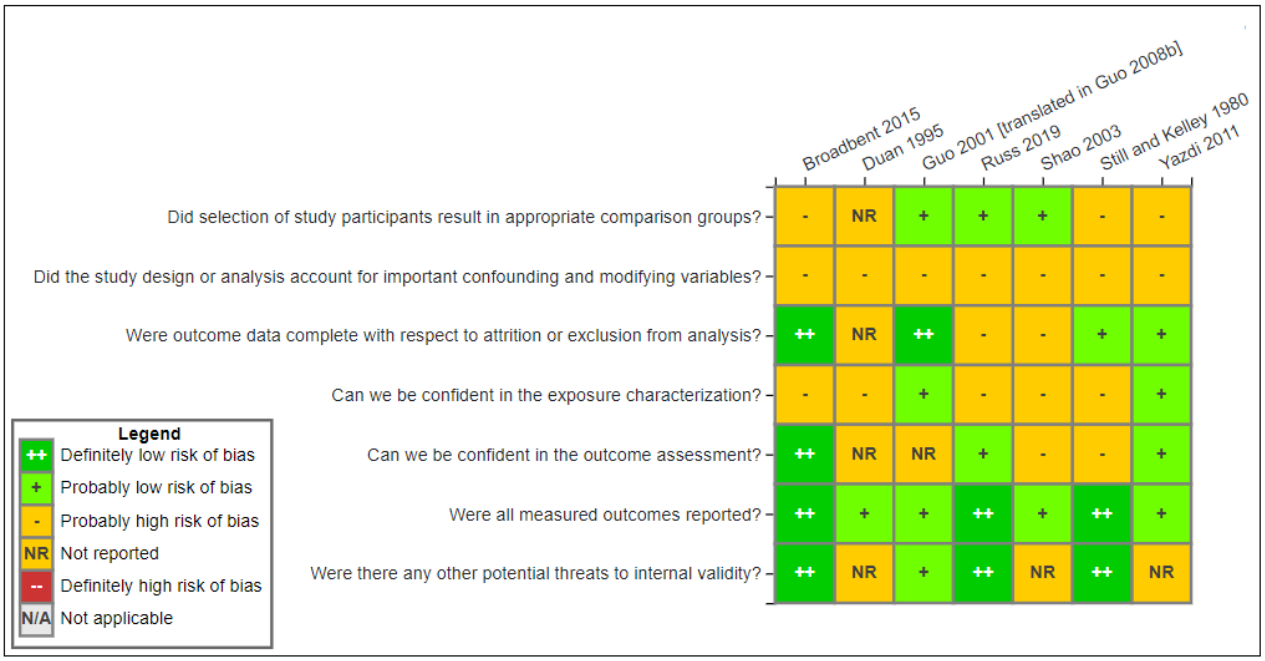
Interactive figure and additional study details in HAWC here.

Figure A3-14. Risk-of-bias Bar Chart for Low Risk-of-bias Adult Cognitive Studies Following Fluoride Exposure



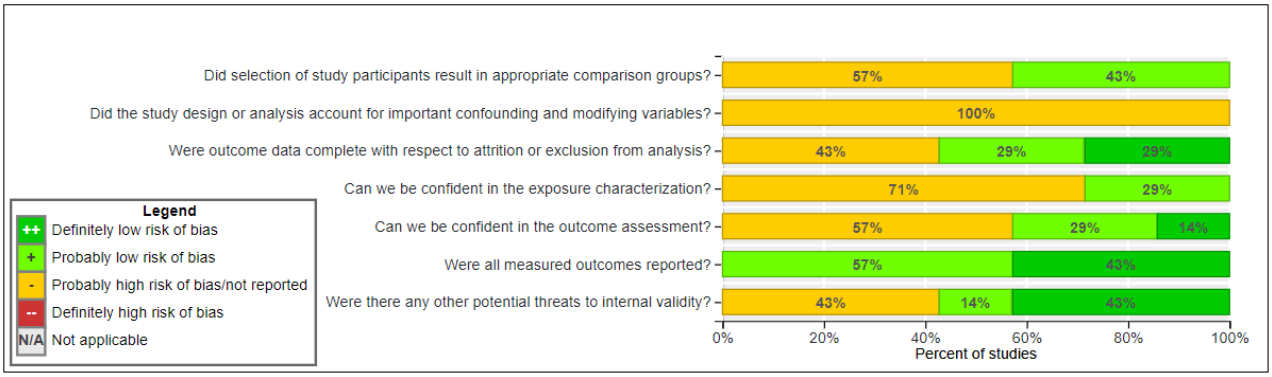
Interactive figure and additional study details in HAWC here.

Figure A3-15. Risk-of-bias Heatmap for High Risk-of-bias Adult Cognitive Studies Following Fluoride Exposure



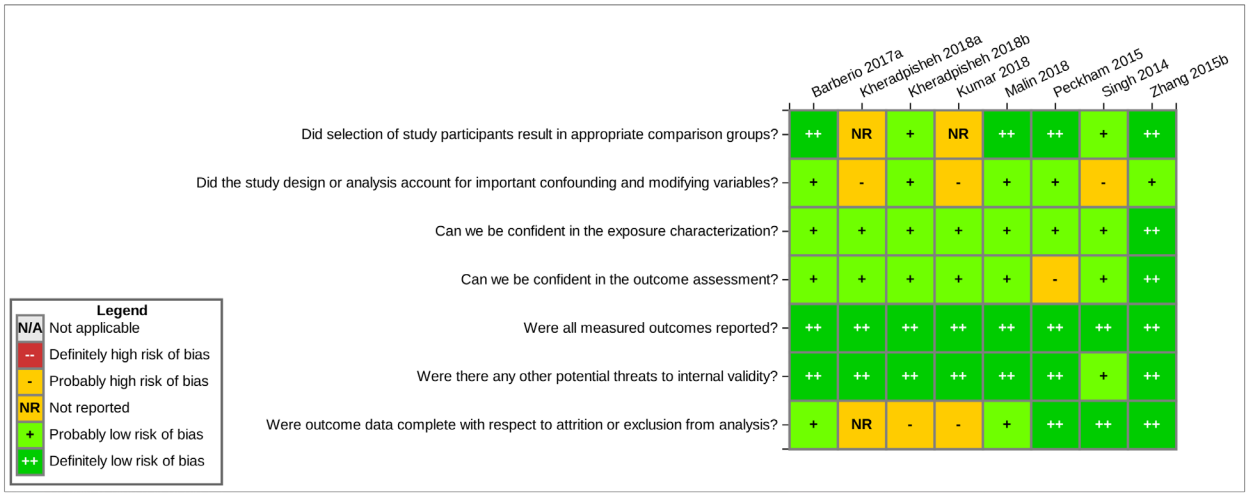
Interactive figure and additional study details in HAWC here.

Figure A3-16. Risk-of-bias Bar Chart for High Risk-of-bias Adult Cognitive Studies Following Fluoride Exposure



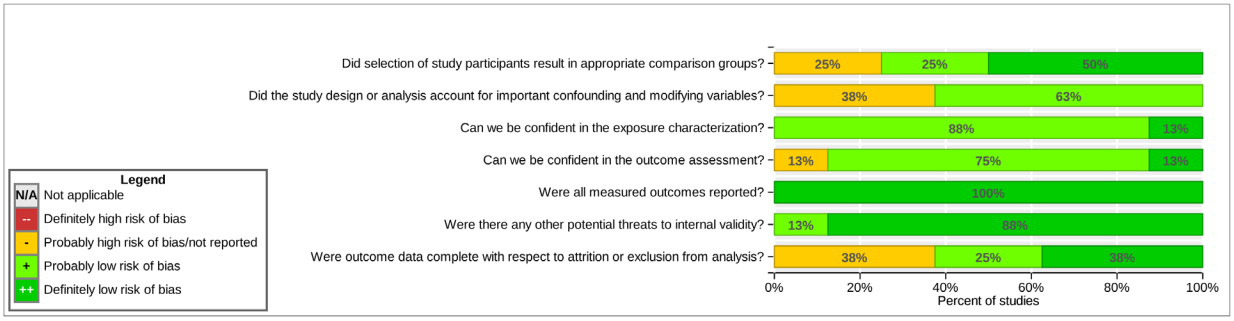
Interactive figure and additional study details in HAWC here.

Figure A3-17. Risk-of-bias Heatmap for Low Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure



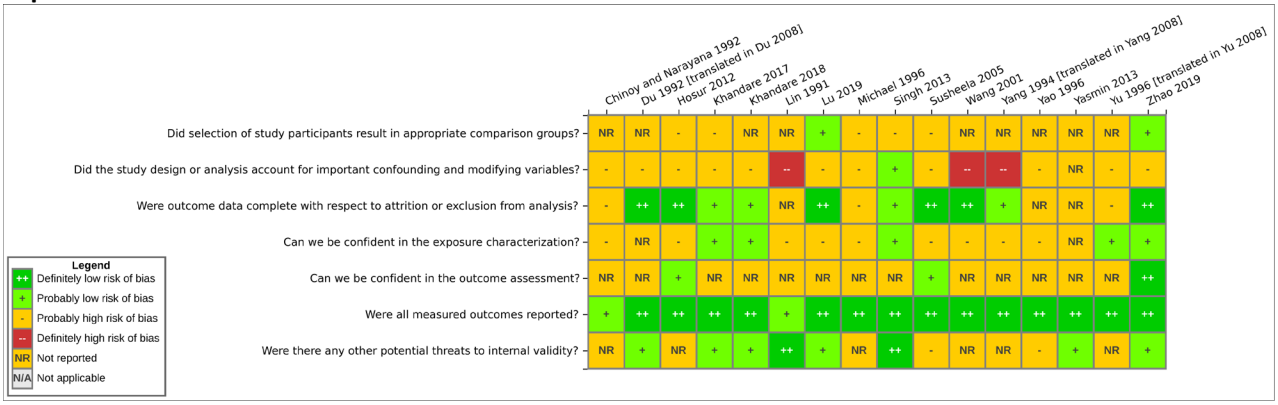
Interactive figure and additional study details in HAWC here.

Figure A3-18. Risk-of-bias Bar Chart for Low Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure



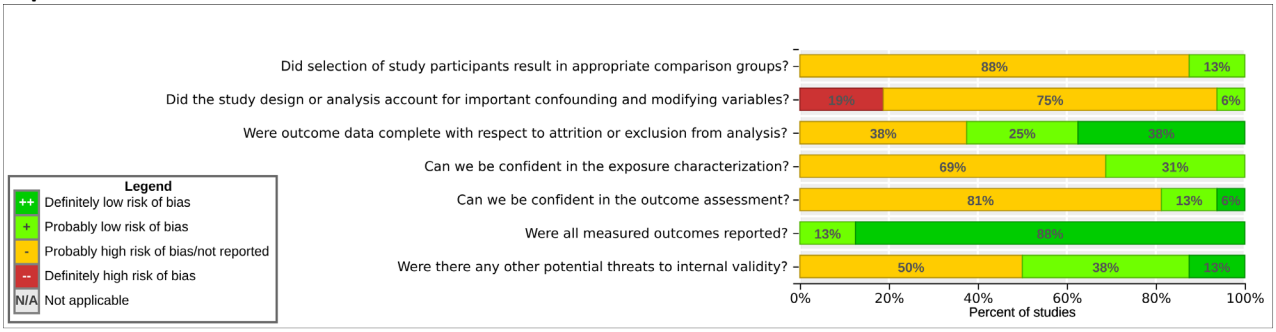
Interactive figure and additional study details in HAWC here.

Figure A3-19. Risk-of-bias Heatmap for High Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure



Interactive figure and additional study details in HAWC here.

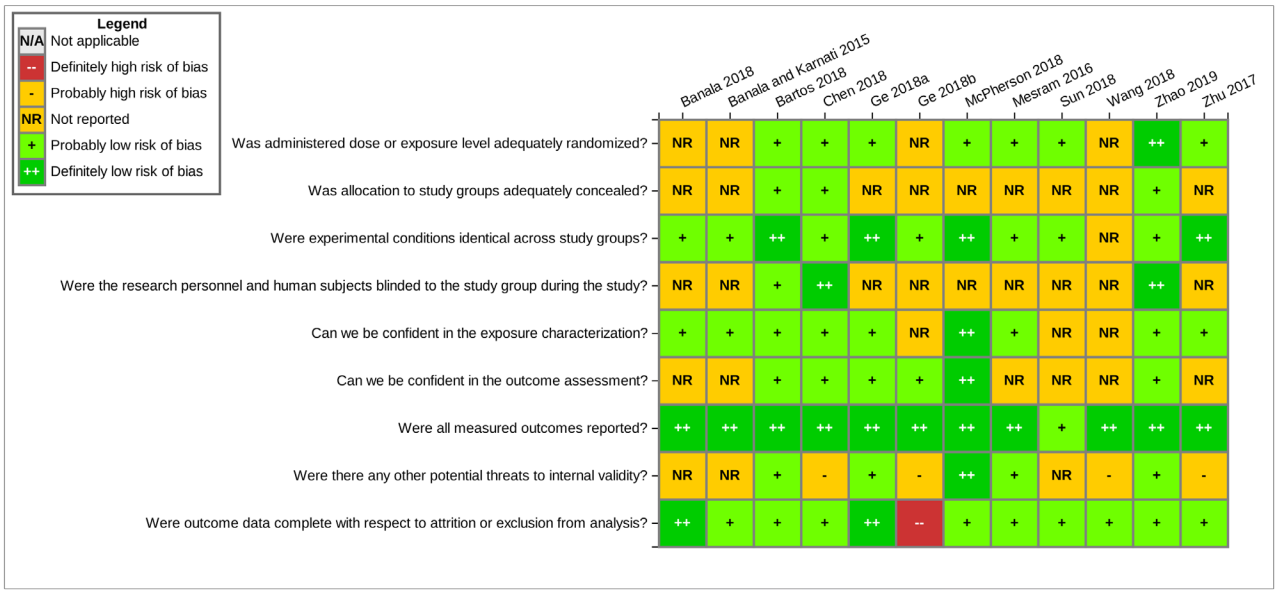
Figure A3-20. Risk-of-bias Bar Chart for High Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure



Interactive figure and additional study details in HAWC here.

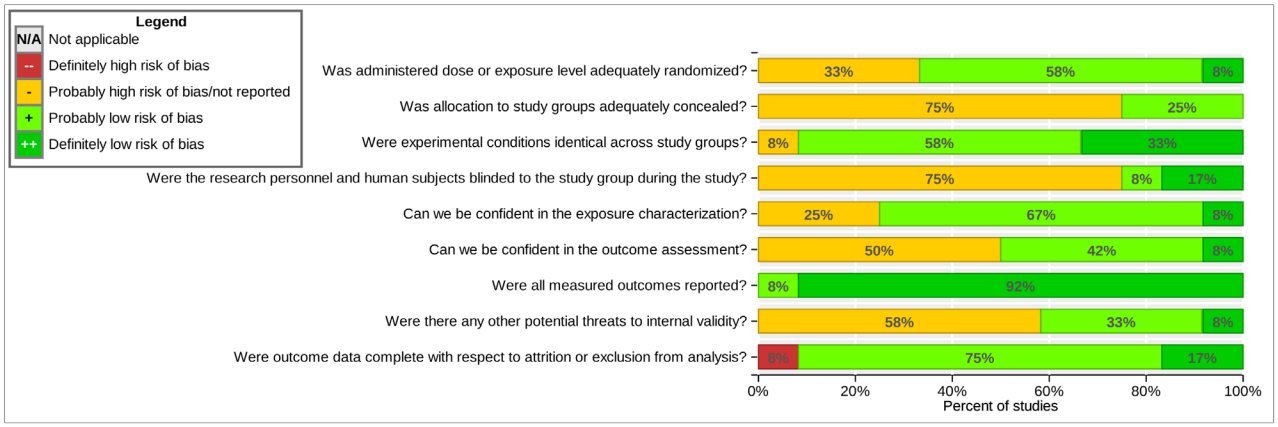
Studies in Non-human Animals

Figure A3-21. Risk-of-bias Heatmap for New Developmental Animal Learning and Memory Studies Following Fluoride Exposure



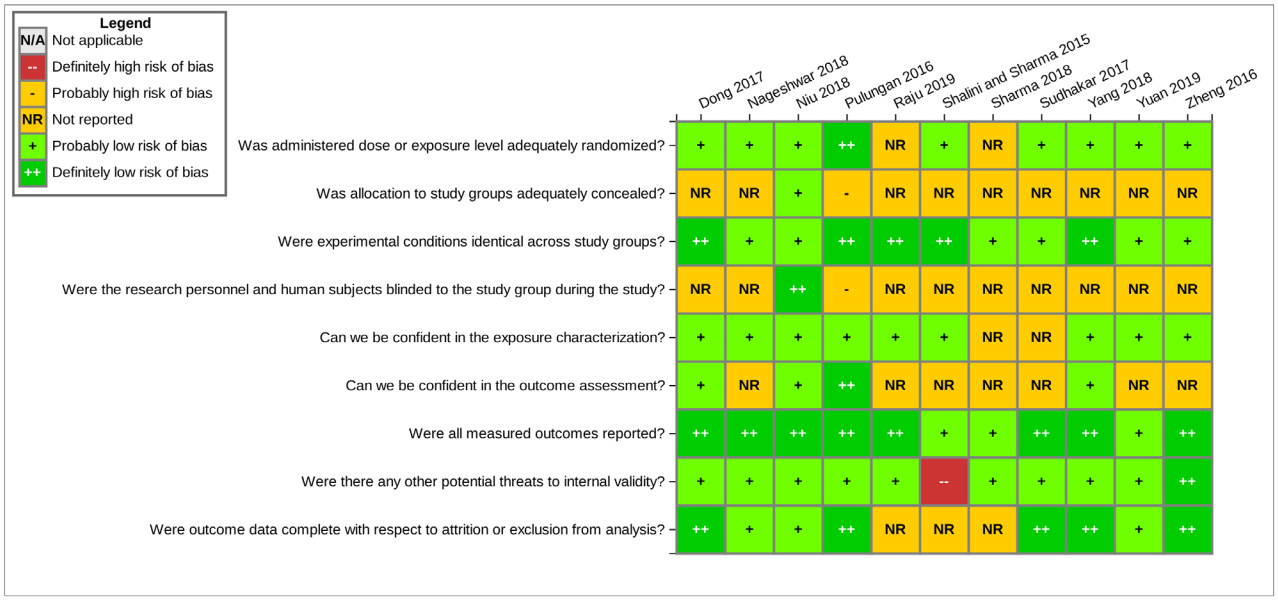
Interactive figure and additional study details in HAWC here.

Figure A3-22. Risk-of-bias Bar Chart for New Developmental Animal Learning and Memory Studies Following Fluoride Exposure



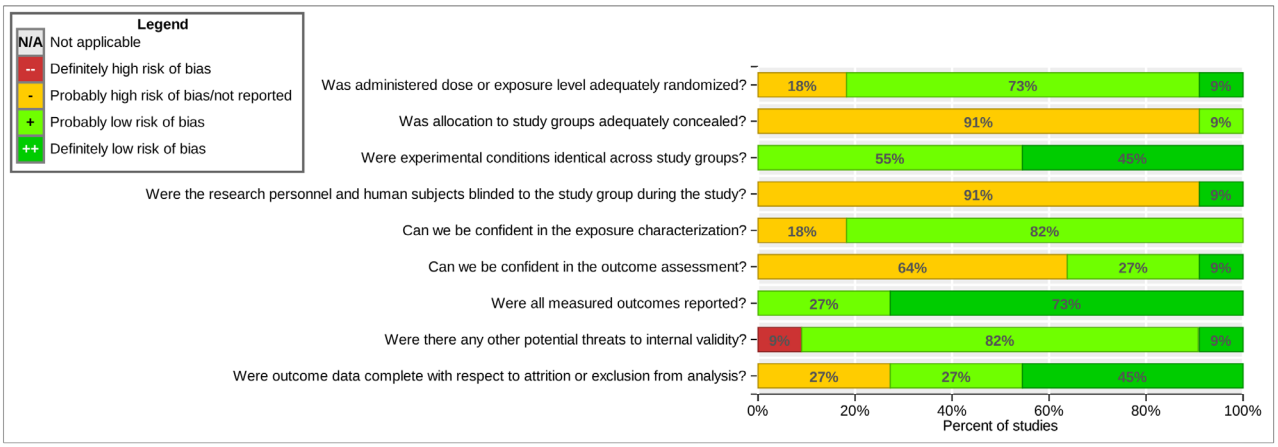
Interactive figure and additional study details in HAWC here.

Figure A3-23. Risk-of-bias Heatmap for New Adult Animal Learning and Memory Studies Following Fluoride Exposure



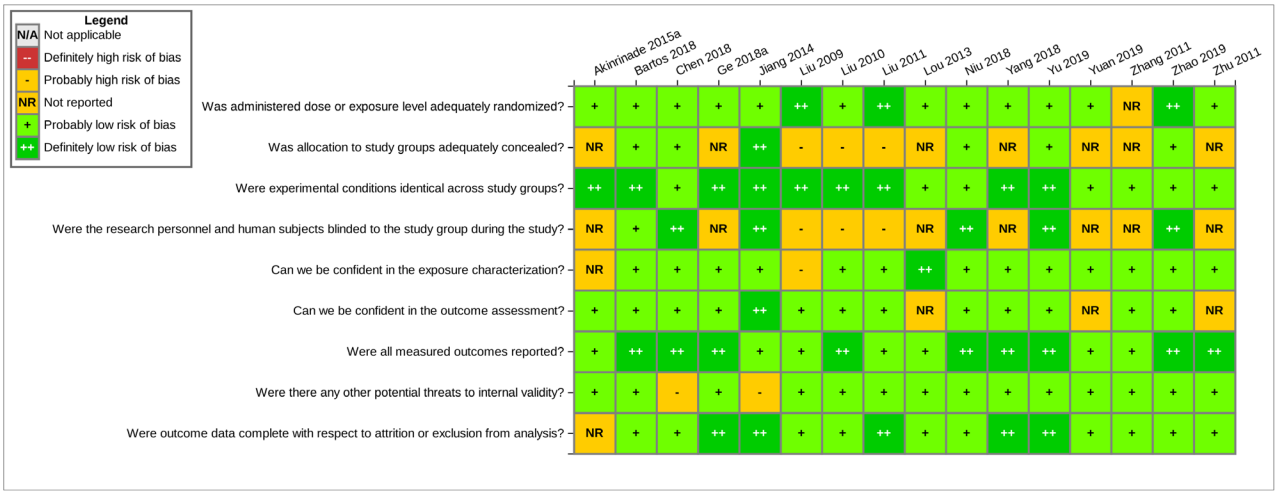
Interactive figure and additional study details in HAWC here.

Figure A3-24. Risk-of-bias Bar Chart for New Adult Animal Learning and Memory Studies Following Fluoride Exposure



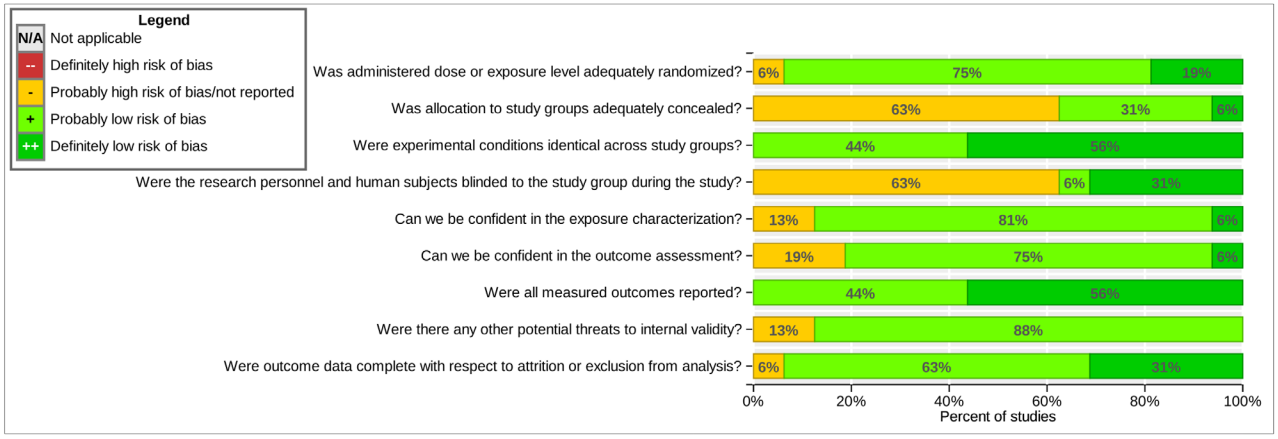
Interactive figure and additional study details in HAWC here.

Figure A3-25. Risk-of-bias Heatmap for Low Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure



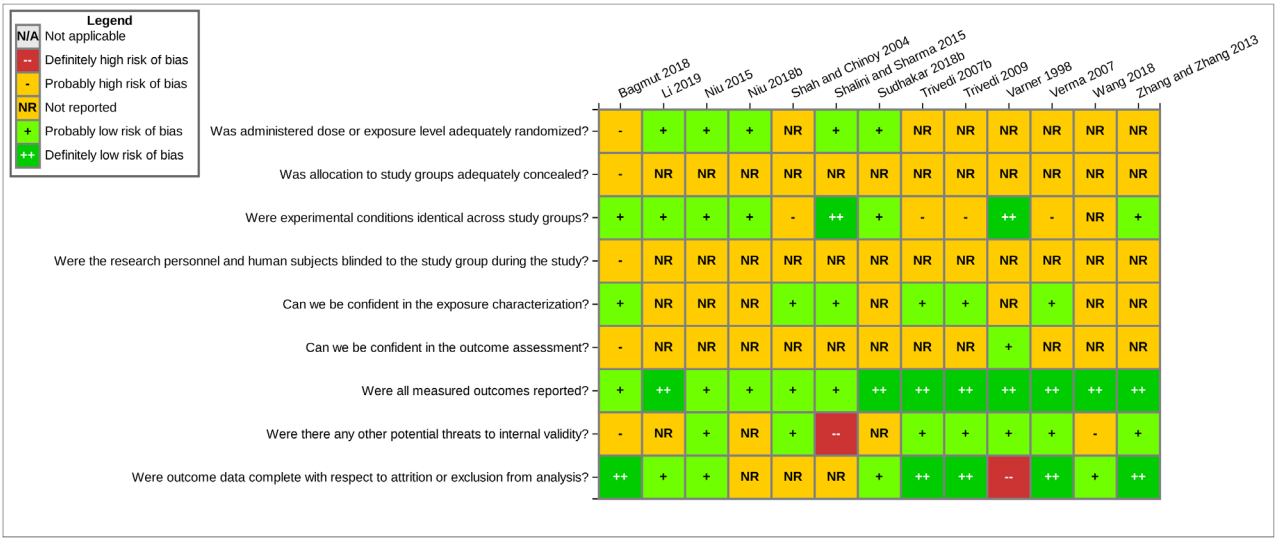
Interactive figure and additional study details in HAWC here.

Figure A3-26. Risk-of-bias Bar Chart for Low Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure



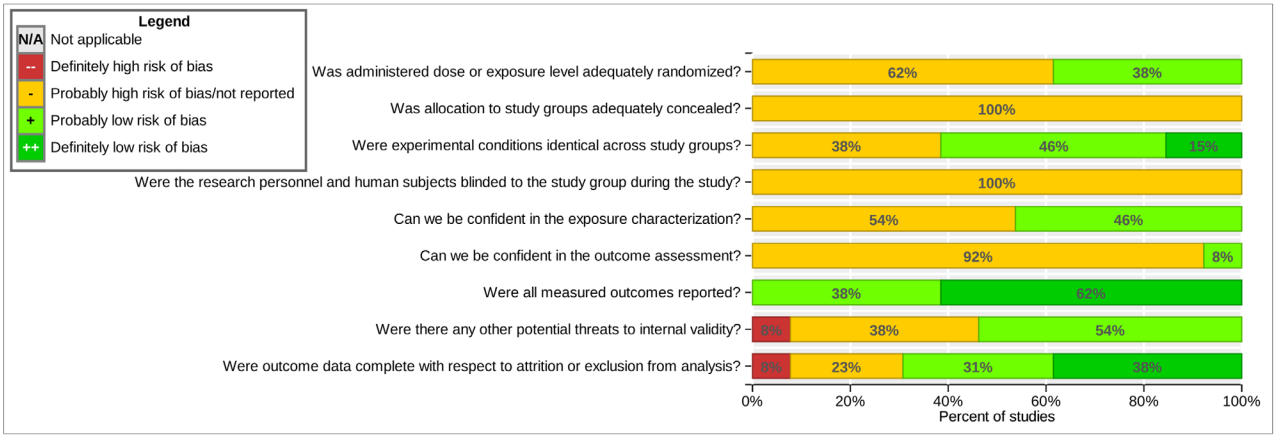
Interactive figure and additional study details in HAWC here.

Figure A3-27. Risk-of-bias Heatmap for High Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure



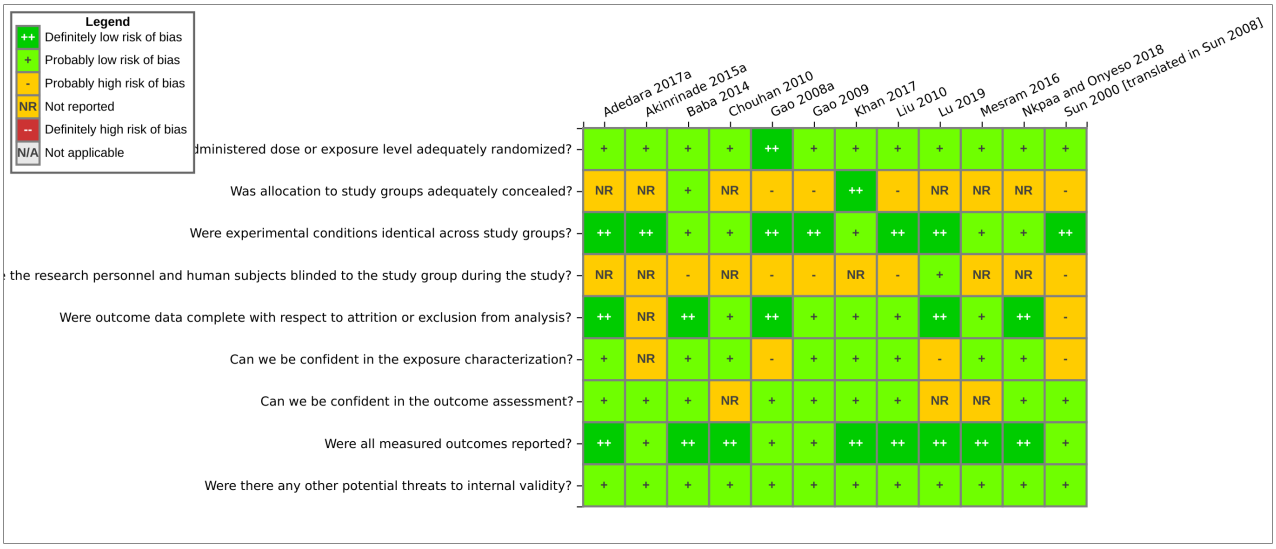
Interactive figure and additional study details in HAWC here.

Figure A3-28. Risk-of-bias Bar Chart for High Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure



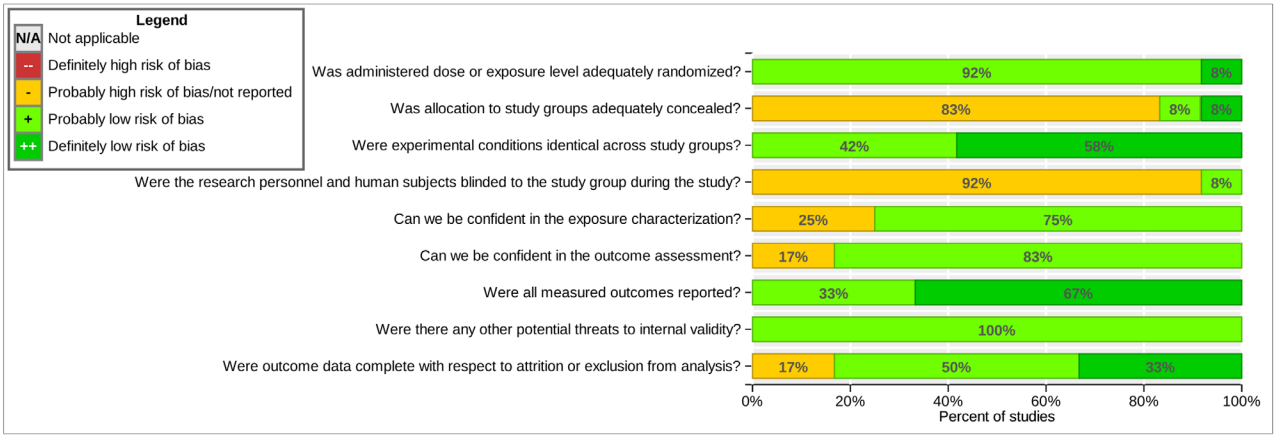
Interactive figure and additional study details in HAWC here.

Figure A3-29. Risk-of-bias Heatmap for Low Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure



Interactive figure and additional study details in HAWC here.

Figure A3-30. Risk-of-bias Bar Chart for Low Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure

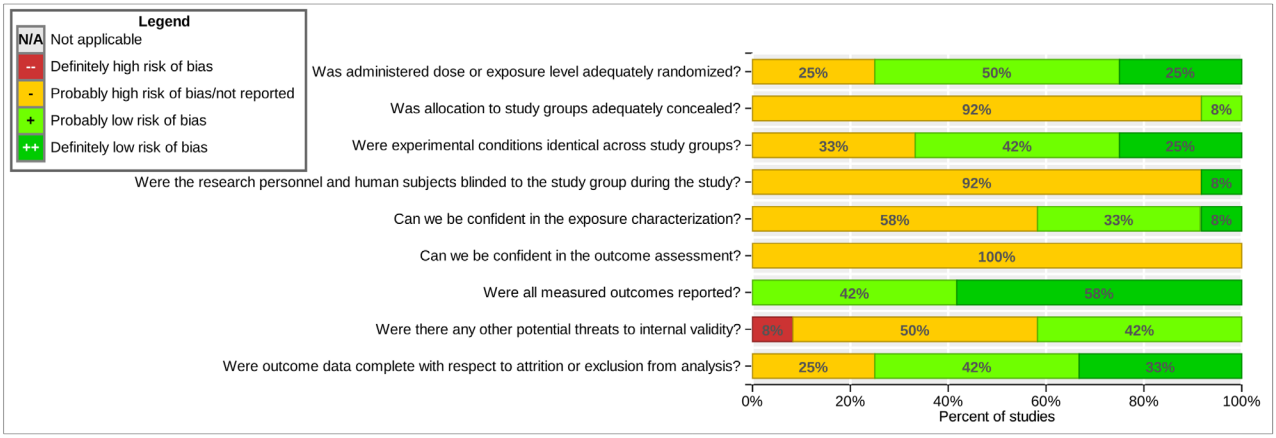


Interactive figure and additional study details in HAWC here.

Figure A3-31. Risk-of-bias Heatmap for High Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure

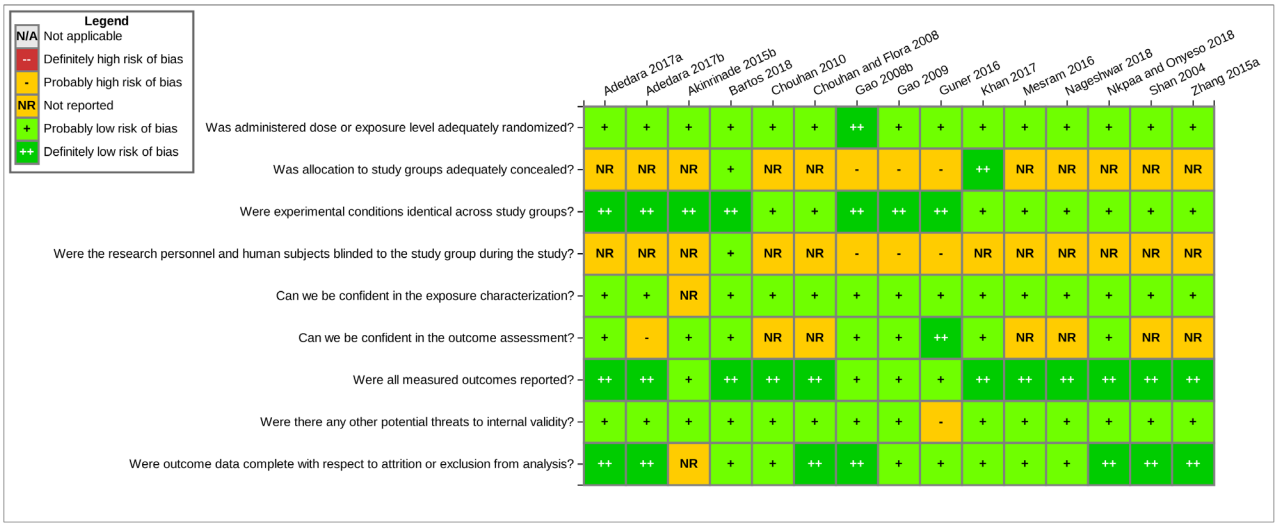
Interactive figure and additional study details in HAWC here.

Figure A3-32. Risk-of-bias Bar Chart for High Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure



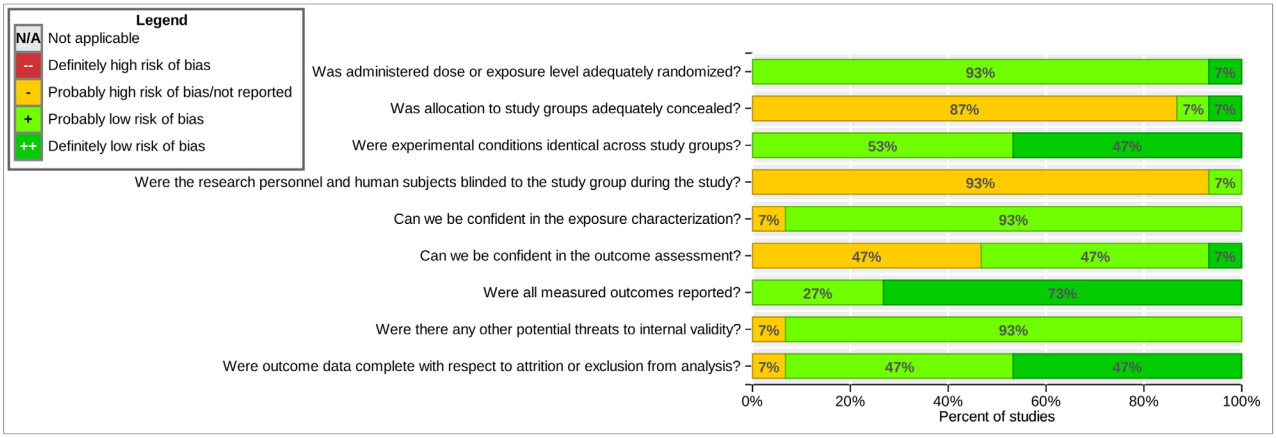
Interactive figure and additional study details in HAWC here.

Figure A3-33. Risk-of-bias Heatmap for Low Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure



Interactive figure and additional study details in HAWC here.

Figure A3-34. Risk-of-bias Bar Chart for Low Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure

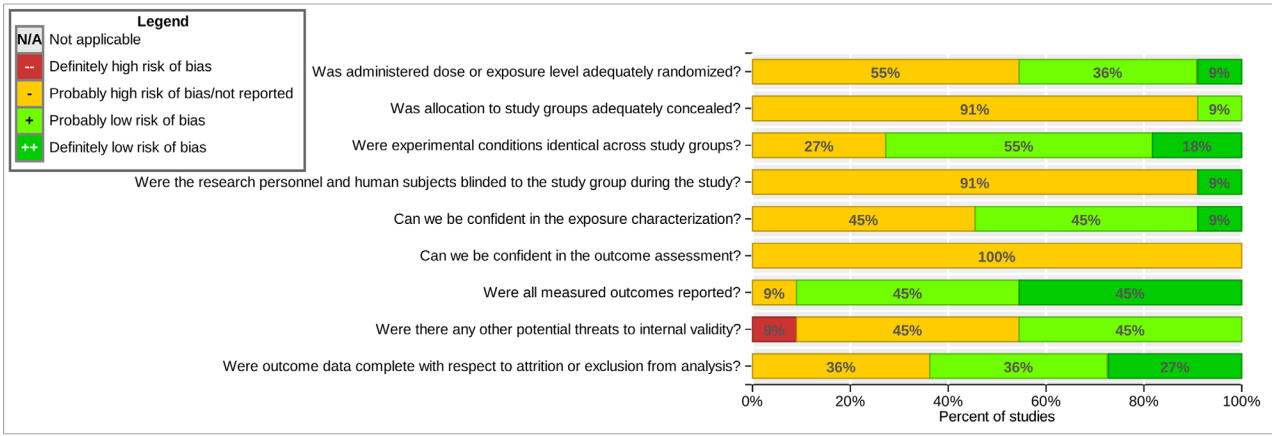


Interactive figure and additional study details in HAWC here.

Figure A3-35. Risk-of-bias Heatmap for High Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure

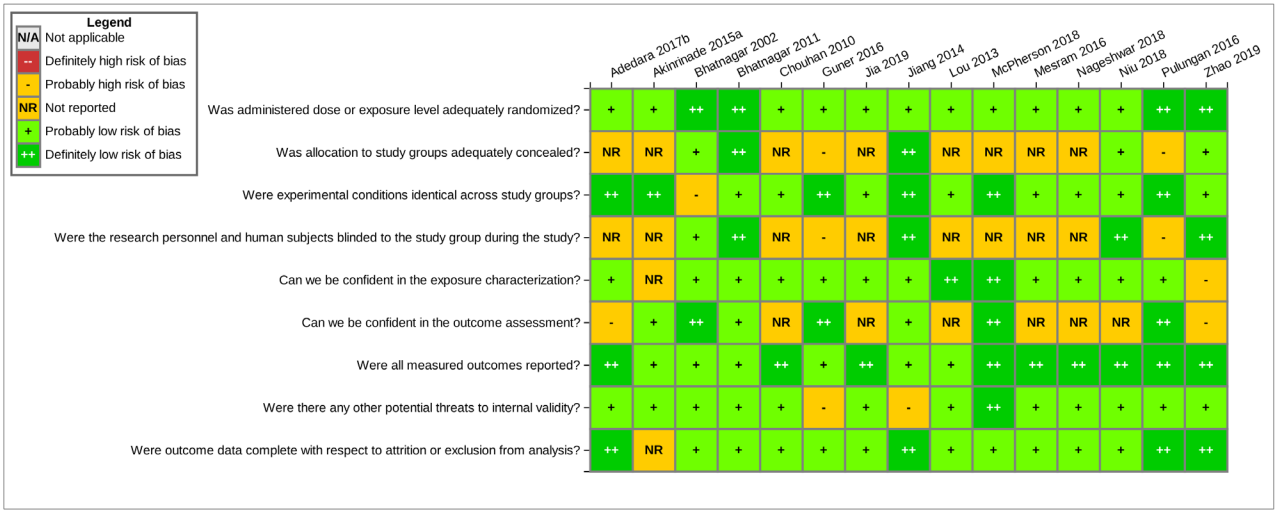
Interactive figure and additional study details in HAWC here.

Figure A3-36. Risk-of-bias Bar Chart for High Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure



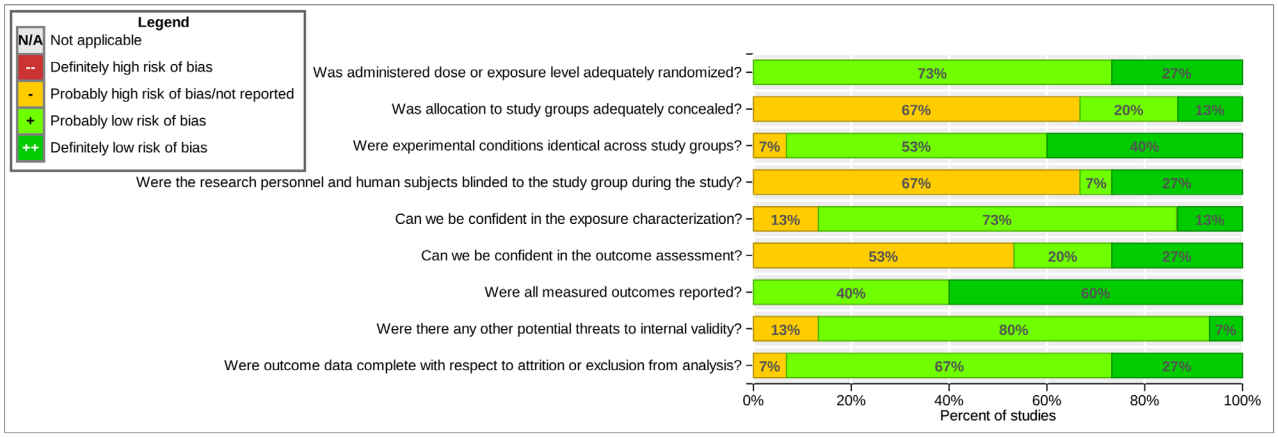
Interactive figure and additional study details in HAWC here.

Figure A3-37. Risk-of-bias Heatmap for Low Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure



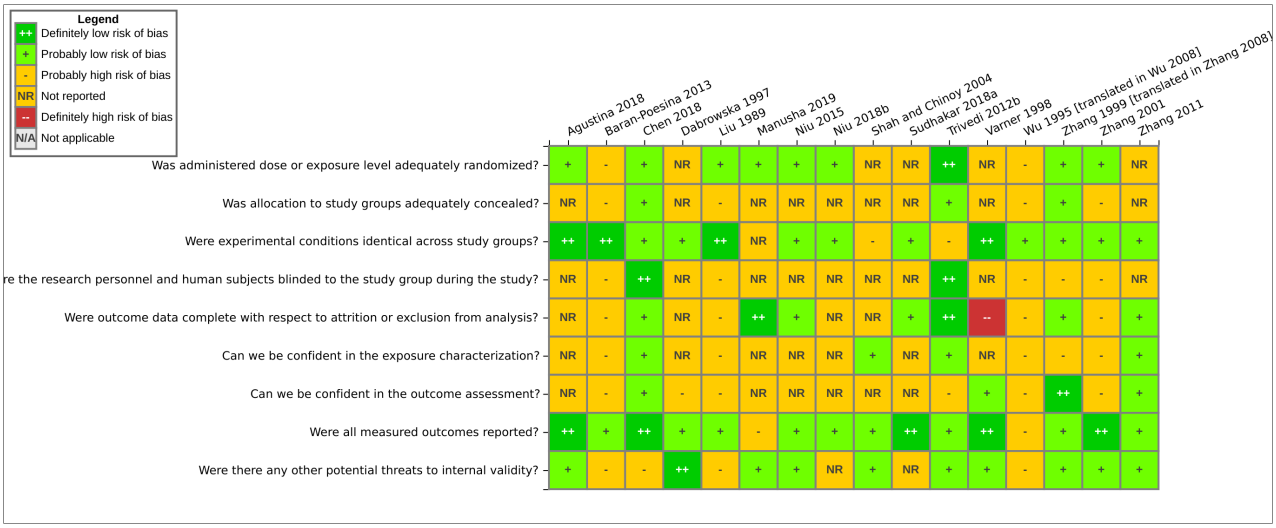
Interactive figure and additional study details in HAWC here.

Figure A3-38. Risk-of-bias Bar Chart for Low Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure



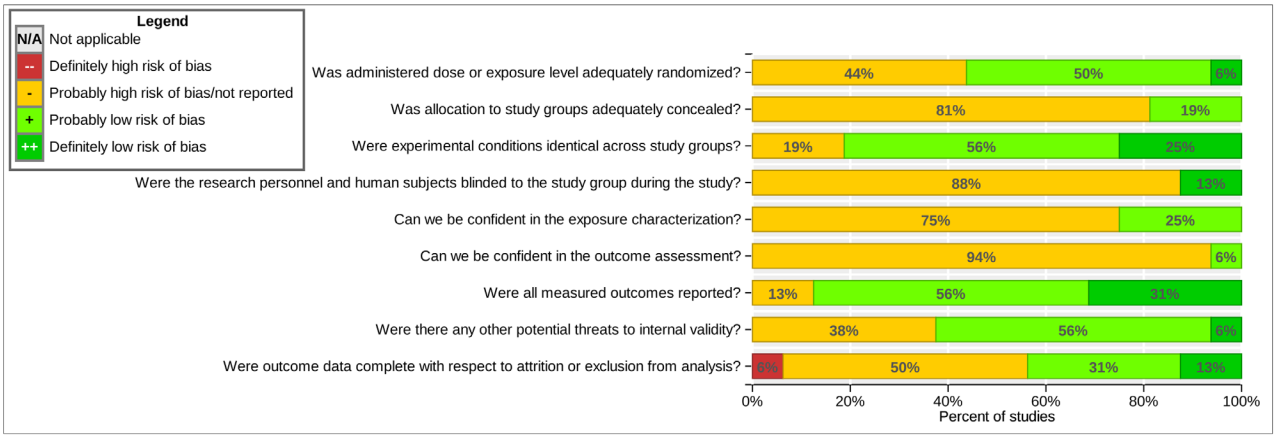
Interactive figure and additional study details in HAWC here.

Figure A3-39. Risk-of-bias Heatmap for High Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure



Interactive figure and additional study details in HAWC here.

Figure A3-40. Risk-of-bias Bar Chart for High Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure



Interactive figure and additional study details in HAWC here.

Appendix 4. Details for Low Risk-of-bias Studies

IQ studies

Bashash *et al.* (2017)

Study Details:

- ***Study design*:** Prospective cohort
- ***Population*:** Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants (pregnant mothers and their children aged 4 or 6–12 years).
- ***Study area:*** Mexico City, Mexico
- ***Sample size:*** 299 mother–child pairs, of whom 211 had data for the IQ analyses.
- ***Data relevant to the review:*** Adjusted and unadjusted associations between IQ scores and maternal or child's urinary fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant association between maternal urinary fluoride and IQ score (adjusted β = −2.50; 95% CI: −4.12, −0.59). No significant associations with children's urinary fluoride.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Study participants were selected from two different cohorts from three hospitals in Mexico City that serve low-to-moderate income populations. One cohort was from an observational study of prenatal lead exposure and neurodevelopment outcomes and the other was from a randomized trial of the effect of calcium on maternal blood lead levels. The authors state that participants had no history of psychiatric disorders, high-risk pregnancies, gestational diabetes, illegal drug use, or continuous prescription drugs, but they do not include any information on smoking habits. Study populations appear to be similar, but there may be some differences because subjects were selected from two different cohorts that were recruited from slightly different time periods.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar despite the subjects coming from different original study populations where different methods were used for recruitment.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Data were collected via questionnaire on maternal age, education, marital status at first prenatal visit, birth order, birth weight, gestational age at delivery, maternal smoking, maternal IQ, and HOME scores. All models were adjusted for gestational age at birth, child's sex, birth weight, birth order, child's age at testing, maternal marital status, smoking history, age at delivery, maternal IQ, education, and cohort, with additional testing for children's urinary fluoride, mercury, lead, and calcium. Sensitivity analyses additionally adjusted for HOME score. Confounders not considered included BMI, iodine deficiency, arsenic, and maternal mental health and

nutrition. Arsenic is assumed not to be a potential co-exposure in this population as the study authors did not discuss it as an issue but did discuss other co-exposures. Arsenic is included in the water quality control program in Mexico City and is not considered a concern in this population.

- *Potentially important study-specific confounders:* All key confounders were addressed.
  - *Direction/magnitude of effect*: Not applicable.
- *Basis for rating:* Probably low risk of bias based on direct evidence that key confounders including other potential co-exposures were addressed and indirect evidence that the methods used to collect the information were valid and reliable and that arsenic is not likely to be an issue in this study population.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although there was a large amount of attrition, the study authors clearly describe all reasons for attrition and also provide characteristics to compare those participants included to those excluded. There were some slight differences between those included and those excluded, but there is nothing to indicate that the attrition would potentially bias the results.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Urinary fluoride concentrations were determined in spot urine samples (2nd morning void) collected from mothers (during at least one trimester) and children ages 6–12 years. Fluoride content was measured using ion-selective electrode-based assays. QC methods were described including between laboratory correlations. All samples were measured in duplicate. Extreme outliers were excluded. Urinary dilution was addressed by using creatinine-adjusted levels.
    - *Direction/magnitude of effect*: Not applicable.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Outcome was assessed using the McCarthy Scales of Children's Abilities (MSCA) in 4-year-old children (translated into Spanish) and the Wechsler Abbreviated Scale of Intelligence (WASI) in 6–12-year-olds. The WASI is a well-established test and the validity of both tests is well documented by the authors. Inter-examiner reliability was evaluated and reported with a correlation of 0.99 (++ for methods). The study report stated that psychologists were blind to the children's fluoride exposure (++ for blinding). Overall rating for methods and blinding = ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)

- *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
- *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.

- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study. Statistical tests of bivariate associations [using Chi-square tests for categorical variables and analysis of variance (ANOVA)] were used to compare the means of the outcomes or exposure within groups based on the distribution of each covariate. Generalized additive models (GAMs) were used to estimate the adjusted association between fluoride exposure and measures of children's intelligence. Residual diagnostics were used to examine model assumptions and identify any potentially influential observations. Results are reported as adjusted effects and 95% CIs. In sensitivity analyses, regression models accounted for clustering at the cohort level by using cohort as a fixed effect in the models. Although using cohort as a random effect would be more appropriate, using individual-level exposure data and accounting for numerous potential confounders in the models likely captured the cohort effect. Additional models with cohort as a random effect were also subsequently made available via personal communication with the study authors and showed similar results to the main model.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, outcome blindly assessed, and the prospective cohort study design.

Choi *et al.* (2015)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** First grade children (ages 6–8 years)
- ***Study area:*** Mianning County in southern Sichuan, China
- ***Sample size:*** 51 first grade children
- ***Data relevant to the review:*** Associations between IQ (digit span for auditory span and working memory and block design for visual organization and reasoning components of WISC-IV only) with continuous urine or drinking water fluoride levels. Study also had information based on dental fluorosis score.
- ***Reported association with fluoride exposure:*** Yes: Compared to the normal/questionable dental fluorosis, the moderate/severe dental fluorosis group was associated with significantly lower total (adjusted β = −4.28; 95% CI: −8.22, −0.33) and backward (adjusted β = −2.13; 95% CI: −4.24,

−0.02) digit span scores. Linear correlations between total digit span and fluoride in urine (adjusted β = −1.67; 95% CI: −5.46, 2.12) and in drinking water (adjusted β = −1.39; 95% CI: −6.76, 3.98) were observed but not significant. Other outcomes not significantly associated with fluoride exposure.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were selected during the same time frame using the same methods. Fifty-one first-grade children residing in Mianning County in southern Sichuan, China were included in this pilot study. It is not specified if the 51 children represented all the first-grade children from this area or if some refused to participate. Children who did not speak Chinese, were not students at the Primary School of Sunshui Village in Mianning County, or those with chronic or acute disease that might affect neurobehavioral function tests were excluded. Demographic characteristics are presented in Table 1 of the study, which indicates that subjects were similar. Potential confounders are adjusted for in the statistical analyses.
  - *Basis for Rating:* Definitely low risk of bias based on direct evidence that the exposure groups were similar and were recruited within the same time frame using the same methods with no evidence of differences in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The parents or guardians completed a questionnaire on demographic and personal characteristics of the children (sex, age at testing, parity, illnesses before age 3, and past medical history) and caretakers (age, parity, education and occupational histories, residential history, and household income). A 20-µL capillary blood sample was collected at the school by a Mianning County Center for Disease Control (CDC) health practitioner and tested for possible iron deficiency which could be used as a covariate of neurodevelopmental performance. Confounders that were not assessed include: maternal BMI, parental mental health, maternal smoking status, maternal reproductive factors, parental IQ, and HOME score. However, the study authors noted that confounding bias appeared to be limited due to the minimal diversity in the social characteristics of the subjects. The study authors indicated that CDC records documented that levels of other contaminants including arsenic and lead were very low in the area. Iodine differences were not specifically addressed, but there is no indication from the information provided that this might be a concern.
  - *Potentially important study-specific confounders:* All key confounders were considered in this study.
    - *Direction/magnitude of effect*: Not applicable.
  - *Basis for rating:* Probably low risk of bias because there is direct evidence that the key confounders are taken into account and indirect evidence that co-exposure to arsenic is likely not an issue in this area and that methods used for collecting the information were valid and reliable.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The majority of results were reported for the 51 children stated to be included in the pilot study. In Table 5 of the study, the N for each dental fluorosis category only totals 43, but the text indicates 8 children did not have a Dean Index because permanent teeth had not erupted.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study used three different measurements of fluoride exposure: well water fluoride concentrations from the residence during pregnancy and onwards, fluoride concentrations from children's first morning urine samples, and degree of children's dental fluorosis. Fluoride concentrations in community well water were measured and recorded by Mianning County CDC; specific methods were not reported, but they likely used standard methods as they were conducted by the CDC and were likely the same as those used to measure the fluoride in urine. Migration of subjects was noted to be limited. Well water fluoride concentrations of the mother's residence during pregnancy and onward were used to characterize a child's lifetime exposure. To provide a measure of the accumulated body burden, each child was given a 330-mL (11.2-oz) bottle of Robust© distilled water (free from fluoride and other contaminants) to drink the night before the clinical examinations, after emptying the bladder and before bedtime. The first urine sample the following morning was collected at home, and the fluoride concentration was determined on a 5-mL sample using an ion-specific electrode at the Mianning CDC. There is no indication that urinary fluoride levels accounted for dilution nor was it clear that the method of administering water to the children and collection methods sufficiently controlled for differences in dilution. One of the investigators, a dentist, performed a blinded dental examination on each child's permanent teeth to rate the degree of dental fluorosis using the Dean Index. The Dean Index is a commonly used index in epidemiological studies and remains the gold standard in the dentistry armamentarium. The Index has the following classifications: normal, questionable, very mild, mild, moderate, and severe. Quality control (QC) procedures are not reported but were likely appropriate.
    - *Direction/magnitude of effect***:** Current levels were used to assess lifetime exposure. This is likely to be a non-differential exposure misclassification and direction of bias is unknown. Because subject migration appears to be limited, it is likely that the current fluoride levels are adequate reflections of past exposure. Dental fluorosis would be an indicator that exposure occurred in the past and there was a fair correlation between degree of dental fluorosis and current urine and water fluoride levels, with both increasing with increasing levels of dental fluorosis.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study authors adopted culture-independent tests considered feasible for children aged 6 to 8 years. The Wide Range Assessment of Memory and Learning (WRAML) was used for the assessment of memory and learning. Three subtests were also used. The Finger Windows subtest assesses sequential visual memory. The Design Memory subtest assesses the ability to reproduce designs from memory following a brief exposure. The Visual Learning subtest assesses the ability to learn the locations of pictured objects over repeated exposures. The Wechsler Intelligence Scale for Children-Revised (WISC-IV) included digit span for auditory span and working memory and block design for visual organization and reasoning. The grooved pegboard test assesses manual dexterity. The tests used have been validated on a western population. Although there is no information provided to indicate that they were validated on the study population, the study authors indicated that the tests were culture-independent (+ for methods). Blinding of the outcome assessors or steps to minimize potential bias was not reported. However, it is unlikely that the assessors had knowledge of the individual exposure as children all came from the same area, and water and urine levels were tested at the CDC. (+ for blinding). Overall = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that all outcomes were assessed blindly using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses were appropriate. Multiple regression models evaluate the associations between exposure indicators and test scores after adjusting for potential confounders. Specific regression models are not described or refenced, just stated to be “standard regression analysis with confounder adjustment.” The distributions of fluoride concentrations in urine and water were skewed and were log10-transformed to approximate a Gaussian distribution (test not specified). Results are reported as adjusted effects and 95% CIs. There is no evidence that residual diagnostics were used to examine model assumptions; however, the impact on the effect estimates is expected to be minimal.
    - *Other potential concerns:* It should be noted that this study was a pilot study and, therefore, had a relatively small sample size (i.e., 51 children).
  - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall: Probably low risk-of-bias ratings in the*** confounding, exposure, and outcome risk-of-bias domains. Study strengths include individual fluoride measurements with blinding at outcome assessment likely. All key confounders and many other confounders were taken into account in the study design or analysis.

Cui *et al.* (2018)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** School children aged 7–12 years from four schools in two districts in China with different fluoride levels
- ***Study area:*** Jinghai and Dagang in Tianjin City, China
- ***Sample size:*** 323 school children
- ***Data relevant to the review:*** IQ scores by urine fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant correlation between IQ score and urinary fluoride (adjusted β = −2.47; 95% CI: −4.93, −0.01).

Risk of Bias:

- **Author contacts:**
  - Authors were contacted in June 2019 to obtain additional information for risk-of-bias evaluation.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Four schools were selected from the same district in China. The schools were selected based on levels of fluoride in the local drinking water and the degree of school cooperation. No details were provided on the number of schools in given areas or the difficulty in getting school cooperation. It was noted that the residents in the four areas had similar living habits, economic situations, and educational standards. Although authors do not provide the specific data to support this, fluoride levels and IQ scores were provided by different subject characteristics. The areas were classified as historically endemic fluorosis and non-fluorosis. Cluster sampling was used to select the grades in each school according to previously set child ages, and classroom was randomly selected with all students within a selected classroom included. Reasons for exclusion do not appear to be related to exposure or outcome.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and recruited within the same time frame using the same methods, with no evidence of differences in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The measurements of all covariates were obtained by structured questionnaires that were completed by children with the help of their parents. Confounders that were assessed include: child's gender, child's ethnicity, child's age, child's BMI, birth (normal vs abnormal), mother's age at delivery, mother's education, income per family member, mother's smoking/alcohol during pregnancy, family member smoking, environmental noise, iodine region (non-endemic vs iodine-excess-

endemic area), factory within 30 m of residence, iodine salt, diet supplements, seafood/pickled food/tea consumption, surface water consumption, physical activity, stress, anger, anxiety/depression, trauma, having a cold 5 times a year, thyroid disease in relatives, mental retardation in relatives, and cancer in relatives. Covariates not considered include parity, maternal and paternal IQ, and quantity and quality of caregiving environment (e.g., HOME score). The authors report that there are no other environmentally toxic substances that may affect intelligence, such as high arsenic or iodine deficiency according to the Tianjin Centers for Disease Prevention and Control.

- *Potentially important study-specific confounders:* All key confounders were considered in this study.
  - *Direction/magnitude of effect*: Not applicable.
- *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key confounders are considered, methods for collecting the information are valid and reliable, and co-exposure to arsenic is likely not an issue in this area.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of the 400 children enrolled, 35 were excluded because they did not have informed consent signed by a guardian or they moved out of the area. Forty-two children were excluded because they did not have a DRD2 genotyping measurement. It is unclear if these children were from the same schools or if they were evenly distributed throughout the study area. It was also unclear if the excluded subjects were similar to those included in the study. In the study, some analyses had fewer than the 323 subjects, but this seems reasonable given the subgroups that were being evaluated.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although children were selected based on area fluoride levels, fluoride in the urine was used in the analysis. Urine was collected from each child the morning of enrollment and analyzed within a week. Fluoride levels were measured using an ion-selective electrode according to the China standard. A brief description of the method was provided, but no QC methods were reported. The study authors did not account for urinary dilution in the spot samples.
    - *Direction/magnitude of effect*: Not accounting for dilution could cause there to be some exposure misclassification. The direction and magnitude would depend on where the differences occurred.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* IQ was measured by professionals using the Combined Raven's Test-The Rural in China method, which is the appropriate test for the study population (++ for methods). Blinding or other methods to reduce bias were not reported. Although it is

unlikely that the outcome assessor would have knowledge of the child's urine fluoride levels, there is potential that they would know if the child was from an endemic or non-endemic area if the IQ tests were conducted at the child's school, and there was no information provided on how the IQ tests were administered. Correspondence with the study author noted the cross-sectional nature of the study with outcome and exposure assessed at the same time making the outcome assessors blind to the exposure. However, there is still potential for knowledge of the area (+ for blinding).

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses were appropriate. Multiple linear regression models were applied to evaluate the relationship between urine fluoride levels and IQ scores, accounting for numerous potential confounders. The urinary fluoride levels were log-transformed due to a skewed distribution. Residual diagnostics were used to examine model assumptions. Model robustness was tested through bootstrap, sensitivity analysis after excluding potential outliers, and cross-validation techniques. Results are reported as adjusted effects and 95% CIs. The analysis did not account for clustering at the school level or at the grade level (students were from four schools in grades selected via a clustered sampling method). There is no evidence that the sampling strategy was otherwise accounted for via sampling weights. The impact of these factors on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for several potential confounders.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key confounders were accounted for in the study design or analysis.

Cui *et al.* (2020)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** School children aged 7–12 years
- ***Study area:*** Tianjin City, China (one randomly selected school from each district based on iodine levels in the water), presumably was an expansion of the Cui *et al.* (2018)
- ***Sample size:*** 498 school children
- ***Data relevant to the review:*** IQ scores by urine fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: A 2-point decrease in IQ was observed in the highest urinary fluoride group compared to the lowest urinary fluoride group (i.e., 110.00 in ≥2.5-mg/L group versus 112.16 in <1.6-mg/L group); however, the results did not achieve statistical significance based on a one-way ANOVA comparing the three different urinary fluoride categories only.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for the 2020 publication. Authors were contacted in June 2019 for additional information on the Cui *et al.* (2018) publication. Information obtained from that correspondence may have been used for additional information in the 2020 publication.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were recruited from 2014 to 2018. One school was selected from each district where water concentrations of water iodine were <10, 10–100, 100–150, 150–300 and >300 µg/L. In each school, classes were randomly sampled for the appropriate age group of 7–12 years old. A table of subject characteristics was provided by IQ. A total of 620 children were recruited, and 122 children who did not have complete information or enough blood sample were excluded. Reasons for exclusion do not appear to be related to exposure or outcome. The characteristics of the 498 included children are presented.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited within the same time frame using the same methods, with no evidence of differences in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* It was noted by the study authors that there were no other environmental poisons except water fluoride. Other studies also conducted in this area of China noted specifically that arsenic was not a concern. Iodine was addressed as that was one of the main points of the study. Twenty-one factors (provided in Table 1 of the study) were selected as confounders, and a homemade questionnaire of unspecified validity was used for obtaining the information. It was noted that child age, stress, and anger were significantly associated with IQ although it is unclear if these varied by fluoride level. However, Cui *et al.* (2018) indicates that stress and anger were not significantly

associated with fluoride, and it is assumed that results would be similar for this study even though more children were included in the current study.

- *Potentially important study-specific confounders:* Age (children 7–12 years old)
  - *Direction/magnitude of effect*: Age is a potential confounder for IQ, even in the narrow age range evaluated in this study. The direction of effects may depend on the number of children in each age group within the different urinary fluoride categories; however, these data were not provided. In general, there were fewer subjects ≤ 9 years of age (i.e., 111) compared to > 9 years of age (i.e., 387) with a significantly higher IQ in the ≤9-year-old age group. Therefore, if exposure were higher in the older subjects, this could bias away from the null.
- *Basis for rating:* Probably high risk of bias because there is indirect evidence that age was not addressed as a confounder and it may be related to both IQ and exposure.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of the 620 (20%) children recruited, 122 were excluded due to incomplete information or inadequate blood sample. No information was provided to indicate if there were similarities or differences in the children included versus the children excluded, but exclusion is unlikely to be related to either outcome or exposure.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children's morning urine was collected with a clean polyethylene tube and fluoride was measured using a fluoride ion-selective electrode following Chinese standard WS/T 89-2015. A brief description was provided, but no QC methods were reported. The study authors do not account for urinary dilution in the spot samples.
    - *Direction/magnitude of effect*: Not accounting for dilution could cause there to be some exposure misclassification. The direction and magnitude would depend on where the differences occurred.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* IQ was measured using the Combined Raven's Test, which is an appropriate test for the study population (++ for methods). Blinding was not mentioned; however, the outcome assessors would not likely have knowledge of the child's urinary fluoride. Subjects appear to have been recruited based on iodine levels and it is, therefore, unlikely that there would be any knowledge of potential fluoride exposure. Correspondence with the study authors for the Cui *et al.* (2018) study also indicated that the outcome assessors would have been blind.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* One-way ANOVA was used to make comparisons between mean IQ by urinary fluoride levels. Consideration of heterogeneity of variances was not reported. There is no adjustment for potential confounders or for clustering of children at the school level. There is no evidence that the sampling strategy was otherwise accounted for (i.e., via sampling weights. The impact of these factors on the effect estimates is expected to be minimal given the use of individual-level data. The primary focus of the study was to evaluate associations between IQ and thyroid hormone or dopamine levels (not between IQ and fluoride levels). It should also be noted that more advanced analyses used for thyroid hormone- and dopamine-IQ associations still lacked adjustment for school and accounting for clustering of children from the same school.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements, but the study is limited by the cross-sectional study design, lack of accounting for urine dilution, and by not addressing age as a potential confounder.

Ding *et al.* (2011)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Elementary school children aged 7–14 years old
- ***Study area:*** Hulunbuir City, Inner Mongolia, China
- ***Sample size:*** 331 school children
- ***Data relevant to the review:*** IQ mean difference based on 10 categories of urine fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant association between urinary fluoride and IQ score (each 1 mg/L increase in urinary fluoride was associated with a lower IQ score of 0.59 points; 95% CI: −1.09, −0.08).

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.

- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study randomly selected 340 7–14-year-olds from four nearby elementary schools in Hulunbuir. Authors stated that the four elementary schools appeared to be very similar in teaching quality. The study authors noted that they followed the principles of matching social and natural factors like economic situation, educational standards, and geological environments as much as possible; however, how this was done is unclear and no table of study subject characteristics by group was provided.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited within the same time frame using the same methods, with no evidence of differences in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* It was noted that none of the four sites had other potential neurotoxins including arsenic in their drinking water. Although they did not provide the specifics, they did provide a reference. In addition, iodine deficiency was noted as not being issue in any of the four areas. Age was the only confounder adjusted in the model. Although dental fluorosis severity by % female was reported, not enough data were provided to determine if it was a confounder that should have been considered in the regression. The study authors note that future studies will include covariates such as parents' educational attainment, mother's age at delivery, and household income.
  - *Potentially important study-specific confounders:* Gender
    - *Direction/magnitude of effect***:** There is not enough information to determine if there is an effect from gender. There were some differences in dental fluorosis level by gender, but it is unclear how this might impact the results or if the distribution of gender differed by age.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that there were differences in gender that were not considered in the study design or analyses.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data were relatively complete (i.e., <5% loss). Of the 340 subjects selected for inclusion, 5 were excluded because they lived in the area for less than a year with an additional 4 not consenting to participate.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analysis was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Spot urine samples were collected and measured using China CDC standards. All samples were analyzed twice using a fluoride ion-selective electrode. Recovery rates were specified as 95–105% with an LOD of 0.05 mg/L. Water samples were collected from small-scale central water supply systems and tube wells with handy pumps and were processed using standard methods similar to the urine samples. Quality assurance validation was reported. A blind professional examiner evaluated the children for dental

fluorosis using the Dean's Index. All urine and water samples were above the LOD. Urine levels were the primary exposure measure used in the analysis. The study authors did not account for urinary dilution in the spot samples. The mean urine fluoride concentration was correlated with the dental fluorosis levels.

- *Direction/magnitude of effect*: Spot urine samples that did not account for dilution could have exposure misclassification. The misclassification is likely non-differential and potential direction of bias is unknown.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* IQ was determined using the Combined Raven's Test-The Rural in China (CRT-RC3) (++ for methods). Although blinding was not reported, it is unlikely that the IQ assessors had knowledge of the children's urine levels or even of the water levels from the four sites as these were sent to a separate lab for testing (+ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses were reasonable (ANOVA and multiple linear regression), but consideration of homogeneity of variance was not reported. The NASEM committee review (NASEM 2021) pointed out a potential concern for the lack of accounting for clustering at the school-level because children were selected from four elementary schools. However, as pointed out in the ***Selection*** domain, the authors stated that they followed the principles of matching social and natural factors like economic situation, educational standards, and geological environments to the extent possible and that the four elementary schools appeared to be very similar in teaching quality. There is no evidence that the sampling strategy was otherwise accounted for (I.e., via sampling weights). The impact of these factors on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for age as a potential confounder.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and that there were no other potential threats to risk of bias.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements, but the study is limited by the cross-sectional study design, lack of accounting for urine dilution, and by not addressing gender as a potential confounder.

Green *et al.* (2019)

Study Details:

- ***Study design*:** Prospective cohort
- ***Population*:** Maternal-Infant Research on Environmental Chemicals (MIREC) participants (pregnant mothers and their children aged 3–4 years)
- ***Study area:*** 10 cities, Canada
- ***Sample size:*** 512 mother−child pairs (238 from non-fluoridated areas, 162 from fluoridated areas; 264 females, 248 males)
- ***Data relevant to the review:*** Adjusted linear regression models evaluating associations between IQ in both genders together and separate, with maternal urinary fluoride across all three trimesters, or with estimated maternal fluoride intake.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower full-scale IQ per 1-mg/L increase in maternal urinary fluoride in boys (adjusted β = −4.49), but not girls (adjusted β = 2.40) and not in both genders combined (adjusted β = −1.95); significantly lower full-scale IQ per 1-mg increases in maternal intake in both genders combined (adjusted β = −3.66 [no sex interaction]); significantly lower full-scale IQ per 1-mg/L increase in drinking water fluoride in both genders combined (adjusted β = −5.29 [no sex interaction]).

Risk of Bias:

- **Author contacts:**
  - Authors were contacted in June 2019 for additional information for the risk if bias evaluation.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Pregnant women were recruited from the same population, during the same timeframe, and using the same methods as the MIREC program. Methods were reported in detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were similar and were recruited with the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study considered several possible covariates including maternal age, pre-pregnancy BMI, marriage status, birth country, race, maternal education, employment, income, HOME score, smoking during pregnancy, secondhand smoke in the home, alcohol consumption during pregnancy, parity, child's gender, child's age at testing, gestational age, birth weight, time of void, and time since last void. The study also conducted secondary analyses to test for lead, mercury, arsenic, and PFOA. There is no indication of any other potential co-exposures in this study population. Iodine deficiency

or excess could not be assessed but is not expected to differentially occur. The study was not able to assess parental IQ or mental health disorders. Methods used to obtain the information included questionnaires and laboratory tests.

- *Potentially important study-specific confounders:* All key confounders were addressed.
  - *Direction/magnitude of effect:* Not applicable.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and direct evidence that key confounders including potential co-exposures were addressed.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of the 610 recruited children, 601 (98.5%) completed testing. Of the 601 mother-child pairs, 512 (85.2%) had all three maternal urine samples and complete covariate data, and 400 (66.6%) had data available to estimate fluoride intake.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Spot urine samples from all three trimesters of pregnancy were evaluated using appropriate methods, and results were adjusted for creatinine and specific gravity. Fluoride intake was estimated based on fluoride water levels and information on consumption of tap water and other water-based beverages (e.g., tea, coffee) was obtained via questionnaire.
    - *Direction/magnitude of effect***:** There is not any specific direction or magnitude of bias expected. Urinary fluoride levels are reflective of a recent exposure. Having measurements from all three trimesters of pregnancy provides a better representation of actual exposure than a single measurement although the potential for missed high exposure is possible. However, the possibility of the occurrence of missed high exposure would be similar in all females and would be non-differential. For the fluoride intake, exposure was based on the fluoride levels in the water at the residence. If women worked outside the home and the majority of intake occurred from areas outside the home (and were different from levels in the home), there is potential to bias toward the null.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The Wechsler Preschool and Primary Scale of Intelligence was normalized for ages 2.5–<4.0 and child sex using the U.S population-based norms. Blinding was not reported, but it is unlikely that the outcome assessors had knowledge of the maternal fluoride level or were aware if the city had fluoridated water.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes were reported.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Multivariate linear regression analyses were used to evaluate the associations between maternal urinary fluoride and fluoride intake and children's IQ scores. Regression diagnostics were used to test assumptions for linearity, normality, and homogeneity. There were no potential influential observations (based on Cook's distance). Sensitivity analyses showed that the effects of maternal urinary fluoride (MUF), fluoride intake, and water fluoride were robust to the exclusion of two very low IQ scores in males (<70). City was accounted for as a covariate in the regression models published. Additional models with city as a random effect were also subsequently made publicly available and showed similar results to the main model.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, prospective cohort design, and addressing potential key confounders.

Rocha-Amador *et al.* (2007)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 6–10 years
- ***Study area:*** Moctezuma (low fluoride, low arsenic) and Salitral (high fluoride, high arsenic) of San Luis Potosí State and 5 de Febrero (high fluoride, high arsenic) of Durango State, Mexico
- ***Sample size:*** 132 children
- ***Data relevant to the review:*** Associations between full-scale IQ, performance IQ, verbal IQ and child's urine or water fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant associations between fluoride and IQ scores (full-scale IQ adjusted βs of –10.2 [water] and −16.9 [urine]; CIs not reported); arsenic also present, but the effect was smaller (full-scale IQ adjusted βs of –6.15 [water] and –5.72 [urine]; CIs not reported).

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***

- *Rating:* Probably low risk of bias (+)
- *Summary:* All children in 1st through 3rd grades in three rural areas in Mexico (n = 480) were screened for study eligibility including age, time at residence, and address. Authors report that the three selected communities were similar in population and general demographic characteristics. Children who had lived in the area since birth and were 6–10 years old were eligible to participate (n = 308). Of the 308 children, 155 were randomly selected and the response rate was 85%, but participation was not reported by area. It was noted, however, that no significant differences in age, gender, or time of residence were observed between participants and non-participants. Timeframe for selection was not mentioned but appears to be similar. Sociodemographic characteristics of subjects was provided in Table 1 of the study. There was a significant difference in SES and transferrin saturation, but these were taken into account in the analysis.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the populations were similar and differences were noted and addressed in the analysis.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study design or analysis accounted for child's age, sex, SES, transferrin saturation, weight, height, blood lead levels, and mother's education. Arsenic levels were highly correlated with fluoride levels; however, arsenic and fluoride were evaluated alone, and arsenic was found to have less of an effect on IQ than fluoride. This provides evidence that arsenic has been addressed as a co-exposure and cannot explain the association between fluoride exposure and decreased IQ. Smoking was not addressed and methods for measuring many of the confounders were not reported.
  - *Potentially important study-specific confounders:* Arsenic
    - *Direction/magnitude of effect*: The presence of arsenic in this study, which also demonstrated an association, would bias the effect away from the null. Although arsenic may contribute to some of the magnitude of the observed effect of fluoride (the exact impact of arsenic on the magnitude cannot be assessed), the presence of arsenic does not fully explain the observed association between fluoride and IQ. The presence of arsenic may affect the magnitude of the association between IQ and fluoride, but it has no impact on the direction of the effect.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and direct evidence that key confounders were addressed.
- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of 155 children randomly selected for study participation, 85% responded to enroll. According to the authors, there were no significant differences in age, gender, or time of residence between responders and non-responders. However, no data are provided to support this, and no breakdown of responders/non-responders by region is provided. Data were provided for the 132 children agreeing to participate.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Urine samples were collected on the same day as psychological evaluations and were analyzed for fluoride according to NIOSH Method 8308 (Fluoride in Urine). For QC, a reference standard was also used (NIST SRM 2671a). Urine samples were also analyzed for arsenic by using the Atomic Absorption Spectrophotometer with hydride system and used a reference standard for QC. Levels were adjusted for urinary creatinine levels to account for dilution in the spot samples. Tap water samples were collected from each child's home on the day of biological monitoring. Fluoride was measured with a sensitive, specific ion electrode. Detailed methods are provided including internal quality controls. It was noted that in the high fluoride group it was common to drink bottled water low in fluoride and to only use the tap water for cooking; therefore, urine was considered the most appropriate measure of exposure. Only children who had lived at the same residence since birth were included.
    - *Direction/magnitude of effect***:** Not applicable.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Neuropsychological profiles were assessed through the WISC-RM (revised for Mexico). This is a well-established test appropriately adjusted for the study population. However, no additional validation is provided (+ for methods). The study report stated that the test assessors were masked to both arsenic and fluoride water levels (++ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* It was reported that an interaction between fluoride and arsenic was measured, but it was only noted in the discussion that the study design precluded testing statistical interaction between fluoride and arsenic. This provides indirect evidence of selective reporting.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that there was selective reporting.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
  - *Statistical analyses:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study. Multivariate linear analyses were used to evaluate the associations between fluoride in water and urine and children's IQ scores Exposures were natural

log-transformed, but rationale was not provided. Regression diagnostics were not used to test model assumptions for linearity, normality, and homogeneity. The analyses did not account for clustering at the community level. The three selected communities were similar in population and general demographic characteristics. Although the analysis used individual-level exposures rather than area-level exposures, if the exposure levels within a certain area are highly correlated (which might be expected), then the results might still be biased. However, the overall impact on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for multiple potential confounders.

        - *Other potential concerns:* None identified.
    - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and outcomes blindly assessed, but it is limited by the cross-sectional study design and not being able to completely rule out the influence of arsenic in the results.

Saxena *et al.* (2012)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 12 years
- ***Study area:*** Madhya Pradesh, India
- ***Sample size:*** 170 children
- ***Data relevant to the review:*** Mean IQ grade (not standard scores; higher IQ grades are associated with lower intelligence) by water fluoride quartiles, continuous water fluoride, or continuous urinary fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant correlations between IQ score and water (r = 0.534) and urinary (r = 0.542) fluoride levels. Significant increase in mean IQ grade (i.e., increase in proportion of children with intellectual impairment) with increasing water fluoride quartile.

Risk of Bias:

- **Author contacts:**
    - Authors were contacted in August of 2017 to obtain additional information for risk-of-bias evaluation.
- ***Population selection:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* There was indirect evidence that subjects were similar and were recruited using the same methods during the same time frame. The study participants were selected from a stratified cluster of geographic areas based on fluoride concentration in groundwater. According to the authors, the selected villages were similar in population and demographic characteristics. Data are provided to show the breakdown in SES,

parental education, height/age, and weight/height and no significant differences were noted. Participation was stated to be voluntary, but participation rates were not provided. It is unclear if the 170 subjects were selected with 100% participation or if the 170 subjects were all that were asked to participate, but it appears that all subjects participated. Timing of the recruitment was not provided but is assumed to occur during the same time frame.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and recruited using the same methods during the same time frame.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There was indirect evidence that key confounders including potential co-exposures were addressed using reasonable methods. A questionnaire, completed with the assistance of parents, was used to collect information on child characteristics (age, sex, height, weight), residential history, medical history (including illness affecting nervous system and head trauma), educational level of the head of the family (in years), and SES of the family. The SES was recorded according to the Pareek and Trivedi classification. The nutritional status of the children was calculated using the Waterlow's classification, which defines two groups for malnutrition using height for age ratio (chronic condition) and weight for height ratio (acute condition). Within both groups, it categorizes the malnutrition as normal, mildly impaired, moderately impaired, or severely impaired. Urinary lead and arsenic were analyzed using the atomic absorption spectrophotometer. Urinary iodine was measured using the Dunn method. Authors do not report which covariates were included in the multivariate regression models; however, there was no difference in reported demographic characteristics. All subjects were the same age, and there was no difference in iodine, lead, or arsenic between the groups. Mean urinary arsenic levels did increase with increasing fluoride even though there was no significant difference by group.
  - *Potentially important study-specific confounders:* All key confounders were considered in this study.
    - *Direction/magnitude of effect***:** Not applicable.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and that key confounders including potential co-exposures were addressed.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results were provided for all 170 children stated to be included in the study.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence of no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* A sample of 200 mL of drinking water was collected at each child's home. The fluoride levels were analyzed by a fluoride ion-selective electrode. Each subject was also asked to collect a sample of their first morning urine. The fluoride content in the urine was determined using a fluoride ion-selective electrode. QA/QC and LOD were not reported and urinary dilution was not assessed. Although only current levels were measured, children who had changed water source since birth were excluded.

        - *Direction/magnitude of effect*: Spot urine samples that did not account for dilution could have exposure misclassification. The misclassification is likely non-differential and not likely to bias in any specific direction. Children who had changed water since birth were excluded, but it was not specifically noted that the fluoride in the water source was stable over the years.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Intelligence is assessed using the Raven's Standard Progressive Matrices and categorized into five grade levels. Although it was not noted that the test was validated to the study population, the test is visual and would be applicable to most populations (+ for methods). There is no mention of blinding by test administrators or evaluators and the exposure groups come from different geographic areas. It was also not reported who measured the levels of fluoride from the home or urine samples. Correspondence with the study authors indicated that the outcome assessors were blind to the children's fluoride status (++ for blinding). Overall rating for methods and blinding = +.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:*
        - *Statistical analyses:* One way analysis of variance (ANOVA), simple linear regression, and multiple linear regression were used to compare mean intelligence grades by water fluoride levels and to assess the association between grades and urinary fluoride. Consideration of heterogeneity of variance (for ANOVA) was not reported. Regression diagnostics were not used to test model assumptions for linearity, normality, and homogeneity. Given the ordinal nature of the intelligence grade variable (score from 1 to 5), ordinal logistic regression would have been a more appropriate method. There was no adjustment for area-level clustering in multivariate analyses (although subjects were selected via stratified cluster sampling from two areas). Although the analysis used individual-level exposures rather than area-level exposures, if the exposure levels within a certain area are highly correlated (which might be expected), then the results might still be biased. However, the overall impact on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for potential confounders.
        - *Other potential concerns:* None identified.

  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate, and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and the addressing of potential key confounders, but it was limited by the cross-sectional study design and lack of addressing dilution in the urine samples.

Seraj *et al.* (2012)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 6–11 years
- ***Study area:*** five villages, Makoo, Iran
- ***Sample size:*** 293 children
- ***Data relevant to the review:*** IQ (mean and distribution) assessed by Raven's Colored Progressive Matrices and presented by fluoride area; beta was also provided for water fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant association between water fluoride and IQ score (adjusted β = −3.865; CIs not reported); significantly higher IQ score in normal area (97.77 ± 18.91) compared with medium (89.03 ± 12.99) and high (88.58 ± 16.01) areas.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were selected from five villages in Makoo. The villages were stated to all be rural with similar general demographic and geographic characteristics and were comparable in terms of SES and parental occupations. Children were 6–11 years old. Age, gender, and education were taken into account in the analysis. No other characteristics were provided or discussed. Participation rates were not reported. There is indirect evidence that the populations were similar, and some possible differences were addressed.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and recruited using the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Age, gender, dental fluorosis intensity, and educational levels (child's and parents') were evaluated as potential confounders. Other potential confounders such as smoking were not discussed. Information was obtained from a detailed questionnaire. Lead was measured, but only found in low levels in the drinking water throughout the study regions. Iodine in the water was also stated to be measured and residents were receiving iodine-enriched salt. Arsenic was not addressed, but there is no evidence that

arsenic levels would vary across villages in this area. Based on water quality maps, co-exposure to arsenic is likely not a major concern in this area.

- *Potentially important study-specific confounders:* Arsenic.
  - *Direction/magnitude of effect*: Conceptually, if there were differential amounts of arsenic in the different villages, co-exposure to arsenic could bias the results with the direction of the bias dependent on where the arsenic was present; however, arsenic is not expected to be a major concern in this study area based on water quality maps.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and that key confounders including potential co-exposures were addressed or were not likely to be an issue in the study area.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Attrition was low if it occurred. It was noted that 293 out of 314 children living in the villages were recruited. It is not clear if 21 children were excluded based on exclusion criteria or if they refused to participate; however, this accounts for less than 10% of the population and results were available for all 293 subjects.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* Exposure was primarily based on area of residence. Fluoride in the groundwater was analyzed by the SPADNS (Sulfophenylazo dihydroxynaphthalene-disulfonate) method, utilizing 4000 UV-Vis spectrophotometer in the environmental health engineering laboratory of the Public Health School of Tehran University of Medical Sciences. Specific details were not provided on methods of collection, samples locations, or if these locations represented the primary sources of drinking water for the subjects. Villages were categorized into normal (0.5–1 ppm), moderate (3.1±0.9 ppm), and high (5.2±1.1 ppm) fluoride based on the mean fluoride content of all seasons presumably for the stated 12-year time period. Subjects were stated to be long-life residents of the village. Dental fluorosis was also measured and increased in severity with fluoride levels; however, all areas had some degree of dental fluorosis. Although authors used an average fluoride level in varying seasons over presumably 12 years, they used a less-established method without reporting reliability or validity, nor did they provide data to indicate that the mean was truly representative of the fluoride levels over time and throughout the village. Although dental fluorosis severity increased with increasing fluoride levels, the data could also indicate potential exposure misclassification.
    - *Direction/magnitude of effect*: The presence of dental fluorosis in all groups indicates that there may have been different exposure in some children at a younger age. Although there were only about 20 children in the “normal” fluoride group with very mild to mild dental fluorosis, this could bias the results toward the null because those children may have experienced a higher level of

fluoride at some point. The other two fluoride groups were exposed to fluoride levels that likely exceeded those in the "normal" fluoride group.

- *Basis for rating:* Probably high risk of bias based on indirect evidence that exposure was assessed using insensitive methods.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Intelligence was evaluated using the Raven's Color Progressive Matrices. This is a well-established method. Although the study authors did not provide data to indicate that the methods were valid in this study population, the test is designed to be culturally diverse. (+ for methods). The study report stated that test administrators were blinded. (++ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that outcomes were blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported. However, because they did not report the method for obtaining the betas in Table 4 of the study, it is not clear if these were adjusted or unadjusted betas.
  - *Basis for rating:* Probably low risk of bias based on direct evidence that all the study's measured outcomes were reported, but the results were not sufficiently reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical methods for comparisons of IQ level by exposure groups were reasonable (ANOVA, post hoc test and Kruskal-Wallis test), but consideration of heterogeneity of variance was not reported. Clustering at the village levels was not accounted for in multivariate analyses which used area-level water fluoride levels. Because the exposure levels within a certain area are highly correlated (which might be expected), the results are likely to be biased. There was adjustment for some potential individual-level confounders, and the children were from five rural areas with similar general demographic and geographic characteristics and were comparable in terms of SES and parental occupations. These factors are expected to mitigate some of the impact of lack of accounting for clustering, and the overall impact on the effect estimates is expected to be minimal.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and outcome. Study strengths include addressing potential key confounders, but it was limited by the cross-sectional study design and the group-level exposure data.

Soto-Barreras *et al.* (2019)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 9–10 years
- ***Study area:*** Chihuahua, Mexico
- ***Sample size:*** 161 children
- ***Data relevant to the review:*** Water fluoride, urinary fluoride, exposure dose, and dental fluorosis index by IQ grade.
- ***Reported association with fluoride exposure:*** No: Results were not presented to evaluate an association between fluoride exposure and IQ, but rather to compare fluoride levels within IQ grades. For this reason, the results for this study are not comparable to other studies that evaluated IQ scores by fluoride exposure levels. No significant differences in measured fluoride levels across IQ grades were observed.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were selected using a multistage cluster sampling. During the first stage, 13 public elementary schools were randomly selected from a pool of 73 using a cluster sample design. Secondly, only fourth grade students were included. Authors stated that they wanted to keep the same grade level, but they were not specific as to why fourth graders were selected as opposed to any other grade. Lastly, only children whose parents or guardians attended and responded to the survey were included. There is no information provided on how the 13 schools selected may be similar or different from the 60 schools not selected. There is no information provided on the number of children in the fourth grade to know participant rates. It was only noted that 245 children were examined, but 161 were included after the exclusion rules were applied. Inclusion and exclusion criteria are presented. Reasons for exclusion do not appear to be related to exposure or outcome. Characteristics of participants and non-participants are not compared; however, characteristics of the 161 included children were provided and any differences were taken into account in the analysis.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposed groups were similar and were recruited using similar methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* No confounders were considered when evaluating fluoride associations with intelligence; they were only applied when evaluating fluoride levels and dental caries. Based on Table 4 of the study, there was no significant association between IQ grade and child's age, sex, parental education, or SES status. No other information was reported or considered. There is no information on potential co-exposures. Based on

water quality maps, the arsenic prediction indicates a greater than 50% probability of exceeding the WHO guidelines for arsenic of 10 µg/L in areas of Chihuahua, Mexico.

- *Potentially important study-specific confounders:* Arsenic.
  - *Direction/magnitude of effect*: The direction and magnitude of effects is unknown. There is potential for arsenic to occur in the study area, but it is not known how it relates to fluoride exposure. If they occur together in the water, it will bias away from the null; however, if they occurred in different areas, there is potential to bias toward the null.
- *Basis for rating:* Probably high risk of bias based on indirect evidence that there is potential for exposure to arsenic that was not sufficiently addressed.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* A total of 161 of 245 children were included in the study. Exclusion criteria are presented and are unrelated to outcome or exposure. For the 161 children, there are no missing outcome data.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+);Probably high risk of bias (-)
  - *Summary:* **Urinary Fluoride (probably low risk of bias):** First morning void urine samples were collected based on NIOSH methods. Water samples were also stated to be collected, but it does not appear that methods followed any particular standard, and there is no indication that subjects were provided with collection containers. Analysis was based on a calibration curve using fluoride ion selective electrode. QC methods were mentioned. Based on results, there were values below detection limits, but LODs or % below LOD were not reported.

    Daily fluoride exposure (probably high risk of bias): Daily fluoride exposure was based on the water fluoride level, drinking water consumption (based on parental report of how many glasses of water consumed), and body weight.
    - *Direction/magnitude of effect*: Spot urine samples that did not account for dilution could have exposure misclassification. The misclassification is likely non-differential and is not likely to bias in any specific direction. Daily exposure was based partially on parental report of water consumption. The direction and magnitude of effect is unknown.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure. The daily fluoride exposure is probably high risk of bias because there is indirect evidence that the exposure was assessed using methods of unknown validity.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Intellectual ability was evaluated using Raven's Colored Progressive Matrices by an independent examiner. Some details were provided, but it was not stated that the tests were assessed blind; however, there is no indication that subjects were from high fluoride areas and the assessor would not have knowledge of the urine or water fluoride

levels. Results for children were converted into a percentile according to age (details not provided) and overall scores were assigned an intellectual grade of I to V as described in the report.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:*
        - *Statistical analyses:* The Kolmogorov-Smirnov test was used to determine variable distribution. The Kruskal Wallis test was used to compare exposure levels between IQ grades with a Dunn's post hoc test. Multivariate logistic regression was used the estimate the association between presence of dental caries and various risk factors. Fluoride levels in drinking water and urine and fluoride exposure dose are compared across intellectual grades. Children were from 13 schools selected via stratified cluster sample design. There was no adjustment for clustering at the school level or for the sampling design. Although the analysis used individual-level exposures rather than area-level exposures, if the exposure levels within a certain school are highly correlated (which might be expected), then the results might still be biased. The large number of clusters (13 schools) makes clustering less of a concern and the impact on the effect estimates is expected to be minimal.
        - *Other potential concerns:* None identified.
    - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements and outcomes blindly assessed, but is limited by the cross-sectional study design, lack of accounting for urine dilution, and by not addressing potential exposures to arsenic in the study area. Although the study is considered to have low potential for bias overall, the focus of the study was to evaluate the relationship between fluoride exposure and lower rates of dental caries. In terms of evaluating an association between fluoride exposure and IQ scores, the study is limited by the way that the data were reported.

Sudhir *et al.* (2009)

Study Details:

- ***Study design*:** Cross-sectional

- ***Population*:** Children aged 13–15 years
- ***Study area:*** Nalgonda district (Andhra Pradesh), India
- ***Sample size:*** 1,000 children
- ***Data relevant to the review:*** Mean IQ grade (not standard scores) or IQ distribution by water fluoride strata (<0.7, 0.7-1.2, 1.3-4.0, and >4.0 ppm).
- ***Reported association with fluoride exposure:*** Yes: Significant increase in mean and distributions of IQ grades (i.e., increase in proportion of children with intellectual impairment) with increasing drinking water fluoride levels.

Risk of Bias:

- **Author contacts:**
  - Authors were contacted in September of 2017 for additional information related to risk-of-bias evaluation, but no response was received.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children were selected from the same general population during the same time frame and were then broken down into nearly equal exposure groups. A cross-sectional study was conducted among 13–15-year-old school children of Nalgonda district, Andhra Pradesh between August and October 2006. Data were collected from the school children who were life-long residents of Nalgonda district, Andhra Pradesh and who consumed drinking water from the same source during the first 10 years of life. A stratified random sampling technique was used. The entire geographical area of Nalgonda district was divided into four strata based on different levels of naturally occurring fluoride in the drinking water supply. Children were randomly selected from schools in the different strata. It was noted that the 1,000 selected children were equally divided among all four strata, however, each group did not have 250 children (but instead 243−267 in each group). Participation rates are not reported. Exclusion criteria included: children who had a history of brain disease and head injuries, children whose intelligence had been affected by congenital or acquired disease, children who had migrated or were not permanent residents, children with orthodontic brackets, and children with severe extrinsic stains on their teeth. Age and gender data are presented in Table 1 of the study, but this information is not presented by the different fluoride groups.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and were recruited using the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Data were collected using a self-administered questionnaire and clinical examination. The self-administered questionnaire requested information on demographic data (appears to cover age and sex), permanent residential address, staple food consumed, liquids routinely consumed, and aids used for oral hygiene maintenance (fluoridated or nonfluoridated). SES was measured using the Kakkar socio-economic status scale (KSESS) with eight closed-ended questions related to parental education, family income, father's occupation, and other factors. All children were asked to fill out the form, and the answers obtained were scored using Kakkar socio-economic status

scoring keys. Based on this scoring, children were divided into three groups—lower class, middle class, or upper class. Age, sex, and SES were not found to be significantly associated with IQ. Other confounders including smoking were not addressed. Co-exposures such as arsenic and lead were not addressed; however, there is no indication that lead is a co-exposure in this population and arsenic is not likely a major concern in this area based on water quality maps.

- *Potentially important study-specific confounders:* Key confounders age, gender, and measures of SES were similar between exposure groups; however, arsenic was not taken into account. Arsenic often occurs in the drinking water along with fluoride in some Indian populations; however, based on water quality maps, this does not appear to be an issue in the Nalgonda district of Andhra Pradesh. Iodine deficiencies are not mentioned.
    - *Direction/magnitude of effect***:** Conceptually, the presence of arsenic would potentially bias away from the null if present with fluoride. Deficiencies in iodine would bias away from the null if present in areas of high fluoride, but toward the null if present in areas of non-high fluoride. Neither of these were considered issues in this study for reasons noted above.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the key confounders are considered, co-exposure to arsenic is likely not an issue in this area, and methods used for collecting the information were valid and reliable.

- ***Attrition:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* Results were available for the 1,000 children selected to participate.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence of no attrition.
- ***Exposure:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Children were placed into one of four strata based on the level of fluoride in drinking water. Collection of water samples was done in the districts. The placement into strata was based on fluoride levels obtained from documented records of District Rural Water Works Department. Once the children were assigned to strata, it was confirmed that the fluoride level of their drinking water was within the strata assigned. This was done using the methodology followed in National Oral Health Survey and Fluoride Mapping 2002–2003. During the initial visits to the schools, the children were interviewed regarding their history of residence and source of drinking water from birth to 10 years. The first child meeting criteria was given a bottle for water collection and the next child was only given a bottle for collection if the water source was different than that of a previous child. Children were asked to collect the sample of water from the source that was used in the initial 10 years of their life and was collected the next day. It was not reported if all bottles were returned. The water samples collected were subjected to water fluoride analysis using an ion-specific electrode, Orion 720A fluoride meter at District Water Works, Nalgonda to confirm the fluoride levels in the water before commencement of clinical examination. LOD and QA/QC details were not reported.
        - *Direction/magnitude of effect***:** There is some potential for exposure misclassification based on recall of the children on the source of water used in

their first 10 years of life. The misclassification is likely non-differential and not likely to bias in any specific direction. Children who had changed water since birth were excluded, but it was not specifically noted that the fluoride in the water source was stable over the years.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Probably high risk of bias (NR)
  - *Summary:* The Raven's standard progressive matrices (1992 edition) was used to assess IQ. This Raven's test is a standard test and although there is no information provided to indicate that the methods were reliable and valid in the study population, this test was created to be culturally fair (+ for methods). Blinding or other methods to reduce potential bias were not reported (NR for blinding). No response was received to an e-mail request for clarification in September 2017. Overall rating for methods and blinding = NR.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that the outcome was not assessed blind and could bias the results.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Chi-square test and Spearman rank correlation were used to assess the association between four different fluoride levels and IQ grades. Area-level exposures were used. Clustering of children within the four areas was not accounted for in the analysis; however, because multiple villages were included in each fluoride exposure level, clustering is less of a concern and the impact on the effect estimates is expected to be minimal.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include verification of exposure measurements and the addressing of potential key confounders, but it was limited by the cross-sectional study design and lack of information on blinding during outcome assessment.

Till *et al.* (2020)

Study Details:

- ***Study design*:** Prospective cohort
- ***Population*:** MIREC participants (pregnant mothers and their children aged 3–4 years)
- ***Study area:*** 10 cities, Canada
- ***Sample size:*** 398 mother−child pairs (247 from non-fluoridated areas, 151 from fluoridated areas; 200 breastfed as infants, 198 formula-fed as infants)
- ***Data relevant to the review:*** Adjusted linear regression models evaluating associations between IQ with water fluoride concentration (with or without adjusting for maternal urine) in formula-fed or breast-fed infants or by fluoride intake from formula.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower performance IQ with water fluoride by breastfeeding status (adjusted βs = −9.26 formula-fed, −6.19 breastfed) and fluoride intake from formula (adjusted β = −8.76); significantly lower full-scale IQ with water fluoride in formula-fed (adjusted β = −4.40); no significant changes in full-scale IQ for water fluoride in breastfed children or fluoride intake from formula-fed children; no significant changes in verbal IQ scores with fluoride exposure.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Pregnant women were recruited between 2008 and 2011 by the MIREC program from 10 cities across Canada. Inclusion and exclusion criteria were provided. Additional details were stated to be available in Arbuckle *et al.* (2013). A total of 610 children were recruited to participate in the developmental follow-up with 601 children completing all testing. The demographic characteristics of women included in the current analyses (n = 398) were not substantially different from the original MIREC cohort (N = 1945) or the subset without complete water fluoride and covariate data (n = 203). A table of characteristics of the study population is provided. Approximately half of the children lived in nonfluoridated cities and half lived in fluoridated cities.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were similar and were recruited with the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariates were selected a priori that have been associated with fluoride, breast feeding, and children's intellectual ability. Final covariates included child's sex and age at testing, maternal education, maternal race, second-hand smoke in the home, and HOME score. City was considered but was excluded from the models. Confounders that were not assessed include: parental mental health, iodine deficiency/excess, parental IQ, and co-exposure to arsenic and lead. Co-exposure to arsenic is less likely an issue in this Canadian population because the population mainly received water from municipal water supplies that monitor for lead and arsenic, and the lack of information

is not considered to appreciably bias the results. In addition, a previous study on this population (Green *et al.* 2019) conducted sensitivity analyses on co-exposures to lead and arsenic. Results from these sensitivity analyses support that co-exposures to lead and arsenic are not likely a major concern in this study population.

- *Potentially important study-specific confounders:* All key confounders were considered in this study.
  - *Direction/magnitude of effect*: Not applicable.
- *Basis for rating:* Probably low risk of bias based on direct evidence that key confounders were addressed and indirect evidence that the methods used to collect the information were valid and reliable and co-exposures were not an issue.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of 610 children, 601 (98.5%) in the MIREC developmental study who were ages 3–4 years completed the neurodevelopment testing. Of the 601 children who completed the neurodevelopmental testing, 591 (99%) completed the infant feeding questionnaire and 398 (67.3%) reported drinking tap water. It was noted that the demographic characteristics were not substantially different from the original MIREC cohort or the 203 subjects without complete water fluoride or covariate data.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Information on breastfeeding was obtained via questionnaire at 30–48 months. Fluoride concentration in the drinking water was assessed by daily or monthly reports provided by water treatment plants. Water reports were first linked with mothers' postal codes and the daily or weekly amounts were averaged over the first 6 months of each child's life. Additional details can be found in Till *et al.* (2018). Maternal urinary exposure was used to assess fetal fluoride exposure. Procedures can be found in Green *et al.* (2019).
    - *Direction/magnitude of effect*: There is not any specific direction or magnitude of bias expected. Urinary fluoride levels are reflective of recent exposure. The possibility of the exposure misclassification would be similar in all subjects and would be non-differential. For the fluoride intake from formula, exposure was based on the fluoride levels in the water at the residence and the proportion of time that the infant was not exclusively breastfed. This exposure misclassification would also be non-differential.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Intelligence was tested using the Wechsler Preschool and Primary Scale of Intelligence III. This is appropriate for both the study population and age group. This is considered a gold standard test. It was not reported whether the evaluators were blind to the child's fluoride exposure status during the assessment. Although it is unlikely that

the assessors had knowledge of the specific drinking water levels or maternal urine levels, there is potential that the outcome assessors had knowledge of the city the child lived in and if the city was fluoridated or non-fluoridated. Correspondence with the study authors on the outcome assessment for Green *et al.* (2019) indicated that it was unlikely that the testers had knowledge of the city's fluoridation. The same is assumed here. Specific measurements included were identified.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Regression diagnostics were used to test assumptions for linearity, normality, and homogeneity. There were two potential influential observations (based on Cook's distance), and sensitivity analyses re-estimated the models without these two variables. Effect modification by breastfeeding status was evaluated. Interestingly, all regression coefficients were divided by 2 to represent change in IQ per 0.5-mg/L change in fluoride. One concern is posed by the lack of accounting for city in the regression models, ideally as a random effect. The authors explored including city as a covariate in the models; however, city was not included either because it was strongly multi-collinear with water fluoride concentration (VIF > 20) (model 1, with water fluoride concentration) or because fluoride intake from formula is a function of water fluoride concentration (assessed at the city level) and was therefore deemed redundant (model 2). However, the models use city-level water fluoride concentrations—and, in sensitivity analyses, adjust for maternal urinary fluoride—which warrants exploration of city as a random effect rather than a fixed effect (as would be the case by having it included as a covariate). Even including individual-level maternal urinary fluoride might not fully account for lack of a city effect, given that the subjects were from six different cities, with half of them fully on fluoridated water. Hence, even individual-level exposures are likely to be correlated at the city level. Based on a previous analysis (Green *et al.* 2019), it is unlikely that exclusion of city from models (as a fixed or random effect) would significantly impact the effect estimates.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, prospective cohort design, and the addressing of potential key confounders.

Trivedi *et al.* (2012)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 12−13 years
- ***Study area:*** Kachchh, Gujarat, India
- ***Sample size:*** 84 children
- ***Data relevant to the review:*** Mean IQ scores and distribution by low and high fluoride villages.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower mean IQ score in the high fluoride villages (92.53 ± 3.13) compared to the low fluoride villages (97.17 ± 2.54) in boys and girls combined (and by gender).

Risk of Bias:

- **Author contacts:**
  - Authors were contacted in September of 2017 to obtain additional information for risk-of-bias evaluation.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There is insufficient information provided on the sampling methods to determine if the populations were similar. Although it was noted that samples were obtained for groundwater quality from March to May of 2011, there is no indication that the children were selected at the same time or during a similar time frame. Correspondence with the author indicates that children were selected within a week of the water collection based on random selection of a school in the village. Study participants were selected from six different villages of the Mundra region of Gujarat, India. Subjects were grouped into high and low villages based on the level of fluoride in the drinking water of those villages. The number of subjects per village were not reported, but it was noted that there were 50 children in the low fluoride group and 34 children in the high fluoride group. It is not clear if the differences in numbers were based on different participation rates or if there were fewer children in the high fluoride villages. Recruitment methods including any exclusion criteria and participation rates were not provided. SES was stated to be low and equal based on questionnaire information, but the results were not provided. It should also be noted that only regular students (having attendance more than 80%) of standard 6^{th} and 7^{th} grades were selected, but it was not noted if attendance varied by village. Correspondence with the study author indicated that there was an average of 20 students per class with an average of 40 students per village. It appears that keeping the requirement for 80% attendance was a limiting factor that caused different numbers of children by area; however, this was applied similarly to both groups.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and recruited using the same methods during the same time frame.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children were stated to be students of the 6th and 7th standard grades. Age was not addressed, but the children would all be of similar age based on the grades included. Results were reported for males and females separately as well as combined. SES and iodine consumption were stated to be analyzed via a questionnaire and were standardized on the basis of the 2011 census of India. Although it was noted in the abstract that the SES was equal (no data provided), the study report did not mention the iodine results. Although the study authors did not address arsenic or lead, they did provide physicochemical analyses for the water samples from the six different villages. While the authors did not specifically analyze lead or arsenic in the water samples, these physicochemical analyses suggest that differential lead or arsenic exposure were unlikely. Moreover, based on water quality maps, arsenic is not expected to be a major concern in this study area. Based on the information from the water quality maps and the physiochemical analysis of the water provided, there is indirect evidence that neither arsenic nor lead were a concern in this study population.
  - *Potentially important study-specific confounders:* Key confounders age, gender, and measures of SES were similar between exposure groups; however, arsenic was not taken into account. Arsenic often occurs in the drinking water along with fluoride in some Indian populations; however, based on water quality maps, arsenic does not appear to be an issue in the study area.
    - *Direction/magnitude of effect***:** Conceptually, the presence of arsenic would potentially bias away from the null if present with fluoride or toward the null if present in the reference group; however, for reasons noted above, arsenic is not considered a concern in this study population.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable, that potential co-exposures were not an issue, and that key confounders were addressed.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results were provided for 84 children, but the methods do not indicate how many children were initially selected to participate nor were any exclusion criteria provided. It was noted in the results that 84 children had their groundwater and urine tested, but it was not noted if analyses were restricted to these children or if exposures were assessed in all the children who had IQ measurements. Correspondence with the study author indicated that the main reason for exclusion was a <80% attendance rate, with fluoride and IQ measured on all 84 children who met the criteria.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence of no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children in villages were grouped based on fluoride levels that were assessed in groundwater (low F villages versus high F villages). The average concentration of these levels was considered to be the levels in the drinking water with confirmation using urinary fluoride levels. The groundwater samples were selected to cover major parts of the taluka and represent overall groundwater quality. Ten samples were

obtained from each village. Fluoride was measured in the groundwater using ion exchange chromatography. Although urine levels were also significantly higher in the high fluoride village, no information was provided on how or when the urinary samples were obtained or how they were measured. However, correspondence with the study author indicated that the groundwater and urine fluoride levels were available for all 84 children indicating that the urine measures were available for the children that had IQ measures. The urine samples were stated to be collected at the same time that the second water sample was collected.

        - *Direction/magnitude of effect*: Fluoride levels were measured in both the drinking water and urine. Although there is some variability in the measurements, there is no overlap between the two groups and the urine and drinking water levels in the children support each other. Any potential exposure misclassification would be non-differential and direction and magnitude are unknown.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Outcome methods were only noted to be reported in Trivedi *et al.* (2007), which was scored as follows: IQ was measured in the children of both areas using a questionnaire prepared by Professor JH Shah, copyrighted by Akash Manomapan Kendra, Ahmedabad, India, and standardized on the Gujarati population with 97% reliability rate in relation to the Stanford-Binet Intelligence Scale (+ for methods). Blinding or other methods to reduce bias are not reported, but correspondence with the study author indicated that the teachers were blind to the status of fluoride. The teachers administered the tests in the presence of a research fellow. It is not completely clear who scored the tests, but it is assumed the teachers. (+ for blinding). Overall rating for methods and blinding = +.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcomes were blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Probably high risk of bias (-)
    - *Summary:*
        - *Statistical analyses:* Mean IQ scores in low and high fluoride villages were compared using a t-test. Consideration of heterogeneity of variances was not reported. Results are reported as means and standard errors of the means, with p-values for significant differences. Area-level exposures were used. There was no accounting for clustering of children within the villages, and comparative

analyses did not account for potential confounders. Urinary fluoride was not considered in the comparative analyses. The lack of individual exposure levels and the lack of accounting for clustering are likely to bias the standard error of the difference in mean IQ levels between the high and low fluoride villages and make the differences appear stronger than they actually are.

- *Basis for rating:* Probably high risk of bias based on indirect evidence that the statistical analyses did not account for clustering, and this lack of accounting could bias the association. There were no other potential threats to risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and the addressing of potential key confounders, but the study was limited by the cross-sectional study design. Another limitation of the study was lack of accounting for clustering, which may bias the standard error of the differences making the effect appear stronger than it actually is; however, this does not change the nearly 5-point difference in IQ score between the two villages.

Wang *et al.* (2012)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 8–13 years (possibly the same study population as Xiang *et al.* (2003a))
- ***Study area:*** Wamiao and Xinhuai villages located in Sihong County, Jiangsu Providence, China
- ***Sample size:*** 526 school children
- ***Data relevant to the review:*** Mean IQ and % low IQ (< 80) by total fluoride intake.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower mean IQ in the endemic versus non-endemic regions, as reported in Xiang *et al.* (2003a); when high exposure group was broken into 4 exposure groups based on fluoride intake, a dose-dependent decrease in IQ and increase in % with low IQ observed; significant correlation between total fluoride intake and IQ ($r = -0.332$); for IQ<80, adjusted OR of total fluoride intake was 1.106 (95% CI: 1.052, 1.163).

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study appears to be the same study population as Xiang *et al.* (2003a) and Xiang *et al.* (2011); however, the study does not cite these studies as providing additional information and numbers of children differ; therefore, it may be a separate analysis on the same villages. The years of testing were not provided so it cannot be determined if study subjects are the same. Two villages, Wamiao and Xinhuai, located 64 km apart in Sihong County, Jiangsu Province were selected for the study. Wamiao is a village in a region with severe endemic fluorosis and Xinhuai is a village in a non-endemic fluorosis region. Neither village has fluoride pollution from coal or industrial sources. Villages were stated to be similar in terms of annual per capita income,

transportation, education, medical conditions, the natural environment, and lifestyle. All primary students ages 8–13 years currently in school in either village were surveyed with exclusions noted. Of 243 children from Wamiao, 236 (97.12%) were included, and of 305 children from Xinhuai, 290 (95.08%) were included. No table of subject characteristics was provided.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited using the same methods within the same time frame, with direct evidence that there was no difference in participation/response rates.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Logistic regression of low IQ rate and total fluoride intake adjusted for age and sex. Both villages had hand-pumped well water for drinking water, but the authors do not mention if arsenic was also present in the drinking water. However, a publication by Xiang *et al. (*2013) on this study area indicates that Xinhuai (the low fluoride area) had significantly higher arsenic levels compared to Wamiao (the endemic fluorosis area), which would bias toward the null. Areas were stated to be similar in annual per capita income, transportation, education, medical conditions, the natural environment, and lifestyle; however, no details were provided. This study did not address other co-exposures, but other studies on populations in these villages (Xiang *et al.* 2011, Xiang *et al.* 2003a) indicate that iodine and lead are not concerns.
  - *Potentially important study-specific confounders:* Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on information provided in Xiang *et al.* (2013), arsenic concentrations were higher in the low fluoride area compared to the high fluoride area. Because there were significant effects on IQ observed in the high fluoride areas, the impact of co-exposure to arsenic is less of a concern. The presence of arsenic in the control village may cause an underestimation of the effect of fluoride, and despite this potential impact, there was still a significant association between fluoride exposure and IQ.
    - *Direction/magnitude of effect***:** Presence of arsenic in this study population would potentially bias toward the null.
  - *Basis for rating:* Probably low of risk bias because there is indirect evidence that the key confounders are take into account, methods used for collecting the information were valid and reliable, and co-exposures to arsenic and lead and iodine deficiency are not attributing to the effect observed in this area. The potential bias toward the null combined with the reporting of an effect increases confidence that there is an effect.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Data are reported for all 526 children noted to be included in the study. There is a slight discrepancy in the reported total number of children from the high-fluoride village and the number of participants from the high-fluoride village between this paper (236 participated of 243 total children) and the 2003 and 2011 publications on the same study population (222 of 238). This discrepancy is not explained but is not expected to appreciably bias the results.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***
  - *Rating:* Probably low risk of bias (+); Probably high risk of bias (-)
  - *Summary:* **Water fluoride (+ probably low risk of bias):** Exposure was based on drinking water levels and fluoride intake. Residents in the Wamiao village were divided into five groups based on fluoride levels in the drinking water. Clean, dry polyethylene bottles were used to collect 50 mL of drinking water from each student's household and fluoride content was measured.

    Total fluoride intake (- probably high risk of bias): Six families from each of the five Wamiao groups were randomly selected as dietary survey households. Intakes of various foods by each person at each meal and intakes of unboiled water, boiled water, and tea were surveyed for four consecutive days. Methods for food collection were described. Five representative households from each village were selected based on geographic location, population distribution, housing structure, and other conditions. Indoor air samples were collected once daily for five consecutive days; outdoor air was sampled at two points once daily for five days. Methods for determining fluoride content in samples were noted to follow specific guidelines. Calculation of total fluoride intake was stated to follow Appendix A of the People's Republic of China Health Industry Standard with some details provided. Although it is assumed the method is valid, it was not detailed how each fluoride determination was made for each subject, and it appears that total fluoride intake was determined based on data from select subjects and not all subjects.
    - *Direction/magnitude of effect***:** There is potential for exposure misclassification based on calculating fluoride intake based on measurements from a few select subjects rather than all subjects. The direction and magnitude of effect cannot be assessed based on the information provided.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure. The intake is probably high risk of bias because there is indirect evidence that the exposure was assessed using methods of unknown validity.

- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* IQ of each child was measured with the Combined Raven's Test for Rural China (CRT-RC) (++ for methods). The test was stated to be administered to the children independently in a school classroom under the supervision of three exam proctors. Testing methods, testing language, and testing conditions were all in strict accordance with the CRT-RC guidebook. Major testing personnel received necessary training by the Psychology Department of East China Normal University. The children undergoing IQ testing and the test scorers were kept double-blinded throughout the testing process. (++ for blinding). Overall rating= ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Logistic regression analysis was used to determine the odds of having low IQ with increasing fluoride intake. Analyses and methods are not well described. There is no mention of what tests were used for the mean IQ comparison by village; however, statistical software (SPSS) was used, suggesting appropriate tests were applied. Simple linear regression analyses were conducted to evaluate associations between total fluoride intake and children's IQ or low IQ rate. There is no evidence that regression diagnostics were used to test model assumptions for linearity, normality, and homogeneity. Clustering at the village level was not accounted for in the analyses. The overall impact of these factors on effect estimates is expected to be minimal given the use of individual-level data and adjustment for potential confounders.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements with blinding at outcome assessment but is limited by the cross-sectional study design and not using individual measurements to calculate fluoride intake. All key confounders were accounted for in the study design or analysis, but there is potential for the presence of arsenic to bias toward the null.

Wang *et al.* (2020b)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** School children aged 7–13 years
- ***Study area:*** Tianjin City, China (possibly a subset of the children from Yu *et al.* (2018))
- ***Sample size:*** 571 school children
- ***Data relevant to the review:*** IQ scores by urine and water fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant associations between IQ score and water fluoride (adjusted β = −1.587 per 1-mg/L increase) and urinary fluoride (adjusted β = −1.214 per 1-mg/L increase) in boys and girls combined based on both quartiles and continuous measures. No significant modification effect of gender.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were from a cross-sectional study conducted in 2015, but no citation was provided on this cohort (presumably the Yu *et al.* (2018) cohort). It was noted that the subjects in that cohort were from districts with historically high or normal fluoride levels. Subjects for this study were selected by using a stratified and multistage random sampling approach. Brief description was provided. The study area consisted of three historically high fluoride areas and four nonendemic areas. A flow diagram was provided for inclusion and exclusion, but this detail was given for all children and not by area. Therefore, it cannot be determined if the participation differed by area. However, there was a 93% recruitment rate, and the 13 excluded due to missing data are not likely excluded due to exposure. Detailed characteristics of the study population are provided. Exclusion criteria included: "children who had congenital or acquired diseases affecting intelligence, or a history of cerebral trauma and neurological disorders, or those with a positive screening test history (like hepatitis B virus infection, Treponema palladium infection and Down's syndrome) and adverse exposures (smoking and drinking) during maternal pregnancy, prior diagnosis of thyroid disease, and children who had had missing values of significant factors (2.2%) were also excluded."
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were recruited using similar methods during the same time frame and that any differences between the exposed groups were accounted for in the statistical analyses.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Study authors noted that the study areas are not exposed to other neurotoxins such as lead, arsenic, or mercury nor were they iodine-deficient. Final models included child's age, child's gender, child's BMI, maternal and paternal education, household income, and low birth weight. Other potential confounders that were considered is unclear as they only noted that the confounders were selected based on current literature. Reasons for exclusion included history of disease affecting intelligence, history of trauma or neurological disorders, positive screening test history, or exposures such as smoking or drinking during pregnancy. Information was obtained by questionnaire or measurements. Variables such as parental BMI, behavioral and mental health disorders, IQ, and quantity and quality of the caregiving environment were not addressed.
  - *Potentially important study-specific confounders:* All key confounders were considered in this study.
    - *Direction/magnitude of effect***:** Not applicable.
  - *Basis for rating:* Probably low risk of bias because there is direct evidence that the key confounders are taken into account, indirect evidence that the methods for collecting the information were valid and reliable, and co-exposure to arsenic is not an issue in this area.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* A detailed chart of the recruitment process is presented. The study had a 93% recruitment rate and only 2.2% of subjects with missing data for certain covariates were excluded.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children provided spot urine samples, presumably at the time of examination. Water samples were randomly collected from public water supplies in each village. Fluoride concentrations were analyzed using fluoride ion-selective electrode according to the national standardized method in China. There is no indication if the urine samples accounted for dilution.
    - *Direction/magnitude of effect***:** Not accounting for dilution could cause there to be some exposure misclassification. The direction and magnitude would depend on where the differences occurred.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Assessments of IQ scores were conducted by graduate students at the School of Public Health, Tongji Medical College at the Huazhong University of Science and Technology. Each team member was assigned a single task, meaning that only one person would have conducted the IQ tests. A Combined Raven's Test for Rural China was used. Therefore, the test was appropriate for the study population (++ for method). It was note that the examiner was trained and blind to the exposure (++ for blinding). Overall = ++
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Logistic and multivariate regression models accounting for potential confounders were used. Results are presented as betas or ORs and 95% CIs. Regression diagnostics were conducted for all models, including

examination of multicollinearity, heteroscedasticity, and influential observations. Mediation and interaction analyses were appropriate. There is no evidence that the stratified and multistage random sampling approach for subject selection was accounted for in the analyses by using sampling weights or accounting for clustering using random effect models; however, selected villages are similar in population and general demographic characteristics. Given the use of individual-level data and adjustment for potential confounders, the impact on the regression coefficients is likely to be minimal.
        - *Other potential concerns:* None identified.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and no other potential threats to risk of bias were identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key confounders were considered in the study design or analysis.

Xiang *et al.* (2003a)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 8–13 years
- ***Study area:*** Wamiao and Xinhuai villages located in Sihong County, Jiangsu Providence, China
- ***Sample size:*** 512 school children
- ***Data relevant to the review:*** Comparison of IQ (mean and distribution) between Wamiao County (a severe endemic fluorosis area) and Xinhuai County (non-endemic fluorosis area); additional breakdown of the Wamiao area into 5 water fluoride exposure groups.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower IQ scores observed with water fluoride levels of 1.53 mg/L or higher. Percent of subjects with IQ scores below 80 was significantly increased at water fluoride levels of 2.46 mg/L or higher. Significant inverse correlation between IQ and urinary fluoride (Pearson correlation coefficient −0.164). Mean IQ scores for children in the non-endemic region (100.41 ± 13.21) were significantly higher than the endemic region (92.02 ± 13.00).

Risk of Bias:

- **Author contacts:**
    - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Two villages, Wamiao and Xinhuai, located 64 km apart in Sihong County, Jiangsu Province were selected for this study, which was conducted between September and December 2002. Wamiao is located in a severe fluorosis endemic area, and Xinhuai is located in a non-endemic fluorosis area. Neither village has fluoride pollution from burning coal or other industrial sources. All eligible children in each village were included; children who had been absent from either village for 2 years or longer or who had a history of brain disease or head injury were excluded. In Wamiao, 93% of the

children (222 out of 238) were included for the study; in Xinhuai, 95% were included (290 out of 305). The children in Wamiao were divided into five subgroups according to the level of fluoride in their drinking water: <1.0 mg/L (group A), 1.0–1.9 mg/L (group B), 2.0–2.9 mg/L (group C), 3.0–3.9 mg/L (group D), and >3.9 mg/L (group E). Children in Xinhuai (0.18–0.76 mg F/L in the drinking water) served as a control group (group F). Demographic characteristics are not presented, and statistical analyses are not adjusted, but mean IQ scores are stratified by child's age, child's gender, family income, and parental education.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited using the same methods within the same time frame, with direct evidence that there was no difference in participation/response rates.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although information was stated to be collected on personal characteristics, medical history, education levels of the children and parents, family SES, and lifestyle, only child's gender, child's age, family income, and parental education were addressed. Other potential co-exposures, such as arsenic, were not addressed. A separate publication in 2003 [(Xiang *et al.* 2003b), letter to the editor], indicated that blood lead levels were not significantly different between the two areas. Although arsenic was not addressed specifically in this publication, Xiang *et al.* (2013) measured both fluoride and arsenic in the Wamiao and Xinhuai areas. Xinhuai (the low fluoride area) had significantly higher arsenic levels compared to Wamiao (the endemic fluorosis area). This is likely to bias toward the null; however, the study observed a significantly lower IQ score in the endemic fluorosis area. Iodine was tested in a subset of the children and found not to be significantly different between the two groups.
  - *Potentially important study-specific confounders:* Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on information provided in Xiang *et al.* (2013), arsenic concentrations were higher in the low fluoride area compared to the high fluoride area. Because there were significant effects on IQ observed in the high fluoride areas, the impact of co-exposure to arsenic is less of a concern. The presence of arsenic in the control village may cause an underestimation of the effect of fluoride, and despite this potential impact, there was still a significant association between fluoride exposure and IQ.
    - *Direction/magnitude of effect***:** Presence of arsenic in this study population would potentially bias towards the null.
  - *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key confounders are taken into account, methods used for collecting the information were valid and reliable, and co-exposures to arsenic and lead and iodine deficiency are not attributing to the effect observed in this area. The potential bias toward the null combined with the reporting of an effect increases confidence that there is an effect.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are complete. IQ results were reported for all 512 children included in the study (222 in the endemic area and 290 in the nonendemic area).

- *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.

- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Exposure was based on drinking water and urinary levels of fluoride. The two study areas were selected to reflect a severe endemic area and a nonendemic area. Drinking water was collected from wells and early-morning spot urine samples were collected from a randomly-selected subsample of children. Both water and urine samples were measured using fluoride ion-selective electrode, but no quality control was discussed. Both absolute and creatinine-adjusted urine results were reported.
    - *Direction/magnitude of effect***:** There is potential for exposure misclassification because only current levels were assessed. Migration of subjects in or out of the area was not assessed, but the study authors noted that, if the children had been absent from the village for 2 or more years, they were excluded. Misclassification would likely be non-differential, which could bias the results in either direction.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* The IQ of each child was measured with the Combined Raven's Test for Rural China (CRT-RC) (++ for methods). The test was stated to be administered to the children independently in a school classroom, in a double-blind manner, under the supervision of an examiner and two assistants, and in accordance with the directions of the CRT-RC manual regarding test administration conditions, instructions to be given, and test environment. (++ for blinding). Overall rating= ++
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* There is no mention of the tests conducted, but data were stated to be analyzed using SAS suggesting appropriate tests were applied. Results provided in the tables indicate that t-tests comparing IQ values between the villages (overall and by gender) were conducted, but it was not reported that heterogeneity of variance was assessed. In addition, correlations between IQ and age, family income, and parents' education level were tested with Pearson's correlation. There is no evidence that a test for trend was conducted

to evaluate the stated "significant inverse concentration-response relationship between the fluoride level in drinking water and the IQ of children."

- A potential concern raised by the NASEM (2020) peer review was the lack of accounting for relationships in exposure between persons from the same village. Given only two villages were included and the analyses consisted of village-level comparisons (no use of individual-level covariate data), it is likely that the standard error of the difference in mean IQ between fluoride in water exposure groups will be biased, making differences appear stronger than they actually are. Without controlling for village effects and given the large differences in fluoride concentrations and IQ levels between villages, the apparent dose-response relationship could be due to a village effect in addition to a fluoride effect. However, the dose-response relationship is apparent within the "exposed" village, diminishing the concern for a village-only effect and likely minimizing the impact on the effect estimates.
- *Other potential concerns:* None identified.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that statistical analyses were appropriate and that there were no other threats to risk of bias.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and outcomes blindly assessed but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key confounders were considered in the study design or analysis, but there is potential for the presence of arsenic to bias toward the null.

Xiang *et al.* (2011)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 8–13 years (same study population as Xiang *et al.* 2003a)
- ***Study area:*** Wamiao and Xinhuai villages located in Sihong County, Jiangsu Providence, China
- ***Sample size:*** 512 school children
- ***Data relevant to the review:*** Mean IQ scores and odds ratio for having an IQ < 80 presented by serum fluoride quartiles.
- ***Reported association with fluoride exposure:*** Yes: Significant linear trend across quartiles of serum fluoride and children's IQ score < 80 (adjusted ORs for Q1 and Q2; Q1 and Q3; and Q1 and Q4, respectively: 1; 2.22 [95% CI: 1.42, 3.47]; and 2.48 [95% CI: 1.85, 3.32]); significant effects observed at ≥ 0.05 mg/L serum fluoride.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study population is the same as that was used in the Xiang *et al.* (2003a) study, but a few more measurements were available and different analyses were conducted. The comparison population is considered the same as previously based on

the study populations being recruited from similar populations, using similar methods, during the same time frame. Demographic characteristics were not provided.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited using the same methods within the same time frame, with direct evidence that there was no difference in participation/response rates.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* As was noted in the 2003 publication, information was collected on personal characteristics, medical history, education levels in the children and parents, family SES, and lifestyle. In the logistic regression model, age and gender were adjusted in the analysis. In the previous report, no significant associations were observed between groups for family income and parents' education. Urinary iodine and blood lead levels were also stated to be measured and were noted not to be significantly different between the groups. Although the iodine levels were reported in the previous publication, the lead levels were not reported nor were the methods. Lead information is reported in a letter to the editor (Xiang *et al.* 2003b) and was not significantly different between the areas. Although arsenic was not addressed specifically in this publication, Xiang *et al.* (2013) measured both fluoride and arsenic in the Wamiao and Xinhuai areas. Xinhuai (the low fluoride area) had significantly higher arsenic levels compared to Wamiao (the endemic fluorosis area). This is likely to bias toward the null; however, the study observed a significantly lower IQ score in the endemic fluorosis area and with increasing serum fluoride.
  - *Potentially important study-specific confounders:* Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on information provided in Xiang *et al.* (2013), arsenic concentrations were higher in the low fluoride area compared to the high fluoride area. Because there were significant effects on IQ observed in the high fluoride areas, the impact of co-exposure to arsenic is less of a concern. The presence of arsenic in the control village may cause an underestimation of the effect of fluoride, and despite this potential impact, there was still a significant association between fluoride exposure and IQ.
    - *Direction/magnitude of effect***:** Presence of arsenic in this study population would potentially bias toward the null.
  - *Basis for rating:* Probably low of risk bias because there is indirect evidence that the key confounders are taken into account, methods used for collecting the information were valid and reliable, and co-exposures to arsenic and lead and iodine deficiency are not attributing to the effects observed in this area. The potential bias toward the null combined with the reporting of an effect increases confidence that there is an effect.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are reported for all 512 children noted to be included in the study.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)

- *Summary:* Fluoride levels were measured in serum with a fluoride ion-selective electrode. A fasting venous blood sample was used. No details are provided on validation (including correlation with drinking water levels) or QA. Children who did not reside in their village for at least 2 years were excluded. Results were provided in quartiles, but they combined the lower two quartiles. After combining the two lower quartiles into one, the exposure levels ranged from <0.05 mg/L (Q1 + Q2) to >0.08 mg/L (Q4).
  - *Direction/magnitude of effect***:** Serum fluoride may not be the best estimate for exposure. There is potential for exposure misclassification because only current levels were assessed. Migration of subjects in or out of the area was not assessed, but the study authors noted that, if the children had been absent from the village for 2 or more years, they were excluded. Misclassification would likely be non-differential, which could bias results in either direction.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* IQ was assessed as part of the 2003 evaluation. IQ was measured with the Combined Raven's Test for Rural China which is appropriate for this population (++ for methods). Although this study does not provide details, the original study article from 2003 provides specific details. The study authors indicate in the 2003 publication that the tests were conducted in a double-blind manner and these are the same results and population (++ for methods). Overall rating=++
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses conducted were appropriate for the study. Chi square tests were used to compare categorical variables, and multiple logistic regression was used to evaluate the association between serum fluoride levels and risk of low IQ. A potential concern raised by the NASEM (2020) peer review was the lack of accounting for relationships in exposure between persons from the same village. Although only two villages were included, in the analyses which consisted of village-level comparisons it is likely that the standard error of the difference in mean IQ between villages will be biased. This is less of a concern for the mean IQ comparisons across quartiles of serum fluoride levels, and for the logistic regression analyses of risk of low IQ and

individual-level serum fluoride levels. Without controlling for village effects and given the large differences in fluoride concentrations and IQ between villages, the apparent dose-response relationship could be due to a village effect in addition to a fluoride effect. However, the dose-response relationship is still present within the "exposed" village, diminishing the concern for a village-only effect and likely minimizing the impact on the effect estimates.

        - *Other potential concerns:* None identified.
    - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements with blinding at outcome assessment but is limited by the cross-sectional study design and use of serum concentrations. All key confounders were considered in the study design or analysis, but there is potential for the presence of arsenic to bias toward the null.

Yu *et al.* (2018)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 7–13 years
- ***Study area:*** Tianjin City, China
- ***Sample size:*** 2,886 school children
- ***Data relevant to the review:*** IQ for normal (≤ 1 mg/L) versus high (> 1 mg/L) water fluoride; betas for IQ score by water and urine fluoride groupings; ORs by IQ category using water and urine fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant difference in mean IQ scores in high water fluoride areas (>1.0 mg/L; 106.4 ± 12.3 IQ) compared to the normal water fluoride areas (≤1.0 mg/L; 107.4 ± 13.0) water fluoride areas. Distribution of IQ scores was also significantly different (p = 0.003). Every 0.5-mg/L increase in water fluoride (between 3.40 and 3.90 mg/L) was associated with a 4.29 lower IQ score (95% CI: −8.09, −0.48).

Risk of Bias:

- **Author contacts:**
    - Authors were contacted in September 2018 to obtain additional information for the risk-of-bias evaluation.
- ***Population selection:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* School children (2,886), aged 7–13 years, were recruited from the rural areas of Tianjin City, China. After exclusion, 1,636 children were assigned to the "normal-fluoride" exposure group and 1,250 were assigned to the "high-fluoride" exposure group based on a cut-off water fluoride level of 1.0 mg/L. A multi-stage random sampling technique, stratified by area, was performed to select representative samples among local children who were permanent residents since birth. Detailed characteristics of the study population are provided. Exclusion criteria included: 1) children who had

congenital or acquired diseases affecting intelligence, 2) children with a history of cerebral trauma and neurological disorders, 3) children with a positive screening test history (like hepatitis B virus infection, Treponema palladium infection and Down's syndrome), and 4) children with adverse exposures (smoking and drinking) during maternal pregnancy. A table of characteristics was provided by fluoride level with differences adjusted in the analysis.

- *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were recruited using similar methods during the same time frame and that any differences between the exposed groups were considered in the statistical analyses.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Demographic data were collected by trained investigators during a face-to-face interview with the recruited children and their parents. Questionnaires were not stated to be validated. The developmental status of the children was further assessed by calculation of BMI, and all measurements were conducted by nurses based on recommended standard methods. Variables that presented differential distribution between the normal-fluoride and high-fluoride exposure groups were adjusted in the linear regression analysis of IQ data and included age, sex, paternal and maternal education levels, and low birth weight. Children exposed to smoking in utero were excluded from the study. Sensitivity analyses were conducted by modifying covariates adjusted in multivariable models among demographics (age and sex); development (BMI); socioeconomics (maternal education, paternal education, and household income); history of maternal disease during pregnancy (gestational diabetes, malnutrition, and anemia); and delivery conditions (hypoxia, dystocia, premature birth, post-term birth, and low birth weight). None of the study sites selected were in areas endemic for iodine deficiency disorders nor were other potential neurotoxins like lead, arsenic, and mercury present. Variables such as parental BMI and behavioral and mental health disorders were not addressed.
  - *Potentially important study-specific confounders:* All key confounders were considered in this study.
    - *Direction/magnitude of effect***:** Not applicable.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that methods of obtaining the information were valid and reliable and direct evidence that all key confounders and co-exposures were addressed.
- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There were 1,636 children assigned to the "normal-fluoride" exposure group based on water fluoride, and 1,250 children were assigned to the "high-fluoride" exposure group. Exclusion from the original group of 2,886 children was adequately described. A total of 2,380 children provided urine samples. There is no indication that the data presented excludes any additional children or urine samples, but results do not indicate a sample size for all results.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* According to the annual surveillance data from the CDC, the drinking water sources and water fluoride concentrations in each village had remained at stable levels over the past decade. During the investigation, water samples were collected randomly from the public water supplies in each village. Spot (early-morning) urine samples from every child and water samples from each village were collected in pre-cleaned, labeled polythene tubes and transported to the lab within 24 hours while frozen. Samples were stored at −80°C until analysis. Concentrations of fluoride ions (mg/L) were analyzed using the national standardized ion-selective electrode method in China; the detection limit was 0.01 mg/L. Samples were diluted with an equal volume of total ionic strength adjusted buffer (TISAB) of pH 5–5.5 for optimal analysis. Double-distilled deionized water was used throughout the experiment. There is no reporting of any QC methods.
    - *Direction/magnitude of effect***:** Spot urine samples may lead to non-differential exposure misclassification. The large population size likely dilutes any potential effects of occasional misclassification. Because the drinking water sources of fluoride had been noted to be stable for the past decade and the children were 13 years or younger, there would only be exposure misclassification if there was a lot of migration between areas.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* IQ scores were measured using the second edition of Combined Raven's Test-The Rural in China (CRT-RC2) for children aged 7–13 years (++ for methods). The test was completed by each participant within 40 minutes according to the instruction manual. For each test, 40 children were randomly allocated to one classroom to take the test independently under the supervision of four trained professionals. There is no mention of whether the evaluators were blinded to the fluoride group of each child (normal vs. high fluoride) or whether there were steps taken to ensure consistency in scoring across the evaluators. It is also not clear if the 40 children randomly assigned to the classroom were specific to the village or if a local center was used. Correspondence with the study authors indicated that the four professionals worked together throughout the examination without knowledge of the child's fluoride exposure (++ for blinding).
  - *Basis for rating:* Definitely low risk of bias based on the direct evidence that the outcome was blindly assessed using instruments that were valid and reliable.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)

- *Summary:*
  - *Statistical analyses:* Statistical analyses used were appropriate for the study. Univariate and multivariable piecewise linear regression models were used to estimate the associations between water fluoride or urinary fluoride levels and IQ scores. Multiple logistic regression analysis was used to evaluate the association between water or urinary fluoride levels and IQ degree using the normal intelligence group as the control. Sensitivity analyses were conducted. There is no evidence that residual diagnostics were used to examine model assumptions or that the complex sampling design (stratified multistage random sampling) was accounted for in the analysis using sampling weights and adjustment for clustering. The impact of these factors on the effect estimates is expected to be minimal given the use of individual-level data and adjustment for and numerous potential confounders.
  - *Other potential concerns:* None identified.
- *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements with blinding at outcome assessment but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key confounders including potential co-exposures were considered in the study design or analysis.

Zhang *et al.* (2015b)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 10–12 years
- ***Study area:*** Tianjin City, China
- ***Sample size:*** 180 children
- ***Data relevant to the review:*** IQ by control and high fluoride groups; IQ correlations with water, serum, or urinary fluoride levels; betas for IQ with urinary fluoride levels (by genotypes)
- ***Reported association with fluoride exposure:*** Yes: S Significant correlation between IQ score and children's serum fluoride (r = –0.47) and urinary fluoride (r = –0.45); significant difference in mean IQ score for high-fluoride area (defined as >1 mg/L in drinking water; 102.33 ± 13.46) compared with control area (<1 mg/L; 109.42 ± 13.30).

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary*: Subjects were similar and recruited during the same time frame using the same methods. Authors recruited schoolchildren from a high fluoride area (1.40 mg/L) and a control area (0.63 mg/L) in Tianjin City, China. In accordance with the principles of

matching social and natural factors such as educational standard, economic situation, geological environments as much as possible, two areas with different fluoride concentrations in the groundwater were selected by a stratified cluster random sampling of this region. A total of 180 5th grade children aged 10 to 12 years from two primary schools located 18 km apart in the Jinnan District were recruited—Gegu Second Primary School (from an endemic fluorosis area) and Shuanggang Experimental Primary School (from a non-endemic fluorosis area). The areas are not affected by other drinking water contaminants, such as arsenic or iodine. All subjects were unrelated ethnic Han Chinese and residents in Tianjin with similar physical and mental health status. The authors excluded subjects with known neurological conditions including pervasive developmental disorders and epilepsy. Descriptive statistics of the study population are presented by exposure group in Table 1 of the study. A number of potential differences are taken into account in the statistical analyses.

- *Basis for rating*: Definitely low risk of bias based on direct evidence that the exposure groups were similar and recruited using similar methods during the same time frame.

- **Confounding:**
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariates included in the statistical models were child's age, child's gender, educational levels of parents, drinking water fluoride (mg/L), and levels of thyroid hormones (T3, T4, and TSH). Authors report that the study areas are not affected by other contaminants such as arsenic or iodine and residents were of similar physical and mental health status. Other important confounders (maternal demographics, smoking, reproductive health) were not considered. Covariate data were obtained from a study questionnaire.
  - *Potentially important study-specific confounders:* All key confounders were considered in this study.
    - *Direction/magnitude of effect*: Not applicable.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and direct evidence that key confounders including potential co-exposures were addressed.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results are complete for the 180 children selected for the study.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Drinking water samples (10 mL) were collected from the tube wells of each child's household. Three fasting venous blood samples were also collected. Urine samples were collected in the early morning before breakfast. Fluoride contents in drinking water (W-F), serum (S-F), and urine (U-F) were measured using an ion analyzer EA940 with a fluoride ion-selective electrode (Shanghai constant magnetic electronic technology Co, Ltd, China) according to the China standard GB 7484-87. All reference solutions for the fluoride determinations were double-deionized water. Parallel samples were set for determination and averages were taken. The quantitation limits of this

method for W-F, S-F, and U-F were 0.2, 0.012, and 0.5 mg/L, respectively. Recovery rates for this method were in the range of 94.3%–106.4%. The intra- and inter-assay coefficients of variation for fluoride were 2.7% and 6.7%, respectively. Dilution of the urinary fluoride was not addressed.
        - *Direction/magnitude of effect*: Not applicable.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* A Combined Raven's Test for Rural China (CRT-RC) was taken to evaluate the IQ of each child (++ for methods). The study report stated that all tests were administered at school by a trained examiner who was masked to participants' drinking water fluoride levels (++ for blinding). Overall rating for methods and blinding=++.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All results outlined in the abstract, introduction, and methods sections were reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:*
        - *Statistical analyses:* Associations between serum and urinary fluoride levels and IQ score were estimated using general linear models and multivariate linear regression by COMT polymorphism. Normality (Kolmogorov-Smirnov test) was evaluated for all continuous variables. There is no evidence that residual diagnostics were used to examine model assumptions or that the complex sampling design (stratified multistage random sampling) was accounted for in the analysis using sampling weights and adjustment for clustering. The impact of these factors on the regression effect estimates is expected to be minimal given the use of individual-level data and adjustment for numerous potential confounders.
        - *Other potential concerns:* None identified.
    - *Basis for rating:* Probably low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, outcomes blindly assessed, and assessment of potential key confounders including potential co-exposures.

Other Neurodevelopmental Studies

Barberio *et al.* (2017b)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Canadian Health Measures Survey (cycles 2 and 3) participants (children aged 3−12 years)
- ***Study area:*** general population of Canada
- ***Sample size:*** 2,221 children (1,120 from Cycle 2, 1,101 from Cycle 3)
- ***Data relevant to the review:*** Associations between learning disability or ADHD (Cycle 2 only) assessed by parent or child self-report and urinary fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant increase in adjusted OR for learning disability with unadjusted urinary fluoride (1.02; 95% CI: 1.00, 1.03) when Cycle 2 and 3 were combined. No significant associations with creatinine-adjusted or specific gravity-adjusted urinary fluoride. No significant association between urinary fluoride and ADHD.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* The comparison groups were selected from Cycles 2 and 3 of the Canadian Health Measures Survey. This is a nationally representative sample of residents living in 10 provinces, with clear exclusion criteria provided. Exclusion only represented about 4% of the target population (all Canadian residents 3−79 years old living in 10 provinces). A table of characteristics of the study population is provided.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the subjects were recruited from the same population using the same methods during the same time frame and exposure groups were similar.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study adjusted for sex, age (3–12 years old), household education, and household income adequacy. Variables to discern fluoride source, including drinking water and dental products, were also considered. Cycle 2 data also included adjustments for: 1) children for whom tap water (vs. bottled or other) was the primary source of drinking water at home or away from home and 2) children who had lived in his or her current home for 3 or more years. Confounders such as parental behavioral and mental health disorders, smoking, and nutrition were not discussed. The study used data from the Canadian Health Measures Survey which consists of a nationally representative sample of Canadians. Most Canadians (~89%) receive water from municipal water supplies, which monitor for levels of lead and arsenic. Therefore, co-exposure to lead and arsenic are less likely an issue in this population and the lack of information is not considered to appreciably bias the results.
  - *Potentially important study-specific confounders:* All key confounders were considered in this study.

        - *Direction/magnitude of effect*: Not applicable.
    - *Basis for rating:* Probably low risk of bias based on direct evidence that key confounders were addressed and indirect evidence that the methods used to collect the information were valid and reliable and that co-exposures were not an issue.
- ***Attrition:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Covariate data were missing for less than 5% of all analyses, apart from household income; household income was reported for only 71–77% of participants and was imputed for the remainder.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Estimates of urinary fluoride (µmol/L) from spot urine were available for a subsample of respondents. Analysis was performed under standardized operating procedures at the Human Toxicology Laboratory of the Institut National de Santé Publique du Québec (accredited under ISO 17025). Fluoride content of urine samples was analyzed using an Orion pH meter with a fluoride ion-selective electrode with limits of detection of 20 µg/L (Cycle 2) and 10 µg/L (Cycle 3). Urinary dilution was addressed by using creatinine-adjusted levels as well as specific gravity-adjusted levels. In Cycle 3 only, estimates of the fluoride concentration of tap water samples collected from randomly selected households were available. The subsample of households selected for tap water sample collection corresponded to the person-level urine fluoride subsample. Analysis of the fluoride concentration of tap water was performed using a basic anion exchange chromatography procedure, with a limit of detection of 0.006 mg/L. QC methods were not addressed.
        - *Direction/magnitude of effect*: There is not any specific direction or magnitude of bias expected. Urinary fluoride levels are reflective of a recent exposure. Having a single concurrent measurement may not be reflective of the exposure associated with the outcome, but if subjects lived in the same area throughout life the exposure may be an adequate representation. Although there is possible exposure misclassification it would be non-differential.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
    - *Rating:* Probably high risk of bias (-)
    - *Summary:* The primary outcome variable, diagnosis of a learning disability by a health professional, was based on a single item from a household survey asked to all respondents: "Do you have a learning disability?". Answer options were: "yes", "no", "don't know", or the participant refused to answer. For Cycle 2, those who indicated having a learning disability were also asked what kind, with the answer options of: "ADD", "ADHD", "dyslexia", or "other". This question was omitted in Cycle 3 and the reason for omission is not described. Parents or guardians answered all questions for children aged 3–11 years, while children 12 years and older answered questions

themselves. The self-reporting of a learning disability did not appear to have been confirmed by medical records or a health professional. (- for methods based on self-report of diagnosis by a health care professional also in Cycle 3 no specific disabilities were described). Blinding was not a concern as spot urine samples were sent to a separate lab and self-reports would not have knowledge of their urine or tap water exposure level (+ for blinding). Overall rating = -.

- *Basis for rating:* Probably high risk of bias based on indirect evidence that the outcome was measured using an insensitive method in the study population.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods sections were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Logistic regression analyses, adjusted and unadjusted for covariates, examined the associations between fluoride exposure and diagnosis of learning disability. Analyses were performed for Cycle 2 only (urinary fluoride and type of learning disability diagnosis), Cycle 3 only (urinary fluoride, water fluoride, and learning disability diagnosis), and Cycles 2 and 3 combined. Analyses used survey weights and bootstrapped weights to ensure proper computation of variance estimates. Results are reported as unadjusted and adjusted ORs with 95% CIs.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include individual exposure measurements and the addressing of potential key confounders but was limited by the cross-sectional study design and insensitive outcome measures.

Bashash *et al.* (2017)

Study Details:

- ***Study design*:** Prospective cohort
- ***Population*:** Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants (pregnant mothers and their children aged 4 or 6–12 years).
- ***Study area:*** Mexico City, Mexico
- ***Sample size:*** 299 mother–child pairs, of whom 287 had data for the general cognitive index (GCI).
- ***Data relevant to the review:*** Adjusted and unadjusted associations between GCI and maternal or child's urinary fluoride concentrations.

- ***Reported association with fluoride exposure:*** Yes: Significant association between maternal urinary fluoride and GCI score (adjusted β = −3.15; 95% CI: −5.42, −0.87). No significant associations with children's urinary fluoride.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Study participants were selected from two different cohorts from three hospitals in Mexico City that serve low-to-moderate income populations. One cohort was from an observational study of prenatal lead exposure and neurodevelopment outcomes and the other was from a randomized trial of the effect of calcium on maternal blood lead levels. The authors state that participants had no history of psychiatric disorders, high-risk pregnancies, gestational diabetes, illegal drug use, or continuous prescription drugs, but they do not include any information on smoking habits. Study populations appear to be similar, but there may be some differences because subjects were selected from two different cohorts that were recruited from slightly different time periods.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar despite the subjects coming from different original study populations where different methods were used for recruitment.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Data were collected via questionnaire on maternal age, education, marital status at first prenatal visit, birth order, birth weight, gestational age at delivery, maternal smoking, maternal IQ, and HOME scores. All models were adjusted for gestational age at birth, child's sex, birth weight, birth order, child's age at testing, maternal marital status, smoking history, age at delivery, maternal IQ, education, and cohort, with additional testing for children's urinary fluoride, mercury, lead, and calcium. Sensitivity analyses additionally adjusted for HOME score. Confounders not considered included BMI, iodine deficiency, arsenic, and maternal mental health and nutrition. Arsenic is assumed not to be a potential co-exposure in this population as the study authors did not discuss it as an issue but did discuss other co-exposures. Arsenic is included in the water quality control program in Mexico City and is not considered a concern in this population.
  - *Potentially important study-specific confounders:* All key confounders were addressed.
    - *Direction/magnitude of effect***:** Not applicable.
  - *Basis for rating:* Probably low risk of bias based on direct evidence that key confounders including other potential co-exposures were addressed and indirect evidence that the methods used to collect the information were valid and reliable and that arsenic is not likely to be an issue in this study population.
- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although there was a large amount of attrition, the study authors clearly describe all reasons for attrition and also provide characteristics to compare those

participants included to those excluded. There were some slight differences between those included and those excluded, but there is nothing to indicate that the attrition would potentially bias the results.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* Urinary fluoride concentrations were determined in spot urine samples (2nd morning void) collected from mothers (during at least one trimester) and children ages 6–12 years. Fluoride content was measured using ion-selective electrode-based assays. QC methods were described including between laboratory correlations. All samples were measured in duplicate. Extreme outliers were excluded. Urinary dilution was addressed by using creatinine-adjusted levels.
        - *Direction/magnitude of effect***:** Not applicable.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* Outcome was assessed using the McCarthy Scales of Children's Abilities (MSCA) in 4-year-old children (translated into Spanish) and the Wechsler Abbreviated Scale of Intelligence (WASI) in 6–12-year-olds. The WASI is a well-established test and the validity of both tests is well documented by the authors. Inter-examiner reliability was evaluated and reported with a correlation of 0.99 (++ for methods). The study report stated that psychologists were blind to the children's fluoride exposure (++ for blinding). Overall rating for methods and blinding = ++.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:*
        - *Statistical analyses:* Statistical analyses used were appropriate for the study. Statistical tests of bivariate associations [using Chi-square tests for categorical variables and analysis of variance (ANOVA)] were used to compare the means of the outcomes or exposure within groups based on the distribution of each covariate. Generalized additive models (GAMs) were used to estimate the adjusted association between fluoride exposure and measures of children's intelligence. Residual diagnostics were used to examine model assumptions and

identify any potentially influential observations. Results are reported as adjusted effects and 95% CIs. In sensitivity analyses, regression models accounted for clustering at the cohort level by using cohort as a fixed effect in the models. Although using cohort as a random effect would be more appropriate, using individual-level exposure data and accounting for numerous potential confounders in the models likely captured the cohort effect. Additional models with cohort as a random effect were also subsequently made available via personal communication with the study authors and showed similar results to the main model.

        - *Other potential concerns:* None identified.
    - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, outcome blindly assessed, and the prospective cohort study design.

Bashash *et al.* (2018)

Study Details:

- ***Study design*:** Prospective cohort
- ***Population*:** ELEMENT participants (pregnant mothers and their children aged 6–12 years)
- ***Study area:*** Mexico City, Mexico
- ***Sample size:*** 210 mother–child pairs
- ***Data relevant to the review:*** Associations between ADHD and other attention/impulsivity scores and maternal urinary fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant associations between maternal urinary fluoride and Conners' Rating Scales-Revised (CRS-R) scores, including Cognitive Problems and Inattention Index (adjusted β = 2.54; 95% CI: 0.44, 4.63), DSM-IV Inattention Index (adjusted β = 2.84; 95% CI: 0.84, 4.84), DSM-IV ADHD Total Index (adjusted β = 2.38; 95% CI: 0.42, 4.34), and ADHD Index (adjusted β = 2.47; 95% CI: 0.43, 4.50).

Risk of Bias:

- **Author contacts:**
    - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Participants were a subset of mother-child dyads enrolled in various longitudinal birth cohort studies of the Early Life Exposure in Mexico to Environmental Toxicants (ELEMENT) project. Subjects were included from two of the four cohorts for which maternal urinary samples were available. Participants in cohort 2A were recruited between 1997 and 1999, and participants in cohort 3 were recruited from 2001 to 2003. Inclusion and exclusion criteria were applied consistently across the two cohorts. A table of subject characteristics was provided in the study and any differences were considered in the analysis. Study populations appear to be similar, but there may be some

differences because subjects were selected from two different cohorts: one from an observational study on prenatal lead exposure and the other from a randomized trial on the effects of calcium on blood lead levels. In addition, they were recruited from slightly different time periods.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposed groups were similar, and any differences were taken into account in the analysis.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Questionnaires were used to collect information on maternal age, maternal education, history of smoking, and marital status during the first pregnancy visit. Child information at birth included birth weight, sex, birth order, and gestational age as calculated by the nurse. Mothers also responded to an SES questionnaire during the visit when the psychometric tests were administered. The Home Observation for Measurement of the Environment (HOME) score was evaluated in a subset of participants. Covariates were selected a priori. Models adjusted for maternal age at delivery, years of education, marital status, smoking history, gestational age at birth, age at outcome assessment, child's sex, birth order, SES, cohort, and calcium intervention. Arsenic is included in the water quality control program in Mexico City and is not considered a concern in this population.
  - *Potentially important study-specific confounders:* None identified, although this study did not specifically address arsenic or other co-exposures. Bashash *et al.* (2017) addressed potential co-exposure to lead and mercury but did not address arsenic. Arsenic was potentially addressed as part of the water quality program in Mexico City.
    - *Direction/magnitude of effect***:** Not applicable.
  - *Basis for rating:* Probably low risk of bias based on direct evidence that key confounders were addressed and indirect evidence that the methods used to collect the information were valid and reliable and that arsenic and other potential co-exposures are not likely to be an issue in this study population.
- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although there was a large amount of attrition from the original cohorts, it was unlikely related to outcome or exposure and there were very little missing data from those included in the study. Of the 231 mothers with a minimum of one maternal urine fluoride measurement and matching outcome identified for the project, only 17 were excluded based on incomplete demographic and outcome information.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Mothers provided at least one spot urine sample during pregnancy. As described in Bashash *et al.* (2017), urinary concentrations were determined on second morning void. Fluoride content was measured using ion-selective electrode-based assay. Bashash *et al.* (2017) describes QC methods. All samples were measured in duplicate

and extreme outliers were excluded. Urinary dilution was addressed by using creatinine-adjusted levels.
        - *Direction/magnitude of effect*: N/A
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* Behaviors associated with ADHD were assessed using the Spanish version of the Conners' Rating Scales-Revised, which has been validated for the evaluation of ADHD. Mothers completed the CRS-R at the same follow-up visit that the child completed the CPT-II tests. All tests were applied under the supervision of an experienced psychologist (++ for methods); however, a limitation of the study noted by the authors was only using parent reports and not teacher reports as they can vary from one another. Blinding was not reported, but it is unlikely that the mothers were aware of their urinary fluoride levels. Although mothers may have had knowledge that they were receiving fluoride through fluoridated salt or naturally occurring fluoride in their water, they would not have knowledge that this was relevant to the study purpose as the ADHD tests were conducted for the original cohort (as was acknowledged by the study authors in the discussion). (++ for blinding). Overall rating = ++.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:*
        - *Statistical analyses:* Bivariate analyses included Chi-square tests for categorical variables and ANOVA for continuous outcomes. Appropriate univariate statistics and transformations were performed before bivariate analyses. Residuals from fully adjusted linear regressions were checked and suggested skewness. Gamma regression with an identity link was used to examine the adjusted association between prenatal fluoride and each neurobehavioral outcome (instead of using log transformation). Generalized additive models were used to visually examine potential non-linearity. Sensitivity analyses examined impact of other potential confounders. Diagnostics were used to assess violations of the model assumptions and to identify remaining influential observations. The Benjamini–Hochberg false discovery rate (FDR) procedure was used to correct for multiple testing.
        - *Other potential concerns:* None identified.

  - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, outcome blindly assessed, and the prospective cohort study design.

Choi *et al.* (2015)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** First grade children (ages 6–8 years)
- ***Study area:*** Mianning County in southern Sichuan, China
- ***Sample size:*** 51 first grade children
- ***Data relevant to the review:*** Associations between learning, memory, visual motor ability, motor ability, and manual dexterity with continuous urine or drinking water fluoride levels. Study also had information based on dental fluorosis score.
- ***Reported association with fluoride exposure:*** No: None of the outcomes were significantly associated with fluoride exposure.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were selected during the same time frame using the same methods. Fifty-one first-grade children residing in Mianning County in southern Sichuan, China were included in this pilot study. It is not specified if the 51 children represented all the first-grade children from this area or if some refused to participate. Children who did not speak Chinese, were not students at the Primary School of Sunshui Village in Mianning County, or those with chronic or acute disease that might affect neurobehavioral function tests were excluded. Demographic characteristics are presented in Table 1 of the study, which indicates that subjects were similar. Potential confounders are adjusted for in the statistical analyses.
  - *Basis for Rating:* Definitely low risk of bias based on direct evidence that the exposure groups were similar and were recruited within the same time frame using the same methods with no evidence of differences in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The parents or guardians completed a questionnaire on demographic and personal characteristics of the children (sex, age at testing, parity, illnesses before age 3, and past medical history) and caretakers (age, parity, education and occupational histories, residential history, and household income). A 20-µL capillary blood sample was collected at the school by a Mianning County Center for Disease Control (CDC) health practitioner and tested for possible iron deficiency which could be used as a

covariate of neurodevelopmental performance. Confounders that were not assessed include: maternal BMI, parental mental health, maternal smoking status, maternal reproductive factors, parental IQ, and HOME score. However, the study authors noted that confounding bias appeared to be limited due to the minimal diversity in the social characteristics of the subjects. The study authors indicated that CDC records documented that levels of other contaminants including arsenic and lead were very low in the area. Iodine differences were not specifically addressed, but there is no indication from the information provided that this might be a concern.

- *Potentially important study-specific confounders:* All key confounders were considered in this study.
  - *Direction/magnitude of effect*: Not applicable.
- *Basis for rating:* Probably low risk of bias because there is direct evidence that the key confounders are taken into account and indirect evidence that co-exposure to arsenic is likely not an issue in this area and that methods used for collecting the information were valid and reliable.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The majority of results were reported for the 51 children stated to be included in the pilot study. In Table 5 of the study, the N for each dental fluorosis category only totals 43, but the text indicates 8 children did not have a Dean Index because permanent teeth had not erupted.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study used three different measurements of fluoride exposure: well water fluoride concentrations from the residence during pregnancy and onwards, fluoride concentrations from children's first morning urine samples, and degree of children's dental fluorosis. Fluoride concentrations in community well water were measured and recorded by Mianning County CDC; specific methods were not reported, but they likely used standard methods as they were conducted by the CDC and were likely the same as those used to measure the fluoride in urine. Migration of subjects was noted to be limited. Well water fluoride concentrations of the mother's residence during pregnancy and onward were used to characterize a child's lifetime exposure. To provide a measure of the accumulated body burden, each child was given a 330-mL (11.2-oz) bottle of Robust© distilled water (free from fluoride and other contaminants) to drink the night before the clinical examinations, after emptying the bladder and before bedtime. The first urine sample the following morning was collected at home, and the fluoride concentration was determined on a 5-mL sample using an ion-specific electrode at the Mianning CDC. There is no indication that urinary fluoride levels accounted for dilution nor was it clear that the method of administering water to the children and collection methods sufficiently controlled for differences in dilution. One of the investigators, a dentist, performed a blinded dental examination on each child's permanent teeth to rate the degree of dental fluorosis using the Dean Index. The Dean

Index is a commonly used index in epidemiological studies and remains the gold standard in the dentistry armamentarium. The Index has the following classifications: normal, questionable, very mild, mild, moderate, and severe. Quality control (QC) procedures are not reported but were likely appropriate.

- *Direction/magnitude of effect*: Current levels were used to assess lifetime exposure. This is likely to be a non-differential exposure misclassification and direction of bias is unknown. Because subject migration appears to be limited, it is likely that the current fluoride levels are adequate reflections of past exposure. Dental fluorosis would be an indicator that exposure occurred in the past and there was a fair correlation between degree of dental fluorosis and current urine and water fluoride levels, with both increasing with increasing levels of dental fluorosis.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study authors adopted culture-independent tests considered feasible for children aged 6 to 8 years. The Wide Range Assessment of Memory and Learning (WRAML) was used for the assessment of memory and learning. Three subtests were also used. The Finger Windows subtest assesses sequential visual memory. The Design Memory subtest assesses the ability to reproduce designs from memory following a brief exposure. The Visual Learning subtest assesses the ability to learn the locations of pictured objects over repeated exposures. The Wechsler Intelligence Scale for Children-Revised (WISC-IV) included digit span for auditory span and working memory and block design for visual organization and reasoning. The grooved pegboard test assesses manual dexterity. The tests used have been validated on a western population. Although there is no information provided to indicate that they were validated on the study population, the study authors indicated that the tests were culture-independent (+ for methods). Blinding of the outcome assessors or steps to minimize potential bias was not reported. However, it is unlikely that the assessors had knowledge of the individual exposure as children all came from the same area, and water and urine levels were tested at the CDC. (+ for blinding). Overall = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that all outcomes were assessed blindly using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*

        - *Statistical analyses:* Statistical analyses were appropriate. Multiple regression models evaluate the associations between exposure indicators and test scores after adjusting for potential confounders. Specific regression models are not described or refenced, just stated to be "standard regression analysis with confounder adjustment." The distributions of fluoride concentrations in urine and water were skewed and were log10-transformed to approximate a Gaussian distribution (test not specified). Results are reported as adjusted effects and 95% CIs. There is no evidence that residual diagnostics were used to examine model assumptions; however, the impact on the effect estimates is expected to be minimal.
        - *Other potential concerns:* It should be noted that this study was a pilot study and, therefore, had a relatively small sample size (i.e., 51 children).
    - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in the confounding, exposure, and outcome risk-of-bias domains. Study strengths include individual fluoride measurements with blinding at outcome assessment likely. All key confounders and many other confounders were taken into account in the study design or analysis.

Li *et al.* (2004) [translated in Li *et al.* 2008a]

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Full term, normal neonates 24−72 hours old from healthy mothers
- ***Study area:*** Zhaozhou County, Heilongiang Province, China
- ***Sample size:*** 91 neonates (46 males and 45 females)
- ***Data relevant to the review:*** Comparison of neurobehavioral capacity between children in the high-fluoride area compared to the control area.
- ***Reported association with fluoride exposure:*** Yes: Significant differences in neurobehavioral assessment total scores between high-fluoride (36.48 ± 1.09) and control (38.28 ± 1.10) groups; significant differences in total neurobehavioral capacity scores as measured by non-biological visual orientation reaction and biological visual and auditory orientation reaction between the two groups (11.34 ± 0.56 in controls compared to 10.05 ± 0.94 in high-fluoride group).

Risk of Bias:

- **Author contacts:**
    - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* There is indirect evidence that the exposure groups were similar. They were recruited during the same time frame using the same methods. From 2002 to 2003, 273 neonates were born in a hospital in Zhaozhou County, China. Ninety-one of 273 full-term neonates (46 males, 45 females) were randomly selected. Mothers ranged in age

from 20 to 31 years, met multiple health criteria, and had not changed residence during pregnancy. Authors report that the two study groups are located in the same area with similar climate, living habits, economic and nutritional conditions, and cultural backgrounds, but do not provide these data in the manuscript. There is no statistically significant difference in the mode of delivery, birth weight, infant length, or sex. Subjects were separated into exposure groups after random selection.

- *Basis for Rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited within the same time frame using the same methods with no evidence of differences in participation/response rates.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* No confounders were specifically controlled in the analysis. The study authors note similarities in characteristics in the two populations (i.e., living habits, economic and nutritional conditions, and cultural backgrounds), but do not provide these data nor do they indicate what specific characteristics were considered. There were no significant differences in infant gender, birth method, gestational age, or infant weight and length. All tests were conducted when children were 1–3 days old. No potential co-exposures were discussed. Although arsenic is considered a potential issue in China, water quality maps indicate that there is a 25–50% probability that the drinking water in that area exceeds the WHO guideline for arsenic of 10 µg/L.
  - *Potentially important study-specific confounders:* Key confounders, including child's age, child's gender, and measures of socioeconomic status (SES), were similar between exposure groups; however, arsenic was not taken into account. Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on water quality maps, arsenic does not appear to be an issue in Zhaozhou County of the Heilongjiang Province. Iodine deficiencies are not mentioned.
    - *Direction/magnitude of effect***:** Conceptually, the presence of arsenic would potentially bias away from the null if it were present with fluoride. Deficiencies in iodine would potentially bias away from the null if it were present in areas of higher fluoride, but toward the null if it were present in areas of lower fluoride. Neither of these are considered a concern in this study for reasons detailed above.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the key confounders are taken into account, co-exposure to arsenic is likely not an issue in this area, and methods used for collecting the information are valid and reliable.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Although authors did not discuss why they only randomly selected 91 of the 273 neonates available, results were available for all 91 subjects.
  - *Basis for rating:* Definitely low risk of bias based on results being available for all subjects.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were split into control and high-fluoride groups based on fluoride levels in their places of residence. Although the levels were provided (1.7–6.0 mg/L for

the high-fluoride group compared to 0.5–1.0 mg/L for the control group), it was not reported how or when these levels were measured. Urine was collected when women were hospitalized, but before labor began. Urine samples were sent to a specific lab for measurement using fluoride ion-selective electrode. It was noted that this procedure strictly followed the internal controls of the laboratory indicating quality control. Level of detection (LOD) was not provided. Urinary fluoride levels were significantly higher in the high-fluoride mothers (3.58 ±1.47 mg/L) compared to the control-group mothers (1.74±0.96 mg/L). There was indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure. Although results were mainly based on exposure area, they were supported by urine data making exposure misclassification less of a concern.

        - *Direction/magnitude of effect*: There is high variability in both water fluoride and urine fluoride in the subjects from the high-exposure area. Although there is no overlap in the water fluoride levels in the exposure areas, there is some overlap in the urine concentrations in the mothers from the two areas. This may reflect the single measurement and pose no specific bias, or it could indicate that some mothers in the high-fluoride area have lower fluoride exposure, which could bias the results toward the null.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.
- ***Outcome:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* A standard neonatal behavioral neurological assessment method was carried out by professionals in the pediatric department working in neonatal section trained specifically for these programs and passing the training exams. (+ for methods). The examinations were carried out 1 to 3 days after delivery. Because urine samples were collected on the day of delivery and sent to a separate laboratory, it is likely that the outcome assessors were blind. Although the subjects were separated by fluoride exposure area, it is not likely that the professionals were aware of the exposure as the tests were conducted in the hospital (+ for blinding).
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed blindly using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* The study authors reported numerous endpoints in sufficient detail; however, because they did not provide a list of endpoints tested there is no direct evidence that all were reported.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that all the study's measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:*
        - *Statistical analyses:* Statistical analyses are described only as a t-test. Consideration of heterogeneity of variance was not reported. Results are

reported as mean and standard deviations of neurological scores. Maternal urinary fluoride levels were only used to compare exposures between exposed and control groups. Infants in the control group were from four villages, and those in the exposed group were from five villages within the same district. Infants were randomly selected before they were assigned to exposed or control groups. In the comparisons, there was no accounting for clustering at the village level. It is likely that the standard error of the difference in mean neurobehavioral assessment scores between the high fluoride group and control group will be biased, making differences appear stronger than they actually are. However, the use of multiple villages per exposure group is likely to mitigate some of the impact of this lack of accounting for clustering, and the overall impact on effect estimates is expected to be minimal.

        - *Other potential concerns:* It should be noted that, although the study states that subjects were randomly selected, it is unclear why only 91 subjects were included and if they were randomly selected to obtain equal groups in the high-fluoride and control groups.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that statistical analyses were appropriate and that there were no other potential threats to risk of bias.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in the confounding, exposure, and outcome risk-of-bias domains. Study strengths include individual fluoride measurements to support the differences in the two areas. Tests were noted to be conducted at the hospital providing indirect evidence that blinding was not a concern during the outcome evaluation. Although there was some potential for bias due to the lack of accounting for arsenic or iodine deficiencies, co-exposure to arsenic is likely not a major concern according to groundwater quality maps.

Riddell *et al.* (2019)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Canadian Health Measures Survey (cycles 2 and 3) participants (children aged 6−17 years)
- ***Study area:*** general population, Canada
- ***Sample size:*** 3,745 children
- ***Data relevant to the review:*** Adjusted odds ratios for ADHD and attention symptoms per 1 unit increase in urinary fluoride, by water fluoride in the tap water, or community fluoridation status.
- ***Reported association with fluoride exposure:*** Yes: Significantly increased odds of ADHD diagnosis (adjusted OR = 6.10; 95% CI: 1.60, 22.8) or hyperactivity/inattentive symptoms (adjusted β = 0.31; 95% CI: 0.04, 0.58) per 1-mg/L increase in tap water fluoride. Also, a significant association between ADHD diagnosis (adjusted OR = 1.21; 95% CI: 1.03, 1.42) or hyperactivity/inattentive symptoms (adjusted β = 0.11; 95% CI: 0.02, 0.58) and community water fluoridation status. No significant associations with urinary fluoride levels.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were part of Cycles 2 and 3 of the Canadian Health Measures Survey. This is a nationally representative sample of residents living in 10 provinces. Specific inclusion criteria were provided. This study was restricted to children 6–17 years of age with different fluoride measurements that consisted of three participant samples. One of the samples was only available in Cycle 3.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were similar and were recruited with the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariates included in all models included child's age at test, child's sex, ethnicity, BMI, parents' education, total household income, exposure to cigarette smoke inside the home, and log-transformed concurrent blood lead levels. Confounders such as parental behavioral and mental health disorders, quantity and quality of caregiving environment, and co-exposure to arsenic were not discussed. The study used data from the Canadian Health Measures Survey which consists of a nationally representative sample of Canadians. Most Canadians (~89%) receive water from municipal water supplies, which monitor for levels of arsenic. Therefore, co-exposure to arsenic is not likely an issue in this population. Rationale for selection of covariates was based on relationship to ADHD diagnosis and to fluoride metabolism based on literature review and consultation with an ADHD expert. There is no information of the source if data for covariates, but this is likely the questionnaires from the Canadian Health Measures Survey, which are considered standardized and validated.
  - *Potentially important study-specific confounders:* All key confounders were considered in this study.
    - *Direction/magnitude of effect***:** Not applicable.
  - *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key confounders are taken into account, co-exposure to arsenic is likely not an issue in this area, and methods used for collecting the information were valid and reliable.
- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There is no information indicating that there were any data excluded due to missing covariates. All exclusions of children were described and reasonable (i.e., drinking bottled water when considered city fluoridation as a measure of fluoride exposure). Outliers were stated to be excluded, but methods for determining this were provided and it was noted that the outliers were 0.27% of the values.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* **Urinary Fluoride:** Spot urine samples were collected under normal non-fasting conditions and analyzed using an Orion pH meter with a fluoride ion-selective electrode after being diluted with an ionic adjustment buffer. Analysis was performed at the Human Toxicology Laboratory of the Institut National de Sante Publique du Quebec. The precision and accuracy of the fluoride analyses, including quality control and quality assurance, were described by Health Canada (2015). The limits of detection were 20 µg/L for Cycle 2 and 10 µg/L for Cycle 3 with no values below detection. Fluoride levels were adjusted for specific gravity.
    Water Fluoride in Tap water: Tap water was collected at the subjects' homes in Cycle 3 only. Samples were analyzed for fluoride concentrations using anion exchange chromatography procedure with a LOD of 0.006 mg/L. Values below the LOD were imputed with LOD/square root 2. Of the 980 samples, 150 (16%) were below detection.
    Chlorinate Water Fluoride status: This was determined by viewing reports on each city's website or contacting the water treatment plant (provided in supplemental material). Children were excluded if they drank bottled water, had a well, had a home filtration system, lived in the current residence for 2 years or less, or lived in an area with mixed city fluoridation.
    - *Direction/magnitude of effect***:** There is not any specific direction or magnitude of bias expected. Urinary fluoride levels are reflective of a recent exposure, but the study authors adjusted to account for dilution. The possibility of exposure misclassification would be similar in all subjects and would be non-differential. There is less potential for exposure misclassification in regard to tap water or chlorinated water fluoride status as children who drank bottled water were excluded and children who had a home filtration system were excluded from the chlorinated water status.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:*
    Strengths and Difficulties Questionnaire (SDQ): The questionnaire was administered to youths under 18 years. Children aged 6–11 years had SDQ ratings provided by parents and guardians, but youths aged 12−17 years completed the questionnaire themselves. Tests consist of 25 items with a 3-point scale. Items were divided into five subscales: emotional problems, conduct problems, hyperactivity-inattention, peer problems, and prosocial behavior. The current study only used the hyperactivity-inattention subscale. Validation of this method was not reported (- for methods).
    ADHD: Ninety percent of youths with ADHD are diagnosed after age 6 years. For children aged 6–11 years, ADHD diagnosis was provided by parents, but youths aged 12−17 years completed the questionnaire themselves. Cycle 2 asked "Do you have a learning disability?" and if yes asked to specify the type (4 options available and described). In Cycle 3, parents were asked directly whether they had ADHD, and children 12 years and older were asked if they had a physician diagnosis of ADHD and, if so, what subtype. (- for methods because different methods were used and only the children 12 years and older in cycle 3 were asked specifically about doctor diagnosis). Both were

measured in both cycles. Blinding is not likely an issue as subjects would not have knowledge of the urine or tap water fluoride levels. However, they would likely have knowledge of the city.
    - *Basis for rating:* Probably high risk of bias based on indirect evidence that the outcome was assessed using insensitive methods that varied based subject age.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods sections were reported in sufficient details.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:*
        - *Statistical analyses:* Robust logistic regression was used to examine the association between fluoride exposure and ADHD diagnosis, adjusting for covariates. Box-Tidewell tests were used to check the linearity of the relationship with the continuous predictors. Linear regression was used for the SDQ scores using Huber-White standard errors. Multicollinearity was evaluated using variance inflation factor (VIF) statistics. Outliers with high studentized residuals, high leverage, or large Cook's distance values were removed from all analyses with urinary fluoride. All regressions were tested for interactions between age and fluoride, and sex and fluoride. Sensitivity analyses were conducted to test the different survey cycles. There is no mention of adjustment for the complex survey design using survey weights or bootstrapped weights to ensure appropriate calculation of the estimated variances; however, the overall impact on effect estimates is expected to be minimal.
        - *Other potential concerns:* None identified.
    - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include individual exposure measurements and the addressing of potential key confounders but was limited by the cross-sectional study design and insensitive outcome measures.

Rocha-Amador *et al.* (2009)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 6–11 years
- ***Study area:*** Durango, Mexico
- ***Sample size:*** 80 children

- ***Data relevant to the review:*** Associations between visuospatial organization and visual memory (using the Rey-Osterrieth Complex Figure Test, children's version) and urinary fluoride levels in the children.
- ***Reported association with fluoride exposure:*** Yes: Significant correlation between urinary fluoride and visuospatial organization (r = −0.29) and visual memory (r = −0.27) scores. No significant correlations with arsenic.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were from the same population and were recruited during the same time frame using the same methods. Although this study compared three sites with antecedents of environmental pollution to mixtures of either F–As, Pb–As, or DDT–PCBs, authors evaluated each contaminant separately. The only area of interest is the area with F and As contamination. The area in Durango state (5 de Febrero) where drinking water is polluted naturally with F and As at levels exceeding 6 and 19 times, respectively, the World Health Organization (WHO) limits (WHO 2008). Children attending public schools were screened through personal interviews for study eligibility. Inclusion criteria were children between 6 and 11 years old, living in the study area since birth, and whose parents signed the agreement to participate. Children with a neurological disease diagnosed by a physician and reported by the mother were excluded from the study. The final sample for the F–As was 80. Participation rates were not reported. Selected demographic characteristics are presented in Table 1 of the study.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the populations were similar and recruited during the same time frame using the same methods.
- ***Confounding:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* Confounding factors in children tested in the analysis included blood lead (PbB), age, gender, and height-for-age z-scores; only age had significant associations and was included in the final analysis. Arsenic was also assessed and analyzed separately from fluoride. Arsenic in urine was analyzed by atomic absorption spectrophotometer coupled to a hydride system (Perkin-Elmer model AAnalyst 100). Although the model did not adjust for arsenic, arsenic in the F–As group was not associated with either endpoint; therefore, arsenic as a co-exposure is not considered a major concern in this study. PbB was analyzed with a Perkin-Elmer 3110 atomic absorption spectrophotometer using a graphite furnace. Authors note that the mean blood lead level in the F–As study area was 5.2 µg/dL and 8% of the children had values above the reference value of 10 µg/dL. PbB was stated not to affect results and was not included in the final analysis. Other confounding data were obtained during the study interview. Father's education was provided and, in the F–As group, was stated to range from 0–16 years, but this was not considered. Maternal education, smoking, and SES were also not

considered. The authors provide an SES score of 5.9 ± 1.4 for the 5 de Febrero region (the fluoride region). It is not clear if this would vary by fluoride or arsenic levels.

- *Potentially important study-specific confounders:* SES.
  - *Direction/magnitude of effect*: There are insufficient data to determine the magnitude or direction of effect. If there is an association between fluoride exposure and SES, the direction of effect would depend on the association.
- *Basis for rating:* Probably high risk of bias based on indirect evidence that the SES was not accounted for in the study design or analysis and may have varied by fluoride levels.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are complete. All 80 participants stated to be the final sample for the site of interest (F–As) were included in all analyses.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Fluoride in urine (FU) was analyzed according to method 8308 (''fluoride in urine'') from the National Institute of Occupational Safety and Health (NIOSH 1984) with a sensitive specific ion electrode. As a quality control check, reference standard ''fluoride in freeze dried urine'' (NIST SRM 2671a) was analyzed. The accuracy was 97.0 +/- 6.0%. Levels of FU and AsU were adjusted for urinary creatinine, which was analyzed by a colorimetric method (Bayer Diagnostic Kit, Sera-Pak1 Plus). However, details on the collection methods were not reported.
    - *Direction/magnitude of effect*: Spot urine samples in a small sample size (i.e., 80 children) may have some exposure misclassification. Adjusting for dilution reduces the potential for misclassification based on differences in dilution. Exposure misclassification would be non-differential.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* IQ is assessed through the Rey-Osterrieth Complex Figure Test (ROCF). This is a less well-established method, although the authors provide citations suggesting it has been validated and standardized for the Mexican population (+ for methods). According to the study report, the neuropsychologist who administered the test was blinded to all exposure types and levels. (++ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.

- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used log-transformed exposure variables (although rationale was not provided). Crude and partial correlations were calculated to evaluate associations between serum fluoride levels and TOCF scores. There is no other description of the regression model, and regression diagnostics to evaluate model assumptions are not presented; however, the overall impact on effect estimates is expected to be minimal.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements and outcomes blindly assessed, but it is limited by the cross-sectional study design, lack of addressing SES in the study population, co-exposure with arsenic, and use of spot samples in a small population.

Valdez Jimenez *et al.* (2017)

Study Details:

- ***Study design*:** Prospective cohort
- ***Population*:** Infants aged 3–15 months
- ***Study area:*** Durango City and Lagos de Moreno, Jalisco, Mexico
- ***Sample size:*** 65 infants
- ***Data relevant to the review:*** The Bayley Scales of Infant Development II was used to assess Mental Development Index Scale and the Psychomotor Development Index scale in children 3 to 15 month and evaluated for associations with first and second trimester maternal urine fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant association between maternal urinary fluoride and MDI score during first trimester (adjusted β = −19.05; SE = 8.9) and second trimester (adjusted β = −19.34; SE = 7.46). No association between maternal fluoride during any trimester and Psychomotor Developmental Index (PDI).

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were recruited from two endemic areas in Mexico. The study authors do not provide information on the similarities or differences between the two areas nor do they indicate if there were different participation rates. However, recruitment methods were the same. Women receiving prenatal care in health centers located in Durango City and Lagos de Moreno, Jalisco, Mexico were recruited in 2013–2014. Participation rates are not likely to be an issue as characteristics were similar

between those who participated and those who did not. Although they did not provide characteristics by area, the characteristics provided do not indicate any differences that may be biased by the selection. Considering the age range for the non-participants, the mean age for non-participants appears to be incorrect (or the age range is incorrect); however, there does not appear to be a difference that would potentially indicate selection bias.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited with the same methods in the same time frame, with no evidence of differences or issues with participation/response rates.

- ***Confounding:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* Questionnaires were used to obtain information about sociodemographic factors, prenatal history, mother's health status before pregnancy (e.g., use of drugs, vaccines, diseases) and the type of water for drinking and cooking. The marginalization index (MI) was obtained from the National Population Council (CONAPO). Two additional surveys were conducted during the 2nd and 3rd trimester of pregnancy to get information about the mother's health, pregnancy evolution, and sources of water consumption. A survey was also conducted to get information about childbirth (type of birth, week of birth, weight and length of the baby at birth, Apgar and health conditions of the baby during the first month of life). This information was corroborated with the birth certificate. Linear regression models included gestational age, children's age, marginality index, and type of drinking water. Bivariate analysis was conducted on the other factors including child's gender prior to conducting multivariable regression models. Some important confounders were not considered, including parental mental health, IQ, smoking, and potential co-exposures. Water quality maps indicate a potential for arsenic to be present in the study area.
  - *Potentially important study-specific confounders:* Arsenic is a potential co-exposure in this area of Mexico.
    - *Direction/magnitude of effect***:** If arsenic were present as a co-exposure it would bias the results away from the null.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that there is a potential for co-exposure with arsenic that was not addressed.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Out of the 90 women selected for inclusion in the study, 65 approved the participation of their infants. The authors provide a table of characteristics between women who consented to their children's cognitive evaluation and those that only participated in biological monitoring. There were no significant differences between the groups. There were fewer women who provided urine during the second and third trimesters. All specified children are included in the relevant analyses.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)

- *Summary:* Fluoride exposure is assessed through morning urine samples and water fluoride levels collected from the children's homes. Sampling methodology is appropriately documented, and water levels were quantified through specific ion-sensitive electrode assays. QC was described and accuracy was >90%. Urinary fluoride was corrected by specific gravity.
  - *Direction/magnitude of effect***:** Not applicable.
- *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Neurodevelopment was assessed with the Bayley Scales of Infant Development II (BSDI-II) that was noted to be reliable and valid for evaluating children from 3 months to 5 years of age. The average age of children assessed was 8 months, with a range of 3–15 months) (++ for methods). The study report stated that a trained psychologist who was blinded about the mother's fluoride exposure evaluated the infants at home (++ for blinding). Overall rating for methods and blinding = ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported. Table 4 of the study only displays data for trimesters 1 and 2. Although 3rd trimester data were collected, they were not reported, likely because data were only available for 29 subjects. No discussion of this was provided.
  - *Basis for rating:* Probably low risk of bias because, although it appears some data were not reported, it is likely because there were insufficient data and not because the authors were selectively reporting the results.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used log10-transformed exposure variables. Normality, homoscedasticity, and linearity assumptions were tested and satisfied for MDI and PDI scores. Bivariate analyses included correlations, t-tests, and ANOVA. Multiple linear regression models by the 1st and 2nd trimester of pregnancy were used to evaluate the association between maternal fluoride exposure and MDI and PDI scores. The best-fit model was selected using a "stepwise method" and the best-fit line was evaluated using "the curve fitting method." It is not further specified or cited what these methods entailed. Best-fit or goodness-of-fit statistics are not reported. It is unclear how a best-fit model could be selected when the authors state that all models adjusted for the same set of covariates regardless of significance, and these covariates also appear in the final model—presumably the best-fit model. It is unlikely that a stepwise method would retain all those covariates unless they were forced in the model. Residual analysis was conducted to assess model validity; however,

there is no description of the results of the residual analysis. Nonetheless, the impact on effect estimates is expected to be minimal.

- *Other potential concerns:* No other potential concerns were identified. In the peer-review report, NASEM (2020) cited the following as potential concerns: "the large difference in numbers of males and females in the offspring (20 males, 45 females), and apparently incorrect probabilities were reported for age differences between participants and nonparticipants, high rates of cesarean deliveries and premature births among participants (degree of overlap not reported), and incorrect comparisons of observed prematurity rates with national expected rates." However, these concerns were taken in consideration in other domains (***Selection, Confounding)***.

Basis for rating: Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.

- ***Basis for classification as low risk-of-bias study overall:*** Definitely low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements and outcome blindly assessed, but it is limited by the cross-sectional study design and lack of accounting for potential co-exposures to arsenic.

Wang *et al.* (2020a)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** School children aged 7–13 years
- ***Study area:*** Tongxu County, China
- ***Sample size:*** 325 school children
- ***Data relevant to the review:*** Associations between ADHD and other measures of learning disability with urine fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant association between psychosomatic problems and urinary fluoride (per 1-mg/L increase; adjusted β = 4.01 [95% CI: 2.74, 5.28]) and increased risk of a T-score > 70 with urinary fluoride (per 1-mg/L increase; adjusted OR = 1.97 [95% CI: 1.19, 3.27]). No significant associations with ADHD or other measures of learning disability.

Risk of Bias:

- **Author contacts:**
  - Authors were contacted in July of 2020 to obtain additional information for risk-of-bias evaluation. No response was received.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were recruited in 2017 from Tongxu County, China. Children were selected from four randomly selected primary schools in the area. Selection was based on specified inclusion rules. It was noted that the living habits and diets of the participants from the four schools were well matched, but details were not provided. The area did not have industrial pollution within 1 km of the living environment of the children, and it was noted that the children were not exposed to other

neurodevelopmental toxicants (lead, cadmium, arsenic, or mercury). A table of subject characteristics was provided in the study, but not by school or exposure. This is a pilot study, and it is not explicitly stated if all eligible subjects participated in the study. There is no information on participation rates or if they varied by school.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposed groups were recruited using similar methods during the same time frame and that any differences between the exposed groups were considered in the statistical analyses.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* It was noted that subjects were well matched in terms of living habits and diets, but there were no specifics provided. It was noted that there was no industrial exposure or exposure to other neurotoxins such as lead, cadmium, arsenic, or mercury. Covariates were collected using a standardized and structured questionnaire completed by the children and their guardians under the direction of investigators, but reliability or validity of the questionnaire was not reported. Information collected included age, gender, weight, height, parental education level, and parental migration (or work as migrant workers). IQ scores evaluated by the Combined Raven's Test-the Rural in China were used to represent basic cognitive function. Models were adjusted for age, BMI, gender, mother and father migration, and urinary creatinine. Adjustments were not made for parental education, race/ethnicity, maternal demographics (e.g., maternal age, BMI), parental behavioral and mental health disorders (e.g., ADHD, depression), smoking (e.g., maternal smoking status, secondhand tobacco smoke exposure), reproductive factors (e.g., parity), iodine deficiency/excess, maternal (and paternal) IQ, quantity and quality of caregiving environment (e.g., HOME score), or SES other than parental migration. There is no evidence to suggest that SES would differ substantially among the four rural schools in the same area of China that were randomly selected.
  - *Potentially important study-specific confounders:* SES.
    - *Direction/magnitude of effect***:** Direction and magnitude is unknown. It was noted that the subjects were matched in terms of living habits and diet and this could be an indication that SES was not different among the groups, but details were not provided.
  - *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key confounders are considered, that the methods for collecting the information were valid and reliable, and that co-exposure to arsenic is not an issue in this area.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are complete. It was noted that there were 325 subjects included and results were available on all subjects.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Spot urine samples were collected from each child in the early morning into cleaned polyethylene tubes. Fluoride concentrations were measured using fluoride ion-selective electrode (with reference to Ma *et al.* (2017); however, that reference cites

Zhou *et al.* (2012)). Therefore, no QC methods or LODs were available. Fluoride concentrations were creatinine-adjusted.

- *Direction/magnitude of effect*: Spot urine samples only account for recent exposure. Although this could cause there to be some exposure misclassification, the number of subjects should help dilute any issues with the non-differential misclassification.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.

- ***Outcome:***
  - *Rating:* Probably high risk of bias (NR)
  - *Summary:* Children's behavior was assessed by the Chinese version of the Conners' Parent Rating Scale-Revised (CPRS-48). The homogeneity reliability of Cronbach α in the Chinese version of CPRS-48 was 0.932; the correlation of Spearman-brown split-half was 0.900; and the retest reliability of total score was 0.594. Raw scores for each subscale are converted into sex- and age-adjusted T-scores within a mean ± standard deviation (SD) of 50 ± 10. The guardians independently completed the CPRS-48 according to the instruction manual under the direction of trained investigators (++ for methods). Blinding is not reported. Although it is unlikely that the outcome assessors were aware of the fluoride levels in the urine, it is unclear if subjects were selected based on areas with endemic fluoride or if parents were aware of fluoride concentrations in the areas. (NR for blinding). Overall rating for methods and blinding = NR.
  - *Basis for rating:* Probably high risk of bias based on no information provided to indicate that the outcome was blindly assessed.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Multiple linear regression models were used to assess the association between urinary fluoride exposure and each behavioral outcome. Logistic regression was used to assess the risk of behavioral problems (T-scores > 70) due to fluoride exposure. Sensitivity analyses were performed, with models adjusting for combinations of age, BMI, gender, mother migrated, father migrated, and urinary creatinine levels. Regression diagnostics to evaluate model assumptions are not described; however, the overall impact on effect estimates is expected to be minimal.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and no other potential threats to risk of bias were identified.

- ***Basis for classification as low risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include individual exposure measurements, but it is limited by the cross-sectional study design and lack of details on blinding of the outcome assessment. All key confounders were considered in the study design or analysis.

Appendix 5. Mechanistic Data from Animal Studies

A number of animal studies were available that presented mechanistic data in several effect categories (see **Figure A5-1**). Limiting the data to studies with at least one exposure at or below 20 ppm fluoride drinking water equivalents (gavage and dietary exposures were back-calculated into equivalent drinking water concentrations for comparison) still provided a sufficient number of studies for evaluation of several mechanistic endpoints while allowing for a more focused look at exposure levels most relevant to human exposures. The following sections summarize the mechanistic data by the effect category. Although there is some evidence of consistency in mechanistic effects, overall, these data are insufficient to increase confidence in the assessment of findings from human epidemiology studies.

Figure A5-1. Number of Animal Mechanistic Studies for Fluoride by Mechanistic Category and Exposure Level*

	Dose Level	
Mechanism	**All**	**≤20 ppm**
Biochemical (brain/neurons)	66	25
Neurotransmitters	53	23
Oxidative stress	78	25
Histopathology	67	30
Apoptosis/cell death	19	7
Inflammation	7	5
Thyroid	50	17
Other	3	2

*Interactive figure and additional study details in Tableau® (https://public.tableau.com/app/profile/ntp.visuals/viz/Animal_Mechanisms_2021/FigureA5-1). The number of studies that evaluated mechanistic effects associated with at least one exposure at or below 20 ppm fluoride is tabulated in the "≤20 ppm" column. The total number of studies per mechanistic category are summarized in the "All" column.

Neurotransmitters

Neurotransmitter and biochemical changes in the brain and neurons were considered to be the mechanistic areas with the greatest potential to demonstrate effects of fluoride on the brain of animals in the lower dose range and provide evidence of changes in the brain that may relate to lower IQ in children (see **Figure A5-2**). Twenty of 23 neurotransmitter studies assessed changes in brain cholinesterase activity associated with fluoride exposure at or below 20 ppm fluoride. Acetylcholine is a major neurotransmitter involved in learning, memory, and intelligence (Chen 2012, Gais and Schonauer 2017). AChE is responsible for the breakdown of acetylcholine in the synapses of nerve cells. Changes in cholinesterase, acetylcholine, or AChE could be related to effects on memory. Evidence of an effect varied among the low risk-of-bias studies that assessed changes in cholinesterase or acetylcholine (n = 11 drinking water studies) (Gao *et al.* 2009, Baba *et al.* 2014, Adedara *et al.* 2017a, Khan *et al.* 2017, Gao *et al.* 2008a, Akinrinade *et al.* 2015a, Sun *et al.* 2000 [translated in Sun *et al.* 2008], Chouhan *et al.* 2010, Mesram *et al.* 2016, Liu *et al.* 2010, Nkpaa and Onyeso 2018), with the majority of studies reporting evidence of an effect that is considered inconsistent with the phenotypic outcome. Decreases in cholinesterase will cause increases in acetylcholine, which can have a positive effect on learning and

memory; however, long-term decreases in cholinesterase can lead to secondary neuronal damage occurring in the cholinergic region of the brain (Chen 2012).

Five of the 11 studies with low risk of bias (Gao *et al.* 2009, Baba *et al.* 2014, Adedara *et al.* 2017a, Khan *et al.* 2017, Nkpaa and Onyeso 2018) found statistically significant decreases in cholinesterase or AChE in brain homogenates (with some brains dissected into specific regions prior to homogenizing) with fluoride concentrations in drinking water at or below 20 ppm, and 4 of the 5 studies found statistically significant decreases in cholinesterase or AChE below 10 ppm. The 5 studies were conducted in rats (Wistar or Sprague-Dawley) with exposure ranging from 28 days to 6 months. An additional 2 out of 11 studies (Gao *et al.* 2008a, Akinrinade *et al.* 2015a) reported decreases in brain homogenate AChE at concentrations at or below 20 ppm fluoride in drinking water, but statistical significance was not reached. These studies were also conducted in rats with exposure for 30 days or 3 months. Gao *et al.* (2008a) reported a dose-dependent decrease in brain homogenate AChE in the low (5 ppm fluoride) and high (50 ppm fluoride) treatment groups compared with the control group, but the decrease was only statistically significant in the high dose group. Similarly, Akinrinade *et al.* (2015a) observed a dose-dependent decrease in percent intensity of AChE immunohistochemistry in the prefrontal cortex associated with 2.1 and 10 ppm sodium fluoride in the drinking water, but neither result was statistically significant. Gao *et al.* (2009) found lower brain homogenate AChE levels in the 5-ppm animals compared with the 50-ppm animals; therefore, the results were not always dose dependent.

Relative to the above-mentioned studies, 2 of the 11 low risk-of-bias studies observed opposite effects on brain cholinesterase levels. Sun *et al.* (2000) [translated in Sun *et al.* 2008] observed a significant increase in brain cholinesterase in Kunming mice associated with fluoride drinking water concentrations from 10 to 100 mg/L, but did not observe a dose response. Chouhan *et al.* (2010) did observe a dose-related increase in AChE levels in brain homogenate of Wistar rats with sodium fluoride concentrations of 1 to 100 ppm for 12 weeks and noted statistically significant results at 1, 50, and 100 ppm but not at 10 ppm.

Mesram *et al.* (2016) did not assess changes in AChE but observed a significant decrease in acetylcholine levels in cerebral cortex homogenate through 30 days of age in rats treated in utero with 20 ppm sodium fluoride, which may suggest an increase in AChE levels. Likewise, Liu *et al.* (2010) did not assess changes in AChE, but measured nicotinic acetylcholine receptors (nAChRs) in brain homogenate of rats following drinking water fluoride exposure, which the authors stated could modulate physiological and pharmacological functions that are involved in learning and memory-related behaviors. Significant decreases in the protein expressions of nAChR subunits at 2.26 ppm fluoride were observed; however, the corresponding receptor subunit mRNAs did not exhibit any changes (Liu *et al.* 2010).

The studies that assessed other neurotransmitters of the brain and neurons were too heterogeneous or limited in number to make any determination on mechanism, even before limiting the review of the data to low risk-of-bias studies. There were only five studies that evaluated dopamine and/or metabolites (Tsunoda *et al.* 2005, Chouhan *et al.* 2010, Reddy *et al.* 2014, Banala *et al.* 2018, Sudhakar and Reddy 2018). Four of the studies observed decreases in dopamine levels in the brain with exposures less than 20 ppm fluoride (Reddy *et al.* 2014, Chouhan *et al.* 2010, Banala *et al.* 2018, Sudhakar and Reddy 2018); however, the fifth study (Tsunoda *et al.* 2005) observed increased dopamine and metabolites at fluoride exposures below 20 ppm (with statistical significance achieved only for the metabolite homovanillic acid in one brain region). No differences from the control group were observed at levels above 20 ppm fluoride. Other neurotransmitters were evaluated at or below 20 ppm fluoride exposure, but generally only in a couple of studies.

Biochemistry (Brain/Neurons)

Similar to above, the endpoints measured in brain biochemistry studies were too heterogeneous or limited in number to make any determination on potential relevance of mechanism, even before limiting the review of the data to low risk-of-bias studies (see **Figure A5-2**). Endpoints related to biochemical changes in the brain or neurons included carbohydrate or lipid changes, RNA or DNA changes, changes in gene expression, or changes in protein expression. For the most part, only a single study was available for any given endpoint. The largest body of evidence on biochemistry was on protein level in various brain regions. Eleven low risk-of-bias studies were identified that evaluated protein levels; however, few studies evaluated the same proteins or areas of the brain. In the few cases where the same protein was evaluated, results were not always consistent. These data are insufficient to increase confidence or support a change to hazard conclusions.

Histopathology

Histological data can be useful in determining whether effects are occurring in the brain at lower fluoride concentrations; however, author descriptions of these effects may be limited thereby making it difficult to directly link histological changes in the brain to learning and memory effects. Histopathology of the brain was evaluated in 31 studies with concentrations at or below 20 ppm fluoride, of which 15 studies were considered low risk-of-bias studies (Adedara *et al.* 2017b, Akinrinade *et al.* 2015a, Bhatnagar *et al.* 2002, Bhatnagar *et al.* 2011, Chouhan *et al.* 2010, Guner *et al.* 2016, Jiang *et al.* 2014, Lou *et al.* 2013, McPherson *et al.* 2018, Mesram *et al.* 2016, Niu *et al.* 2018, Pulungan *et al.* 2016, Nageshwar *et al.* 2018, Zhao *et al.* 2019, Jia *et al.* 2019). In all but one low risk-of-bias study [Pulungan *et al.* (2016); gavage], animals were exposed to fluoride via drinking water. All low risk-of-bias studies were conducted in rodents, and all but three studies were conducted in rats (Wistar [seven studies]; Sprague-Dawley [four studies]; Long-Evans hooded [one study]). Overall, the low risk-of-bias studies that evaluated histopathology in the brain had low potential for bias for key questions regarding randomization and exposure characterization; however, eight studies were rated as probably high risk of bias for the key risk-of-bias question regarding outcome assessment based on lack of reporting of blinding of outcome assessors and/or inadequate description of outcome measures or lesions. Moreover, low image quality in some of the studies hampered the ability to verify the quality of the data. Further technical review of the 15 low risk-of-bias studies was conducted by a board-certified pathologist. Based on confidence in the results for each study, the technical reviewer further categorized the low risk-of-bias studies as studies with higher or low confidence in the outcome assessment, which is reflected in the following summary of the brain histopathology results. Main limitations of the histopathology data identified by the pathologist included lack of information on methods of euthanasia and fixation. Perfusion fixation is generally considered the best practice for lesions of the central nervous system in addition to complete fixation of the brain prior to its removal from the skull (Garman *et al.* 2016). Four of the low risk-of-bias studies reported that they used this method (Bhatnagar *et al.* 2002, Bhatnagar *et al.* 2011, McPherson *et al.* 2018, Pulungan *et al.* 2016). Two of the low risk-of-bias studies handled the brains before fixation was complete, which can produce artifacts that can resemble dead neurons (Zhao *et al.* 2019, Nageshwar *et al.* 2018). Fixation and brain removal details were inadequately described in the remaining low risk-of-bias studies.

Although there was heterogeneity in the endpoints reported (e.g., cell size, shape, and counts; nuclei fragmentation; increased vacuolar spaces) and some variation in the consistency of the evidence based on the area of the brain evaluated, the majority of the low risk-of-bias studies (11 of 14 drinking water studies) found some histological change in the brain of rats or mice treated with fluoride at concentrations at or below 20 ppm, of which 8 studies reported histological changes in the brain at or below 10 ppm. Histological changes in the hippocampus (one of the areas of the brain most evaluated for histological changes) associated with fluoride exposures at or below 20 ppm were reported in three

of four low risk-of-bias studies with higher confidence in the outcome assessment (Bhatnagar *et al.* 2002, Bhatnagar *et al.* 2011, Guner *et al.* 2016) and in three of four low risk-of-bias studies with lower confidence in the outcome assessment (Jiang *et al.* 2014, Niu *et al.* 2018, Nageshwar *et al.* 2018). McPherson *et al.* (2018) was the only drinking water study (with higher confidence in the histopathology outcome assessment) that did not observe any histological changes in hippocampus at 10 or 20 ppm fluoride in male Long-Evans hooded rats exposed in utero through adulthood (>PND80). Although there are too few studies to definitively explain the inconsistency in results, McPherson *et al.* (2018) also did not observe any associations between fluoride exposure and impairments to learning and memory, which is inconsistent with the majority of developmental exposure studies that observed learning and impairments associated with fluoride exposure for other strains of rats. Similarly, histological changes in the cortex were reported in three of the four low risk-of-bias drinking water studies with higher confidence in the outcome assessment (Chouhan *et al.* 2010, Bhatnagar *et al.* 2011, Akinrinade *et al.* 2015a) and in three of four low risk-of-bias studies with lower confidence in the outcome assessment (Lou *et al.* 2013, Mesram *et al.* 2016, Nageshwar *et al.* 2018).

Histological changes were also consistently reported in other areas of the brain in studies with higher confidence in the outcome assessment, including the amygdala, caudate putamen, cerebellum, and hypothalamus, although each of these areas of the brain were only evaluated in one low risk-of-bias study (Bhatnagar *et al.* 2011, Guner *et al.* 2016). Pulungan *et al.* (2016), one of two low risk-of-bias studies with higher confidence in the outcome assessment that did not report histological changes in the brain, observed a decreasing trend in the number of pyramidal cells in the prefrontal cortex with increasing dose, but this was not changed at concentrations below 20 ppm (study administered sodium fluoride via gavage; the 5-mg/kg-day dose was considered to be equivalent to 15.3 ppm fluoride in drinking water) nor were any of the results statistically significant.

Oxidative Stress

Oxidative stress is considered a general mechanistic endpoint that cannot be specifically linked to neurodevelopmental or cognitive effects in humans; however, like histopathology, it may help in identifying changes in the brain occurring at lower concentrations of fluoride. Oxidative stress in the brain was evaluated in 25 studies that examined concentrations at or below 20 ppm fluoride, of which 15 studies had low potential for bias (Adedara *et al.* 2017a, Adedara *et al.* 2017b, Akinrinade *et al.* 2015b, Chouhan *et al.* 2010, Gao *et al.* 2008b, Guner *et al.* 2016, Mesram *et al.* 2016, Nkpaa and Onyeso 2018, Shan *et al.* 2004, Zhang *et al.* 2015a, Chouhan and Flora 2008, Gao *et al.* 2009, Khan *et al.* 2017, Bartos *et al.* 2018, Nageshwar *et al.* 2018). All of the low risk-of-bias studies were conducted in rats (mainly Wistar or Sprague-Dawley) and administered fluoride via drinking water with exposure durations ranging from 28 days to 7 months. Although there was heterogeneity in the endpoints reported (i.e., varying measures of protein oxidation, antioxidant activity, lipid peroxidation, and reactive oxygen species [ROS]) and some variation in the consistency of the evidence based on the endpoint, the majority of the studies (13 of 15 studies) (Adedara *et al.* 2017a, Adedara *et al.* 2017b, Akinrinade *et al.* 2015b, Gao *et al.* 2008b, Gao *et al.* 2009, Guner *et al.* 2016, Mesram *et al.* 2016, Nkpaa and Onyeso 2018, Shan *et al.* 2004, Zhang *et al.* 2015a, Khan *et al.* 2017, Nageshwar *et al.* 2018, Bartos *et al.* 2018) found evidence of oxidative stress in the brains of rats treated with fluoride at concentrations at or below 20 ppm, of which 10 studies reported oxidative stress in the brain below 10 ppm fluoride. The most consistent evidence of oxidative stress in the brain was reported through changes in antioxidant activity. Eleven of the 12 low risk-of-bias studies that evaluated antioxidant activity reported an effect at concentrations at or below 20 ppm (Adedara *et al.* 2017a, Adedara *et al.* 2017b, Akinrinade *et al.* 2015b, Gao *et al.* 2008b, Gao *et al.* 2009, Guner *et al.* 2016, Mesram *et al.* 2016, Nkpaa and Onyeso 2018, Khan *et al.* 2017, Bartos *et al.* 2018, Nageshwar *et al.* 2018). Decreases in antioxidant activity using measures of superoxide dismutase (SOD) activity were reported in seven of

eight low risk-of-bias studies (Adedara *et al.* 2017a, Adedara *et al.* 2017b, Akinrinade *et al.* 2015b, Mesram *et al.* 2016, Nkpaa and Onyeso 2018, Khan *et al.* 2017, Nageshwar *et al.* 2018) and, among these seven studies, all that also measured changes in catalase (CAT) activity (n = 6 studies) also reported decreased activity (Adedara *et al.* 2017a, Adedara *et al.* 2017b, Mesram *et al.* 2016, Nkpaa and Onyeso 2018, Khan *et al.* 2017, Nageshwar *et al.* 2018). A decrease in total antioxidant capacity (T-AOC) as a measure of antioxidant activity was also consistently reported in two low risk-of-bias studies (Gao *et al.* 2008b, Gao *et al.* 2009), and a decrease in glutathione peroxidase (GPx) activity was reported in two of three low risk-of-bias studies (Adedara *et al.* 2017b, Nkpaa and Onyeso 2018).

Relative to the above-mentioned studies, 2 of the 15 low risk-of-bias studies (Chouhan and Flora 2008, Chouhan *et al.* 2010) did not observe statistically significant effects on oxidative stress in the brain with concentrations at or below 20 ppm fluoride; however, the measure of oxidative stress evaluated in Chouhan and Flora (2008) and Chouhan *et al.* (2010) (glutathione [GSH] to oxidized glutathione [GSSG] ratio as an indication of antioxidant activity and ROS levels) were not evaluated in any other low risk-of-bias study. Chouhan and Flora (2008) observed a dose-dependent increase in ROS levels associated with 10, 50, and 100 mg/L sodium fluoride in the drinking water; however, results were not statistically significant at any dose. In Chouhan *et al.* (2010), the levels of ROS were significantly higher at 50 ppm sodium fluoride in drinking water, but statistical significance was not met at doses below 20 ppm fluoride (1 and 10 ppm sodium fluoride) or at 100 ppm sodium fluoride; yet, hydrogen peroxide levels as a measure of ROS were found to be significantly increased at 15 ppm sodium fluoride in drinking water in studies conducted by another group of authors (Adedara *et al.* 2017a, Adedara *et al.* 2017b).

Apoptosis/Cell Death

Seven low risk-of-bias studies were identified that evaluated apoptosis with concentrations at or below 20 ppm fluoride. Results from these studies were inconsistent and were insufficient for evaluating fluoride-induced apoptosis. These data are insufficient to increase confidence or support a change to hazard conclusions.

Inflammation

Five low risk-of-bias studies were identified that evaluated potential effects of fluoride on inflammation with concentrations at or below 20 ppm. The inflammation markers were too heterogeneous or limited in number to make any determination on potential relevance of mechanism, even before limiting the review of the data to low risk-of-bias studies. These data are insufficient to increase confidence or support a change to hazard conclusions.

Thyroid

Seventeen studies were identified that evaluated potential effects of fluoride on the thyroid with concentrations at or below 20 ppm (see **Figure A5-1**). These animal thyroid data are not further described because this endpoint has been directly evaluated in a number of human studies that have failed to identify consistent evidence to suggest that thyroid effects are a requisite mechanism by which fluoride causes neurodevelopmental or cognitive effects in humans.

Figure A5-2. Number of Low Risk-of-bias Animal Studies that Evaluated Biochemical, Neurotransmission, and Oxidative Stress Effects at or Below 20 ppm by Mechanism Subcategory and Direction of Effect*

Mechanism	Mechanism Subcategory	Direction of Effect ↑	↓	NS	Grand Total
Biochemistry (brain/neurons)	Carbohydrate/lipid-related		1	1	2
	Gene expression		1		1
	Protein levels associated with brain function	8	7	7	11
	Other biochemical	2			2
Neurotransmitters	Cholinesterase	2	7	3	11
	Dopamine and metabolites		1		1
	Other neurotransmitters	2	2	1	2
Oxidative stress	Antioxidant activity	1	10	4	12
	Lipid peroxidation byproduct	7		4	11
	Protein oxidation	2		1	2
	ROS	2		2	4

*Interactive figure and additional study details in Tableau® (https://public.tableau.com/app/profile/ntp.visuals/viz/Fluoride_Animal_SelectMechanisms_2021/FigureA5-2). This figure displays study counts for low risk-of-bias studies, as these counts are most relevant to the text in this section. Counts for high risk-of bias studies or all studies combined can be accessed in the interactive figure in Tableau®. Study counts are tabulated by significance—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with at least one endpoint in the mechanistic subcategory with significantly increasing results at fluoride exposure levels of ≤20 ppm. These columns are not mutually exclusive (i.e., a study may report on multiple endpoints with varying results within a single mechanistic subcategory and therefore may be reflected in the counts for the "↑", "↓", and NS columns, but would only be counted once in the Grand Total column). Endpoints, species, strain, sex, and exposure duration are available for each study in the interactive figure in Tableau®.

National Toxicology Program
U.S. Department of Health and Human Services

DRAFT NTP MONOGRAPH ON THE

SYSTEMATIC REVIEW OF FLUORIDE EXPOSURE AND NEURODEVELOPMENTAL AND COGNITIVE HEALTH EFFECTS*

Revised September 16, 2020

*The September 6, 2019 draft monograph was peer reviewed by a committee convened by the National Academy of Sciences, Engineering, and Medicine (NASEM). This current draft incorporates changes in response to that review and is being submitted to the same NASEM committee for an additional round of peer review.

Office of Health Assessment and Translation
Division of the National Toxicology Program
National Institute of Environmental Health Sciences
National Institutes of Health

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOREWORD

The National Toxicology Program (NTP), established in 1978, is an interagency program within the Public Health Service of the U.S. Department of Health and Human Services. Its activities are executed through a partnership of the National Institute for Occupational Safety and Health (part of the Centers for Disease Control and Prevention), the Food and Drug Administration (primarily at the National Center for Toxicological Research), and the National Institute of Environmental Health Sciences (part of the National Institutes of Health), where the program is administratively located. NTP offers a unique venue for the testing, research, and analysis of agents of concern to identify toxic and biological effects, provide information that strengthens the science base, and inform decisions by health regulatory and research agencies to safeguard public health. NTP also works to develop and apply new and improved methods and approaches that advance toxicology and better assess health effects from environmental exposures.

NTP conducts literature-based evaluations to determine whether exposure to environmental substances (e.g., chemicals, physical agents, and mixtures) may be associated with adverse health effects. These evaluations result in hazard conclusions or characterize the extent of the evidence and are published in the NTP Monograph series, which began in 2011. NTP Monographs serve as an environmental health resource to provide information that can used to make informed decisions about whether exposure to a substance may be of concern for human health.

NTP conducts these health effects evaluations following pre-specified protocols that apply the general methods outlined in the "Handbook for Conducting a Literature-Based Health Assessment Using the OHAT Approach for Systematic Review and Evidence Integration."[1] The protocol describes project-specific procedures tailored to each systematic review in a process that facilitates evaluation and integration of scientific evidence from published human, experimental animal, and mechanistic studies.

The key feature of the systematic review approach is the application of a transparent framework to document the evaluation methods and the basis for scientific judgements. This process includes steps to comprehensively search for studies, select relevant evidence, assess individual study quality, rate confidence in bodies of evidence across studies, and then integrate evidence to develop conclusions for the specific research question. Draft monographs undergo external peer review prior to being finalized and published.

NTP Monographs are available free of charge on the NTP website and cataloged in PubMed, a free resource developed and maintained by the National Library of Medicine (part of the National Institutes of Health). Data for these evaluations are included in the Health Assessment and Workspace Collaborative.

For questions about the monographs, please email NTP or call 984-287-3211.

[1] OHAT is the abbreviation for Office of Health Assessment and Translation, which is within the Division of the National Toxicology Program at the National Institute of Environmental Health Sciences.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

TABLES:

FIGURES:

DATA FIGURES:

RISK-OF-BIAS FIGURES:

META-ANALYSES TABLES:

Attachment A

Attachment B

META-ANALYSES FIGURES:

Attachment A

Attachment B

ABSTRACT

Background: The overall objective of this evaluation was to undertake a systematic review of published literature to reach conclusions concerning the potential for exposure to fluoride to affect neurodevelopment and cognition. The review only addresses whether exposure to fluoride could present a potential hazard (i.e., has the potential to cause harm, at any exposure level, including exposures that are higher than those typically encountered from consuming fluoridated drinking water in the United States). Benefits of fluoride with respect to oral health are not addressed in this monograph.

Previous reviews of epidemiological studies, including a 2006 evaluation by the National Research Council (NRC), found support for an association between consumption of high levels of naturally occurring fluoride in drinking water and adverse neurological effects in humans and recommended further investigation (NRC 2006). Most of the evidence reviewed was from dental and skeletal fluorosis-endemic regions that have higher levels of naturally occurring fluoride than the fluoride concentrations historically added to water in community water fluoridation programs (0.8–1.2 mg/L). For community water systems that add fluoride, the Public Health Service now recommends a fluoride concentration of 0.7 mg/L.

NTP previously published a systematic review of the evidence from experimental animal studies of the effects of fluoride on learning and memory (NTP 2016). The systematic review found a low-to-moderate level of evidence that learning and memory deficits occur in non-human mammals exposed to fluoride. Studies in animals generally used fluoride drinking water concentrations that far exceeded the levels used in water fluoridation, and the lack of studies at lower fluoride concentrations was identified as a data gap. The evidence for effects on learning and memory was strongest (moderate) in animals exposed as adults, and evidence was weaker (low) in animals exposed during development. Since the publication of the NTP (2016) systematic review of the animal evidence, additional animal studies have been published, many examining the effects of perinatal exposures. In addition, the number of studies examining cognitive and neurobehavioral effects of fluoride in humans has grown considerably since the NRC (2006) review, including several recent prospective cohort studies evaluating prenatal fluoride exposures.

Objective: To conduct a systematic review of the human, experimental animal [extending (NTP 2016) report], and mechanistic literature to evaluate the evidence and develop hazard conclusions about whether fluoride exposure is associated with neurodevelopmental and cognitive effects.

Method: A systematic review protocol was developed and utilized following the Office of Health Assessment and Translation (OHAT) approach for conducting literature-based health assessments.

Results: The literature search and screening process identified 159 published human studies, 339 published experimental animal studies, and 60 in vitro/mechanistic studies relevant to the objective. Ninety-two of the 159 human studies evaluated the association between fluoride exposure and neurodevelopmental or cognitive effects, and the remaining human studies evaluated thyroid effects or other potential mechanistic data. The majority of the experimental animal studies were mechanistic studies, which were not assessed in the NTP (2016) report. Since the NTP (2016) systematic review (through April 2019), 35 experimental animal studies evaluating effects on learning and memory and/or motor activity and sensory effects of fluoride were identified.

Supported by a meta-analysis, the human body of evidence provides a consistent and robust pattern of findings that higher fluoride exposure (e.g., >1.5 mg/L in drinking water) is associated with adverse effects on neurocognitive development, including lower intelligence quotient (IQ) in children. There is a moderate level of evidence from cognitive neurodevelopmental studies in children based on five prospective cohort studies and 14 cross-sectional studies where exposure was identified as occurring prior to outcome. The evidence for cognitive effects in adults is limited, coming from two cross-sectional studies, and is inadequate to evaluate whether fluoride exposure in adults is associated with cognitive effects. The assessment of the new animal data focuses on evaluating a deficiency identified during the prior NTP (2016) review concerning the difficulty in distinguishing potential effects of fluoride on motor and sensory functions from effects specifically on learning and memory functions. Further examination of the animal data, including studies carried out at the NTP, has not resolved this issue. Because of this and other deficiencies related to overall study quality, the animal body of evidence is now considered inadequate to inform conclusions on whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans. While the animal data provide evidence of effects of fluoride on neurodevelopment, the human evidence base is primarily focused on cognitive neurodevelopmental effects, and these human data are the primary basis of conclusions.

Conclusions: Because the majority of available studies evaluated cognitive neurodevelopmental effects in children, the focus of the hazard conclusions is on cognitive neurodevelopmental effects, primarily IQ. When focusing on findings from studies with exposures in ranges typically found in drinking water in the United States (0.7 mg/L for optimally fluoridated community water systems)[2] that can be evaluated for dose response, effects on cognitive neurodevelopment are inconsistent, and therefore unclear. However, when considering all the evidence, including studies with exposures to fluoride levels higher than 1.5 mg/L in water, NTP concludes that fluoride is *presumed to be a cognitive neurodevelopmental hazard to humans*. This conclusion is based on a moderate level of evidence that shows a consistent and robust pattern of findings in human studies across several different populations demonstrating that higher fluoride exposure (e.g., >1.5 mg/L in drinking water) is associated with lower IQ and other cognitive effects in children. Limited and weaker evidence is considered to provide an inadequate level of evidence that fluoride is associated with cognitive effects in adults. The evidence from animal studies is inadequate to inform conclusions on cognitive effects, and the mechanisms underlying fluoride-associated cognitive neurodevelopmental effects are not well characterized.

[2]As of April 2020, 1.08% of persons living in the United States (~ 3.5 million people) were served by community water systems (CWS) containing ≥ 1.1 mg/L naturally occurring fluoride. CWS supplying water with ≥ 1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population (~ 1.9 million people), and systems supplying water with ≥ 2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people) (https://www.cdc.gov/fluoridation/data-tools/reporting-system.html).

INTRODUCTION

The NTP's Office of Health Assessment and Translation (OHAT) conducted a systematic review to evaluate the evidence that exposure to fluoride is associated with neurodevelopmental or cognitive effects. This review was initiated in response to a nomination from the Fluoride Action Network. There are numerous human and animal studies reporting neurodevelopmental and cognitive health effects of exposure to excess fluoride. As noted by the National Research Council (NRC) in their 2006 report, although the studies lacked sufficient detail to fully assess their quality and relevance to the U.S. populations, the consistency of the results suggesting that fluoride may be neurotoxic warrants additional research (NRC 2006).

Fluoride salts are added to community water systems and dental products in the United States (e.g., toothpaste, mouth rinses, and supplements) for the prevention of dental caries. Approximately 67% of the U.S. population receives fluoridated water through a community drinking water system (CDC 2013). In other countries fluoride supplementation has been achieved by fluoridating food products such as salt, or milk. Fluoride supplementation has been recommended to prevent bone fractures (Jones *et al.* 2005). Fluoride also can occur naturally in drinking water. Other sources of human exposure include other foods and beverages, industrial emissions, pharmaceuticals, and pesticides (e.g., cryolite, sulfuryl fluoride). Soil ingestion is another source of fluoride exposure in young children (US EPA 2010).

The U.S. Public Health Service (PHS) first recommended communities add fluoride to drinking water in 1962. PHS guidance is advisory, not regulatory, which means that while PHS recommends community water fluoridation as a public health intervention, the decision to fluoridate water systems is made by state and local governments.[3] For community water systems that add fluoride, PHS now recommends a fluoride concentration of 0.7 milligrams/liter (mg/L) (equal to 0.7 parts per million [ppm]). Under the Safe Drinking Water Act, the U.S. Environmental Protection Agency (EPA) sets maximum exposure level standards for drinking water quality. The current enforceable drinking water standard for fluoride, or the maximum contaminant level goal (MCLG, a concentration at which no adverse health effects are expected), is 4.0 mg/L. This is the maximum amount of fluoride contamination (naturally occurring not from water fluoridation) that is allowed in water from public water systems; it is set to protect against increased risk of skeletal fluorosis, a condition characterized by pain and tenderness of the major joints. EPA also has a non-enforceable secondary drinking water standard of 2.0 mg/L, which is recommended to protect children against the tooth discoloration and/or pitting that can be caused by severe dental fluorosis during the formative period prior to eruption of the teeth. Although the secondary standard is not enforceable, EPA does require that public water systems notify the public if the average levels exceed it (NRC 2006). As of April 2020, 1.08% of persons living in the United States (~ 3.5 million people) were served by community water systems (CWS) containing ≥ 1.1 mg/L naturally occurring fluoride. CWS supplying water with ≥ 1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population (~ 1.9 million people), and systems supplying water with ≥ 2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people) (https://www.cdc.gov/fluoridation/data-tools/reporting-system.html).

Controversy over community water fluoridation stems from concerns about the potential harmful effects of fluoride and the ethics of water fluoridation. Commonly cited health concerns related to

[3] For many years, most fluoridated community water systems used fluoride concentrations ranging from 0.8 to 1.2 mg/L (US DHHS 2015).

fluoride are bone fractures and skeletal fluorosis, lower intelligence quotient (IQ) and other neurological effects, cancer, and endocrine disruption. Effects on neurological function, endocrine function (e.g., thyroid, parathyroid, pineal), metabolic function (e.g., glucose metabolism), and carcinogenicity were assessed in the 2006 NRC report Fluoride in Drinking Water: A Scientific Review of EPA's Standards (NRC 2006). The NRC review considered adverse effects of water fluoride, focusing on a range of concentrations (2–4 mg/L) above the current 0.7-mg/L recommendation for community water fluoridation (NRC 2006). The NRC report concluded that the current MCLG should be lowered to protect against severe enamel fluorosis and to reduce the risk of bone fractures associated with skeletal fluorosis (NRC 2006). Other than severe fluorosis, the NRC did not find sufficient evidence of negative health effects at fluoride levels below 4.0 mg/L; however, the NRC concluded that the consistency of the results of IQ deficits in children exposed to fluoride at 2.5 to 4 mg/L in drinking water from a few epidemiological studies of Chinese populations appeared significant enough to warrant additional research on the effects of fluoride on intelligence. The NRC report noted several challenges to evaluating the literature, citing deficiencies in reporting quality, lack of consideration of all sources of fluoride exposure, incomplete consideration of potential confounding, selection of inappropriate control subject populations in epidemiological studies, absence of demonstrated clinical significance of reported endocrine effects, and incomplete understanding of the biological relationship between histological, biochemical, and molecular alterations with behavioral effects (NRC 2006).

In 2016, NTP conducted a systematic review of the evidence from experimental animal studies on the potential effects of fluoride exposure on learning and memory (NTP 2016). The NTP (2016) systematic review found a low-to-moderate level of evidence that learning and memory deficits occur in experimental animals exposed to fluoride. Based on the findings in NTP (2016), NTP decided to conduct additional animal studies before carrying out a full systematic review to incorporate human, animal, and potentially relevant mechanistic evidence in order to reach hazard identification conclusions for fluoride and learning and memory effects. As the NTP (2016) report on the experimental animal evidence focused on learning and memory and developed confidence ratings for bodies of evidence by life stage of exposure (i.e., exposure during development or adulthood), this report also evaluates two different age groups in humans (i.e., children and adults) with a focus on cognitive neurodevelopmental effects in children and cognitive effects in adults in order to address potential differences in the health impact based on timeframe of exposure (i.e., during development or during adulthood). This evaluation has been conducted separately from the 2016 experimental animal assessment, but like the 2016 assessment, it has assessed mainly learning and memory effects in experimental animal studies to determine whether the findings inform the assessment of cognitive neurodevelopmental effects in children and cognitive effects in adults. The September 6, 2019 draft of this monograph was reviewed by a committee convened by the National Academy of Sciences, Engineering, and Medicine (NASEM). The current document incorporates changes in response to that review.

OBJECTIVE AND SPECIFIC AIMS

Objective

The overall objective of this evaluation is to undertake a systematic review to develop NTP hazard identification conclusions on the association between exposure to fluoride and neurodevelopmental and cognitive effects based on integrating levels of evidence from human and non-human animal studies with consideration of the degree of support from mechanistic data.

Specific Aims

- Identify literature that assessed neurodevelopmental and cognitive health effects, especially outcomes related to learning, memory, and intelligence, following exposure to fluoride in human, animal, and relevant in vitro/mechanistic studies.
- Extract data on potential neurodevelopmental and cognitive health effects from relevant studies.
- Assess the internal validity (risk of bias) of individual studies using pre-defined criteria.
- Assess effects on thyroid function to help evaluate potential mechanisms of impaired neurological function.
- Summarize the extent and types of health effects evidence available.
- Describe limitations of the systematic review, limitations of the evidence base, identify areas of uncertainty, as well as data gaps and research needs for neurodevelopmental and cognitive health effects of fluoride.

Dependent on the extent and nature of the available evidence:

- Synthesize the evidence using a narrative approach or meta-analysis (if appropriate) considering limitations on data integration such as study design heterogeneity.
- Rate confidence in the body of evidence for human and animal studies separately according to one of four statements: High, Moderate, Low, or Very Low/No Evidence Available.
- Translate confidence ratings into level of evidence of health effects for human and animal studies separately according to one of four statements: High, Moderate, Low, or Inadequate.
- Combine the level of evidence ratings for human and animal data to reach one of five possible hazard identification conclusions: Known, Presumed, Suspected, Not classifiable, or Not identified to be a hazard to humans.

METHODS

Problem Formulation and Protocol Development

The research question and specific aims stated above were developed and refined through a series of problem formulation steps including:

(1) receipt of nomination from the public in June 2015 to conduct analyses of fluoride and developmental neurobehavioral toxicity;

(2) analysis of the extent of evidence available and the merit of pursuing systematic reviews, given factors such as the extent of new research published since previous evaluations and whether these new reports address or correct the deficiencies noted in the literature (NRC 2006, OEHHA 2011, SCHER 2011);

(3) request for information in a Federal Register notice (dated October 7, 2015);

(4) consideration of comments providing a list of studies to review through Federal Register notice and public comment period from October 7, 2015 to November 6, 2015;

(5) release of draft concept titled *Proposed NTP Evaluation on Fluoride Exposure and Potential for Developmental Neurobehavioral Effects* in November 2015;

(6) presentation of draft concept at the NTP Board of Scientific Counselors (BSC) meeting on December 1−2, 2015;

(7) consideration of comments on NTP's draft concept from the NTP BSC meeting in December 2015; and

(8) consideration of input on the draft protocol from review by technical advisors.

NTP published a systematic review of the animal evidence on the effects of fluoride on learning and memory (NTP 2016). NTP has conducted additional studies in animals to assess the effect of fluoride exposure on learning and memory. The results from this experimental animal work were published (McPherson *et al.* 2018) and are incorporated into the current review, which considers the epidemiological, animal, and mechanistic evidence in its conclusions. The protocol used to conduct this systematic review was posted in June 2017 with updates posted in May 2019 and September 2020 (https://ntp.niehs.nih.gov/go/785076).[4] A brief summary of the methods is presented below.

NASEM Review

The September 6, 2019 draft of this monograph was peer reviewed by a committee convened by NASEM. The current draft reflects clarifications and changes in response to that review (NASEM 2020), including the addition of meta-analyses of the IQ studies in children.

PECO Statements

PECO (**P**opulation, **E**xposure, **C**omparators and **O**utcomes) statements were developed as an aid to identify search terms and inclusion/exclusion criteria as appropriate for addressing the overall research question (effects on neurodevelopmental or cognitive function and thyroid associated with fluoride exposure) for the systematic review (Higgins and Green 2011).The PECO statements are listed below for human, animal, and in vitro/mechanistic studies (see **Table 1**, **Table 2**, and **Table 3**).

Using the PECO statements, the evaluation searched for evidence of neurodevelopmental or cognitive function, and thyroid effects associated with fluoride exposure from human studies, controlled exposure animal studies, and mechanistic/in vitro studies. Mechanistic data can come from a wide variety of studies that are not intended to identify a disease phenotype. This source of experimental data includes in vitro and in vivo laboratory studies directed at cellular, biochemical, and molecular mechanisms that attempt to explain how a substance produces particular adverse health effects. The mechanistic data were first organized by general categories (e.g., biochemical effects in the brain and neurons, neurotransmitters, oxidative stress, etc.) to evaluate the information available. Categories focused on were those with more robust data at levels of fluoride more relevant to human exposure. The intent was not to develop a mechanism for fluoride induction of learning and memory effects, but to evaluate whether a plausible series of mechanistic events exists to support effects observed in the low-dose

[4] NTP conducts systematic reviews following pre-specified protocols that describe the review procedures selected and applied from the general methods outlined in the OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673). The protocol describes project-specific procedures tailored to each systematic review.

region (below approximate drinking water equivalent concentrations of 20 ppm) that may strengthen the hazard conclusion.

Table 1. Human PECO (Population, Exposure, Comparator and Outcome) Statement	
PECO Element	**Evidence**
Population	Humans without restriction as to age or sex, geographic location, or life stage at exposure or outcome assessment
Exposure	Exposure to fluoride based on administered dose or concentration, biomonitoring data (e.g., urine, blood, other specimens), environmental measures (e.g., air, water levels), or job title or residence. Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; CASRN 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable populations not exposed to fluoride or exposed to lower levels of fluoride (e.g., exposure below detection levels)
Outcomes	Neurodevelopmental outcomes including learning, memory, intelligence, other forms of cognitive behavior, other neurological outcomes (e.g., anxiety, aggression, motor activity), and biochemical changes in the brain or nervous system tissue; or measures of thyroid function, biochemical changes, or thyroid tissue

Table 2. Animal PECO Statement	
PECO Element	**Evidence**
Population	Non-human mammalian animal species (whole organism)
Exposure	Exposure to fluoride based on administered dose or concentration, and biomonitoring data (e.g., urine, blood, other specimens). Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; CASRN 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable animals that were untreated or exposed to vehicle-only treatment
Outcomes	Neurodevelopmental outcomes including learning, memory, intelligence, other forms of cognitive behavior, other neurological outcomes (e.g., anxiety, aggression, motor activity), and biochemical changes in the brain or nervous system tissue; or measures of thyroid function, biochemical changes, or thyroid tissue

Table 3. In Vitro/Mechanistic PECO Statement	
PECO Element	**Evidence**
Population	Human or animal cells, tissues, or biochemical reactions (e.g., ligand binding assays)
Exposure	Exposure to fluoride based on administered dose or concentration. Relevant forms are those used as additives for water fluoridation: • Fluorosilicic acid (also called hydrofluorosilicate; CASRN 16961-83-4) • Sodium hexafluorosilicate (also called disodium hexafluorosilicate or sodium fluorosilicate; CASRN 16893-85-9) • Sodium fluoride (CASRN 7681-49-4) • Other forms of fluoride that readily dissociate into free fluoride ions (e.g., potassium fluoride, calcium fluoride, ammonium fluoride)
Comparators	Comparable cells or tissues that were untreated or exposed to vehicle-only treatment
Outcomes	Endpoints related to neurological and thyroid function, including neuronal electrophysiology; mRNA, gene, or protein expression; cell proliferation or death in brain or thyroid tissue/cells; neuronal signaling; synaptogenesis, etc.

Literature Search

Main Literature Search

Search terms were developed to identify all relevant published evidence on developmental neurobehavioral toxicity or thyroid-related health effects potentially associated with exposure to fluoride by reviewing Medical Subject Headings for relevant and appropriate neurobehavioral and thyroid-related terms, and by extracting key neurological and thyroid-related health effects and

developmental neurobehavioral terminology from reviews and a sample of relevant primary data studies. A combination of relevant subject headings and keywords were subsequently identified. A test set of relevant studies was used to ensure the search terms retrieve 100% of the test set. Six electronic databases were searched (see **Main Literature Database Search**) using a search strategy tailored for each database (specific search terms used for the PubMed search are presented in **Appendix 1**; the search strategy for other databases are available in the protocol (https://ntp.niehs.nih.gov/go/785076). A search of PubChem indicated that sodium fluoride was not found in either the Tox21 or ToxCast databases; therefore, these databases were not included in the search. No language restrictions or publication year limits were imposed. These six databases were searched in December 2016 and the search was regularly updated during the review process through April 1, 2019.

Evaluations must include cut-off dates for the literature search to enable synthesis and development of conclusions. Following the NASEM committee peer review in November 2019 (NASEM 2020), an additional search was conducted on May 1, 2020, where only primary human epidemiology studies were prioritized during screening. The review of the 2020 search results focused on the human studies because they formed the basis of the conclusions in the September 6, 2019 draft. A supplemental literature search of Chinese-language databases (described below) was also conducted.

Publications identified in these searches are categorized as "references identified through database searches" in **Figure 4.** Studies identified from other sources or manual review that might impact conclusions were considered under "references identified through other sources" in **Figure 4**. Literature searches for this systematic review were conducted independently from the literature search conducted for NTP (2016). The current literature search strategy was based on the search terms used for NTP (2016) and refined for the current evaluation, including the addition of search terms to identify human studies. Although the review process identified studies prior to 2015, the current assessment did not evaluate the studies published prior to 2015 and relied on the NTP (2016) assessment. The focus of the literature searches for this systematic review was to identify and evaluate relevant animal studies that were published since completion of the literature searches for the NTP (2016) assessment in addition to the human and mechanistic data that were not previously evaluated.

Supplemental Chinese Database Literature Search

Following NASEM committee peer review in November 2019 (NASEM 2020), additional searches were developed for non-English-language databases to systematically search for studies that were previously identified from other resources (e.g., Chinese-language studies from the Fluoride Action Network website). Non-English-language databases with the greatest potential to contain relevant non-English publications that were not previously identified through database searches were selected. Multiple non-English language databases were explored before finding two databases (CNKI and Wanfang) that covered studies previously identified from other sources. These two Chinese electronic databases were searched in May 2020 with no language restrictions or publication year limits. Search terms from the main literature search were refined to focus on human epidemiology studies. The CNKI and Wanfang databases have character limits in the search strings; therefore, key terms were prioritized using text analytics to identify the most prevalent terms from neurodevelopmental or cognitive human epidemiology studies previously identified as relevant. Search strings were designed to capture known relevant studies that were previously identified from searching other resources without identifying large numbers of non-relevant studies [the search strategy for both databases are available in the protocol (https://ntp.niehs.nih.gov/go/785076)]. Publications retrieved were compared to publications retrieved from the main literature search and duplicates were removed. The remaining relevant publications are categorized as "references identified through database searches" in **Figure 4**. New animal and

mechanistic references retrieved were scanned for evidence that might extend the information currently in the September 6, 2019 draft. Although additional studies were identified, data that would materially advance the animal and mechanistic findings were not identified; therefore, these studies were not extracted nor were they added to the draft. Newly-retrieved human references were reviewed to identify studies that might impact conclusions with priority given to identifying and translating null studies that may have been missed using previous approaches. Null studies that were identified were translated and included.

Databases Searched

Main Literature Database Search

- BIOSIS (Thomson Reuters)
- EMBASE
- PsycINFO (APA PsycNet)
- PubMed (NLM)
- Scopus (Elsevier)
- Web of Science (Thomson Reuters, Web of Science indexes the journal Fluoride)

Supplemental Chinese Database Literature Search

- CNKI
- Wanfang

Searching Other Resources

The reference lists of all included studies; relevant reviews, editorials and commentaries; and the Fluoride Action Network website (http://fluoridealert.org/) were manually searched for additional relevant publications. Following NASEM committee peer review in November 2019 (NASEM 2020), the Fluoride Action Network website was again searched for relevant references and contacted to identify null or no effect studies.

Unpublished Data

Unpublished data were eligible for inclusion provided the owner of the data was willing to have the data made public and peer reviewed (see protocol for more details https://ntp.niehs.nih.gov/go/785076). No unpublished data were identified during the literature search.

Study Selection

Evidence Selection Criteria

In order to be eligible for inclusion, studies had to satisfy eligibility criteria that reflect the PECO statement in **Table 1**, **Table 2**, and **Table 3**. The following additional exclusion criteria were applied (see protocol for additional details; https://ntp.niehs.nih.gov/go/785076):

(1) Case studies and case reports.

(2) Articles without original data (e.g., reviews, editorials, or commentaries). Reference lists from these materials, however, were reviewed to identify potentially relevant studies not identified from the database searches. New studies identified were assessed for eligibility for inclusion.

(3) Conference abstracts or reports.

Screening Process

References retrieved from the literature search were independently screened by two trained screeners at the title and abstract level to determine whether a reference met the evidence selection criteria. Screening procedures following the evidence selection criteria in the protocol were pilot-tested with experienced contract staff overseen by NTP. For citations with no abstract or non-English abstracts, articles were screened based on title relevance (title would need to indicate clear relevance); number of pages (articles ≤ 2 pages were assumed to be conference reports, editorials, or letters unlikely to contain original data); and/or PubMed Medical Subject Headings (MeSH). Using this approach, literature was manually screened for relevance and eligibility against the evidence selection criteria using a structured form in SWIFT-Active Screener. While the human screeners review studies, SWIFT-Active Screener aids in the process by employing a machine-learning software program used to priority-rank studies for screening (Howard *et al.* 2020). SWIFT-Active Screener also refines a statistical model that continually ranks the remaining studies according to their likelihood for inclusion. In addition, SWIFT-Active Screener employs active learning to continually incorporate user feedback during title and abstract screening to predict the total number of included studies, thus providing a statistical basis for a decision about when to stop screening (Miller *et al.* 2016). Title and abstract screening was stopped once the statistical algorithm in SWIFT-Active Screener estimated that 98% of the predicted number of relevant studies were identified.

Studies that were not excluded during the title and abstract screening were further screened for inclusion with a full-text review by two independent reviewers using DistillerSR® by Evidence Partners, a web-based, systematic-review software program with structured forms and procedures to ensure standardization of the process. Screening conflicts were resolved through discussion and consultation with technical advisor(s), if necessary. During full-text review, studies that were considered relevant were tagged to the appropriate evidence streams (i.e., human, animal, and/or in vitro). Studies tagged to human or animal evidence streams were also categorized by outcome as primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence); secondary neurobehavioral outcomes (anxiety, aggression, motor activity, or biochemical); or related to thyroid effects. In vitro data were tagged as being related to neurological effects or thyroid effects. Translation assistance was sought to assess the relevance of non-English studies. Following full-text review, the remaining studies were "included" and used for the evaluation.

Screening of the May 2020 Literature Search Update

Following the NASEM committee peer review in November 2019 (NASEM 2020), an additional search was conducted on May 1, 2020, where only primary human epidemiology studies were identified. The study screening and selection process was focused on the human studies with primary outcomes for the evaluation because they form the basis of the conclusions. Animal in vivo, human secondary outcome-only, and human and animal mechanistic references were identified as part of the screening process. These studies were then scanned for evidence that might extend the information in the September 6, 2019 draft. Studies from the May 2020 literature search update will be listed in an appendix; however, data from the studies were not extracted unless it was believed they would materially advance the findings.

Data Extraction

Data were collected (i.e., extracted) from included studies by one member of the evaluation team and checked by a second member of the team for completeness and accuracy. Any discrepancies in data extraction were resolved by discussion or consultation with a third member of the evaluation team.

Data extraction was completed using the Health Assessment Workspace Collaborative (HAWC), an open source and freely available web-based interface application.[5] Data extraction elements are listed separately for human, animal, and in vitro studies in the protocol (https://ntp.niehs.nih.gov/go/785076). Data for primary and secondary outcomes as well as thyroid hormone level data were extracted from human studies. Studies evaluating only goiters or thyroid size were not extracted. All primary outcomes and functional neurological secondary outcomes (e.g., motor activity) were extracted from animal studies identified since the NTP (2016) report. For animal mechanistic data, studies were tiered based on exposure dose (with preference given to fluoride drinking water equivalent exposures, which were calculated using the method described in the NTP (2016) report) of 20 ppm or less as deemed most relevant to exposures in humans), exposure duration or relevant time window (i.e., developmental), exposure route (with preference given to oral exposures over injection exposures), and commonality of mechanism (e.g., inflammation, oxidative stress, changes in neurotransmitters, and histopathological changes were considered pockets of mechanistic data). Data were not extracted from in vitro studies; however, these studies were evaluated for biological plausibility of the human and animal results. Thyroid data were also reviewed but not extracted. The data extraction results for included studies are publicly available and can be downloaded in Excel format through HAWC (https://hawcproject.org/assessment/405/). Methods for transforming and standardizing dose levels and results from behavioral tests in experimental animals are detailed in the protocol (https://ntp.niehs.nih.gov/go/785076).

In 2016, NTP conducted a systematic review of the evidence from experimental animal studies on the potential effects of fluoride exposure on learning and memory (NTP 2016). The literature searches for the current assessment identified and evaluated relevant animal studies published since the 2016 assessment and also included human and mechanistic data that were not previously evaluated. Although literature search activities for the current assessment identified experimental animal studies prior to 2015, the current assessment did not re-evaluate studies published prior to 2015, but relied on the NTP (2016) assessment.

Quality Assessment of Individual Studies

Risk of bias was assessed for individual studies using a tool developed by OHAT that outlines a parallel approach to evaluating risk of bias from human, animal, and mechanistic studies to facilitate consideration of risk of bias across evidence streams with common terms and categories. The risk-of-bias tool is comprised of a common set of 11 questions that are answered based on the specific details of individual studies to develop risk-of-bias ratings for each question. Study design determines the subset of questions used to assess risk of bias for an individual study (see **Table 4**).

Assessors were trained with an initial pilot phase undertaken to improve clarity of rating criteria and to improve consistency among assessors. Studies were independently evaluated by two trained assessors who answered all applicable risk-of-bias questions with one of four options in **Table 5** following pre-specified criteria detailed in the protocol (https://ntp.niehs.nih.gov/go/785076). The criteria describe aspects of study design, conduct, and reporting required to reach risk-of-bias ratings for each question

[5] HAWC (Health Assessment Workspace Collaborative): A Modular Web-based Interface to Facilitate Development of Human Health Assessments of Chemicals (https://hawcproject.org/portal/).

and specify factors that can distinguish among ratings (e.g., what separates “definitely low” from “probably low” risk of bias).

Key Risk-of-bias Questions

In the OHAT approach, some risk-of-bias questions or elements are considered potentially more important when assessing studies because there is more empirical evidence that these areas of bias have a greater impact on estimates of the effect size or because these issues are generally considered to have a greater effect on the credibility of study results in environmental health studies (Rooney *et al.* 2016). There were three Key Questions for observational human studies: confounding, exposure characterization, and outcome assessment. There were also three Key Questions for experimental animal studies: randomization, exposure characterization, and outcome assessment. In addition, for animal developmental studies, failure to consider the litter as the unit of analysis was also a key risk-of-bias concern. When there was not enough information to assess the potential bias for a risk-of-bias question and authors did not respond to an inquiry for further information, a conservative approach was followed, and the studies were rated probably high risk of bias for that question.

Table 4. OHAT Risk-of-bias Questions and Applicability by Study Design

Risk-of-bias Questions	Experimental Animal*	Human Controlled Trials**	Cohort	Case-Control	Cross-Sectional***	Case Series
1. Was administered dose or exposure level adequately randomized?	X	X				
2. Was allocation to study groups adequately concealed?	X	X				
3. Did selection of study participants result in the appropriate comparison groups?			X	X	X	
4. Did study design or analysis account for important confounding and modifying variables?			X	X	X	X
5. Were experimental conditions identical across study groups?	X					
6. Were research personnel blinded to the study group during the study?	X	X				
7. Were outcome data complete without attrition or exclusion from analysis?	X	X	X	X	X	
8. Can we be confident in the exposure characterization?	X	X	X	X	X	X
9. Can we be confident in the outcome assessment (including blinding of outcome assessors)?	X	X	X	X	X	X
10. Were all measured outcomes reported?	X	X	X	X	X	X
11. Were there no other potential threats to internal validity?	X	X	X	X	X	X

*Experimental animal studies are controlled exposure studies. Non-human animal observational studies can be evaluated using the design features of observational human studies such as cross-sectional study design.

**Human Controlled Trials are studies in humans with controlled exposure (e.g., Randomized Controlled Trials, non-randomized experimental studies)

***Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).

Any discrepancies in ratings between assessors were resolved by a senior technical specialist and through discussion when necessary to reach the final recorded risk-of-bias rating for each question along with a statement of the basis for that rating. Members of the evaluation team were consulted for assistance if additional expertise was necessary to reach final risk-of-bias ratings based on specific aspects of study design or performance reported for individual studies. Study procedures that were not reported were assumed not to have been conducted, resulting in an assessment of "probably high" risk of bias. Authors were queried by email to obtain missing information and responses received were used to update risk-of-bias ratings.

Table 5. The Four Risk-of-bias Rating Options	
Answers to the risk-of-bias questions result in one of the following four risk-of-bias ratings	
++	***Definitely Low*** *risk of bias*: There is direct evidence of low risk-of-bias practices
+	***Probably Low*** *risk of bias*: There is indirect evidence of low risk-of-bias practices **OR** it is deemed that deviations from low risk-of-bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias
– NR	***Probably High*** *risk of bias*: There is indirect evidence of high risk-of-bias practices (indicated with "-") **OR** there is insufficient information provided about relevant risk-of-bias practices (indicated with "NR" for not reported). Both symbols indicate probably high risk of bias.
––	***Definitely High*** *risk of bias*: There is direct evidence of high risk-of-bias practices

Organizing and Rating Confidence in Bodies of Evidence

Health Outcome Categories for Neurodevelopmental and Cognitive Effects

After data were extracted from all studies, the health effects results within the category of neurodevelopmental or cognitive effects were grouped across studies to develop bodies of evidence or collections of studies with data on the same or related outcomes. The grouping of health effect results was not planned a priori. The vast majority of the human studies evaluated intelligence quotient (IQ) in children as the single outcome; therefore, the discussion of cognitive neurodevelopmental effects in children focuses on IQ studies with supporting information from data on other endpoints. Cognitive function in adults was evaluated separately. Consistent with the NTP (2016) assessment, the primary focus within the animal study body of evidence was on animal studies with endpoints related to learning and memory.

Considerations for Pursuing a Narrative or Quantitative Evidence Synthesis

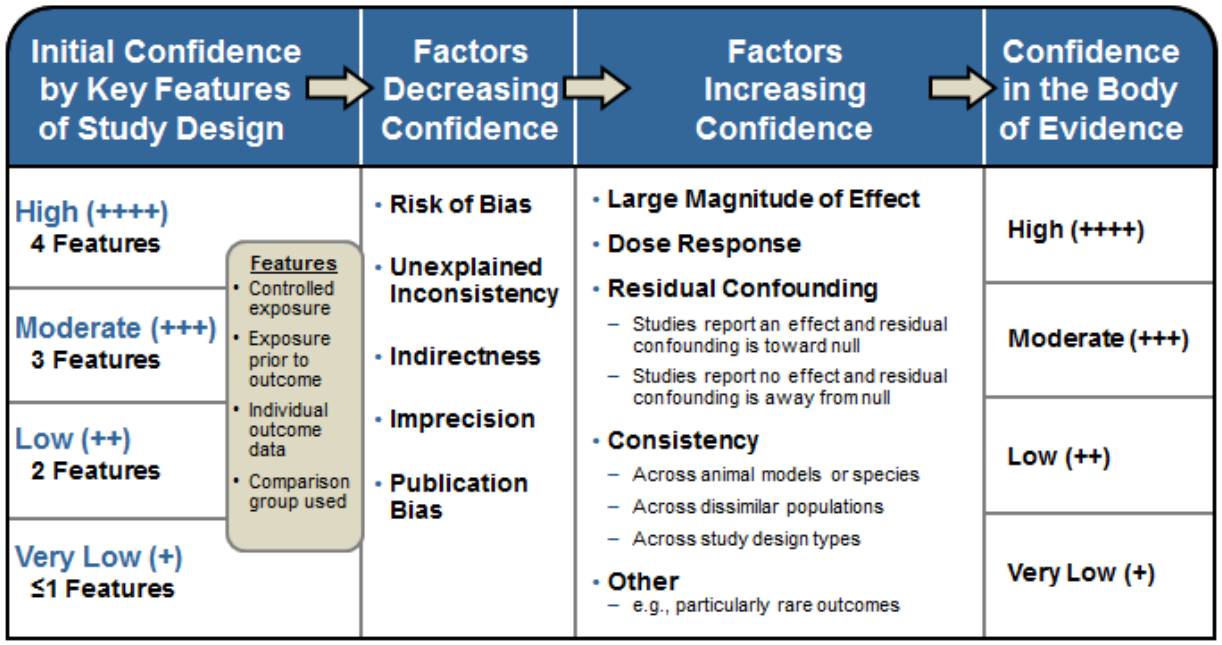
Heterogeneity within the available evidence was used to determine which type of evidence integration was appropriate—a quantitative synthesis (meta-analysis) or narrative approach for evidence integration. Choi *et al.* (2012) and Duan *et al.* (2018) conducted meta-analyses and found that high fluoride exposure was associated with lower IQ scores. Choi *et al.* (2012) was able to determine a risk ratio for living in an endemic fluorosis area but was unable to develop a dose-response relationship. Duan *et al.* (2018) suggested a significant non-linear dose-response relationship between fluoride dose and intelligence with the relationship stated to be most evident with exposures from drinking water

containing above 4 mg/L (or 4 ppm). Duan *et al.* (2018) found similar results as Choi *et al.* (2012) for the standardized mean difference; however, the majority of the available studies in both analyses compare populations with high fluoride exposure to those with lower fluoride exposure (with the lower exposure levels frequently in the range of drinking water fluoridation in the United States). After evaluating the available data, NTP determined that a narrative review—not a meta-analysis or other quantitative assessment—was appropriate for evidence integration due to heterogeneity in dose among the available human evidence, and because a hazard conclusion could be reached without conducting a meta-analysis. However, in the November 2019 review of the September 6, 2019 draft monograph (NASEM 2020), NASEM recommended that a meta-analysis be conducted. In response, NTP performed a meta-analysis of IQ studies in children. The meta-analysis protocol can be found with the revised systematic review protocol posted in September 2020 (https://ntp.niehs.nih.gov/go/785076).

Confidence Rating: Assessment of Body of Evidence

The quality of evidence for neurodevelopmental and cognitive function outcomes was evaluated using the GRADE system for rating the confidence in the body of evidence (Guyatt *et al.* 2011, Rooney *et al.* 2014). More detailed guidance on reaching confidence ratings in the body of evidence as "high," "moderate," "low," or "very low" is provided in the OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673, see STEP 5). In brief, available human and animal studies on a particular health outcome were initially grouped by key study design features, and each grouping of studies was given an initial confidence rating by those features. Starting at this initial rating (see column 1 of **Figure 1**), potential downgrading of the confidence rating was considered for factors that decrease confidence in the results (see column 2 of **Figure 1** [risk of bias, unexplained inconsistency, indirectness or lack of applicability, imprecision, and publication bias]); and potential upgrading of the confidence rating was considered for factors that increase confidence in the results (see column 3 of **Figure 1** [large magnitude of effect, dose response, consistency across study designs/populations/animal models or species, consideration of residual confounding, and other factors that increase our confidence in the association or effect]). Consideration of consistency across study designs, human populations, or animal species is not included in the GRADE guidance (Guyatt *et al.* 2011); however, it is considered in the modified version of GRADE used by OHAT (Rooney *et al.* 2014, NTP 2015).

Figure 1. Assessing Confidence in the Body of Evidence

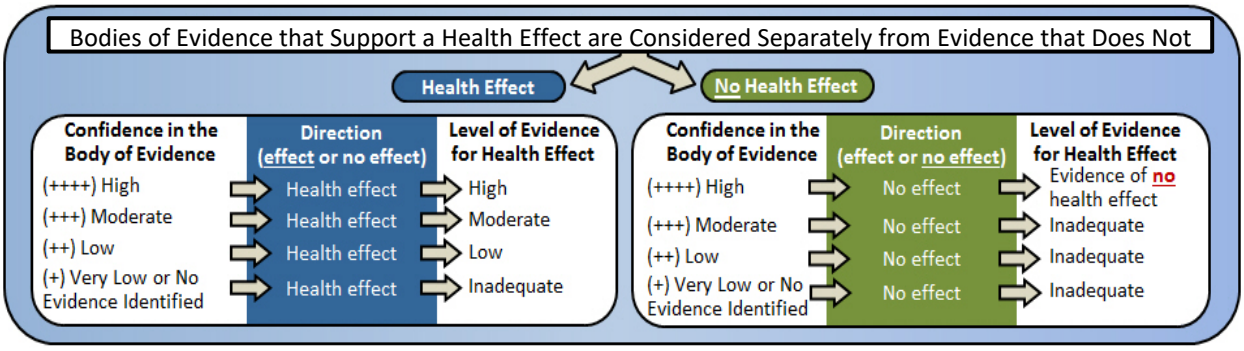


Confidence ratings were assessed by the evaluation team for accuracy and consistency, and discrepancies were resolved by consensus and consultation with technical advisors as needed. Confidence ratings for the primary outcomes are summarized in evidence profile tables for each outcome.

Preparation of Level of Evidence Conclusions

The confidence ratings were translated into level of evidence of health effects for each type of health outcome separately according to one of four statements: (1) High, (2) Moderate, (3) Low, or (4) Inadequate (see **Figure 2**). The descriptor “evidence of no health effect” is used to indicate confidence that the substance is not associated with a health effect. Because of the inherent difficulty in proving a negative, the conclusion “evidence of no health effect" is only reached when there is high confidence in the body of evidence.

Figure 2. Translate Confidence Ratings into Evidence of Health Effect Conclusions



Evidence Descriptors	Definition
High Level of Evidence	There is high confidence in the body of evidence for an association between exposure to fluoride and the health outcome(s).
Moderate Level of Evidence	There is moderate confidence in the body of evidence for an association between exposure to fluoride and the health outcome(s).
Low Level of Evidence	There is low confidence in the body of evidence for an association between exposure to fluoride and the health outcome(s), or no data are available.
Inadequate Evidence	There is insufficient evidence available to assess if exposure to fluoride is associated with the health outcome(s).
Evidence of No Health Effect	There is high confidence in the body of evidence that exposure to fluoride is not associated with the health outcome(s).

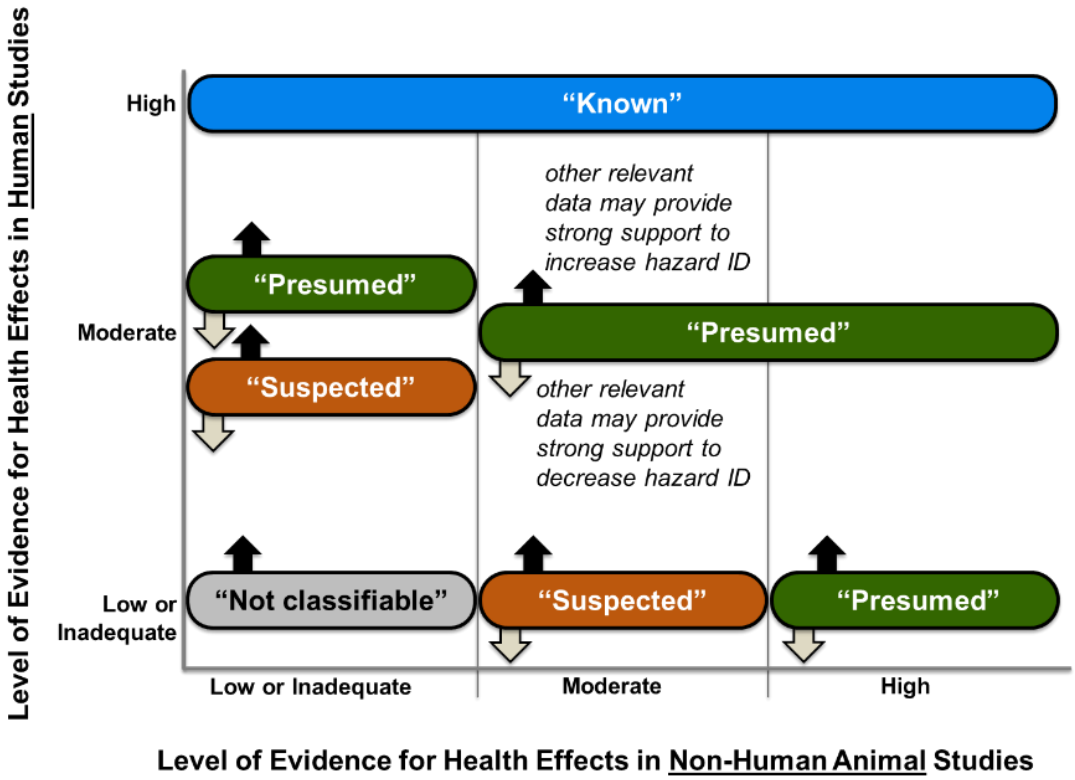
Integrate Evidence to Develop Hazard Identification Conclusions

Finally, the levels of evidence ratings for human and animal data were integrated with consideration of in vitro/mechanistic data to reach one of five possible hazard identification categories: (1) Known, (2) Presumed, (3) Suspected, (4) Not classifiable, or (5) Not identified to be a neurodevelopmental hazard to humans (see **Figure 3**).

Consideration of Human and Animal Data

Initial hazard identification conclusions were attempted by integrating the highest level-of-evidence conclusion for neurodevelopmental effects in children and cognitive effects in adults for the human and the animal evidence streams. The level-of-evidence conclusion for human data for neurodevelopmental or cognitive effects were considered with the level of evidence for non-human animal data to reach one of four initial hazard identification conclusions: Known, Presumed, Suspected, or Not classifiable. When either the human or animal evidence stream was characterized as "Inadequate Evidence," then conclusions were based on the remaining evidence stream alone (which is equivalent to treating the missing evidence stream as "Low" in **Figure 3**). If a moderate level-of-evidence conclusion for human data was reached with "Inadequate or Low Evidence" for the animal evidence stream, a hazard identification conclusion of either "suspected to be a hazard to humans" or "presumed to be a hazard to humans" could be reached based on scientific judgement as to the robustness of the body of evidence that supports moderate confidence in the human data and consideration of the potential impact of additional studies (NTP 2019).

Figure 3. Hazard Identification Scheme for Neurodevelopmental or Cognitive Effects



Consideration of Mechanistic Data

There is no requirement to consider mechanistic or mode-of-action data to reach a hazard identification conclusion regarding neurodevelopmental or cognitive health effects. However, when available, this and other relevant supporting types of evidence may be used to raise (or lower) the category of the hazard identification conclusion. Mechanistic data can come from a wide variety of studies that are not intended to identify a disease phenotype. This source of experimental data includes in vitro and in vivo laboratory studies directed at cellular, biochemical, genetic, and molecular mechanisms that attempt to explain how a chemical produces particular adverse effects.

For the evaluation of toxicity associated with fluoride exposure, NTP was interested in mechanistic or in vitro measures that comprise a coherent biological process that may support the plausibility of corresponding neurological outcomes reported from in vivo studies in animals or humans. The PECO statement in **Table 3** provides the specific endpoints considered including neuronal electrophysiology; mRNA, gene, or protein expression; cell proliferation or death in brain or thyroid tissue/cells; neuronal signaling; or synaptogenesis. In general, the mechanisms for fluoride-associated neurodevelopmental or cognitive effects are not well understood at this time, and mechanistic events identified in studies of animals receiving high fluoride exposures may not reflect biological processes occurring in humans at lower exposure levels. Mechanistic data from in vivo studies were used when feasible to examine the biological plausibility of the primary health outcomes considered in developing a hazard conclusion.

The factors outlined for increasing or decreasing confidence that the mechanistic data support biological plausibility are conceptually similar to those used to rate confidence in bodies of evidence for human or

animal in vivo studies are listed below and described in depth in the protocol (https://ntp.niehs.nih.gov/go/785076). Four factors were considered that contribute to increased confidence: potency, dose response, consistency in terms of cellular events observed at the same or lower doses than in vivo health effects, and consistency across cellular targets on the same functional pathway. Three factors were considered that contribute to decreased confidence: unexplained inconsistency across studies of the same endpoint, indirectness/applicability of the pathway for human health or concentrations for human exposure, and publication bias. Evaluations of the strength of evidence provided by mechanistic data were made on an outcome-specific basis based on discussion by the evaluation team and consultation with technical advisors as needed.

- If mechanistic data provided strong support for biological plausibility of the relationship between exposure and the health effect, the hazard identification conclusion may be upgraded (indicated by black "up" arrows in **Figure 3**) from that initially derived by considering the human and non-human animal evidence together.
- If mechanistic data provided strong opposition for biological plausibility of the relationship between exposure and the health effect, the hazard identification conclusion may be downgraded (indicated by gray "down" arrows in **Figure 3**) from that initially derived by considering the human and non-human animal evidence together.

Although it is envisioned that strong evidence for a relevant neurological effect from mechanistic data alone could indicate a potential that the substance is a neurodevelopmental hazard to humans, for this evaluation the mechanistic data were only considered to inform the biological plausibility of observed outcomes from in vivo exposure studies in humans or animals because of a general lack of understanding of the mechanistic basis for neurological outcomes.

RESULTS AND EVIDENCE SYNTHESIS

Literature Search Results

Literature Search Results Counts and Title and Abstract Screening

The electronic database searches retrieved 25,524 unique references in total (20,883 references during the initial search conducted in December 2016, 3,733 references during the literature search updates [including the final updated search conducted for the primary epidemiology studies on May 1, 2020], and 908 references from the supplemental Chinese database searches); 15 additional references were identified by technical advisors or from reviewing reference lists in published reviews and included studies. As a result of title and abstract screening, 1,038 references were moved to full-text review, 11,478 were excluded during manual title and abstract screening for not satisfying the PECO criteria, and an additional 13,023 were not screened and excluded based on the SWIFT algorithm.

Full-text Review

Among the 1,038 references that underwent full-text review, 499 references were excluded during the full-text review with reasons for exclusion documented at this stage; 332 references were excluded for not satisfying the PECO criteria; and 167 references from the May 2020 searches (main literature search update and supplemental Chinese database searches) were excluded for not including information that would materially advance the human, animal in vivo, or mechanistic findings (see the Literature Search Section for a description of the methodology). These screening results are outlined in a study selection diagram that reports numbers of studies excluded for each reason at the full text review stage (see **Figure 4**) [using reporting practices outlined in Moher *et al.* (2009)]. After full-text review, 539 studies were considered relevant with primary neurological outcomes, secondary neurological outcomes, and/or outcomes related to thyroid function (see **Appendix 2**). A few studies assessed data for more than one evidence stream (human, non-human mammal, and/or in vitro), and several human and animal studies assessed more than one type of outcome (e.g., primary and secondary outcomes). The number of included studies is summarized below. There are:

- 159 human studies (78 primary only; 13 secondary only; 5 primary and secondary; 6 primary and thyroid; 2 secondary and thyroid; and 55 thyroid only);
- 339 non-human mammal studies (7 primary only; 186 secondary only; 67 primary and secondary; 6 primary, secondary, and thyroid; 4 secondary and thyroid; and 69 thyroid only); and,
- 60 in vitro/mechanistic studies (48 neurological and 12 thyroid).

One publication contained human, experimental non-human mammal, and in vitro data. Three publications contained both human and experimental non-human mammal data. Fourteen publications contained data relevant to both experimental non-human mammal studies and in vitro studies.

Evaluation of SWIFT-Active Screener Results

During the initial title and abstract screening of 20,883 references using SWIFT-Active Screener, approximately 38%[6] of the 20,883 studies were manually screened in duplicate to identify an estimated 98.6% of the predicted number of relevant studies using the statistical algorithm in SWIFT-Active Screener (13,023 references were not screened). SWIFT-Active Screener predicted that there were 739 relevant studies during the initial title and abstract screening, of which 729 studies were identified and moved to full-text review. The SWIFT-Active statistical algorithm predicted that 10 relevant studies at the title and abstract level (10 represents 1.4% × 739 predicted relevant studies; or 739 predicted relevant studies minus 729 identified relevant studies during screening) were not identified by not screening the remaining 13,023 studies.

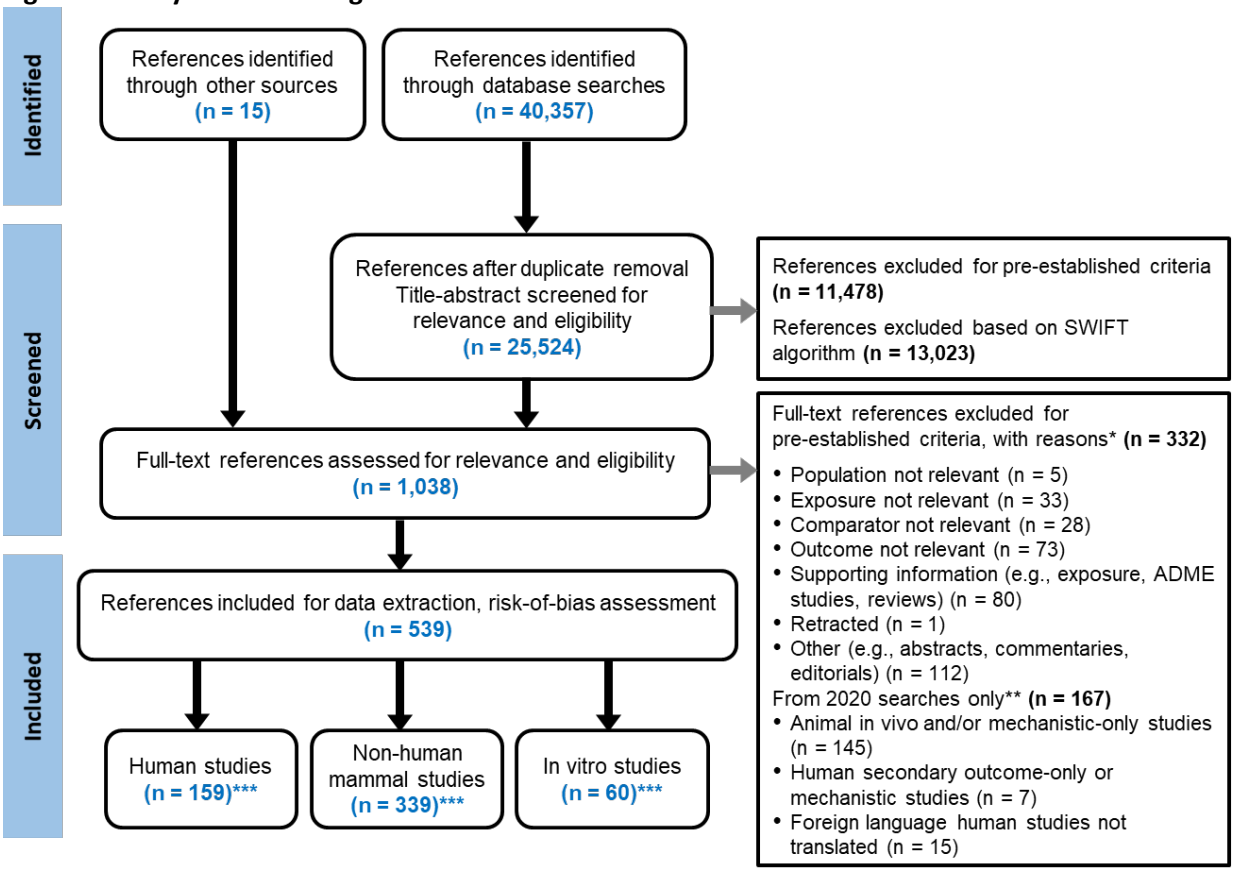
To further consider the impact of using SWIFT-Active Screener for this systematic review, NTP evaluated the SWIFT-Active screening results to gain a better understanding of the relevance of the last group of studies that were screened before 98% predicted recall was satisfied. The goal was to determine the likelihood of having missed important studies by not screening all of the literature. To do this, NTP evaluated subsets of studies screened in SWIFT-Active for trends and followed those studies through to full-text review for a final determination of relevance and potential impact (i.e., whether the studies had data on primary outcomes). Based on this evaluation, NTP estimates that the use of Swift-Active Screener may have resulted in missing 1–2 relevant human studies and 1–2 relevant animal studies with primary neurodevelopmental or cognitive outcomes.

Supplemental Chinese Database Searches and Human Epidemiology Studies

Following the NASEM committee peer review in November 2019 (NASEM 2020), supplemental searches were conducted in non-English language databases (CNKI and Wanfang). One focus of the screening of these supplemental search results was to identify null or no-effect studies that evaluated primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence) that may have been missed in previous approaches or may have been absent from the Fluoride Action Network website. Of the 908 references that were identified in the supplemental Chinese database searches, 16 relevant studies with primary neurological outcomes were identified (which were not identified through the main literature searches). Among these 16 studies, Kang *et al.* (2011) was the only null study with primary neurological outcomes that was identified through the supplemental Chinese database searches. NTP had the study translated to English, and the study was included. Note that Kang *et al.* (2011) is also identified by the Fluoride Action Network as a null study, but their website does not include an English translation of the study. The other 15 relevant studies contained results that would likely add to the body of evidence showing a negative association between fluoride exposure and primary neurological outcomes. Because this body of evidence is already so large, and because time was a factor in the revision of the monograph, these studies were not translated or included as this information would likely not materially advance the human findings.

[6] Howard *et al.* (2020) evaluated the performance of the SWIFT-Active Screener methods for estimating total number of relevant studies using 26 diverse systematic review datasets that were previously screened manually by reviewers. The authors found that on average, 95% of the relevant articles were identified after screening 40% of the total reference list when using SWIFT-Active Screener. In the document sets with 5,000 or more references, 95% of the relevant articles were identified after screening 34% of the available references, on average, using SWIFT-Active Screener.

Figure 4. Study Selection Diagram



* Studies may have been excluded for more than one reason; the first reason identified by the screener was recorded.

** Animal in vivo, human secondary outcome-only, and human and animal mechanistic references from the 2020 database searches were scanned for evidence that might strengthen the information in the September 6, 2019 draft monograph. Although 145 additional animal in vivo and/or mechanistic studies and 7 additional human secondary outcome-only or mechanistic-only studies were identified, information that would materially advance the human, animal in vivo, and mechanistic findings was not identified; therefore, these studies were not included. Additionally, 15 human primary outcome studies from the 2020 Chinese database search were excluded based on English abstracts and google translations because information that would materially advance the human findings was not identified; 1 null publication from the 2020 Chinese database search (Kang *et al.* 2011) was identified, translated, and included.

*** One publication contained human, experimental non-human mammal, and in vitro data. Three publications contained both human and experimental non-human mammal data. Fourteen publications contained data relevant to both experimental non-human mammal studies and in vitro studies.

Neurodevelopmental and Cognitive Health Effects Results

All the neurodevelopmental and cognitive data were initially considered and evaluated, with more in-depth analysis where similar endpoints were evaluated across multiple studies (e.g., IQ). Hazard conclusions were developed separately for two different age groups (i.e., children and adults) to address potential differences in the health impact based on exposure during development compared to adulthood. Although the data cover a wide array of endpoints (see **Figure 5**), the hazard conclusion covers a single category for each age group. The largest bodies of evidence were for IQ (n = 71 studies), learning and memory (n = 8 studies), as well as other cognitive development effects (e.g., total

neurobehavioral scores and total mental capacity index in children and cognitive impairment in adults; n = 14 studies)[7]. Due to heterogeneity in the endpoints examined and the limited number of human or animal studies, congenital neurological malformations and neurological complications of fluorosis were not evaluated because the body of evidence was inadequate to evaluate these potential effects. These health outcomes are not further discussed in this assessment. To the extent possible, human and animal data were grouped into similar categories (e.g., IQ in humans was considered comparable to learning and memory in animals). NTP had previously assessed animal data related to effects on learning and memory associated with fluoride exposure (NTP 2016). Therefore, to update the conclusions of the NTP (2016) systematic review, only more recent animal studies were evaluated in this assessment. Although the previous NTP (2016) report was conducted through January 14, 2016, the current assessment included studies published from 2015 through April 2019 and considered studies from the NTP (2016) report. Thirty-five animal studies have been identified that met these criteria, including 23 studies with learning and memory endpoints and 12 studies with only motor and sensory endpoints. Consistent with the NTP (2016) assessment, only learning and memory studies have been considered in the development of hazard identification conclusions. The additional motor and sensory studies have been considered, along with information on motor and sensory effects reported in the learning and memory studies, to provide evidence of possible indirectness related to the learning and memory assessments.

Risk-of-bias Considerations

Risk-of-bias ratings for each individual study for all risk-of-bias questions are available in **Appendix 3**. The risk of bias of individual studies in the body of evidence was considered in developing confidence ratings. The key risk-of-bias questions (i.e., confounding, exposure characterization, and outcome assessment for human studies and randomization, exposure characterization, and outcome assessment for experimental animal studies) are discussed in the consideration of the body of evidence. For this assessment, the key risk-of-bias questions, if not addressed appropriately, are considered to potentially have the greatest impact on the results. In addition, for developmental studies in animals, controlling for potential litter effects (i.e., adjusting for similarities in responses between littermates) was also a key risk-of-bias concern. The other risk-of-bias questions were also taken into consideration and were used to identify any other risk-of-bias concerns that may indicate serious issues with the studies. No study was excluded based on concerns for risk of bias; however, confidence conclusions were considered with and without higher risk-of-bias studies (i.e., studies rated probably high risk of bias for at least two key risk-of-bias questions or definitely high for any single question) to assess the impact of the higher risk-of-bias studies. The remaining studies (i.e., other than the higher risk-of-bias studies) were considered lower risk of bias. Based on NASEM recommendations (NASEM 2020), **Appendix 4** was created for the lower risk-of-bias studies to describe strengths and limitations of the studies identified during the assessment and to clarify why they are considered to pose lower risk of bias.

Human Neurodevelopmental and Cognitive Data

While there were several neurodevelopmental and cognitive endpoints assessed (see **Figure 5**), most of the available studies evaluated intelligence (e.g., IQ) in children. Other measures of neurodevelopment or cognitive function in children were also assessed, including general cognitive index (GCI), mental capacity, mental development index (MDI), or neonatal behavioral neurological assessment (NBNA). However, because the majority of studies evaluated intelligence, the discussion focuses primarily on IQ in children with separate discussions on other measures of cognitive function and neurobehavioral

[7]Some studies are included in more than endpoint category (e.g., IQ and other cognitive developmental effects); therefore, these counts are not mutually exclusive.

effects in children and cognitive effects in adults. The available body of literature that evaluates the association between fluoride exposure and neurodevelopmental and cognitive effects is relatively robust (n = 92) and confidence considerations in the body of evidence and hazard conclusions are focused on the studies with the least potential for bias (n = 31). Studies with higher potential for bias (n = 61) have also been evaluated and determined to have little impact on the confidence and hazard conclusions. A subgroup analysis within the meta-analysis (described in detail in **Appendix 5**) demonstrated that results were robust to the exclusion of higher risk-of-bias studies (see **Appendix 5**, **Figure A5-6**). All evaluated studies can be found in **Appendix 2**.

This section is organized to present and explain NTP's two confidence ratings in the bodies of evidence from epidemiological studies that fluoride exposure is associated with cognitive neurodevelopmental effects in children and cognitive effects in adults. These confidence ratings were determined as described in **Figure 1**.

Summary: There is moderate confidence in the body of evidence that fluoride exposure is associated with cognitive neurodevelopmental effects in children, and low confidence in the body of evidence that fluoride exposure is associated with cognitive effects in adults. The moderate confidence rating is supported by consistent evidence from the available studies of an association between high-fluoride exposure (mainly greater than the WHO Drinking Water Quality Guideline [>1.5 mg/L] (WHO 2011), but also high exposure via fluoridated salt and food) and lower IQ or cognitive function in children. There is also a recent study of lower IQ in children living in areas where drinking water fluoride concentrations are <1.5 mg/L. Specifically, a study conducted in Canada observed significantly lower IQ scores in boys and girls associated with higher estimated total maternal consumption of fluoride during pregnancy from drinking water and other water-based beverages including black and green tea. When looking at maternal urinary fluoride concentrations, the significant negative association with IQ scores was seen in in boys but not girls (Green *et al.* 2019). Another study conducted in Mexico with similar maternal urinary fluoride concentrations during pregnancy as seen in Green *et al.* (2019) observed significantly lower IQ scores in boys and girls associated with higher perinatal exposure to fluoride (Bashash *et al.* 2017). Although the body of evidence in children supports lower IQ with increased fluoride exposure, there is a lack of evidence of an association between exposure to fluoride and cognitive effects in adults (Jacqmin *et al.* 1994, Li *et al.* 2016). The body of evidence available to examine the association between exposure to fluoride and cognitive effects in adults is limited to two lower risk-of-bias cross-sectional studies; due to the limited number of studies and a lack of an observed effect, this body of evidence is considered inadequate to evaluate whether fluoride exposure is associated with cognitive effects in adults (see **Table 7**).

Most of the available epidemiological studies that evaluated the association between fluoride exposure and cognitive neurodevelopmental effects assessed IQ and other measures of cognitive function in children (see **Figure 5**). Confidence conclusions are based on those studies with the lowest potential for bias (n = 28; 26 in children and 2 in adults) (see **Table 6**). Most of these studies measured fluoride levels in drinking water or urine. All but two of the studies were conducted in infants or children. The two studies in adults were conducted in older adult populations (≥60 years old; one in France and the other in a fluorosis-endemic area of China) to evaluate the effects of fluoride on cognitive impairment.

The studies in children were conducted in multiple populations. Of the 26 studies in children:

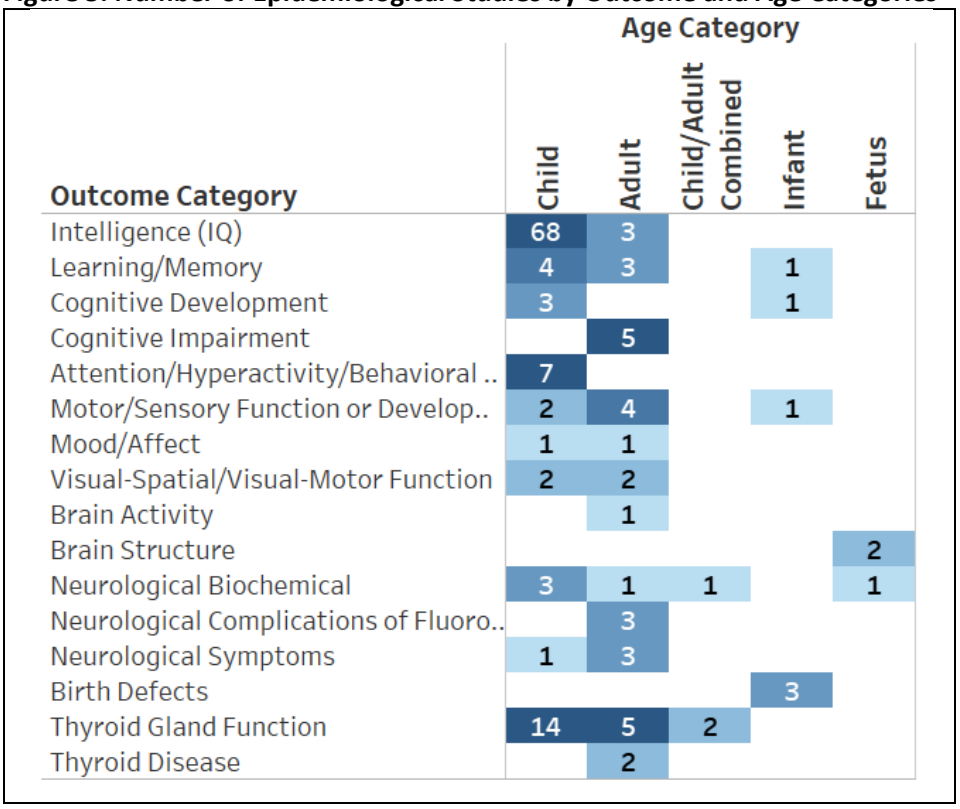
- 12 were conducted in 6 areas of China based on 8 study populations (1 study with both IQ and other neurodevelopmental outcomes, 9 studies with IQ only, and 2 studies with other neurodevelopmental outcomes);
- 6 were conducted in 4 areas of Mexico based on 5 study populations (1 study with both IQ and other neurodevelopmental outcomes, 2 studies with IQ only, and 3 studies with other neurodevelopmental outcomes);
- 4 were conducted in Canada using 2 separate cohorts (2 studies with IQ only and 2 studies with other neurodevelopmental outcomes);
- 3 were conducted in 3 areas of India (all IQ studies); and
- 1 was conducted in Iran (IQ study).

The IQ studies used many different tests to measure IQ. The IQ tests used often differed by population as not all IQ tests are appropriate for all populations (e.g., western vs. Asian populations). In some cases, different IQ tests were used to study similar populations. Overall, studies used IQ or cognitive tests appropriate for the population and were age appropriate. Other neurodevelopmental outcomes assessed in some studies included neurobehavioral effects (in infants), learning and memory impairment, and learning disabilities such as attention deficit hyperactivity disorder (ADHD). The different tests conducted and the populations on which the tests were conducted are indicated in **Table 6**.

The lower risk-of-bias studies (i.e., studies not meeting criteria for higher risk of bias) showing associations with cognitive neurodevelopmental effects in children include 5 prospective cohort studies from 3 study populations (Bashash *et al.* 2017, Valdez Jimenez *et al.* 2017, Green *et al.* 2019, Bashash *et al.* 2018, Till *et al.* 2020) and 21 cross-sectional studies from 16 study populations (Li *et al.* 2004 [translated in Li *et al.* 2008a], Choi *et al.* 2015, Rocha-Amador *et al.* 2007, Rocha-Amador *et al.* 2009, Saxena *et al.* 2012, Seraj *et al.* 2012, Xiang *et al.* 2003a, Xiang *et al.* 2011, Zhang *et al.* 2015b, Ding *et al.* 2011, Barberio *et al.* 2017b, Yu *et al.* 2018, Cui *et al.* 2018, Cui *et al.* 2020, Wang *et al.* 2020b, Wang *et al.* 2020a, Wang *et al.* 2012, Soto-Barreras *et al.* 2019, Sudhir *et al.* 2009, Trivedi *et al.* 2012, Riddell *et al.* 2019) (see **Figure D1** through **Figure D12**). One limitation of the 21 cross-sectional studies was the lack of direct evidence that exposure to fluoride occurred prior to the development of the neurodevelopmental outcomes. However, several studies from different study populations (n = 5) indicated that a large portion of the exposed children had dental fluorosis (ranging from 43–100%) at the time of the assessment (Choi *et al.* 2015, Ding *et al.* 2011, Seraj *et al.* 2012, Yu *et al.* 2018, Sudhir *et al.* 2009). Because dental fluorosis occurs when fluoride is consumed during enamel formation usually during the first 6–8 years of life, the presence of dental fluorosis suggests that exposures to fluoride occurred prior to the outcome assessment. Ten studies from seven study populations (including Yu *et al.* (2018), Wang *et al.* (2012) listed above) excluded subjects that had not lived in the study area for a specified period of time, sometimes since birth (Rocha-Amador *et al.* 2007, Rocha-Amador *et al.* 2009, Saxena *et al.* 2012, Yu *et al.* 2018, Wang *et al.* 2020b, Wang *et al.* 2012, Xiang *et al.* 2011, Xiang *et al.* 2003a, Soto-Barreras *et al.* 2019, Sudhir *et al.* 2009). Another study evaluated fluoride exposure in mothers and included urine levels just prior to birth and assessed children a few days after birth (Li *et al.* 2004 [translated in Li *et al.* 2008a]). Because these areas were generally known to be fluoride-endemic areas for long periods of time, it can generally be assumed that in these 14 cross-sectional studies from 11 study populations, exposure occurred prior to the outcome. These exposure concerns were not an issue for the prospective studies because fluoride levels were measured prenatally. Therefore, the moderate confidence in the body of evidence in children is primarily based on the consistency of findings across different populations in the 5 lower risk-of-bias prospective cohort studies and the 14

cross-sectional studies where exposure is considered to have occurred prior to the outcome with initial and final ratings of moderate confidence.

Figure 5. Number of Epidemiological Studies by Outcome and Age Categories*



*Interactive figure and additional study details in (https://public.tableau.com/profile/ntp.visuals#!/vizhome/Fluoride_Epi_2020Update/Figure5)

Table 6. Studies on Neurodevelopmental and Cognitive Function in Humans[a,b]					
Study	**Study Design (Location/Subjects) [n]**	**Exposure Measures and Summary Statistics**	**Assessment Timing**	**Outcome and Methods**	**Neurological Outcome Summary**
Children-IQ Studies					
China					
Choi *et al.* (2015)	Cross-sectional Mianning County/1st grade children [51]	Drinking water GM: 2.20 mg/L Children's urine GM: 1.64 mg/L Severity of fluorosis (Dean Index)	Children (ages 6–8 years)	IQ: WISC-IV (square root block design and digit span)	Compared to normal/ questionable fluorosis, moderate/severe fluorosis significantly associated with lower total (adjusted β = −4.28; 95% CI: −8.22, −0.33) and backward digit span scores (adjusted β = −2.13; 95% CI: −4.24, −0.02); linear correlation between fluoride in urine (adjusted β = −1.67; 95% CI: −5.46, 2.12) and in drinking water (adjusted β = −1.39; 95% CI: −6.76, 3.98) with total digit span was observed but not significant; other outcomes not significantly associated with fluoride exposure Adjusted for child's age, child's gender, parity, illness before 3 years old, household income last year, and caretaker's age and education
Cui *et al.* (2018)	Cross-sectional Tianjin City (districts Jinghai and Dagang)/school children [323]	Children's urine Range (log-transformed): −1.2–2.2	Children (ages 7–12 years)	IQ: Combined Raven's Test for Rural China	Significant correlation between IQ score and urinary fluoride (adjusted β = −2.47) Adjusted for child age, mother's education, family member smoking, stress, and anger
Cui *et al.* (2020)	Cross-sectional Tianjin City (all districts) /school children [498]	Children's urine <1.6–≥2.5 mg/L	Children (ages 7–12 years)	IQ: Combined Raven's Test	No significant difference in IQ score in the three urinary fluoride exposure groups based on a one-way ANOVA <1.6 mg/L = 112.16 ± 11.50 1.6-2.5 mg/L = 112.05 ± 12.01 ≥2.5 mg/L = 110 ± 14.92 No statistical adjustment for confounders

Table 6. Studies on Neurodevelopmental and Cognitive Function in Humans[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurological Outcome Summary
Zhang *et al.* (2015b)	Cross-sectional Tianjin City (Jinnan District)/school children [180]	Drinking water Mean: 0.63 (control), 1.40 (endemic fluorosis) mg/L (SD not reported) Children's urine Mean (SD): 1.1 (0.67) (control), 2.4 (1.01) (endemic fluorosis) mg/L Children's serum Mean (SD): 0.06 (0.03) (control), 0.18 (0.11) (endemic fluorosis) mg/L	Children (ages 10–12 years)	IQ: Combined Raven's Test for Rural China	Significant correlation between IQ score and serum fluoride (r = −0.47) and urinary fluoride (r = −0.45); significant difference in IQ score for high-fluoride area (>1 ppm; 102.33 ± 13.46) compared with control area (109.42 ± 13.30) Adjusted for child's age and gender, if applicable
Yu *et al.* (2018)*	Cross-sectional Tianjin City (7 towns) /children [2,886]	Drinking water Mean (SD): 0.50 (0.27) (normal), 2.00 (0.75) (high) mg/L Children's urine Mean (SD): 0.41 (0.49) (normal), 1.37 (1.08) (high) mg/L	Children (ages 7–13 years)	IQ: Combined Raven's Test for Rural China	Significant difference in mean IQ scores in high water fluoride area (>1.0 mg/L; 106.4 ± 12.3 IQ) compared to the normal area (≤1.0 ppm; 107.4 ± 13.0 IQ); distribution of the IQ scores also significantly different (p = 0.003); every 0.5-mg/L increase in water fluoride was associated with a 4.29 lower IQ score (95% CI: −8.09, −0.48) between 3.40 and 3.90 mg/L Adjusted for child's age, child's gender, maternal education, paternal education, and low birth weight
Wang *et al.* (2020b)	Cross-sectional Tianjin City (villages not specified)/school children [571]	Drinking water Mean (SD): 1.39 (1.01) mg/L Children's urine Mean (SD): 1.28 (1.30) mg/L	Children (ages 7–13 years)	IQ: Combined Raven's Test for Rural China	Significant associations between IQ and water and urine fluoride concentrations in boys and girls combined based on both quartiles and continuous measures (water: −1.587 per 1-mg/L increase; urine: −1.214 per 1-mg/L increase); there was no significant modification effect of gender Adjusted for child's age, child's gender, BMI, maternal education, paternal education, household income, and low birth weight

Table 6. Studies on Neurodevelopmental and Cognitive Function in Humans[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurological Outcome Summary
Ding *et al.* (2011)	Cross-sectional Inner Mongolia (Hulunbuir City)/ elementary school children [331]	Drinking water Mean (SD): 1.31 (1.05) mg/L Children's urine Range: 0.1–3.55 mg/L	Children (ages 7–14 years)	IQ: Combined Raven's Test for Rural China	Significant association between urinary fluoride and IQ score (each 1-mg/L increase was associated with a lower IQ score of 0.59 points (95% CI: −1.09, −0.08); dose response relationship between fluoride and dental fluorosis ($p < 0.0001$) Adjusted for child's age
Xiang *et al.* (2003a)	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children [512]	Drinking water Mean (SD): 0.36 (0.15) (control), 2.47 (0.79) (high fluoride) mg/L Children's urine Mean (SD): 1.11 (0.39) (control), 3.47 (1.95) (high fluoride) mg/L Village of residence (non-endemic v. endemic fluorosis)	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significant dose-related effect of fluoride on IQ score based on quintile levels with significantly lower IQ scores observed at water fluoride levels 1.53 mg/L or higher; Pearson correlation coefficient of -0.164 with urinary fluoride; IQ scores for children in non-endemic region (100.41 ± 13.21) significantly higher than endemic region (92.02 ± 13.00); calculated a lower-bound confidence limit benchmark concentration (BMCL) of 1.85 mg/L
Xiang *et al.* (2011)	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children (same population as Xiang *et al.* (2003a)) [512]	Children's serum Mean (SD): 0.041 (0.009) (control), 0.081 (0.019) (high fluoride) mg/L	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significant trend on association between quartiles of serum fluoride and children's IQ score < 80 (adjusted ORs for Q1 and Q2; Q1 and Q3; and Q1 and Q4, respectively: 1; 2.22 [95% CI: 1.42, 3.47]; and 2.48 [95% CI: 1.85, 3.32]); significant effects at ≥0.05 ppm fluoride Adjusted for child's age and gender
Wang *et al.* (2012)	Cross-sectional Wamiao and Xinhuai villages (Sihong County)/school children (same population as Xiang *et al.* (2003a)) [526]	Drinking water Mean (SD): 0.36 (0.11) (control), 2.45 (0.80) (high fluoride) mg/L Children's total fluoride intake Mean (SD): 0.78 (0.13) (control), 3.05 (0.99) (high fluoride) mg/day Village of residence (non-endemic v. endemic fluorosis)	Children (ages 8–13 years)	IQ: Combined Raven's Test for Rural China	Significantly lower mean IQ in the high fluoride village (92.02 ± 13.00) compared to the control village (100.41 ± 13.21); when high exposure group was broken into 4 exposure groups, a dose-dependent decreasing IQ and increase in % with low IQ observed; significant correlation between total fluoride intake and IQ ($r = −0.332$); OR for IQ<80 per increase in total fluoride intake = 1.106; 95% CI 1.052–1.163). Adjusted for child's age and gender

Table 6. Studies on Neurodevelopmental and Cognitive Function in Humans[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurological Outcome Summary
Mexico					
Rocha-Amador *et al.* (2007)	Cross-sectional Moctezuma and Salitral in San Luis Potosi State and 5 de Febrero of Durango State /elementary school children [132]	Drinking water Mean (SD): 0.8 (1.4), 5.3 (0.9), 9.4 (0.9) mg/L (3 rural areas) Children's urine Mean (SD): 1.8 (1.5), 6.0 (1.6), 5.5 (3.3) mg/L (3 rural areas)	Children (ages 6–10 years)	IQ: WISC-Revised Mexican Version	Significant associations between fluoride and IQ scores (full IQ adjusted βs of −10.2 with water and −16.9 with urine; CIs not reported); arsenic also present, but the effect was smaller (full IQ adjusted βs of −6.15 with water and −5.72 with urine; CIs not reported) Adjusted for blood lead, mother's education, SES, height-for-age z-scores, and transferrin saturation
Bashash *et al.* (2017)	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [299] IQ analysis [211]	Maternal urine during pregnancy Mean (SD): 0.90 (0.35) mg/L Children's urine Mean (SD): 0.82 (0.38) mg/L	Children (ages 6–12 years)	IQ: WASI-Spanish Version	Significant effect between maternal urinary fluoride and offspring IQ score (adjusted β = −2.50; 95% CI: −4.12, −0.59); associations with children's urine not significant Adjusted for gestational age, weight at birth, child's gender, parity (being the first child), age at outcome measurement, and maternal characteristics including smoking history (ever smoked during the pregnancy vs. nonsmoker), marital status (married vs not married), age at delivery, education, IQ, and cohort

Table 6. Studies on Neurodevelopmental and Cognitive Function in Humans[a,b]					
Study	**Study Design (Location/Subjects) [n]**	**Exposure Measures and Summary Statistics**	**Assessment Timing**	**Outcome and Methods**	**Neurological Outcome Summary**
Soto-Barreras *et al.* (2019)	Cross-sectional Chihuahua/school children [161]	Drinking water Range: 0.05–2.93 mg/L Children's urine Range: 0.11–2.10 mg/L	Children (ages 9–10 years)	IQ: Raven's Colored Progressive Matrices	No significant differences in fluoride exposure level (urine fluoride [p = 0.559], exposure dose [p = 0.389], or fluorosis index [p = 0.851]) between the different IQ grades No statistical adjustment for confounders
Canada					
Green *et al.* (2019)	Cohort (prospective) 10 cities/Maternal-Infant Research on Environmental Chemicals (MIREC) [512] Non-Fluoridated [238] Fluoridated [162] Boys [248] Girls [264]	Maternal urine during pregnancy Mean (SD): 0.51 (0.36) mg/L (0.40 [0.27] mg/L in non-fluoridated areas and 0.69 [0.42] mg/L in fluoridated areas) Maternal fluoride intake during pregnancy Mean (SD): 0.54 (0.44) mg/day (0.30 [0.26] and 0.93 [0.43] mg/day, respectively) Drinking water Mean (SD): 0.31 (0.23) mg/L (0.13 [0.06] and 0.59 [0.08] mg/L, respectively)	Children (ages 3−4 years)	IQ: full scale, performance, and verbal using Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	Significantly lower full-scale IQ (adjusted β = −4.49; 95% CI: −8.38, −0.60) and performance IQ (adjusted β = −4.63; 95% CI: −9.01, −0.25) per 1-mg/L increase in maternal urine in boys, but not girls (adjusted β = 2.40; 95% CI: −2.53, 7.33 and adjusted β = 4.51; 95% CI: −1.02, 10.05, respectively); significantly lower full-scale IQ (adjusted β = −3.66; 95% CI: −7.16, −0.15) per 1-mg increase in maternal fluoride intake (no sex interaction); significantly lower full-scale IQ (adjusted β = −5.29; 95% CI: −10.39, −0.19) per 1-mg/L increase in water fluoride concentration (no sex interaction); no significant changes observed in verbal IQ Adjusted for city, HOME score, maternal education, race, child's gender, and prenatal secondhand smoke exposure

Table 6. Studies on Neurodevelopmental and Cognitive Function in Humans[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurological Outcome Summary
Till *et al.* (2020)	Cohort (prospective) 10 cities/ MIREC [398] Non-Fluoridated [247] Fluoridated [151] Breastfed as infants [200] Formula-fed as infants [198]	Maternal urine during pregnancy Mean (SD) breastfed: 0.42 (0.28) mg/L in non-fluoridated areas and 0.70 (0.39) mg/L in fluoridated areas formula-fed: 0.38 (0.27) mg/L in non-fluoridated areas and 0.64 (0.37) mg/L in fluoridated areas Infant fluoride intake Mean (SD) breastfed: 0.02 (0.02) mg/day in non-fluoridated areas and 0.12 (0.07) mg/day in fluoridated areas formula fed: 0.08 (0.04) mg/day in non-fluoridated areas and 0.34 (0.12) mg/day in fluoridated areas Drinking water Mean (SD) breastfed: 0.13 (0.06) mg/L in non-fluoridated areas and 0.58 (0.08) mg/L in fluoridated areas formula fed: 0.13 (0.05) mg/day in non-fluoridated areas and 0.59 (0.07) mg/L in fluoridated areas	Children (ages 3–4 years)	IQ: full-scale, performance, and verbal using Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)	Significantly lower performance IQ with water fluoride (−9.26 formula-fed, −6.19 breastfed) and fluoride intake from formula (−8.76); significantly lower full scale IQ with water fluoride in formula-fed (−4.40); lower full-scale IQ for water fluoride in breastfed (−1.34) and fluoride intake from formula (−2.69) were not significant; no significant changes in verbal IQ scores with fluoride exposure Adjusted for maternal education, maternal race, child's age at IQ testing, child's sex, HOME total score, and second-hand smoke status in the child's house (separate analysis also adjusted for mother's urinary fluoride)
India					
Sudhir *et al.* (2009)	Cross-sectional Nalgonda District (Andhra Pradesh)/school children [1,000]	Drinking water Level 1: <0.7 ppm Level 2: 0.7–1.2 ppm Level 3: 1.3–4.0 ppm Level 4: >4.0 ppm	Children (ages 13–15 years)	IQ: Raven's Standard Progressive Matrices	Significant increase in mean and distributions of IQ grades (i.e., increase in intellectually impaired children) with increasing drinking water fluoride levels No statistical adjustment for confounders

Table 6. Studies on Neurodevelopmental and Cognitive Function in Humans[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurological Outcome Summary
Saxena *et al.* (2012)	Cross-sectional Madhya Pradesh/school children [170]	Drinking water ≥1.5 mg/L (high fluoride group) Children's urine Range: 1.7–8.4 mg/L	Children (age 12 years)	IQ: Raven's Standard Progressive Matrices	Significant correlation between water (r = 0.534; p = 0.000) and urinary fluoride (r = 0.542; p = 0.000) levels and IQ score; no significant differences in the levels of urinary lead or arsenic in children from the different groups Confounders included in the analysis were not reported
Trivedi *et al.* (2012)	Cross-sectional Kachchh, Gujarat/school children (6th and 7th grades) [84]	Drinking water Mean (SE): 0.84 (0.38) (low), 2.3 (0.87) (high) Children's urine Mean (SE): 0.42 (0.23) (low), 2.69 (0.92) (high)	Children (age 12–13 years)	IQ: questionnaire prepared by Professor JH Shah (97% reliability rating)	Significantly lower IQ score in the high fluoride (92.53 ± 3.13) compared to the low fluoride (97.17 ± 2.54) areas in boys and girls combined (as well as separately) No statistical adjustment for confounders
Iran					
Seraj *et al.* (2012)	Cross-sectional Makoo/school children [293]	Drinking water Mean (SD): 0.8 (0.3) (normal), 3.1 (0.9) (medium), 5.2 (1.1) (high) mg/L	Children (ages 6–11 years)	IQ: Raven's Colored Progressive Matrices	Significant correlation between water fluoride and IQ score (adjusted β = −3.865; CIs not reported); significantly higher IQ score in normal area (97.77 ± 18.91) compared with medium (89.03 ± 12.99) and high (88.58 ± 16.01) areas Adjusted for child's age, child's gender, child's education level, mother's education level, father's education level, and fluorosis intensity
Children-Other Neurodevelopmental Studies					
China					
Choi *et al.* (2015)	Cross-sectional Mianning County/1st grade children [51]	Drinking water GM: 2.20 mg/L Children's urine GM: 1.64 mg/L Severity of fluorosis (Dean Index)	Children (ages 6–8 years)	Learning and memory: Neuropsychological tests including WRAML Visual motor ability: WRAVMA Motor ability: Finger tapping task Manual dexterity: Grooved pegboard test	Outcomes unrelated to the IQ test not significantly associated with fluoride exposure Adjusted for child's age, child's gender, parity, illness before 3 years old, household income last year, and caretaker's age and education

Table 6. Studies on Neurodevelopmental and Cognitive Function in Humans[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurological Outcome Summary
Li *et al.* (2004) [translated in Li *et al.* 2008a]	Cross-sectional Zhaozhou County, Heilongjiang Province/neonates [91]	Drinking water Range: 0.5–1.0 mg/L (control); 1.7–6.0 mg/L (high) Maternal urine during pregnancy Mean (SD): 1.74 (0.96) mg/L (control); 3.58 (1.47) mg/L (high)	Neonates (24–72 hours after delivery)	Neurodevelopmental: Neonatal behavioral neurological assessment (NBNA)	Significant differences in neurobehavioral assessment total scores between high-fluoride (36.48 ± 1.09) and control groups (38.28 ± 1.10); significant differences in total score of behavioral capability that includes measures of non-biological visual orientation reaction and biological visual and auditory orientation reaction between the two groups (11.34 ± 0.56 in controls compared to 10.05 ± 0.94 in high-fluoride group) No statistical adjustment for confounders
Wang *et al.* (2020a)	Cross-sectional Tongxu County/school children [325]	Children's urine Mean (SD): 1.54 (0.89) mg/L	Children (ages 7–13 years)	ADHD and behavior measures: Conners' Parent Rating Scale-Revised (Chinese version) (CPRS-48)	Significant association between psychosomatic problems and urinary fluoride level (per 1-mg/L increase β=4.01; 95% CI 2.74, 5.28; OR for T-score >70=1.97; 95% CI 1.19, 3.27) Adjusted for child's age, child's gender, child's BMI, urinary creatinine, mother migrated and father migrated
Mexico					
Rocha-Amador *et al.* (2009)	Cross-sectional Durango/elementary school children [80]	Children's urine GM (SD): 5.6 (1.7) mg/L	Children (ages 6–11 years)	Visuospatial organization and visual memory: Rey-Osterrieth Complex Figure Test, children's version	Significant correlation between urinary fluoride and visuospatial organization (r = −0.29) and visual memory scores (r = −0.27); no significant correlation with arsenic Adjusted for age
Valdez Jimenez *et al.* (2017)	Cohort (Prospective) Durango City and Lagos de Moreno/infants [65]	Drinking water Range: 0.5–12.5 mg/L (all trimesters) Maternal urine Range: 0.16–8.2 mg/L (all trimesters)	Infants (ages 3–15 months)	Mental development index (MDI): Bayley Scales of Infant Development II (BSDI-II)	Significant correlation between maternal urinary fluoride and MDI score during first trimester (adjusted β = −19.05; SE = 8.9) and second trimester (adjusted β = −19.34; SE = 7.46) Adjusted for gestational age, child's age, marginality index, and type of drinking water

Table 6. Studies on Neurodevelopmental and Cognitive Function in Humans[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurological Outcome Summary
Bashash *et al.* (2017)	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [299] GCI analysis [287]	Maternal urine during pregnancy Mean (SD): 0.90 (0.35) mg/L Children's urine Mean (SD): 0.82 (0.38) mg/L	Children (age 4 years)	General cognitive index (GCI): McCarthy Scales of Children's Abilities (MSCA)	Significant effect between maternal urinary fluoride and offspring GCI score (adjusted β = −3.15; 95% CI: −5.42, −0.87); associations with children's urine not significant Adjusted for gestational age, weight at birth, child's gender, parity (being the first child), age at outcome measurement, and maternal characteristics including smoking history (ever smoked during the pregnancy vs. nonsmoker), marital status (married vs not married), age at delivery, IQ, education, and cohort
Bashash *et al.* (2018)	Cohort (prospective) Mexico City/Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants [210]	Maternal urine during pregnancy Mean 0.85 (95% CI: 0.81, 0.90) mg/L	Children (ages 6–12 years)	ADHD: Conners' Rating Scales-Revised (CRS-R)	Significant associations between maternal urinary fluoride and CRS-R scores including Cognitive Problems + Inattention Index (adjusted β = 2.54; 95% CI: 0.44, 4.63), DSM-IV Inattention Index (adjusted β = 2.84; 95% CI: 0.84, 4.84), DSM-IV ADHD Total Index (adjusted β = 2.38; 95% CI: 0.42, 4.34), and ADHD Index (adjusted β = 2.47; 95% CI: 0.43, 4.50) Adjusted for gestational age, birth weight, child's gender, parity, age at outcome measurement, and maternal characteristics including smoking history (ever smoked vs. nonsmoker), marital status (married vs. not married), education, socioeconomic status, and cohort

Table 6. Studies on Neurodevelopmental and Cognitive Function in Humans[a,b]					
Study	**Study Design (Location/Subjects) [n]**	**Exposure Measures and Summary Statistics**	**Assessment Timing**	**Outcome and Methods**	**Neurological Outcome Summary**
Canada					
Barberio *et al.* (2017b)	Cross-sectional General population/ Canadian Health Measures Survey (Cycles 2 and 3) [2,221]	Children's urine Mean Cycle 2: 32.06 (95% CI: 29.65, 34.46) µmol/L Mean Cycle 3: 26.17 (95% CI: 22.57, 29.76) µmol/L	Children (ages 3–12 years)	Learning disability, ADHD (Cycle 2 only): Parent or child self-report	Significant increase in adjusted OR for learning disability (adjusted OR = 1.02; 95% CI: 1.00, 1.03) only when Cycle 2 and 3 were combined using unadjusted urinary fluoride (associations no longer significant once adjusted for creatinine and specific gravity); no significant associations found between urinary fluoride and ADHD (only evaluated in Cycle 2) Adjusted for child's age, child's gender, household income adequacy, and highest attained education in the household
Riddell *et al.* (2019)	Cross-sectional General population/ Canadian Health Measures Survey (Cycles 2 and 3) [3,745]	Children's urine Mean (SD): 0.61 (0.39) mg/L; non-fluoridated water-0.46 (0.32) mg/L, fluoridated water-0.82 (0.54) Drinking water Mean (SD): 0.23 (0.24) mg/L; non-fluoridated water-0.04 (0.06) mg/L, fluoridated water-0.49 (0.22)	Children (ages 6–17 years)	Hyperactivity/inattention: Strengths and Difficulties Questionnaire (SDQ) ADHD: parent or self-reported physician diagnosis	Significantly increased risk of ADHD with fluoride in tap water (adjusted OR = 6.10 per 1-mg/L increase; 95% CI: 1.60, 22.8) or community water fluoridation status (1.21; 95% CI: 1.03, 1.42), but not with urinary fluoride; similar results observed with attention symptoms based on the SDQ scores Adjusted for child's age, child's gender, child's BMI, ethnicity, parental education, household income, blood lead, and smoking in the home

Table 6. Studies on Neurodevelopmental and Cognitive Function in Humans[a,b]

Study	Study Design (Location/Subjects) [n]	Exposure Measures and Summary Statistics	Assessment Timing	Outcome and Methods	Neurological Outcome Summary
Adult Studies					
Jacqmin *et al.* (1994)	Cross-sectional France (Gironde and Dordogne)/elderly adults [3,490]	Drinking water Range: 0.03–2.03 mg	Adults (ages ≥ 65 years)	Cognitive function: Mini-Mental State (MMS) Examination	No significant increase in the prevalence of cognitive impairment with increasing fluorine quartiles No statistical adjustment for confounders
Li *et al.* (2016)	Cross-sectional China (Inner Mongolia)/adults [511]	Drinking water intake and urinary fluoride Mean (SD) levels reported for a subset of subjects with normal scores (2.23 [2.23] mg and 1.46 [1.04] mg/L, respectively) and subjects with cognitive impairment (3.62 [6.71] mg and 2.47 [2.88] mg/L, respectively)	Adults (ages ≥ 60 years)	Cognitive function: MMS Examination	Results suggested that degree of fluoride exposure was consistent with severity of skeletal fluorosis, and fluoride exposure may be a risk factor for cognitive impairment; however, neither water fluoride intake (adjusted ORs = 0.94 [95% CI: 0.85, 1.04] and 0.86 [95% CI: 0.69, 1.06] in the moderate and severe cognitive impairment groups, respectively) nor urinary fluoride levels (adjusted ORs = 1.12 [95% CI: 0.89, 1.42] and 1.25 [95% CI: 0.87, 1.81] in the moderate and severe cognitive impairment groups, respectively) were significantly correlated with cognitive impairment Adjusted for sex, age, education, marital status (married vs. not married), alcohol consumption (non-drinkers, light drinkers, moderate to heavy drinkers), smoking history (never smoker, ex-smoker, light smoker, heavy smoker), and serum homocysteine levels

*Three additional publications based on subsample (i.e., 50–60 children) of the larger Yu *et al.* (2018) cohort were identified (Zhao *et al.* 2020, Zhou *et al.* 2019, Zhao *et al.* 2019); however, these publications focused on mechanistic considerations and are not included in the study totals for IQ because the main study by Yu *et al.* (2018) is considered a better representation of the IQ results.

[a]Includes lower risk-of-bias studies.

[b]Definitions: **ADHD**: attention-deficit/hyperactivity disorder; **GCI**: General Cognitive Index; **GM**: geometric mean; **HOME**: Home Observation Measurement of the Environment; **IQ**: intelligence quotient; **MSCA**: McCarthy Scales of Children's Abilities; **WASI**: Wechsler Abbreviated Scale of Intelligence (Spanish version); **WISC-IV**: Wechsler Intelligence Scale for Children-Revised; **WRAML**: Wide Range Assessment of Memory and Learning; **WRAVMA**: Wide Range Assessment of Visual Motor Ability.

Overall Risk-of-bias Discussion of the Body of Evidence

The confidence rating for the body of evidence in humans was based on studies with the lowest potential for bias (i.e., studies rated probably low or definitely low risk of bias for at least two of the three key risk-of-bias questions). Each of these 28 studies (including 26 studies in children and 2 in adults) had little or no risk-of-bias concerns, and confidence in the body of evidence was not

downgraded for risk of bias. However, the remaining studies in the human body of evidence were rated as probably high or definitely high risk of bias for at least two key risk-of-bias questions or had other major concerns. Risk-of-bias ratings for individual studies for all questions are available in **Figure A3-1** and **Figure A3-3**. Among the studies with lower potential for bias (see **Figure A3-1** and **Figure A3-2**), the key risk-of-bias question with the most potential for bias was the potential for confounding. Potential confounding was a concern for 5 of the 26 lower risk-of-bias studies in children (see *Confounding* for further discussion). Among the studies with higher overall potential for bias, there were a number of risk-of-bias concerns, including potential confounding, poor exposure characterization, poor outcome assessment, and, in many cases, potential concern with participant selection (see **Figure A3-3** and **Figure A3-4**). Many of the studies (n = 32) included in the entire human body of evidence were initially published in a foreign language (mainly Chinese) and were either translated and published in volume 41 of the journal *Fluoride* (n = 19) or were translated by the Fluoride Action Network (n = 13) (http://fluoridealert.org/researchers/translations/complete_archive/). Most of these studies were considered to have high potential for bias due to lack of information across many key risk-of-bias questions. Therefore, in order to assess if the lack of information relevant to key risk-of-bias concerns was the result of a loss in translation, the original Chinese publications and the translated versions of the five studies that had the most potential for being included in the lower risk-of-bias group of studies were reviewed to determine if any of the risk-of-bias concerns could be addressed (An *et al.* 1992, Chen *et al.* 1991 [translated in Chen *et al.* 2008], Du *et al.* 1992 [translated in Du *et al.* 2008], Guo *et al.* 1991 [translated in Guo *et al.* 2008a], Li *et al.* 2009). For all five studies, the translations were determined to be accurate and there was no impact on the key risk-of-bias concerns.

Confounding

Potential confounding variables and/or effect modifiers that were considered key for all studies, populations, and outcomes included child's age, child's sex, and socioeconomic status (e.g., maternal education, household income, marital status, crowding). Additional potential confounding variables and/or effect modifiers considered important for this evaluation depending on the study population and outcome included race/ethnicity; maternal demographics (e.g., maternal age, body mass index [BMI]); parental behavioral and mental health disorders (e.g., ADHD, depression); smoking (e.g., maternal smoking status, secondhand tobacco smoke exposure); reproductive factors (e.g., parity); nutrition (e.g., BMI, growth, anemia); iodine deficiency/excess; minerals and other chemicals in water associated with neurotoxicity (e.g., arsenic, lead); maternal and paternal IQ; and quantity and quality of caregiving environment (e.g., Home Observation Measurement of the Environment [HOME] score). To be assigned a rating of probably low risk of bias for the key risk-of-bias question regarding confounding, studies were not required to address every potential confounder listed; however, studies were required to address the three key covariates for all studies, the potential for co-exposures if applicable (e.g., arsenic and lead, both of which could affect cognitive function), and any other potential confounders considered important for the specific study population and outcome. For example, studies of populations in China, India, and Mexico, where there is concern for exposures to high fluoride and high arsenic, were required to address arsenic, and smoking needed to be addressed in studies of adults when dementia was evaluated. In order to identify areas of China, India, and Mexico where arsenic is a concern, groundwater quality maps were evaluated (https://www.gapmaps.org/Home/Public#) (Podgorski and Berg 2020). If no arsenic measurements were available for the area, the arsenic groundwater quality predictions from the global arsenic 2020 map were used (Podgorski and Berg 2020). If an area had less than 50% probability of having arsenic levels greater than 10 µg/L (the WHO guideline concentration), the area was considered not to have an issue with arsenic that needed to be addressed by the study authors.

Among studies with lower risk-of-bias concerns, 21 of the 26 studies were considered to have lower potential for bias due to confounding. Relative to confounders considered key for all studies and populations (i.e., age, sex, and SES), one study did not address age, one study did not address sex, and two studies did not account for indicators of SES (e.g., parental education, household income). Nine of the 26 lower risk-of-bias studies accounted for maternal or family member smoking. Potential confounding related to co-exposure to arsenic was not accounted for in five lower risk-of-bias studies and was the main potential concern in these studies; however, three of these studies (Xiang *et al.* 2011, Wang *et al.* 2012, Xiang *et al.* 2003a) were still considered low risk of bias for confounding due to the fact that arsenic was observed in the low fluoride areas (which would bias the effect toward the null), but an effect was still observed. The other two studies did not address arsenic and were in areas that had potential for arsenic exposure to occur (Soto-Barreras *et al.* 2019, Valdez Jimenez *et al.* 2017). Seven studies did not consider co-exposures to lead; however, for all of these studies this co-exposure was considered unlikely to have an impact in these study populations as there was no evidence that lead was prevalent or occurring in relation to fluoride.

Although there is variability in the potential confounders considered and differences in populations evaluated, the consistency of the results among the lower risk-of-bias studies indicates that confounding is not a major concern in this body of evidence. Even though 5 of 26 lower risk-of-bias studies in children are considered to have higher potential for bias due to confounding that could not be ruled out for that specific population and outcome (see **Figure 6**), results were consistent across multiple populations; all but two of the lower risk-of-bias studies in children reported an association between higher fluoride exposure and lower IQ or another cognitive effect. Seven of the lower risk-of-bias studies confirmed the robustness of the results by conducting sensitivity analyses (Bashash *et al.* 2018, Bashash *et al.* 2017, Green *et al.* 2019, Yu *et al.* 2018, Wang *et al.* 2020a, Wang *et al.* 2020b, Till *et al.* 2020). None of the sensitivity analyses adjusting for additional confounders found meaningful shifts in the association between fluoride exposure and IQ or other measures of cognitive function. Bashash *et al.* (2017) found that adjusting for HOME score increased the association between maternal urinary fluoride and children's IQ. Bashash *et al.* (2018) examined several potential confounders in sensitivity analyses involving subsets of participants, including HOME scores, child contemporaneous fluoride exposure measured by child urinary fluoride adjusted for specific gravity, and maternal lead and mercury exposures. The authors reported that no sensitivity analyses indicated appreciable changes in the fluoride-related association with behaviors related to ADHD, nor did they find evidence of effect modification between sex and maternal urinary fluoride. Green *et al.* (2019) found that adjusting for lead, mercury, manganese, perfluorooctanoic acid, and arsenic concentrations did not substantially alter the associations with IQ. Sensitivity analyses by Yu *et al.* (2018) that adjusted for covariates (including age, sex, and socioeconomic status) did not find differences in the results compared to the primary analyses. Both Wang *et al.* (2020a) and Wang *et al.* (2020b) found the results of the sensitivity analysis to be the same as the results from the preliminary analysis. Till *et al.* (2020) found that adjusting for maternal urinary fluoride levels had little effect on the results.

As previously mentioned, most of the higher risk-of-bias studies in the human body of evidence did not address the potential confounders of greatest concern. Many of these studies conducted only simple statistical analyses without accounting for any potential confounders (50 of 61 higher risk-of-bias studies), and many studies did not report whether the study subjects were from areas of similar socioeconomic status or environmental conditions (n = 20 higher risk-of-bias studies). Potential confounding related to important co-exposures (e.g., arsenic and lead) was often not addressed in higher risk-of-bias studies. In studies where there was high exposure to fluoride via drinking water with high naturally-occurring fluoride or from the use of coal-containing fluoride, most researchers did not

account for potential exposures to arsenic, which is commonly found in coal and drinking water in fluoride-endemic areas of China and Mexico. In general, researchers did not account for potential exposures to lead; however, studies reporting lead levels in fluoride-endemic areas, including areas in China, often reported low levels of lead (Xiang *et al.* 2011, Choi *et al.* 2012, Seraj *et al.* 2012, Choi *et al.* 2015, Yu *et al.* 2018, Saxena *et al.* 2012, Xiang *et al.* 2003b). Therefore, lead is not assumed to be a common exposure in fluoride-endemic areas. Most of the studies did not account for smoking or socioeconomic status, nor did they provide information to lessen the risk-of-bias concern (e.g., list of study characteristics indicating no significant differences between comparison groups). However, as noted for the lower risk-of-bias studies, given the consistency of the evidence, confounding among higher risk-of-bias studies is likely less of a concern for the body of evidence as a whole than for any individual study.

Figure 6. Potential Confounders Considered in Lower Risk-of-bias Studies Conducted in Children

Study (Location)[1]	Potential Confounding Factors Considered[2]																Notes	Reported Effect of Fluoride[4]
	Subject Characteristics				Other Exposures					Socioeconomic Factors		Parental Characteristics				Other[3]		
	Age	Sex	Race/Ethnicity	Health Factors[5]	Arsenic	Smoking	Iodine	Lead	Other[3]	SES	Caregiving Environment (e.g., HOME score)	Demographics[5]	Reproductive Factors[5]	Health Factors[5]	IQ			
Overall RoB Rating for Confounding: Probably Low																		
Barberio 2017b (Canada)	√	√	–	–	√	–	–	√	–	√	–	–	–	–	–	–	–	Yes
Bashash 2017 (Mexico)	√	√	–	–	√	√	–	√	√	√	√	√	√	–	√	√	Other exposures: Hg, Ca Demographics: maternal age Reproductive: parity, birth order, birth weight, gestational age at delivery Other: cohort	Yes
Bashash 2018 (Mexico)	√	√	–	–	√	√	–	√	√	√	√	√	√	–	–	√	Other exposures: Hg, Ca Demographics: maternal age Reproductive: parity, birth order, birth weight, gestational age at delivery Other: cohort	Yes
Choi 2015 (China)	√	√	–	√	√	–	–	√	–	√	–	√	√	√	–	√	Health: subject Fe deficiency, illnesses before age 3, medical history of subject and caretakers Demographics: parental age Reproductive: parity Other: residential history	Yes
Cui 2018 (China)	√	√	√	√	√	√	√	–	–	√	–	√	√	√	–	√	Health: subject BMI, stress/anger/anxiety/depression, psychological trauma, having a cold, in relatives: thyroid diseases, cancer, mental retardation Demographics: maternal age Reproductive: abnormal birth Other: alcohol consumption, proximity to factory, physical activity, various dietary factors, environmental noise	Yes
Green 2019 (Canada)	√	√	√	–	√	√	–	√	√	√	√	√	√	–	–	√	Other exposures: Hg, Mn, PFOA Demographics: parental age, pre-pregnancy BMI Reproductive: parity, weeks of gestation, birth weight, maternal chronic condition during pregnancy Other: alcohol consumption, birth country, voiding interval in urine sampling, breastfeeding duration	Yes
Li et al., 2004 (translated in Li 2008a) (China)	√	√	–	–	√	–	–	–	–	√	–	√	√	–	–	–	Demographics: living habits, cultural background Reproductive: gestational age, birth weight, birth method Other: nutritional conditions	Yes
Riddell 2019 (Canada)	√	√	√	√	√	√	–	√	–	√	–	–	–	–	–	–	Health: subject BMI	Yes
Rocha-Amador 2007 (Mexico)	√	√	–	√	√	–	–	√	–	√	–	–	–	–	–	–	Health: subject height and weight by age, transferrin saturation	Yes
Saxena 2012 (India)	√	√	–	√	√	–	√	√	–	√	–	–	–	–	–	√	Health: subject height for age ratio, weight for height ratio, medical history Other: residential history	Yes
Seraj 2012 (Iran)	√	√	–	–	√	–	√	√	–	√	–	–	–	–	–	√	Other: fluorosis intensity	Yes
Sudhir 2009 (India)	√	√	–	–	√	–	–	√	–	√	–	–	–	–	–	√	Other: staple food consumed, liquids routinely consumed, aids used for oral hygiene maintenance (fluoridated or nonfluoridated)	Yes
Till 2020 (Canada)	√	√	√	–	√	√	–	–	–	√	√	–	–	–	–	√	Other: city	Yes
Trivedi 2012b (India)	√	√	–	–	√	–	√	–	–	√	–	–	–	–	–	–		Yes
Wang 2020a (China)	√	√	–	√	√	–	–	√	√	√	–	√	–	–	–	–	Other exposures: cadmium and mercury Health: subject BMI Demographics: mother and father migration, living habits Other: diet, industrial pollution within 1 km of living environment	Yes
Wang 2020b (China)	√	√	–	√	√	√	√	√	–	√	–	–	√	–	–	√	Health: subject BMI, exclusions based on diseases affecting intellegence, history of trauma or neurological disorders, positive screening test history Reproductive: low birth weight Other: alcohol consumption	Yes
Wang 2012 (China)	√	√	–	√	–	–	√	√	–	√	–	–	–	√	–	√	Health: medical conditions Other: transportation, natural environment, and lifestyle	Yes
Xiang 2003 (China)	√	√	–	–	–	–	√	√	–	√	–	–	–	–	–	–		Yes
Xiang 2011 (China)	√	√	–	–	–	–	√	√	–	√	–	–	–	–	–	–		Yes
Yu 2018 (China)	√	√	–	√	√	√	√	√	√	√	–	–	√	–	–	√	Health: subject BMI Reproductive: disease history during pregnancy, delivery conditions Other: dental fluorosis prevalence, consanguineous marriage	Yes
Zhang 2015b (China)	√	√	–	√	√	–	√	√	√	√	–	–	–	–	–	√	Health: subject physical and mental health status Other: thyroid hormone levels, residential history, having knowledge of fluorosis, COMT genotype	Yes

Study (Location)[1]	Potential Confounding Factors Considered[2]																Notes	Reported Effect of Fluoride[4]
	Subject Characteristics				Other Exposures					Socioeconomic Factors		Parental Characteristics				Other[3]		
	Age	Sex	Race/Ethnicity	Health Factors[3]	Arsenic	Smoking	Iodine	Lead	Other[3]	SES	Caregiving Environment (e.g., HOME score)	Demographics[3]	Reproductive Factors[3]	Health Factors[3]	IQ			
Overall RoB Rating for Confounding: Probably High																		
Cui 2020 (China)	–	√	–	√	√	√	√	–	–	√	–	√	√	√	–	√	Health: stress/anger/anxiety, having a cold, in relatives: mental retardation Demographics: maternal age Reproductive: abnormal birth, smoking and drinking during pregnancy Other: thyroid horomone levels	No
Ding 2011 (China)	√	–	–	–	√	–	√	√	–	–	–	–	–	–	–	–	–	Yes
Rocha-Amador 2009 (Mexico)	√	√	–	√	√	–	–	√	–	–	–	–	–	–	–	–	Health: subject height and weight by age	Yes
Soto-Barreras 2019 (Mexico)	√	√	–	–	–	–	–	–	–	√	–	–	–	–	–	–		No
Valdez Jimenez 2017 (Mexico)	√	√	–	–	–	–	–	–	–	√	–	√	√	√	–	√	Demographics: maternal age Health: pre-pregnancy history of drugs, vaccines, diseases Reproductive: prenatal history, parity, type of birth, week of birth, weight and length at birth, gestational age, Apgar and health conditions of the baby during the first month of life Other: infant feeding type (breastfeeding, formula)	Yes

Notes:

[1]Includes all lower risk-of-bias studies in children. Studies are organized as those with an overall risk-of-bias rating for confounding as probably low (green) followed by those with an overall risk-of-bias rating for confounding as probably high (yellow).

[2]Potential confounding factors and/or effect modifiers represented here are those considered important for this evaluation. See study details provided in HAWC for information on additional confounders.

Factors outlined in blue (subject age, subject sex, arsenic, SES) are considered key confounders.

A √ indicates that a factor was considered (and may or may not have been adjusted for in final model). For 'Other Exposures', a √ might also be used when a co-exposure was not expected to be an issue because there is no evidence to indicate that the co-exposure was prevalent or occurring in relation to fluoride. See risk-of-bias explanations in HAWC for details. A hyphen (–) indicates that the factor was not considered.

[3]See the "Notes" column for additional details.

[4]Extent of reported effects varies by study. "Yes" indicates that study authors reported one or more significant effects on IQ or other cognitive functions associated with fluoride exposure.

Exposure assessment

Exposure was assessed using a variety of methods in the human body of evidence. Studies provided varying levels of details on the methods used and employed different exposure characterization methods to group study subjects into exposed and reference groups. Exposure metrics included spot urine (from children or mothers during at least one trimester), serum, individual drinking water, intake from infant formula, estimated total exposure dose, municipal drinking water (with residence information), area of residence (endemic versus a non-endemic fluorosis area with or without individual validation of exposure), burning coal (with or without fluoride), and occupation type. Analytical methods to measure fluoride in biological or water samples also varied, some of which included atomic absorption, ion selective electrode methods, colorimetric methods, or the hexamethyldisiloxane microdiffusion method. Individual-level measures of exposure were generally considered more accurate than group-level measures; however, using group-level measures (e.g., endemic versus non-endemic area) in an analysis was less of a concern if the study provided water or urine fluoride levels from some individuals to verify that there were differences in the fluoride exposure between groups. Studies that provided results by area but also reported individual urinary or serum fluoride concentrations or other biochemical measures, including dental fluorosis in the children or urinary levels in mothers during pregnancy, were considered to have probably low risk of bias.

In general, there were few or no risk-of-bias concerns regarding exposure assessment in the lower risk-of-bias studies. Many of the lower risk-of-bias studies used individual urine or water measures with appropriate analyses. Urinary fluoride levels include all ingested fluoride and are considered a valid measure to estimate fluoride exposure (Villa *et al.* 2010, Watanabe *et al.* 1995); however, some concerns exist. Urinary fluoride is thought to reflect recent exposure but can be influenced by the timing of exposure (e.g., when water was last consumed, when teeth were last brushed). When compared to 24-hour urine samples, spot urine samples are more prone to these influences and can also be affected by differences in dilution; however, many studies attempted to account for dilution either using urinary creatinine or specific gravity. Good correlations between 24-hour samples and urinary fluoride concentrations from spot samples adjusted for urinary dilution have been described (Zohouri *et al.* 2006). Despite potential issues with spot urine samples, if authors made appropriate efforts to reduce the concern for bias, studies that used this metric were generally considered to have probably low risk of bias for exposure.

Although there are concerns related to using maternal urine samples, many studies provide evidence to suggest that urinary fluoride is a reasonable measure of exposure. Using three methods to account for urine dilution, Till *et al.* (2018) reported that adjusted risk estimates did not differ from unadjusted estimates. Analyzing the same study population as Till *et al.* (2018), Green *et al.* (2019) found that adjusting for time of urine collection or time of collection since last void during pregnancy did not substantially affect associations with IQ results in either boys or girls. In addition, adjusting the maternal urinary fluoride for creatinine did not substantially alter the association observed (Green *et al.* 2019). To provide a more accurate and sensitive measurement of maternal urinary fluoride than a single measurement provides, Green *et al.* (2019) only included participants with valid fluoride measurements at each trimester in their analysis. Several other studies also measured urinary fluoride multiple times throughout pregnancy (Bashash *et al.* 2018, Bashash *et al.* 2017, Green *et al.* 2019, Valdez Jimenez *et al.* 2017). Other studies demonstrated correlations between the urinary fluoride and fluoride in the drinking water, fluorosis, or estimated dose based on water (Green *et al.* 2019, Saxena *et al.* 2012, Zhang *et al.* 2015b, Ding *et al.* 2011, Choi *et al.* 2015, Yu *et al.* 2018). Till *et al.* (2018) demonstrated that there was a linear association between urinary fluoride concentrations in pregnant women and drinking water fluoride concentrations regardless of method to correct for urine dilution or whether or not

adjustments were made for dilution. Bashash *et al.* (2017) excluded exposure outliers but found that doing so did not change the results in a meaningful way. Taken together, these studies suggest that urinary fluoride is a reasonable measure of exposure despite some of the potential issues.

A frequent critical limitation among the higher risk-of-bias studies was lack of information regarding exposure or poor exposure characterization. Many of the higher risk-of-bias studies only compared subjects living in two regions with differing levels of fluoride exposure, and while most of them did provide some differentiation in levels of fluoride between the areas, limited or no individual exposure information was reported. Among studies that provided drinking water levels of fluoride in two areas being compared, sufficient information to determine if the individual study subjects were exposed to these levels was often not reported. Some studies also lacked information on fluoride analysis methods and timing of the exposure measurements. In some cases (n = 3), study areas that were considered endemic for dental and/or skeletal fluorosis were compared to non-endemic areas, or high-fluoride areas were compared to low-fluoride areas, with no other information provided on fluoride levels in the areas (Sun *et al.* 1991, Li *et al.* 2003 [translated in Li *et al.* 2008c], Ren *et al.* 1989 [translated in Ren *et al.* 2008]). While living in an area endemic for fluorosis could be an indicator of exposure, these studies did not specify if the study subjects themselves had fluorosis. Another study used only dental fluorosis as a measure of fluoride exposure in subjects that were all from an endemic area with similar drinking water fluoride levels (Li *et al.* 2010).

Outcome assessment

Studies included in this evaluation used a wide variety of methods to measure IQ and other cognitive effects. Measures of IQ were generally standardized tests of IQ; however, for these standardized methods to be considered low potential for bias they needed to be conducted in the appropriate population or modified for the study population. Because results of these tests can be subjective, it was important that the outcome assessors were blind to the fluoride exposure when evaluating the results of the tests. If the study reported that the assessor was blind to the exposure, this was assumed to mean that the outcome assessor did not have any knowledge of the exposure, including whether the study subjects were from high-fluoride communities.

The lower risk-of-bias studies have few concerns regarding outcome assessment. Four studies (Barberio *et al.* 2017b, Riddell *et al.* 2019, Sudhir *et al.* 2009, Wang *et al.* 2020a) had concerns for potential bias in the outcome assessment and that was due to either the use of self-reported of outcomes or the lack of accounting for blinding at the time of the outcome assessment in cases where there was potential concern. The remainder of the studies used appropriate measures of IQ or other cognitive effects for the study population. Seventeen of the studies reported blinding of the outcome assessors or correspondence with the study authors indicated that it was not likely an issue. For the remainder of the studies, it was assumed that the outcome assessors were most likely blind because exposure was assessed via urine or drinking water obtained at the same time as the outcome assessment.

Among the studies with higher risk of bias, the main limitation in the outcome assessment was the lack of reporting on whether the outcome was assessed without knowledge of exposure. Although there is little concern that the children's knowledge of their own exposure would bias the way they took the IQ tests, there is potential for bias if the tests were administered by an interviewer, or if the scoring of results could be subjective (e.g., drawing tests), and the interviewer or scorer had knowledge of the children's exposure. Most of the studies did not provide sufficient information on the person scoring or administering the tests or other information on the assessment methods to alleviate concerns for potential interviewer or reviewer bias. In some cases, the outcomes were not considered sensitive

measures (e.g., Seguin Form Board Test to test for IQ), or the test was not considered appropriate for the study population (e.g., a test validated in a western population was used on a rural Chinese population).

IQ in Children

The results from 17 studies (3 prospective cohort and 14 cross-sectional studies from 13 different study populations) with lower potential for bias that evaluated IQ in children (Bashash *et al.* 2017, Choi *et al.* 2015, Ding *et al.* 2011, Rocha-Amador *et al.* 2007, Saxena *et al.* 2012, Seraj *et al.* 2012, Xiang *et al.* 2003a, Xiang *et al.* 2011, Zhang *et al.* 2015b, Yu *et al.* 2018, Green *et al.* 2019, Cui *et al.* 2018, Wang *et al.* 2020b, Wang *et al.* 2012, Sudhir *et al.* 2009, Till *et al.* 2020, Trivedi *et al.* 2012) provide consistent evidence that exposure to fluoride is associated with lower IQ scores (see **Figure D1** through **Figure D7**); however, the analyses performed and the specific results varied by study. Consistent results between increased fluoride levels and lower IQ scores were seen across studies using different exposure measures [e.g., single serum samples (Xiang *et al.* 2011, Zhang *et al.* 2015b), single spot urine samples in children (Xiang *et al.* 2003a, Rocha-Amador *et al.* 2007, Ding *et al.* 2011, Saxena *et al.* 2012, Zhang *et al.* 2015b, Cui *et al.* 2018, Yu *et al.* 2018, Wang *et al.* 2020b), and prenatal maternal urinary measures (Bashash *et al.* 2017, Green *et al.* 2019)] (see **Figure D6** and **Figure D7**). The consistency also occurs across different study designs and study populations. There were two studies with lower potential for bias (Cui *et al.* 2020, Soto-Barreras *et al.* 2019) that did not provide evidence of an association between fluoride exposure and IQ, but evaluating the association between fluoride and IQ levels was not the primary focus in either of these studies.

The three prospective cohort studies all found an association between increasing fluoride exposure and lower IQ in children. Two of the studies (Green *et al.* 2019, Till *et al.* 2020) were based on the same study population but evaluated fluoride exposure differently. Bashash *et al.* (2017) observed a significant inverse association between children's IQ and maternal urinary fluoride during pregnancy (measured during all three trimesters and included if at least one measurement was available; an increase of 0.5 mg/L of maternal urinary fluoride was associated with a 2.5-point lower IQ score [95% CI: −4.12, −0.59]) in boys and girls combined (see **Figure D7**); however, the association between IQ level and children's urinary fluoride levels, while inverse, was not significant (single spot urine sample; an increase of 0.5 mg/L of child urinary fluoride was associated with a 0.89-point lower IQ score [95% CI: −2.63, 0.85]) (Bashash *et al.* 2017). Green *et al.* (2019) also observed a significantly lower IQ for boys associated with maternal urinary fluoride averaged across trimesters (a significant 4.49-point lower IQ score [95% CI: −8.38, −0.60] in IQ per 1-mg/L increase in maternal urinary fluoride); results were not significant in girls (2.40-point increase [95% CI: −2.53, 7.33] in IQ) or in boys and girls combined (1.95-point lower IQ score per 1-mg/L increase; 95% CI: −5.19, 1.28). Other measures of prenatal exposure (maternal fluoride intake or water fluoride concentrations) were associated with lower IQ scores in boys and girls combined although the authors did not report boys and girls separately, as they found no significant effect measure modification between child sex and fluoride exposure in these analyses (Green *et al.* 2019). Specifically, when evaluating the association between estimated maternal fluoride intake based on maternal water and beverage consumption during pregnancy and IQ in children, a 1-mg increase in daily maternal consumption of fluoride during pregnancy was associated with a significantly lower IQ score of 3.66 points in boys and girls combined (95% CI: −7.16, −0.15). Similarly, water fluoride concentrations for pregnant women from fluoridated areas (mean water fluoride levels of 0.59 ± 0.08 mg/L) versus pregnant women from non-fluoridated areas (mean water fluoride levels of 0.13 ± 0.06 mg/L) were associated with a significant 5.29-point lower IQ score per 1-mg/L increase in fluoride in both boys and girls combined (95% CI: −10.39, −0.19) (Green *et al.* 2019). Using the same study population as Green *et al.* (2019), but using fluoride intake from formula or water concentrations in

formula-fed versus breastfed infants, Till *et al.* (2020) observed a significantly lower performance IQ scores regardless of the comparison used. They did not observe any effect on verbal IQ, and full-scale IQ was only significantly lower in formula-fed infants using water fluoride concentrations as the exposure measure. All other comparisons showed negative associations but were not significant.

Cross-sectional studies also demonstrated a consistent association between fluoride and lower IQ scores. Rocha-Amador *et al.* (2007) observed significant negative correlations between IQ and both water and children's single spot urinary fluoride levels in a population in Mexico (adjusted β = −10.2 per log fluoride increase [CIs not reported] and −16.9 per log fluoride increase [CIs not reported], respectively) (see **Figure D7**). The authors also observed a significant inverse association between IQ and children's drinking water and single spot urinary arsenic levels (adjusted β = −6.15 [CIs not reported] and −5.72 [CIs not reported], respectively). Because fluoride and arsenic were highly correlated in the study area, the authors were not able to adjust for exposure to arsenic when evaluating the effects of fluoride exposure (Rocha-Amador *et al.* 2007). Ding *et al.* (2011) reported a negative dose-response relationship between children's single spot urinary fluoride levels and IQ (see **Figure D4**); after adjusting for age, using multiple linear regression, they found a 0.59-point lower IQ score (95% CI: −1.09, −0.08) per 1-mg/L increase in urinary fluoride (p-value < 0.0001) (see **Figure D7**). Cui *et al.* (2018) observed a significant association between log-transformed children's single spot urine fluoride and lower IQ scores (2.47-point lower IQ scores [95% CI: −4.93, −0.01] per unit increase in urinary fluoride), and the association was the strongest in subjects with the TT polymorphism in the dopamine receptor D2 gene which, according to the authors, probably results in a reduced D2 receptor density (12.31-point lower IQ score [95% CI: −18.69, −5.94] per unit increase in urinary fluoride) (Cui *et al.* 2018).

Although Green *et al.* (2019) observed a significant negative association between maternal fluoride levels and IQ scores in boys but not girls in a Canadian population, sex differences were not observed in a cross-sectional study conducted using children spot urine fluoride concentrations in China (Wang *et al.* 2020b). Wang *et al.* (2020b) evaluated boys and girls combined and separately and observed significant decreasing trends in the sexes both combined and alone by urinary fluoride quartiles. When evaluated as a continuous variable, spot urinary fluoride levels (per 1-mg/L increase) were significantly associated with lower IQ scores in both girls (−1.379 [95%CI: −2.628, −0.129]) and boys (−1.037 [95% CI: −2.040, −0.035]), as well as, combined (−1.214 [95%CI: −1.987, −0.442]). Green *et al.* (2019) did not find any sex difference when using water fluoride concentrations, but Wang *et al.* (2020b) found that based on water fluoride quartiles there was a significant trend in girls and in boys and girls combined but not in boys alone. While there was a decreasing trend in boys, the results did not achieve statistical significance (p = 0.077). However, when water fluoride levels were evaluated as a continuous variable (per 1 mg/L increase), there were significant associations between lower IQ scores in both girls (−1.649 [95%CI: −3.201, −0.097] and boys (−1.422 [95%CI: −2.792, −0.053], as well as, combined (−1.587 [95%CI: −2.607, −0.568]).

Other cross-sectional studies also observed consistent results across populations, but there were some slight variations based on exposure measurement, level of exposure, or based on the outcome measured. Choi *et al.* (2015) conducted a pilot study with 51 children in an area of China with a wide range of fluoride concentrations in the drinking water. Aside from observing no association between the square root block design test score and fluoride exposure from drinking water, the authors observed consistent negative associations between IQ measures and fluoride in children's single spot urine or drinking water and significant associations between specific tasks from an omnibus IQ test (i.e., significantly lower WISC-IV backward and total digit span scores) and fluoride exposure based on moderate or severe dental fluorosis in children (see **Figure D7**). While observing no association between

IQ and low children's single spot urinary fluoride levels (0.01–1.60 mg/L), Yu *et al.* (2018) observed significant negative associations (p values not reported) between IQ and median children's urinary fluoride levels (1.60–2.50 mg/L)—with an IQ score 2.67 points lower (95% CI: −4.67, −0.68) for every 0.5-mg/L increment of urinary fluoride—and high children's urinary fluoride levels at 2.50–5.54 mg/L with an IQ score of 0.84 points lower (95% CI: −2.18, 0.50) for every 0.5-mg/L increment of urinary fluoride (see **Figure D7**). The authors also reported a significant negative association between drinking water fluoride levels at 3.40–3.90 mg/L (4.29-point lower IQ score [95% CI: −8.09, −0.48] for every 0.5-mg/L increment of water fluoride); a 0.04-point lower IQ score (95% CI: −0.33, 0.24) was observed for 0.5-mg/L increments of water fluoride at levels of 0.20–3.40 mg/L. When comparing water fluoride concentrations of >1 mg/L to ≤1 mg/L, there was an increased risk (adjusted OR = 1.25; 95% CI: 0.69, 2.26) for marginal intelligence (i.e., IQ score = 70–79) and a decreased risk (adjusted OR = 0.47; 95% CI: 0.32, 0.71) of excellent intelligence (i.e., IQ score ≥ 130) (see **Figure D4**). Similar results were observed using children's urinary fluoride levels (adjusted OR for marginal intelligence = 1.44; 95% CI: 0.72, 2.91; adjusted OR for excellent intelligence = 0.49; 95% CI: 0.26, 0.93) (Yu *et al.* 2018).

Two lower risk-of-bias studies (Cui *et al.* 2020, Soto-Barreras *et al.* 2019) did not observe a significant association between fluoride and IQ in children; however, both studies only performed simple comparisons between IQ and fluoride exposure. Cui *et al.* (2020) studied children in the same region as Cui *et al.* (2018)—and possibly included some of the same subjects but over a longer timeframe—and did not observe a significant change in IQ score with increasing urinary fluoride levels. Although there was a 2-point drop in IQ between the lowest fluoride exposure group (i.e., spot urine fluoride < 1.6 mg/L) and the highest fluoride exposure group (i.e., ≥ 2.5 mg/L), the difference in IQ (112.16 ± 11.50 versus 110.00 ± 14.92) was not significant (p = 0.58). However, this study did not account for age in the analysis, even though they reported a significant difference in IQ score based on age (p < 0.001). Soto-Barreras *et al.* (2019) also did not find an association between various fluoride exposure metrics (water, spot urine, exposure dose, and fluorosis index) and IQ grade. However, this study only compared fluoride exposure levels within the 5 IQ grades and did not adjust for any potential confounders.

The results from 47 studies with higher potential for bias that evaluated IQ in children provide consistent supporting evidence of decrements in IQ associated with exposures to fluoride. Forty-one of the 47 studies reported an association between high fluoride exposure and lower IQ scores in children.

Meta-analysis

In response to the recommendations of the NASEM review of the September 6, 2019 draft monograph (NASEM 2020), a two-part meta-analysis was conducted. The first part was an update to two previous meta-analyses (Choi *et al.* 2012, Duan *et al.* 2018) of group-level exposures from studies that reported a comparison of the mean IQ score between two or more exposure groups. The second part was a new meta-analysis and included studies with more precise individual-level exposures (e.g., urine, water, fluoride intake). The meta-analysis protocol can be found with the revised systematic review protocol posted in September 2020 (https://ntp.niehs.nih.gov/go/785076).

Group-level exposures

For the group-level exposure meta-analysis, a comparison on the mean outcome measure (IQ score) was conducted across two exposure groups ("exposed" and "reference"). If there were more than two exposure groups, the highest exposure group was designated the exposed group and the lowest exposure group was designated the reference group. For studies that had more than one exposed group (n = 17), a sensitivity analysis was performed to evaluate the impact of using any exposed group compared to the reference group. Using mean IQ levels with measures of uncertainty (95% confidence

interval [CI], standard error [SE], and sample size [N]) for exposed versus reference groups, the standardized mean difference (SMD) and corresponding 95% CI were calculated for each study. Given the heterogeneity of the studies, random effects models were used to obtain the pooled effect estimate, calculated as a weighted SMD with a corresponding 95% CI. These methods are consistent with both the Choi *et al.* (2012) and Duan *et al.* (2018) meta-analyses. More detailed methods are provided in **Appendix 5** and in the meta-analysis protocol, which can be found with the revised systematic review protocol posted in September 2020 (https://ntp.niehs.nih.gov/go/785076).

Characteristics of the 46 studies that compared mean IQ scores between groups of children with different levels of fluoride exposure are shown in **Table A-1**. One study was conducted in New Zealand, 1 study was conducted in Mexico, 4 studies were conducted in Iran, 9 studies were conducted in India, and the remaining 31 studies were performed in China (**Table A-1**). Five study populations were exposed to fluoride from coal burning (Guo *et al.* 1991 [translated in Guo *et al.* 2008a], Li *et al.* 1994 [translated in Li *et al.* 2008b], Li *et al.* 1995, Li *et al.* 2009, Bai *et al.* 2014); otherwise, it is assumed that study populations were exposed to fluoride through drinking water. Measures of fluoride exposure included water fluoride (n = 28), dental fluorosis (n = 7), and other non-drinking water sources of exposure to fluoride (e.g., fluoride exposure from coal burning [n = 11]). Thirteen studies presented results for males and 12 studies reported results for females; 9 studies examined children < 10 years old and 11 studies examined children ≥ 10 years old. The CRT-RC was used to measure children's IQ in 23 studies. Other measures of IQ included Wechsler intelligence tests (Ren *et al.* 1989 [translated in Ren *et al.* 2008], Wang *et al.* 1996 [translated in Wang *et al.* 2008b], An *et al.* 1992, Broadbent *et al.* 2015), the Binet IQ test (Guo *et al.* 1991 [translated in Guo *et al.* 2008a], Xu *et al.* 1994), the Raven's test (Yao 1997, Yao *et al.* 1996, Seraj *et al.* 2006, Seraj *et al.* 2012, Eswar *et al.* 2011, Poureslami *et al.* 2011, Shivaprakash *et al.* 2011, Khan *et al.* 2015, Sebastian and Sunitha 2015, Mondal *et al.* 2016), the Raymond B Cattell test (Karimzade *et al.* 2014), the Japan IQ test (Sun *et al.* 1991, Zhang *et al.* 1998), the Index of Mental Capacity (Li *et al.* 1994 [translated in Li *et al.* 2008b]), the Sequin Form Board test (Nagarajappa *et al.* 2013), and other tests using a doctor-prepared questionnaire (Trivedi *et al.* 2012, Trivedi *et al.* 2007). This meta-analysis includes 27 studies that were also included in Choi *et al.* (2012) and 25 studies that were also included in Duan *et al.* (2018). Also included in this meta-analysis were an additional 3 studies published since the Duan *et al.* (2018) publication and 11 studies that were not captured in either of the previous meta-analyses. Overall, the updated group-level results were highly consistent with these previous meta-analyses (Choi *et al.* 2012, Duan *et al.* 2018) (**Table A5-1**).

The random-effects pooled SMD estimated from the 46 studies included in the meta-analysis was −0.50 (95% CI: −0.61, −0.39) (**Table A5-1**, **Figure A5-1)**. There was evidence of heterogeneity (I^2 = 89%, p < 0.001, **Table A5-1**) and publication bias (funnel plot and Egger's p < 0.001, Begg's p = 0.08; **Figure A5-2**, **Figure A5-3**). Eliminating publication bias through trim-and-fill analysis supports the results with an adjusted pooled effect estimate of – 0.42 (95% CI: −0.54, 0.30) (**Figure A5-4**). Among the 46 studies, all but two showed SMD estimates that indicated an inverse association, ranging from −5.34 (95% CI: −6.34, −4.34) to −0.04 (95% CI: −0.45, 0.36). The studies with a positive association (Broadbent *et al.* 2015) reported an SMD estimate of 0.01 (95% CI: −0.19, 0.22) to 0.13 (95% CI: −0.16, 0.42). Three studies (Aravind *et al.* 2016, Kundu *et al.* 2015, Razdan *et al.* 2017) were excluded from the main analysis due to uncertainties about the way the intelligence assessment for children was performed, but sensitivity analyses that included these studies did not reveal any substantial changes in the pooled SMD estimate (−0.57 [95% CI: −0.69, −0.45]) (see **Figure A-35**).

Several subgroup analyses, discussed in **Appendix 5** and outlined in the meta-analysis protocol (found with the revised systematic review protocol posted in September 2020

[https://ntp.niehs.nih.gov/go/785076]), included risk of bias, gender, age group, country, outcome assessment type, and exposure assessment type. Among the lower risk-of-bias studies (n = 9), the random-effects pooled SMD was −0.31 (95% CI: −0.52, −0.10) with an I^2 of 87% and heterogeneity test p-value < 0.001 (**Table A5-1** and **Figure A5-6**). There was no evidence of publication bias (funnel plot and Egger's p = 0.72, **Figure A5-7** and **Figure A5-8**). Among the higher risk-of-bias studies (n = 37), the random-effects pooled SMD was −0.56 (95% CI: −0.68, −0.43) with an I^2 of 88% and heterogeneity test p-value < 0.001 (**Table A5-1** and **Figure A5-6**). There was evidence of publication bias among the higher risk-of-bias studies (funnel plot and Egger's p < 0.001, **Figure A5-7** and **Figure A5-8**); eliminating publication bias through trim-and-fill analysis supports the results with an adjusted pooled SMD estimate of −0.35 (95% CI: −0.50, −0.21) (**Figure A5-7** and **Figure A5-9**).

Subgroup analyses by gender, age group, country, outcome assessment type, and exposure assessment type further support the consistent and robust pattern of results (**Table A5-1**). Except for the subgroup analysis of the four studies from Iran, heterogeneity remained at an I^2 of ≥ 70% when the analyses were restricted by subgroup. Sensitivity analyses that removed an outlier (Khan *et al.* 2015) or compared all exposed groups versus the reference (i.e., exposed groups were combined if a study reported more than one exposed group) also did not appreciably change the results (**Figure A-45**, **Table A5-1**, and **Figure A-25**).

Individual-level exposures

The individual-level exposure meta-analysis included 6 studies with individual-level exposures that reported effect estimates as beta coefficients and included a 95% CI or SE. Characteristics of the studies with individual-level exposures are shown on **Table B-1**. All studies included in this meta-analysis were considered lower risk of bias. Adjusted effect estimates were used, and if results from multiple models were reported within a single study, the most adjusted results were selected. (For more details, see the meta-analysis protocol, which can be found with the revised systematic review protocol posted in September 2020 [https://ntp.niehs.nih.gov/go/785076].) To ensure consistent units across studies, units of fluoride exposure were transformed to 1 mg/L. For Bashash *et al.* (2017), Yu *et al.* (2018), and Till *et al.* (2020), units of exposure were transformed from 0.5 mg/L to 1 mg/L. For Cui *et al.* (2018), units of exposure were transformed from 1 log mg/L to 1 mg/L. Cui *et al.* (2018) reported an association between IQ and log transformed exposure. A sensitivity analysis was performed to evaluate the impact of using Cui *et al.* (2018), since the relationship between IQ and exposure evaluated in this study was not linear (as was the case among the other studies included). Yu *et al.* (2018) reported estimates from piecewise linear regression models and provided three ranges for urinary fluoride exposure (low 0.01–1.60 mg/L, medium 1.60–2.50 mg/L, high 2.50–5.54 mg/L) and two ranges for water fluoride (low 0.20–3.40 mg/L and high 3.40–3.90 mg/L). Since these piecewise effect estimates are likely correlated, the study-specific pooled effect estimates were used for urine and water fluoride exposures for the overall effect meta-analysis. A sensitivity analysis was performed to evaluate the impact of using pooled estimates rather than piecewise estimates from Yu *et al.* (2018).

For studies with overlapping populations (i.e., multiple studies that used the same cohort), results were selected considering the following factors: most appropriate exposure metric, exposure range, exposure period, number of subjects, and statistical adjustment for potential confounders. In the overall effect analysis, for studies reporting multiple measures of fluoride exposure, the results associated with measured or estimated individual-level exposures, biomarker levels (such as urinary fluoride), or fluoride intake levels were prioritized over water fluoride concentrations (revised protocol posted in September 2020 [https://ntp.niehs.nih.gov/go/785076]); however, subgroup analyses by exposure metric (urinary fluoride, fluoride intake, and water fluoride) were also performed. Yu *et al.* (2018) and

Wang *et al.* (2020b) used the same study cohort of children recruited in 2015 from the rural areas of Tianjin City, China. Since Wang *et al.* (2020b) (n = 571) used a subset of the original study sample from Yu *et al.* (2018) (n = 2,668), only results from Yu *et al.* (2018) were included in the meta-analysis. A sensitivity analysis was performed to evaluate the impact of using the effect estimate from Wang *et al.* (2020b) rather than the pooled effect estimate from Yu *et al.* (2018). Green *et al.* (2019) and Till *et al.* (2020) used the same Maternal-Infant Research on Environmental Chemicals (MIREC) cohort that reported drinking tap water in 10 Canadian cities with the studies overlapping for 398 mother-child pairs. Both studies reported effect estimates for maternal urinary fluoride (MUF) and water fluoride concentrations. In the Green *et al.* (2019) study, 512 mother-child pairs had MUF data (and all covariates) compared to 398 pairs in Till *et al.* (2020). Water fluoride levels were available for 420 pairs in Green *et al.* (2019) compared to 398 pairs in Till *et al.* (2020). Both studies reported effect estimates adjusted for maternal education, maternal race, child's sex, HOME total score, and secondhand smoke status in the child's home. In addition, Till *et al.* (2020) adjusted for child's age at IQ testing (the age range for all children was 3–4 years old). Because of the larger sample size and covariate adjustments were similar, results from Green *et al.* (2019) were included in the main analysis. However, because of the more adjusted estimates from Till *et al.* (2020) compared to Green *et al.* (2019), a sensitivity analysis was performed using the water fluoride result for formula-fed children and the MUF result from Till *et al.* (2020). For fluoride from intake, the estimates from both studies were used since they represent total fluoride intake (Green *et al.* 2019) and infant fluoride intake from formula (Till *et al.* 2020).

The overall pooled effect estimate based on 6 lower risk-of-bias studies with individual-level measures of exposure shows that a 1-mg/L increase in urinary fluoride was associated with a statistically significant lower IQ score of 1.40 (95% CI: −2.33, −0.47) points. Studies with individual-level urinary fluoride measures had evidence of moderate heterogeneity (I^2 = 46%, p = 0.101; **Table A5-2**, **Figure A5-16**). Eliminating publication bias through trim-and-fill analysis supports the conclusion that a 1-mg/L increase in individual-level exposure to urinary fluoride was associated with lower IQ, with an adjusted pooled effect estimate of −0.82 (95% CI: −1.81, 0.17) (**Figure A5-19**). Fluoride intake and water fluoride were also significantly associated with an IQ score of 3.31 points lower (95% CI: −6.12, −0.50) and 4.77 points lower (95% CI: −9.10, −0.45), respectively (**Table A5-2**); however, the results for both metrics were based on two studies each and should be interpreted with caution.

No substantial changes in the pooled effect estimates were seen in sensitivity analyses to evaluate the following scenarios: using the piecewise estimates from Yu *et al.* (2018) (−1.37, 95% CI: −2.38, −0.37) (**B-1**); using effect estimates from Wang *et al.* (2020b) rather than Yu *et al.* (2018) (−1.24, 95% CI: −1.94, −0.54) (**Figure B-6**); and using the water fluoride result for formula-fed children and MUF result from Till *et al.* (2020) rather than effect estimates from Green *et al.* (2019) (−1.50, 95% CI: −2.44, −0.57) (**Figure B-11**).

Other Neurodevelopmental or Cognitive Effects in Children

Among the studies with lower potential for bias, the results from three prospective cohort studies (Bashash *et al.* 2017, Valdez Jimenez *et al.* 2017, Bashash *et al.* 2018) and six cross-sectional studies (Li *et al.* 2004 [translated in Li *et al.* 2008a], Rocha-Amador *et al.* 2009, Choi *et al.* 2015, Barberio *et al.* 2017b, Wang *et al.* 2020a, Riddell *et al.* 2019) based on seven study populations provide mostly consistent results for associations of fluoride exposure with cognitive impairment in children other than decrements in IQ, such as hand-eye coordination, neurobehavioral assessment, behavioral capacity, and learning disabilities (see **Figure D8** through **Figure D10**). Because IQ cannot be assessed in infants, other neurodevelopmental tests were conducted. Two studies (Li *et al.* 2004 [translated in Li *et al.* 2008a],

Valdez Jimenez *et al.* 2017), based in China and Mexico, evaluated neonates (within 3 days of birth) or infants (3–15 months) (see **Figure D8** and **Figure D10**).

In neonates, the high fluoride group (based on a single maternal urine fluoride level just prior to birth [3.58 ± 1.47 mg/L] compared to controls [1.74 ± 0.96 mg/L]) had significant lower ($p < 0.05$) total neurobehavioral assessment scores (38.28 ± 1.10 in controls compared to 36.48 ± 1.09 in high fluoride group) and total behavioral capacity scores (11.34 ± 0.56 in controls compared to 10.05 ± 0.94 in high fluoride group) as measured by a standard neonatal behavioral neurological assessment (NBNA) method (Li *et al.* 2004 [translated in Li *et al.* 2008a]). In infants, the Mental Development Index (MDI)—which measures functions including hand-eye coordination, manipulation, understanding of object relations, imitation and early language development—was significantly negatively correlated with maternal urinary fluoride in both the first and second trimesters (adjusted βs = −19.05 with standard error of 8.9 for first trimester and −19.34 with standard error of 7.46 for second trimester) (Valdez Jimenez *et al.* 2017). This study did not find an association between maternal fluoride during any trimester and Psychomotor Developmental Index (PDI), which measures gross motor development (adjusted βs = 6.28 and 5.33 for first and second trimesters, respectively; no variance provided) (Valdez Jimenez *et al.* 2017). The General Cognitive Index (GCI) of the McCarthy Scales of Children's Abilities (MSCA) in 4-year-old children was significantly negatively associated with maternal creatinine-adjusted urinary fluoride levels during pregnancy (collected during each trimester) even after adjusting for maternal bone lead (adjusted β = −3.15 [95% CI: −5.42, −0.87] in a model adjusting for main covariates (e.g., gestational age, weight at birth, sex, maternal smoking, and indicators of socioeconomic status); adjusted β = −5.63 [95% CI: −8.53, −2.72] in a model limited to a subset of cases who had data on maternal bone lead and adjusted for main covariates and maternal bone lead) (Bashash *et al.* 2017) (see **Figure D10**). Choi *et al.* (2015), however, evaluated cognitive function endpoints in addition to IQ and found no significant associations between concurrent water or urinary fluoride levels and Wide Range Assessment of Visual Motor Ability (WRAVMA) scores, finger tapping, and the grooved pegboard test although there were some significant associations based on degree of fluorosis (see **Figure D10**). Another study using construction and memory scores in children 6–11 years old observed statistically significant lower scores with increasing concurrent child single spot urinary fluoride even after adjusting for age ($p < 0.05$; −0.29 and −0.27 for copy and immediate recall, respectively [CIs not reported]); however, scores were not significantly associated with urinary arsenic levels (−0.05 and 0.02 for copy and immediate recall, respectively [CIs not reported]) (Rocha-Amador *et al.* 2009) (see **Figure D9**).

Barberio *et al.* (2017b) evaluated learning disabilities in children 3–12 years of age, including ADHD, attention deficit disorder (ADD), and dyslexia, as part of the Canadian Health Measures Survey and found a small but significantly increased risk in self-reported (children 12 years of age) or parent- or guardian-reported (children 3–11 years of age) learning disabilities associated with higher spot urinary fluoride levels in children (adjusted OR = 1.02; 95% CI: 1.00, 1.03) (see **Figure D11**); however, significant associations were not observed in analyses using creatinine- or specific gravity-adjusted urinary fluoride (Barberio *et al.* 2017b). Barberio *et al.* (2017b) also reported no associations between single spot urinary fluoride and ADHD in children ages 3 to 12 years. Riddell *et al.* (2019) used the same Canadian Health Measured Survey, but evaluated children 6–17 years old. Riddell *et al.* (2019) found a significantly increased risk for ADHD diagnosis with both tap water fluoride (adjusted OR per 1-mg/L increase = 6.10; 95% CI: 1.60, 22.8) and community water fluoridation status (adjusted OR = 1.21; 95% CI: 1.03, 1.42). A similar increase in hyperactivity-inattention symptoms score based on the Strengths and Difficulties Questionnaire was observed with both tap water fluoride (adjusted β per 1-mg/L increase = 0.31; 95% CI: 0.04, 0.58) and community fluoridation status (adjusted β = 0.11; 95% CI: 0.02, 0.20). As was observed with Barberio *et al.* (2017b), Riddell *et al.* (2019) did not observe an association with either

ADHD diagnosis (adjusted OR per 1-mg/L increase = 0.96; 95% CI: 0.63, 1.46) or hyperactivity-inattention symptoms (adjusted β = 0.31; 95% CI: −0.04, 0.66) and specific-gravity-adjusted spot urinary fluoride concentrations. Bashash *et al.* (2018) evaluated behaviors associated with ADHD in children ages 6–12 years using the Conners' Rating Scales-Revised (CRS-R) and observed significant associations between maternal urinary fluoride (measured during each trimester) and ADHD-like symptoms, particularly those related to inattention (an increase in 0.5 mg/L of maternal urinary fluoride was associated with a 2.84-point increase [95% CI: 0.84, 4.84] in the DSM-IV Inattention Index and a 2.54-point increase [95% CI: 0.44, 4.63] in the Cognitive Problems and Inattention Index). These two scales contributed to the global ADHD Index and the DSM-IV ADHD Total Index, which were also associated with higher levels of prenatal fluoride exposure (an increase of 0.5 mg/L in maternal urinary fluoride was associated with a 2.38-point increase [95% CI: 0.42, 4.34] in the DSM-IV ADHD Total Index and a 2.47-point increase [95% CI: 0.43, 4.50] in the ADHD Index) (see **Figure D10**). Significant associations were not observed between maternal urinary fluoride concentrations during pregnancy and child performance on measures of hyperactivity nor were there any significant results in children using the Connors' Continuous Performance Test (CPT-II, 2nd Edition), a computerized test of sustained attention and inhibitory control (Bashash *et al.* 2018). Wang *et al.* (2020a) also used a Connors' Parent Rating Scale (Chinese version), but only found a significant association between spot urinary fluoride concentrations (model adjusted for creatinine) and psychosomatic problems (adjusted OR for T-score > 70 = 1.97; 95% CI: 1.19, 3.27 and adjusted β = 4.01; 95% CI: 2.74, 5.28). No associations were found between spot urinary fluoride and ADHD index or other behavioral measures.

Higher risk-of-bias studies (n = 5) also provide some evidence of associations of fluoride exposure with neurodevelopmental or cognitive effects in children other than effects on IQ, but the results are inconsistent with heterogeneous outcomes (Li *et al.* 1994 [translated in Li *et al.* 2008b], Shannon *et al.* 1986, Malin and Till 2015, Morgan *et al.* 1998, Mustafa *et al.* 2018).

Cognitive Effects in Adults

Results from two lower risk-of-bias studies in adults did not find consistent evidence for an association between cognitive impairment (based on the Mini-Mental State Examination) and exposure to fluoride (Jacqmin *et al.* 1994, Li *et al.* 2016). Jacqmin *et al.* (1994) did not find an association between drinking water fluoride and cognitive impairment in populations in France (n = 3,490) and found prevalence rates of cognitive impairment to be the same regardless of fluoride exposure (see **Figure D12**). In an analysis of 38 cognitively-impaired cases and 38 controls matched for several confounders including age, gender, education, alcohol consumption, and smoking, Li *et al.* (2016) did find significantly higher urinary fluoride levels and skeletal fluorosis scores in the cognitively-impaired group compared with the control group; however, the authors found no significant correlation between cognitive impairment and total daily water fluoride intake (adjusted ORs = 0.94 [95% CI: 0.85, 1.04] and 0.86 [95% CI: 0.69, 1.06] in the moderate and severe cognitive impairment groups, respectively) or urinary fluoride levels (adjusted ORs = 1.12 [95% CI: 0.89, 1.42] and 1.25 [95% CI: 0.87, 1.81] in the moderate and severe cognitive impairment groups, respectively) in subjects from fluorosis-endemic areas of China (n = 511).

Higher risk-of-bias studies (n = 7) provide some evidence of cognitive impairment in adults associated with exposure to fluoride. In aluminum factory workers (exposed to gaseous and particular fluoride emissions during the production of aluminum metal), significant decreases in IQ (Duan *et al.* 1995), diminished performance on several neurobehavioral core battery tests (NCTBs) (Guo *et al.* 2001 [translated in Guo *et al.* 2008b]), and impaired psychomotor performance and memory were observed (Yazdi *et al.* 2011). One study conducted on adult subjects with fluorosis (dental and skeletal) from a fluorosis-endemic area compared with healthy subjects from a non-endemic area observed significant

differences for some cognitive function tests (i.e., tests of speech fluency, recognition, and working memory) but not others and generally did not observe a significant change in IQ except in the operation scores (Shao 2003). One prospective cohort study evaluated exposure to fluoride in children at age of 5 years based on whether or not the children resided in areas with community water fluoridation or used fluoride toothpaste or fluoride tablets, and found no clear differences in IQ scores of the subjects at age 38 years (Broadbent *et al.* 2015). One additional study suggested that populations living in areas with higher drinking water fluoride had lower levels of dementia (Still and Kelley 1980); however, the study was not focused on effects of fluoride, but rather if fluoride was able to reduce the risk associated with aluminum by competing with aluminum and reducing the aluminum bioavailability. Therefore, the study was considered inadequate to evaluate the effects of fluoride on dementia (Still and Kelley 1980). A more recent study in Scotland evaluated dementia rates associated with aluminum and fluoride drinking water concentrations and observed an increase in dementia only in the highest quartile of fluoride (56.3 µg/L) compared to the lowest quartile (<44.4 µg/L), but found a significant increase with all quartiles of aluminum compared with the reference group (Russ *et al.* 2019). In addition to studies that reported on cognitive impairment and exposure to fluoride, two studies were identified that reported effects on motor and sensory function (Rotton *et al.* 1982) and a higher prevalence of self-reported headaches, insomnia, and lethargy (Sharma *et al.* 2009).

Mechanistic Data in Humans

Eight lower risk-of-bias studies were available that evaluated mechanistic data in humans associated with fluoride exposure that was considered potentially relevant to neurological effects, including effects on thyroid hormones in children (Singh *et al.* 2014, Zhang *et al.* 2015b, Kumar *et al.* 2018), adults (Kheradpisheh *et al.* 2018b, Kheradpisheh *et al.* 2018a, Malin *et al.* 2018), or children and adults combined (Barberio *et al.* 2017a). In addition, some studies evaluated self-reported thyroid conditions in children and adults combined (Barberio *et al.* 2017a) and thyroid diseases in adults (Kheradpisheh *et al.* 2018b, Peckham *et al.* 2015) (see **Figure A3-5** and **Figure A3-6**). Although the lower risk-of-bias studies provide some evidence of mechanistic effects (primarily changes in thyroid stimulating hormone [TSH] levels in children), the studies were too heterogeneous or limited in number to make any determination on mechanism (see **Figure 7**).

Among the seven lower risk-of-bias studies that reported on changes in thyroid hormones, three studies were conducted in children (Zhang *et al.* 2015b, Singh *et al.* 2014, Kumar *et al.* 2018) and reported increases in TSH levels. Zhang *et al.* (2015b) reported significant increases in TSH in children from a fluorosis-endemic area (median fluoride drinking water concentration = 1.40 mg/L; interquartile range = 1.23–1.57 mg/L) compared with a non-fluorosis-endemic area (median fluoride drinking water concentration = 0.63 mg/L; interquartile range = 0.58–0.68 mg/L), while 3,5,3'-triiodothyronine (T_3) or thyroxine (T_4) were not significantly different between the two groups. Similarly, Singh *et al.* (2014) observed significantly higher TSH levels in children without dental fluorosis who lived in a fluorosis-endemic area (fluoride drinking water concentrations of 1.6–5.5 mg/L) compared with children without dental fluorosis who lived in a non-fluorosis-endemic area (fluoride drinking water concentrations of 0.98–1.00 mg/L). Higher TSH levels in children with dental fluorosis from the fluorosis-endemic area compared with children without dental fluorosis from the non-fluorosis-endemic area were observed but did not reach statistical significance. Significant differences in T_4 or T_3 were not observed between groups (Singh *et al.* 2014). Kumar *et al.* (2018) also observed a significant increase in TSH levels in children from a fluorosis endemic area (1.5–5.8 mg/L fluoride) compared with a control area (0.94–1.08 mg/L fluoride). There were also decreases in T_3 and T_4, but results were not statistically significant.

Barberio *et al.* (2017a) evaluated fluoride effects on TSH levels in children and adults combined and found no relationship between fluoride exposure (measures in urine and tap water) and TSH levels. In the one study that evaluated thyroid hormone levels in adults but not children, Kheradpisheh *et al.* (2018b) found a significant increase in TSH associated with higher fluoride concentrations in drinking water in both adults with and without thyroid diseases such as hypothyroidism, hyperthyroidism, thyroid nodules, or thyroid cancer. Significant increases in T_3 were associated with higher fluoride in drinking water in adults without thyroid diseases, but increases in T_3 were not significant in adults with thyroid diseases. A significant association between T_4 and higher fluoride in drinking water was not observed in adults with or without thyroid diseases (Kheradpisheh *et al.* 2018b).

Other than changes in hormone levels, there is limited evidence of fluoride-related mechanistic effects in the three lower risk-of-bias studies that evaluated thyroid-related effects. Barberio *et al.* (2017a) found no relationship between fluoride exposure and self-reported thyroid conditions in children and adults (children were older than 12). Kheradpisheh *et al.* (2018b) also found no association between fluoride exposure and hypothyroidism in an adult population in Iran. One study found a significantly higher prevalence of hypothyroidism in areas with higher fluoride concentrations in drinking water (>0.7 mg/L) compared with areas with lower fluoride drinking water concentrations (≤0.7 mg/L) (Peckham *et al.* 2015).

Several higher risk-of-bias studies were available that evaluated potential mechanistic data in humans associated with fluoride exposure, including effects on thyroid hormones mostly in children (n = 11 studies); catecholamines in adults (Michael *et al.* 1996) or in subjects of unknown ages (Chinoy and Narayana 1992); acetylcholinesterase (AChE) or serotonin levels in children (Singh *et al.* 2013, Lu *et al.* 2019); brain histopathology or biochemistry in aborted fetuses (Du *et al.* 1992 [translated in Du *et al.* 2008], Yu *et al.* 1996 [translated in Yu *et al.* 2008]); and mitochondrial fission/fusion molecules in children (Zhao *et al.* 2019). Similar to the lower risk-of-bias studies, the higher risk-of-bias studies provide some evidence of mechanistic effects (primarily changes in TSH levels in children); however, the data are insufficient to identify a clear mechanism by which fluoride causes neurodevelopmental or cognitive effects in humans.

Among higher risk-of-bias studies (see **Figure A3-7** and **Figure A3-8**), varying results were reported in 11 studies that evaluated fluoride exposure and effects on thyroid hormones, and a few of these studies (Lin *et al.* 1991, Yang *et al.* 1994 [translated in Yang *et al.* 2008], Wang *et al.* 2001) were complicated by high or low iodine in the high fluoride area. When considering fluoride effects on each of the hormones individually, similar to results from lower risk-of-bias studies, the most consistent evidence of fluoride-associated effects on a thyroid hormone was reported as changes in TSH levels in children, although there was some variation in the direction of effect. Six of the nine higher risk-of-bias studies that evaluated changes in TSH levels in children reported increases in TSH levels with higher fluoride (Lin *et al.* 1991, Susheela *et al.* 2005, Wang *et al.* 2001, Yang *et al.* 1994 [translated in Yang *et al.* 2008], Yao *et al.* 1996, Yasmin *et al.* 2013). Two of the nine higher risk-of-bias studies reported decreases in TSH levels in children with higher fluoride (Khandare *et al.* 2017, Khandare *et al.* 2018). One of the nine studies found no significant alterations in TSH levels in children from fluorosis-endemic areas (Hosur *et al.* 2012) (see **Figure 8**).

When considering fluoride-associated effects on TSH, T_3, and T_4 levels together, studies that evaluated changes in all three thyroid hormones reported varying combinations of increases, decreases, or no changes in levels across the three hormones, although among the eight lower and higher risk-of-bias studies that evaluated the effects of fluoride exposure on TSH, T_3, and T_4 levels and reported increases

in TSH levels in children, seven of the eight studies found no alterations in T_3 levels (one study found an increase in T_3), and six of the eight studies found no alterations in T_4 levels (two studies found an increase in T_4). Studies also displayed variation by age in fluoride-associated effects on TSH, T_3, and T_4. Due to the dynamic relationship between the thyroid gland, the pituitary gland, and the production and clearance of TSH, T_3, and T_4, the variations in results are not unexpected and do not eliminate the possibility of a mechanistic link between thyroid effects and neurodevelopmental or cognitive effects; however, the data do not support a clear indication that thyroid effects are a mechanism by which fluoride causes these effects in humans.

In addition to evaluating thyroid hormone levels, a few higher risk-of-bias studies evaluated other mechanistic data associated with fluoride exposure; however, the data are insufficient to identify a clear mechanism by which fluoride might cause neurodevelopmental or cognitive effects in humans. Serum epinephrine and norepinephrine were significantly increased in a fluoride-endemic region (not reported whether subjects were children or adults) compared to a non-endemic region (Chinoy and Narayana 1992). Serum adrenaline and noradrenaline were significantly increased in adults in a fluoride-endemic area (fluoride in the drinking water ranged from 1.0–6.53 ppm) compared to a control area (fluoride in the drinking water ranged from 0.56–0.72 ppm) (Michael *et al.* 1996). Serum AChE was significantly reduced in children from a high fluoride region compared to a lower fluoride region (Singh *et al.* 2013). Serum serotonin was significantly increased in children from Turkey who were drinking water containing 2.5 mg/L of fluoride compared to children drinking bottled water or water containing <0.5 mg/L of fluoride (Lu *et al.* 2019). Aborted fetuses from high fluoride areas in China were found to have histological changes in the brain and significant changes in neurotransmitter levels compared to a control area (Du *et al.* 1992 [translated in Du *et al.* 2008], Yu *et al.* 1996 [translated in Yu *et al.* 2008]).

There are also two more recent lower risk-of-bias studies that evaluated polymorphisms in dopamine-related genes; however, a determination on mechanism cannot be made at this time due to the limited number of studies. For children (10–12 years old) with a Val158Met polymorphism in the *COMT* gene (i.e., catechol-O-methyltransferase), which results in slower degradation and greater availability of dopamine within the brain, a stronger association between increasing urinary fluoride levels and decreasing IQ was reported (Zhang *et al.* 2015b). For children (7–12 years old) with a dopamine receptor-2 (DRD2) Taq 1A polymorphism (which is involved in reduced D2 receptor density and availability) and the TT (variant) genotype, a significant inverse relationship between log urine fluoride and IQ was observed; however, this significant relationship was not observed in children with the CC (wild-type) or CT (hybrid) genotypes (Cui *et al.* 2018).

Figure 7. Number of Lower Risk-of-bias Studies that Evaluated Thyroid Hormones in Children and Adults by Endpoint and Direction of Effect*

Endpoint	Direction of Effect ↑	NS	Grand Total
serum T3	1	3	4
serum T4		4	4
serum TSH	5	2	7
self-reported thyroid condition		1	1
thyroid disease		1	1
hypothyroidism prevalence	1		1

*Interactive figure and additional study details in Tableau® (https://public.tableau.com/profile/ntp.visuals#!/vizhome/Fluoride_EpiThyroid_UPDATE/Figures6and7). This figure displays study counts for lower risk-of-bias studies in both children and adults, as these counts are most relevant to the summary of fluoride-related mechanistic effects in lower risk-of-bias studies. Counts for higher risk-of bias studies and studies by age (i.e., children, adults, or children/adults combined) can also be accessed in the interactive figure in Tableau®. Study counts are tabulated by significance (unless if study footnotes in Tableau indicate that statistical significance was not tested)—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with significantly increased results.

Figure 8. Number of Higher Risk-of-bias Studies that Evaluated Thyroid Hormones in Children by Endpoint and Direction of Effect*

Endpoint	Direction of Effect ↑	↓	NS	Grand Total
serum T3	1	1	6	8
serum T4	3		5	8
serum TSH	6	2	1	9

*Interactive figure and additional study details in Tableau® (https://public.tableau.com/profile/ntp.visuals#!/vizhome/Fluoride_EpiThyroid_UPDATE/Figures6and7). This figure displays study counts for higher risk-of-bias studies in children, as these counts are most relevant to the summary of fluoride-related effects on thyroid hormones in higher risk-of-bias studies. Counts for lower risk-of bias studies, studies in adults, or all studies combined, can also be accessed in the interactive figure in Tableau®. Study counts are tabulated by significance (unless if study footnotes in Tableau indicate that statistical significance was not tested)—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with significantly increased results.

Animal Learning and Memory Data

[Note: An earlier version of the monograph underwent NASEM committee peer review in November 2019 (NASEM 2020). In this earlier review the committee criticized this section primarily over concerns that the NTP's risk-of-bias evaluations failed to adequately capture a number of important threats to internal validity that are specific to neurobehavioral outcomes in animal tests. The committee found

several examples of studies cited in the monograph where national or international guidelines for performance or statistical analyses were not followed or descriptions of methods were insufficient to evaluate their adequacy. Additionally, the committee took issue with a conclusion of the prior peer reviewed systematic review of the experimental animal literature (NTP 2016) concerning the degree to which motor activity deficits might compromise neurobehavioral assessments, affecting the directness of applicability of deficits in animal learning and memory to support the plausibility of IQ deficits in exposed children.

The NTP generally does not take issue with the NASEM peer review comments and acknowledges that further efforts to disentangle the potential for motor activity deficits to influence tests of learning and memory in the fluoride literature are warranted. The NTP agrees with the comments of the NASEM committee concerning the overall poor quality of the experimental animal database, with many studies suffering from major reporting deficiencies. NTP also found these to be general issues with the experimental animal database and were identified as deficiencies that led to the inadequate conclusion. However, the following experimental animal study section remains largely unchanged from the initial version of the monograph reviewed by NASEM in November 2019. The reasons for this are: (1) because a more critical risk-of-bias assessment would result in fewer relevant animal studies judged to be of high quality; (2) because the highest quality experimental animal study reviewed for this monograph (McPherson *et al.* 2018) did not find effects of fluoride on learning, memory or motor activity in the critical $\leq$20 ppm in drinking water concentration range; and (3) because of the availability of a large number of human epidemiology studies directly addressing neurobehavioral and cognitive effects of fluoride in children, a decision was made to focus efforts to address comments on the critical human epidemiology evaluation in this revised and updated monograph. NTP acknowledges the helpful comments of the NASEM committee on the following section and refers readers to the NASEM review (NASEM 2020) when considering the information provided. For the purpose of this updated review the NTP considers that the experimental animal data remain *inadequate* to inform conclusions on whether fluoride exposure is associated with cognitive effects (including cognitive neurodevelopmental effects) in humans. NTP is aware of a number of additional relevant experimental animal studies published since the literature cutoff date for the monograph. These additional studies have not been formally reviewed and may shed further light on these issues].

In 2016, NTP conducted a systematic review of the available experimental animal studies to develop level-of-evidence conclusions on the association between fluoride exposure and neurobehavioral effects, specifically effects related to learning and memory impairment (NTP 2016). As previously discussed, the evaluation of the animal body of evidence in this assessment is an extension of the NTP (2016) systematic review and is consistent with the methodology and format used in that report.

NTP (2016) identified two main issues with the animal body of evidence related to effects of fluoride exposure on learning and memory: indirectness and concerns for risk of bias. The concern related to indirectness was based on the fact that many learning and memory tests rely on a motor response (e.g., latency to achieve the desired effect). Changes in motor function or activity levels associated with fluoride exposures could complicate the interpretation of the results on learning and memory test performance depending on the outcome measured. The directness of the measure as an indicator of learning and memory (i.e., the ability to rule out impaired motor or sensory function) was considered when addressing confidence in the data. Concerns in these studies related to risk of bias included the following factors: lack of randomization, lack of blinding or other methods to reduce potential bias at outcome, lack of exposure information, lack of control for litter effects, lack of expected response in the

control animals, and lack of reporting of other key study information such as sample size or sex of the animals.

Since the NTP (2016) report was published, additional experimental animal studies were identified that evaluated learning and memory impairment associated with fluoride exposure, including 12 developmental exposure studies (Banala and Karnati 2015, Mesram *et al.* 2016, Zhu *et al.* 2017, Sun *et al.* 2018, Wang *et al.* 2018, Ge *et al.* 2018b, McPherson *et al.* 2018, Zhao *et al.* 2019, Ge *et al.* 2018a, Chen *et al.* 2018, Bartos *et al.* 2018, Banala *et al.* 2018); 5 Morris water maze study in adults (Zheng *et al.* 2016, Niu *et al.* 2018, Dong *et al.* 2017, Sharma *et al.* 2018, Yang *et al.* 2018); and 7 other maze studies in adults (Pulungan *et al.* 2016, Shalini and Sharma 2015, Sharma *et al.* 2018, Sudhakar *et al.* 2017, Nageshwar *et al.* 2018, Yuan *et al.* 2019, Raju *et al.* 2019). In addition, 12 studies were identified that evaluated motor activity/coordination or sensory effects without evaluating learning and memory impairment (Adedara *et al.* 2017a, Nageshwar *et al.* 2017, Nkpaa and Onyeso 2018, Sudhakar *et al.* 2018b, Ahmad *et al.* 2017, Kinawy and Al-Eidan 2018, Manusha *et al.* 2019, Sudhakar *et al.* 2018a, Agustina *et al.* 2018, Lu *et al.* 2019, Jia *et al.* 2019, Li *et al.* 2019).

Although Adedara *et al.* (2017a) and Nkpaa and Onyeso (2018) evaluated exploration, the authors concluded that the track plots in the open field novel environment test were consistent with impaired locomotor activity in the fluoride-treated animals. The additional studies reviewed did not address the concern of indirectness and most included risk-of-bias concerns; however, a few of these more recent studies are notable in that they provide results on learning and memory effects that could possibly be distinguished from effects on motor activity. Bartos *et al.* (2018) used a step-down inhibitory avoidance test to evaluate short-term and long-term memory in rat offspring. Although the authors did not discuss activity in the animals, this test would be expected to result in increased latency in animals if there was decreased activity with fluoride exposure. The fluoride-treated female offspring, however, had decreased latency indicating diminished memory of the foot shock. Chen *et al.* (2018) also evaluated female rat offspring (treatment continued until the offspring were 6 months old) and observed an effect of fluoride on latency to reach the platform and the number of platform crossings in the Morris Water Maze; however, swimming speed was measured, and no changes were observed. The tracks during the spatial probe test were also very different in the two higher exposed groups (i.e., 50 and 100 mg/L NaF), suggesting that the animals did not know the location of the platform. It is not clear if litter effects were addressed in the study.

After further evaluation of the data available in NTP (2016) and in this update, it is concluded that the animal data are inadequate to evaluate the effects of fluoride on learning and memory primarily due to the inability to separate the learning and memory effects from the effects on motor activity or motor coordination. The majority of the studies that evaluated effects of fluoride on learning and memory did not also evaluate a motor activity component to determine if the learning and memory effects could be attributed to motor activity or coordination deficits. Of the studies that did evaluate both learning and memory and motor activity/coordination, studies mainly found an association between fluoride exposure and both types of neurological outcomes or found no effect of fluoride exposure on either type of neurological outcome irrespective of the dose range or duration of dosing. In addition, studies that found effects on motor activity/coordination or learning and memory often did not provide sufficient indicators of general health of the animals to reliably attribute impaired performance on a task to a specific acquisition of learning and memory or motor activity/coordination. The few studies that provided this information used different test methods or results were inconsistent. Thus, it is difficult to conclude that evidence from experimental animal studies is meaningful when considering the specific question of fluoride's potential influence on human IQ or cognitive function, particularly at

human -relevant exposure levels. Based on this consideration, the experimental animal body of evidence does not contribute to confidence in conclusions derived from human epidemiological studies with respect to effects on human IQ. Although the evidence supports an association between fluoride exposure and neurodevelopmental effects, the data are not sufficient to support the primary effect evaluated in children (i.e., IQ) nor is it sufficient to support a conclusion on cognitive effects in adults especially in the absence of additional adult human data.

Mechanistic Data in Animals

There are a wide variety of studies in animals that evaluate mechanistic effects potentially related to neurological changes following oral fluoride exposure (see **Figure 9**). Categories of mechanistic endpoints with the largest amount of available data include changes in biochemical components of the brain or neurons, neurotransmitters, oxidative stress, histopathology, and thyroid function. Limiting the data to studies with at least one exposure at or below 20 ppm fluoride drinking water equivalents (gavage and dietary exposures were back calculated into equivalent drinking water concentrations for comparison) still provided a sufficient number of studies for evaluation of these mechanistic endpoints. Neurotransmitter and biochemical changes in the brain and neurons were considered to be the mechanistic areas with the greatest potential to demonstrate effects of fluoride on the brain of animals in the lower dose range and provide evidence of changes in the brain that may relate to lower IQ in children (see **Figure 10**). Histological data can be useful in determining whether effects are occurring in the brain at lower fluoride concentrations; however, author descriptions of these effects may be limited thereby making it difficult to directly link histological changes in the brain to learning and memory effects. Oxidative stress is considered a general mechanistic endpoint that cannot be specifically linked to neurodevelopmental or cognitive effects in humans; however, like histopathology, it may help in identifying changes in the brain occurring at lower concentrations of fluoride. Although any effects in the brain or neurological tissue at lower concentrations of fluoride may support reduced IQ in humans, it may be difficult to distinguish the potential effects of fluoride on learning and memory functions from other neurological or general health outcomes.

Figure 9. Number of Animal Mechanistic Studies for Fluoride by Mechanistic Category and Exposure Level*

Mechanism	Dose Level	
	All	≤20 ppm
Biochemical (brain/neurons)	66	25
Neurotransmitters	53	23
Oxidative stress	78	25
Histopathology	67	30
Apoptosis/cell death	19	7
Inflammation	7	5
Thyroid	50	17
Other	3	2

*Interactive figure and additional study details in Tableau® (https://public.tableau.com/profile/ntp.visuals#!/vizhome/Animal_Mechanisms_All_June2019/Figure8). The number of

studies that evaluated mechanistic effects associated with at least one exposure at or below 20 ppm fluoride is tabulated in the "≤20 ppm" column. The total number of studies per mechanistic category are summarized in the "All" column.

The following sections summarize the mechanistic data by category of mechanistic endpoint. Although there is some evidence of consistency in mechanistic effects, overall these data are insufficient to increase confidence or support a change to hazard conclusions.

Neurotransmitters

Twenty of 23 neurotransmitter studies assessed changes in brain cholinesterase activity associated with fluoride exposure at or below 20 ppm fluoride. Acetylcholine is a major neurotransmitter involved in learning, memory, and intelligence (Chen 2012, Gais and Schonauer 2017). AChE is responsible for the breakdown of acetylcholine in the synapses of nerve cells. Changes in cholinesterase, acetylcholine, or AChE could be related to effects on memory. Evidence of an effect varied among the lower risk-of-bias studies that assessed changes in cholinesterase or acetylcholine (n = 11 drinking water studies) (Gao *et al.* 2009, Baba *et al.* 2014, Adedara *et al.* 2017a, Khan *et al.* 2017, Gao *et al.* 2008a, Akinrinade *et al.* 2015a, Sun *et al.* 2000 [translated in Sun *et al.* 2008], Chouhan *et al.* 2010, Mesram *et al.* 2016, Liu *et al.* 2010, Nkpaa and Onyeso 2018), with the majority of studies reporting evidence of an effect that is considered inconsistent with the phenotypic outcome. Decreases in cholinesterase will cause increases in acetylcholine, which can have a positive effect on learning and memory; however, long-term decreases in cholinesterase can lead to secondary neuronal damage occurring in the cholinergic region of the brain (Chen 2012).

Five of the 11 studies with lower risk of bias (Gao *et al.* 2009, Baba *et al.* 2014, Adedara *et al.* 2017a, Khan *et al.* 2017, Nkpaa and Onyeso 2018) found statistically significant decreases in cholinesterase or AChE in brain homogenates (with some brains dissected into specific regions prior to homogenizing) with fluoride concentrations in drinking water at or below 20 ppm, and 4 of the 5 studies found statistically significant decreases in cholinesterase or AChE below 10 ppm. The 5 studies were conducted in rats (Wistar or Sprague-Dawley) with exposure ranging from 28 days to 6 months. An additional 2 out of 11 studies (Gao *et al.* 2008a, Akinrinade *et al.* 2015a) reported decreases in brain homogenate AChE at concentrations at or below 20 ppm fluoride in drinking water, but statistical significance was not reached. These studies were also conducted in rats with exposure for 30 days or 3 months. Gao *et al.* (2008a) reported a dose-dependent decrease in brain homogenate AChE in the low (5 ppm fluoride) and high (50 ppm fluoride) treatment groups compared with the control group, but the decrease was only statistically significant in the high dose group. Similarly, Akinrinade *et al.* (2015a) observed a dose-dependent decrease in percent intensity of AChE immunohistochemistry in the prefrontal cortex associated with 2.1 and 10 ppm sodium fluoride in the drinking water, but neither result was statistically significant. Gao *et al.* (2009) found lower brain homogenate AChE levels in the 5-ppm animals compared with the 50-ppm animals; therefore, the results were not always dose dependent.

Relative to the above-mentioned studies, 2 of the 11 lower risk-of-bias studies observed opposite effects on brain cholinesterase levels. Sun *et al.* (2000) [translated in Sun *et al.* 2008] observed a significant increase in brain cholinesterase in Kunming mice associated with fluoride drinking water concentrations from 10 to 100 mg/L, but did not observe a dose response. Chouhan *et al.* (2010) did observe a dose-related increase in AChE levels in brain homogenate of Wistar rats with sodium fluoride concentrations of 1 to 100 ppm for 12 weeks and noted statistically significant results at 1, 50, and 100 ppm but not at 10 ppm.

Mesram *et al.* (2016) did not assess changes in AChE but observed a significant decrease in acetylcholine levels in cerebral cortex homogenate through 30 days of age in rats treated in utero with 20 ppm

sodium fluoride, which may suggest an increase in AChE levels. Likewise, Liu *et al.* (2010) did not assess changes in AChE, but measured nicotinic acetylcholine receptors (nAChRs) in brain homogenate of rats following drinking water fluoride exposure, which the authors stated could modulate physiological and pharmacological functions that are involved in learning and memory-related behaviors. Significant decreases in the protein expressions of nAChR subunits at 2.26 ppm fluoride were observed; however, the corresponding receptor subunit mRNAs did not exhibit any changes (Liu *et al.* 2010).

The studies that assessed other neurotransmitters of the brain and neurons were too heterogeneous or limited in number to make any determination on mechanism, even before limiting the review of the data to lower risk-of-bias studies. There were only five studies that evaluated dopamine and/or metabolites (Tsunoda *et al.* 2005, Chouhan *et al.* 2010, Reddy *et al.* 2014, Banala *et al.* 2018, Sudhakar and Reddy 2018). Four of the studies observed decreases in dopamine levels in the brain with exposures less than 20 ppm fluoride (Reddy *et al.* 2014, Chouhan *et al.* 2010, Banala *et al.* 2018, Sudhakar and Reddy 2018); however, the fifth study (Tsunoda *et al.* 2005) observed increased dopamine and metabolites at fluoride exposures below 20 ppm (with statistical significance achieved only for the metabolite homovanillic acid in one brain region). No differences from the control group were observed at levels above 20 ppm fluoride. Other neurotransmitters were evaluated at or below 20 ppm fluoride exposure, but generally only in a couple of studies.

Biochemistry (brain/neurons)

Similar to above, the endpoints measured in brain biochemistry studies were too heterogeneous or limited in number to make any determination on potential relevance of mechanism, even before limiting the review of the data to lower risk-of-bias studies (see **Figure 10**). Endpoints related to biochemical changes in the brain or neurons included carbohydrate or lipid changes, RNA or DNA changes, changes in gene expression, or changes in protein expression. For the most part, only a single study was available for any given endpoint. The largest body of evidence on biochemistry was on protein level in various brain regions. Eleven lower risk-of-bias studies were identified that evaluated protein levels; however, few studies evaluated the same proteins or areas of the brain. In the few cases where the same protein was evaluated, results were not always consistent. These data are insufficient to increase confidence or support a change to hazard conclusions.

Histopathology

Histopathology of the brain was evaluated in 31 studies with concentrations at or below 20 ppm fluoride, of which 15 studies had a lower potential for bias (Adedara *et al.* 2017b, Akinrinade *et al.* 2015a, Bhatnagar *et al.* 2002, Bhatnagar *et al.* 2011, Chouhan *et al.* 2010, Guner *et al.* 2016, Jiang *et al.* 2014, Lou *et al.* 2013, McPherson *et al.* 2018, Mesram *et al.* 2016, Niu *et al.* 2018, Pulungan *et al.* 2016, Nageshwar *et al.* 2018, Zhao *et al.* 2019, Jia *et al.* 2019). In all but one lower risk-of-bias study [Pulungan *et al.* (2016); gavage], animals were exposed to fluoride via drinking water. All lower risk-of-bias studies were conducted in rodents, and all but three studies were conducted in rats (Wistar [seven studies]; Sprague-Dawley [four studies]; Long-Evans hooded [one study]). Overall, the lower risk-of-bias studies that evaluated histopathology in the brain had low potential for bias for key questions regarding randomization and exposure characterization; however, eight studies were rated as probably high risk of bias for the key risk-of-bias question regarding outcome assessment based on lack of reporting of blinding of outcome assessors and/or inadequate description of outcome measures or lesions. Moreover, low image quality in some of the studies hampered the ability to verify the quality of the data. Further technical review of the 15 lower risk-of-bias studies was conducted by a board-certified pathologist. Based on confidence in the results for each study, the technical reviewer further categorized the lower risk-of-bias studies as studies with higher or lower confidence in the outcome

assessment, which is reflected in the following summary of the brain histopathology results. Main limitations of the histopathology data identified by the pathologist included lack of information on methods of euthanasia and fixation. Perfusion fixation is generally considered the best practice for lesions of the central nervous system in addition to complete fixation of the brain prior to its removal from the skull (Garman *et al.* 2016). Four of the lower risk-of-bias studies reported that they used this method (Bhatnagar *et al.* 2002, Bhatnagar *et al.* 2011, McPherson *et al.* 2018, Pulungan *et al.* 2016). Two of the lower risk-of-bias studies handled the brains before fixation was complete, which can produce artifacts that can resemble dead neurons (Zhao *et al.* 2019, Nageshwar *et al.* 2018). Fixation and brain removal details were inadequately described in the remaining lower risk-of-bias studies.

Although there was heterogeneity in the endpoints reported (e.g., cell size, shape, and counts; nuclei fragmentation; increased vacuolar spaces) and some variation in the consistency of the evidence based on the area of the brain evaluated, the majority of the lower risk-of-bias studies (11 of 14 drinking water studies) found some histological change in the brain of rats or mice treated with fluoride at concentrations at or below 20 ppm, of which 8 studies reported histological changes in the brain at or below 10 ppm. Histological changes in the hippocampus (one of the areas of the brain most evaluated for histological changes) associated with fluoride exposures at or below 20 ppm were reported in three of four lower risk-of-bias studies with higher confidence in the outcome assessment (Bhatnagar *et al.* 2002, Bhatnagar *et al.* 2011, Guner *et al.* 2016) and in three of four lower risk-of-bias studies with lower confidence in the outcome assessment (Jiang *et al.* 2014, Niu *et al.* 2018, Nageshwar *et al.* 2018). McPherson *et al.* (2018) was the only drinking water study (with higher confidence in the histopathology outcome assessment) that did not observe any histological changes in hippocampus at 10 or 20 ppm fluoride in male Long-Evans hooded rats exposed in utero through adulthood (>PND80). Although there are too few studies to definitively explain the inconsistency in results, McPherson *et al.* (2018) also did not observe any associations between fluoride exposure and impairments to learning and memory, which is inconsistent with the majority of developmental exposure studies that observed learning and impairments associated with fluoride exposure for other strains of rats. Similarly, histological changes in the cortex were reported in three of the four lower risk-of-bias drinking water studies with higher confidence in the outcome assessment (Chouhan *et al.* 2010, Bhatnagar *et al.* 2011, Akinrinade *et al.* 2015a) and in three of four lower risk-of-bias studies with lower confidence in the outcome assessment (Lou *et al.* 2013, Mesram *et al.* 2016, Nageshwar *et al.* 2018).

Histological changes were also consistently reported in other areas of the brain in studies with higher confidence in the outcome assessment, including the amygdala, caudate putamen, cerebellum, and hypothalamus, although each of these areas of the brain were only evaluated in one lower risk-of-bias study (Bhatnagar *et al.* 2011, Guner *et al.* 2016). Pulungan *et al.* (2016), one of two lower risk-of-bias studies with higher confidence in the outcome assessment that did not report histological changes in the brain, observed a decreasing trend in the number of pyramidal cells in the prefrontal cortex with increasing dose, but this was not changed at concentrations below 20 ppm (study administered sodium fluoride via gavage; the 5-mg/kg-day dose was considered to be equivalent to 15.3 ppm fluoride in drinking water) nor were any of the results statistically significant.

Oxidative stress

Oxidative stress in the brain was evaluated in 25 studies that examined concentrations at or below 20 ppm fluoride, of which 15 studies had lower potential for bias (Adedara *et al.* 2017a, Adedara *et al.* 2017b, Akinrinade *et al.* 2015b, Chouhan *et al.* 2010, Gao *et al.* 2008b, Guner *et al.* 2016, Mesram *et al.* 2016, Nkpaa and Onyeso 2018, Shan *et al.* 2004, Zhang *et al.* 2015a, Chouhan and Flora 2008, Gao *et al.* 2009, Khan *et al.* 2017, Bartos *et al.* 2018, Nageshwar *et al.* 2018). All of the lower risk-of-bias studies

were conducted in rats (mainly Wistar or Sprague-Dawley) and administered fluoride via drinking water with exposure durations ranging from 28 days to 7 months. Although there was heterogeneity in the endpoints reported (i.e., varying measures of protein oxidation, antioxidant activity, lipid peroxidation, and reactive oxygen species [ROS]) and some variation in the consistency of the evidence based on the endpoint, the majority of the studies (13 of 15 studies) (Adedara *et al.* 2017a, Adedara *et al.* 2017b, Akinrinade *et al.* 2015b, Gao *et al.* 2008b, Gao *et al.* 2009, Guner *et al.* 2016, Mesram *et al.* 2016, Nkpaa and Onyeso 2018, Shan *et al.* 2004, Zhang *et al.* 2015a, Khan *et al.* 2017, Nageshwar *et al.* 2018, Bartos *et al.* 2018) found evidence of oxidative stress in the brains of rats treated with fluoride at concentrations at or below 20 ppm, of which 10 studies reported oxidative stress in the brain below 10 ppm fluoride. The most consistent evidence of oxidative stress in the brain was reported through changes in antioxidant activity. Eleven of the 12 lower risk-of-bias studies that evaluated antioxidant activity reported an effect at concentrations at or below 20 ppm (Adedara *et al.* 2017a, Adedara *et al.* 2017b, Akinrinade *et al.* 2015b, Gao *et al.* 2008b, Gao *et al.* 2009, Guner *et al.* 2016, Mesram *et al.* 2016, Nkpaa and Onyeso 2018, Khan *et al.* 2017, Bartos *et al.* 2018, Nageshwar *et al.* 2018). Decreases in antioxidant activity using measures of superoxide dismutase (SOD) activity were reported in seven of eight lower risk-of-bias studies (Adedara *et al.* 2017a, Adedara *et al.* 2017b, Akinrinade *et al.* 2015b, Mesram *et al.* 2016, Nkpaa and Onyeso 2018, Khan *et al.* 2017, Nageshwar *et al.* 2018) and, among these seven studies, all that also measured changes in catalase (CAT) activity (n = 6 studies) also reported decreased activity (Adedara *et al.* 2017a, Adedara *et al.* 2017b, Mesram *et al.* 2016, Nkpaa and Onyeso 2018, Khan *et al.* 2017, Nageshwar *et al.* 2018). A decrease in total antioxidant capacity (T-AOC) as a measure of antioxidant activity was also consistently reported in two lower risk-of-bias studies (Gao *et al.* 2008b, Gao *et al.* 2009), and a decrease in glutathione peroxidase (GPx) activity was reported in two of three lower risk-of-bias studies (Adedara *et al.* 2017b, Nkpaa and Onyeso 2018).

Relative to the above-mentioned studies, 2 of the 15 lower risk-of-bias studies (Chouhan and Flora 2008, Chouhan *et al.* 2010) did not observe statistically significant effects on oxidative stress in the brain with concentrations at or below 20 ppm fluoride; however, the measure of oxidative stress evaluated in Chouhan and Flora (2008) and Chouhan *et al.* (2010) (glutathione [GSH] to oxidized glutathione [GSSG] ratio as an indication of antioxidant activity and ROS levels) were not evaluated in any other lower risk-of-bias study. Chouhan and Flora (2008) observed a dose-dependent increase in ROS levels associated with 10, 50, and 100 mg/L sodium fluoride in the drinking water; however, results were not statistically significant at any dose. In Chouhan *et al.* (2010), the levels of ROS were significantly higher at 50 ppm sodium fluoride in drinking water, but statistical significance was not met at doses below 20 ppm fluoride (1 and 10 ppm sodium fluoride) or at 100 ppm sodium fluoride; yet, hydrogen peroxide levels as a measure of ROS were found to be significantly increased at 15 ppm sodium fluoride in drinking water in studies conducted by another group of authors (Adedara *et al.* 2017a, Adedara *et al.* 2017b).

Apoptosis/cell death

Seven lower risk-of-bias studies were identified that evaluated apoptosis with concentrations at or below 20 ppm fluoride. Results from these studies were inconsistent and were insufficient for evaluating fluoride-induced apoptosis. These data are insufficient to increase confidence or support a change to hazard conclusions.

Inflammation

Five lower risk-of-bias studies were identified that evaluated potential effects of fluoride on inflammation with concentrations at or below 20 ppm. The inflammation markers were too heterogeneous or limited in number to make any determination on potential relevance of mechanism,

even before limiting the review of the data to lower risk-of-bias studies. These data are insufficient to increase confidence or support a change to hazard conclusions.

Thyroid

Seventeen studies were identified that evaluated potential effects of fluoride on the thyroid with concentrations at or below 20 ppm (see **Figure 9**). These animal thyroid data are not further described because this endpoint has been directly evaluated in a number of human studies that have failed to identify consistent evidence to suggest that thyroid effects are a requisite mechanism by which fluoride causes neurodevelopmental or cognitive effects in humans.

Figure 10. Number of Lower Risk-of-bias Animal Studies that Evaluated Biochemical, Neurotransmission, and Oxidative Stress Effects at or Below 20 ppm by Mechanism Subcategory and Direction of Effect*

Mechanism	Mechanism Subcategory	Direction of Effect ↑	↓	NS	Grand Total
Biochemistry (brain/neurons)	Carbohydrate/lipid-related		1	1	2
	Gene expression		1		1
	Protein levels associated with brain function	8	7	7	11
	Other biochemical	2			2
Neurotransmitters	Cholinesterase	2	7	3	11
	Dopamine and metabolites		1		1
	Other neurotransmitters	2	2	1	2
Oxidative stress	Antioxidant activity	1	10	4	12
	Lipid peroxidation byproduct	7		4	11
	Protein oxidation	2		1	2
	ROS	2		2	4

*Interactive figure and additional study details in Tableau® (https://public.tableau.com/profile/ntp.visuals#!/vizhome/Fluoride_Animal_SelectMechanisms_UPDATE/Figure9). This figure displays study counts for lower risk-of-bias studies, as these counts are most relevant to the text in this section. Counts for higher risk-of bias studies or all studies combined can be accessed in the interactive figure in Tableau®. Study counts are tabulated by significance—statistically significant increase (↑), statistically significant decrease (↓), or not significant (NS). For example, the "↑" column displays numbers of unique studies with at least one endpoint in the mechanistic subcategory with significantly increasing results at fluoride exposure levels of ≤20 ppm. These columns are not mutually exclusive (i.e., a study may report on multiple endpoints with varying results within a single mechanistic subcategory and therefore may be reflected in the counts for the "↑", "↓", and NS columns, but would only be counted once in the Grand Total column). Endpoints, species, strain, sex, and exposure duration are available for each study in the interactive figure in Tableau®.

In Vitro/Mechanistic Data on Neurodevelopmental or Cognitive Effects

Although in vitro data were collected as part of the systematic review process, NTP determined that the information on neurological effects obtained from these studies is too general, and results cannot necessarily be attributed to effects on learning and memory or other cognitive functions at this time. The in vitro data may help support specific mechanisms identified from in vivo mechanistic data; however, as described above, no specific mechanism has been determined for fluoride effects on learning and memory or other neurodevelopmental or cognitive outcomes.

Evidence Synthesis for Neurodevelopmental or Cognitive Effects

There is consistent evidence that exposure to fluoride is associated with cognitive neurodevelopmental effects in children. There is moderate confidence in the human data in children from 5 well-conducted

prospective studies, supported by 14 cross-sectional studies where exposure was identified as likely occurring prior to outcome. The human body of evidence in adults is considered inadequate to evaluate whether fluoride exposure is associated with cognitive effects due to low confidence in the human data in adults, a limited number of studies, and a lack of evidence of an effect (i.e., there is not sufficient evidence of an effect, but the confidence in the data is not high enough to conclude that there is no effect). The animal data are inadequate to evaluate for learning and memory effects primarily due to the uncertainty in distinguishing effects on cognitive outcomes from secondary effects on the nervous system or general health including motor activity issues; however, these data do provide evidence of other neurodevelopmental effects. There is also evidence from mechanistic studies of adverse neurological effects of fluoride in humans and animals of unknown relationship to cognition.

The initial moderate confidence is based on 19 studies where exposure occurred prior to outcome and that evaluated individual-based outcomes and used a comparison group. Factors considered for upgrading or downgrading the confidence are as follows:

- **Risk of bias**: Only studies that were considered to have lower risk of bias were included in the moderate confidence rating; therefore, there is no downgrade for risk-of-bias concerns.

- **Unexplained inconsistencies:** The data are relatively consistent and there was no downgrade for this factor. In terms of IQ data, 17 studies observed significant effects associated with fluoride, and 2 studies found no significant association but neither of these studies adjusted for confounders. Consistency among neurodevelopmental effects other than IQ was also considered; however, the conclusions are based on the IQ data so these other neurodevelopmental effects would not impact a potential adjustment in confidence. Studies measuring neurodevelopmental effects other than IQ did not show consistent effects. It is not known whether fluoride exposure would be expected to be associated with neurodevelopmental outcomes in addition to IQ or other cognitive measures.

- **Indirectness:** IQ in humans is a direct measure of effect and therefore no adjustment in confidence is warranted.

- **Imprecision:** The meta-analysis indicates that there was no reason to downgrade due to imprecision.

- **Publication bias:** While the meta-analysis that estimated the pooled SMD among 46 included studies (both higher and lower risk-of-bias) indicated that there was potential for publication bias, a subgroup analysis indicated that there was no publication bias among the lower risk-of-bias studies (see **Figure A5-8**). Among the higher risk-of-bias studies, the trim-and-fill analysis estimated that, in the absence of publication bias, the negative direction of effect and statistical significance remained (**Figure A5-9**). For the meta-analysis that calculated a pooled effect estimate among the studies with individual-level measures, the funnel plot indicated publication bias; however, the trim-and-fill analysis estimated that once adjusted for publication bias, the negative direction of effect remained (**Appendix 5**, **Figure A5-16** and **Figure A5-18**). Therefore, no downgrade was applied for publication bias.

- **Large magnitude of effect:** While some individual studies indicate a large magnitude of effect, the overall pooled effect estimate from the meta-analysis of studies with individual-level

measures does not demonstrate a large magnitude of effect (**Appendix 5**). Therefore, the overall data would not support an upgrade due to a large magnitude of effect.

- **Dose-response:** Linear dose-response models provide the best fit to the data in studies examining individual-level measures of fluoride exposure and IQ. A meta-analysis of studies that compared mean IQ scores between groups of children with different levels of fluoride exposure showed a significantly lower mean SMD at higher concentrations of fluoride (>1.5 mg/L) from water (SMD: −0.14; 95% CI: −0.19, −0.08; n = 31 studies) and urine (SMD: −0.18; 95% CI: −0.31, −0.05; n = 22 studies) (**Appendix 5**); however, the dose-response relationship at fluoride concentrations below 1.5 mg/L fluoride in urine or drinking water is less certain. The overall dose-response could be used to upgrade the confidence in the body of evidence.

- **Residual confounding:** Xiang *et al.* (2003a), Xiang *et al.* (2011), and Wang *et al.* (2012) studied the same population where arsenic occurred in the area with low fluoride, but did not occur in the area with high fluoride. This would have biased the results toward the null, but there was a significantly lower IQ scores in the area with high fluoride. The remaining studies do not provide enough information to consider residual confounding as an impactful factor for the body of evidence. Therefore, the overall data would not support an upgrade due to residual confounding.

- **Consistency:** There is consistent evidence across study populations and study designs that fluoride is associated with lower IQ scores at higher concentrations of fluoride. There is uncertainty and less of a consistent pattern at concentrations below 1.5 mg/L. There is also a lack of consistency observed with and among other types of neurodevelopmental effects. The consistency in the overall results of the data set could increase the confidence.

Summary judgement on potential upgrades or downgrades in the confidence: Although the OHAT approach for evidence integration allows for the initial confidence in the body of evidence to be increased based on consistency or dose response, the NTP judgement is that the magnitude of effect and the overall strength and quality of the human literature base provides a moderate confidence in the body of evidence that fluoride causes cognitive neurodevelopmental effects in children.

The moderate confidence in the body of evidence in children translates to a moderate level of evidence that fluoride is associated with lower IQ and other cognitive neurodevelopmental effects in children.

The limited and weaker evidence of cognitive effects in adults is considered to provide an inadequate level of evidence that fluoride is associated with cognitive effects in adults. The animal body of evidence is also considered to provide an inadequate level of evidence for cognitive effects in adults.

Integration of these level-of-evidence conclusions supports an initial hazard conclusion of *presumed to be a cognitive neurodevelopmental hazard to humans* because of the extent and consistency of effect in the available data in children. Because most of the available studies evaluated intelligence in children, the primary focus in human data was on IQ and other cognitive neurodevelopmental effects, which is the primary basis for the hazard conclusion. A separate conclusion on other neurodevelopmental effects was not reached based on limited information in humans.

The moderate level of evidence in the human data in children supports a hazard conclusion of *presumed* instead of *suspected* due to the relatively large and consistent body of evidence, especially in relation to measures of IQ (17 of 19 lower risk-of-bias studies that assessed IQ reported an association between

higher fluoride and lower IQ scores) across multiple populations. A conclusion of *presumed* is supported by a statistically significant effect observed in the meta-analysis. Furthermore, the *presumed* hazard conclusion is supported by the low expectation that new studies would decrease the hazard conclusion.

Effects in children

- **Human body of evidence:** Moderate Confidence = Moderate Level of Evidence
- **Animal body of evidence:** Overall poor quality of studies and few studies that specifically assess effects on learning and memory after exposure during developmental periods separately from other neurological effects including motor activity = Inadequate Level of Evidence
- **Initial hazard conclusion (Moderate Human x Inadequate Animal)** = Presumed to be a Cognitive Neurodevelopmental Hazard to Humans
- **Final hazard conclusion (after consideration of biological plausibility)** = Presumed to be a Cognitive Neurodevelopmental Hazard to Humans

Effects in adults

- **Human body of evidence:** Low Confidence with no discernible effect = Inadequate Level of Evidence
- **Animal body of evidence:** Overall poor quality of studies and few studies that specifically assess effects on learning and memory after exposure in adulthood separately from other neurological effects including motor activity = Inadequate Level of Evidence
- **Initial hazard conclusion (Inadequate Human x Inadequate Animal)** = Not classifiable
- **Final hazard conclusion (after consideration of biological plausibility)** = Not classifiable

Table 7. Neurodevelopmental and Cognitive Function Evidence Profile for Fluoride

INITIAL CONFIDENCE for each body of evidence (# of studies)	Factors decreasing confidence "---" if no concern; "↓" if serious concern to downgrade confidence					Factors increasing confidence "---" if not present; "↑" if sufficient to upgrade confidence				FINAL CONFIDENCE RATING
	Risk of Bias	Unexplained Inconsistency	Indirectness	Imprecision	Publication Bias	Large Magnitude	Dose Response	Residual Confounding	Consistency Species/Model	
*Human IQ or cognitive function tests in children**										
Initial Moderate (5 prospective cohort studies[a]; 14 cross-sectional studies[b])	---	---	---	---	---	---	---	---	---	**Moderate**
Initial Low (7 cross-sectional studies)[c]	---	---	---	---	---	---	---	---	---	**Low**
*Human IQ or cognitive function tests in adults***										
Initial Low (2 cross-sectional studies)[d]	---	---	---	---	---	---	---	---	---	**Low**
Animal learning and memory or cognitive function										
Inadequate to assess effects in human										

References:

Human: Barberio *et al.* (2017b)[c]; Bashash *et al.* (2017)[a]; Bashash *et al.* (2018)[a]; Choi *et al.* (2015)[b]; Cui *et al.* (2018)[c]; Cui *et al.* (2020)[c]; Das and Mondal (2016); Ding *et al.* (2011)[b]; Green *et al.* (2019)[a]; Jacqmin *et al.* (1994)[d]; Li *et al.* (2004) [translated in Li *et al.* 2008a] [b]; Li *et al.* (2016)[d]; Riddell *et al.* (2019)[c]; Rocha-Amador *et al.* (2007)[b]; Rocha-Amador *et al.* (2009)[b]; Saxena *et al.* (2012)[b]; Seraj *et al.* (2012)[b]; Soto-Barreras *et al.* (2019)[b]; Sudhir *et al.* (2009)[b]; Till *et al.* (2020)[a]; Trivedi *et al.* (2012)[c]; Valdez Jimenez *et al.* (2017)[a]; Wang *et al.* (2012)[b]; Wang *et al.* (2020b)[b]; Wang *et al.* (2020a)[c]; Xiang *et al.* (2003a)[b]; Xiang *et al.* (2011)[b]; Yu *et al.* (2018)[b]; Zhang *et al.* (2015b)[c]

*This includes learning disabilities, neonatal behavioral neurological assessment, mental development index, memory score for copy, and immediate recall.

**This includes Mini-Mental State Examination scores, psychomotor performance, and memory.

DISCUSSION

The overall objective of this evaluation was to undertake a systematic review of published literature to reach conclusions concerning the potential for exposure to fluoride to affect neurodevelopment and cognition. This review only addresses whether exposure to fluoride could present a potential hazard (i.e., has the potential to cause harm at any exposure level, including exposures that are higher than typically encountered from consuming fluoridated drinking water in the United States). Benefits of fluoride with respect to oral health are not addressed in this monograph.

Given this context, when focusing on human epidemiology studies with exposures in ranges typically found in the water distribution systems in the United States (0.7 mg/L for optimally fluoridated community water systems)[8] that can be evaluated for dose response, effects on cognitive neurodevelopment are inconsistent and therefore unclear. However, given the totality of the data, including studies with exposures to fluoride levels higher than the WHO safe water guideline of 1.5 mg/L in water (WHO 2011), the NTP concludes that fluoride is *presumed to be a cognitive neurodevelopmental hazard to humans.* This conclusion is based on a moderate level of evidence that shows a consistent and robust pattern of findings in human studies across several different populations demonstrating that higher fluoride exposure (e.g., >1.5 mg/L in drinking water) is associated with lower IQ and other cognitive effects in children. Limited and weaker evidence is considered to provide an inadequate level of evidence that fluoride is associated with cognitive effects in adults. The primary focus of the human data was on IQ and cognitive neurodevelopmental effects; therefore, the conclusion was based on these data. After further evaluation of the experimental animal data available in NTP (2016) and in this systematic review, NTP concludes that in terms of evaluating the effects of fluoride on learning and memory to support the cognitive effects observed in humans, the animal data are inadequate. The animal data do provide evidence for effects of fluoride on neurodevelopment; however, other neurodevelopmental outcomes were not further evaluated because of the limited information in humans. Biological plausibility of effects from mechanistic studies was considered but did not significantly influence the conclusions. Although multiple categories of mechanistic data were evaluated and provide some evidence of adverse effects in the brain, a coherent series of mechanistic events to account for fluoride-associated cognitive neurodevelopmental deficits is not sufficiently understood for these findings to contribute to the overall confidence assessment.

The human body of evidence provides a consistent and robust pattern of findings that higher fluoride exposure is associated with lower IQ scores in children. The moderate level of evidence is based on 5 lower risk-of-bias prospective cohort studies and 14 lower risk-of-bias cross-sectional studies that are considered to have sufficient evidence of fluoride exposure occurring prior to the outcome. The evaluation of the animal body of evidence in this assessment is an extension of the NTP (2016) systematic review on the association between fluoride exposure and neurobehavioral effects related to learning and memory in animals, which identified a concern related to indirectness. This concern was that many of the learning and memory tests rely on a motor response (e.g., latency to achieve the desired effect). The review of animal data published since the 2016 review focused on addressing this

[8]As of April 2020, 1.08% of persons living in the United States (~ 3.5 million people) were served by CWS containing ≥ 1.1 mg/L naturally occurring fluoride. CWS supplying water with ≥ 1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population (~ 1.9 million people), and systems supplying water with ≥ 2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people) (https://www.cdc.gov/fluoridation/data-tools/reporting-system.html).

indirectness concern. Further examination of the literature has not provided clarification of this issue. Due to the inability to separate these effects from effects on general health and other effects on the nervous system, the animal body of evidence is now considered inadequate to contribute to the evaluation of cognitive effects in humans. Although the animal data are not considered sufficient to specifically support the IQ changes observed in children, the data do support possible neurodevelopmental effects.

The NTP conclusion that fluoride is *presumed to be a cognitive neurodevelopmental hazard to humans* is supported by the extent, consistency, and robustness of the effect in the available data in children. Seventeen of the 19 lower risk-of-bias studies reported an association between higher fluoride exposure and lower IQ scores in children across multiple populations. Meta-analyses conducted at the recommendation of NASEM based on their review of the September 6, 2019 draft monograph provide further support for the hazard conclusion of *presumed* (NASEM 2020). The random-effects pooled SMD estimate from the 46 studies included in the group-level meta-analysis was consistent with two previous meta-analyses reporting statistically significant associations between higher fluoride exposure and lower IQ in children. A risk-of-bias subgroup analysis demonstrated that the significant negative association remained when the meta-analysis was restricted to 9 lower risk-of-bias studies. Further subgroup analyses by gender, age group, country, outcome assessment type, and exposure assessment type support a consistent and robust pattern of results. A second meta-analysis of the individual-level data from six lower risk-of-bias studies also provided evidence of a statistically significant negative association between fluoride exposure and lower IQ in children (overall pooled effect estimate per 1-mg/L increase in urinary fluoride was associated with a 1.40-point lower IQ score [95% CI: −2.33, −0.47]). Given the evidence, there is a low expectation that new studies would change the hazard conclusion.

There are few studies in humans and numerous studies in animals that evaluate mechanistic effects related to fluoride exposure. There are sufficient mechanistic data to determine that fluoride exposure at lower concentrations has effects on the nervous system; however, for the cognitive neurodevelopmental outcome evaluated, there are insufficient data to support a specific mechanism or mode of action. Due to the large number of mechanistic studies conducted in animals, evaluation of the mechanistic data in animals focused on studies that had exposures more relevant to humans (i.e., ≤20 ppm in the drinking water). Changes in AChE, which could potentially be related to cognitive effects such as IQ, were evaluated in one study of children and several studies of animals (measured in both the blood and in areas of the brain); however, the majority of these studies, including the study of children, reported results inconsistent with the phenotypic outcome. Animal studies that evaluated changes in other neurotransmitters and other biochemical measures provide some evidence of effects in the brain, but the data are limited due to the heterogeneity of the outcomes measured. Most consistently, studies evaluating histopathology and oxidative stress demonstrated that effects can occur in the brains of animals at or below 20 ppm, which, without supporting a specific mechanism or mode of action relevant to learning and memory impairments, provides evidence of an association between exposure to lower concentrations of fluoride and neurological effects in animals. Therefore, the evidence of neurological effects at exposure levels more relevant to humans that is demonstrated in the mechanistic data supports the NTP conclusion that fluoride is *presumed to be a cognitive neurodevelopmental hazard to humans;* however, it does not provide enough evidence to increase confidence in the human body of evidence or support a higher hazard identification conclusion.

Generalizability to the U.S. Population

For many years, fluoride concentrations were adjusted to levels between 0.8 and 1.2 mg/L in fluoridated community water systems in the United States. The U.S. Public Health Service recommended an adjustment downward to a fluoride concentration of 0.7 mg/L because of evidence of an increase in dental fluorosis in children (US DHHS 2015). In April 2020, the CDC Water Fluoridation Reporting System estimated that the majority (i.e., 97.5%) of fluoride concentrations in water for U.S. children and adolescents (≤19 years old) are below 1.2 mg/L (https://www.cdc.gov/fluoridation/data-tools/reporting-system.html).

NTP's conclusion that fluoride is *presumed to be a cognitive neurodevelopmental hazard to humans* is based on consistent evidence from 26 lower risk-of-bias studies that evaluated fluoride exposure and effects on children's IQ and other cognitive effects. Although there are many studies that evaluated associations between fluoride in the drinking water and IQ in children, no studies evaluating IQ were conducted in the United States. Generalizing the results from the IQ studies in this evaluation to the U.S. population can be difficult, in part because many studies were conducted in areas with fluoride drinking water concentrations that are much higher than drinking water fluoride concentrations in the United States. Among the human body of evidence evaluated for this assessment (including lower and higher risk-of-bias studies), there are 33 studies that evaluated associations between fluoride in drinking water and IQ in children and compared a reference or low exposure group to higher fluoride-exposed groups. Of these 33 studies, only 10 studies included fluoride exposure groups with fluoride concentrations < 1.5 mg/L (i.e., fluoride exposure groups that would potentially be relevant to levels observed in the United States) (Xu *et al.* 1994, Xiang *et al.* 2003a, Qin *et al.* 1990 [translated in Qin *et al.* 2008], Kang *et al.* 2011, Broadbent *et al.* 2015, Sebastian and Sunitha 2015, Sudhir *et al.* 2009, Zhang *et al.* 2015b, Zhang *et al.* 1998, Wang *et al.* 2020b). Of these 10 studies, 4 were considered to have lower risk of bias (Zhang *et al.* 2015b, Xiang *et al.* 2003a, Wang *et al.* 2020b, Sudhir *et al.* 2009).

In addition to the four studies mentioned above, several other studies that evaluated fluoride exposure on a continuous basis could be used to assess generalizability to the United States. This includes studies that examined fluoride exposure levels below 1.5 mg/L for which a dose response could be assessed. **Table 8** provides a summary of children's IQ studies that evaluated lower fluoride exposures (<1.5 mg/L) in drinking water and/or urine (assuming, for comparison purposes, an approximate 1-to-1 equivalence between drinking water fluoride and urinary fluoride concentrations) and provided information to evaluate dose response in the lower fluoride exposure range (e.g., three or more fluoride exposure groups or dose-response curve provided). Based on review of these studies (discussed further below), there is uncertainty if IQ changes in children occur at lower fluoride levels.

Among studies with lower risk of bias for which a dose response could be assessed, four of nine studies that examined fluoride exposure levels below 1.5 mg/L (**Table 8)** (Green *et al.* 2019, Zhang *et al.* 2015b, Xiang *et al.* 2003a, Wang *et al.* 2020b) applied regression models to individual exposure outcome measures and observed a linear association between urinary fluoride levels and lower IQ in children even at the lower fluoride concentrations. However, two of these studies (Xiang *et al.* 2003a, Zhang *et al.* 2015b) did not find an association between IQ and drinking water levels below 1.5 mg/L. Xiang *et al.* (2003a) observed a significantly lower IQ in endemic versus nonendemic villages, but when they grouped children from the endemic villages by exposure level, they did not observe a significantly lower IQ score for children exposed to lower mean exposure levels of fluoride (0.75 mg/L). Although a significant difference in IQ might not be expected due to the fact that there were only nine children in this group, the difference was less than one point in IQ. Zhang *et al.* (2015b) used a simple correlation

and did not observe a significant relationship between fluoride levels in the drinking water (with concentrations up to 1.57 mg/L) and IQ. Sudhir *et al.* (2009) observed a significant increase in IQ grade (which is associated with lower IQ) at concentrations of 0.7–1.2 mg/L. The other four of nine studies do not provide a clear dose response at the lower fluoride levels. Bashash *et al.* (2017) concluded that there was no clear association between IQ scores and maternal urinary fluoride below 0.8 mg/L. Yu *et al.* (2018) observed a correlation between lower IQ in children and fluoride exposure only with concentrations in drinking water above 3.4 mg/L or with urinary fluoride concentrations of 1.6 mg/L or higher. The study authors did note a decreased probability of having an IQ above 130 (i.e., 40% fewer people with high IQ for every 0.5-mg/L increase in fluoride) with water fluoride levels between 0.20 and 1.40 mg/L, but this was not observed with higher levels of fluoride. Although Cui *et al.* (2018) noted that IQ decreased in a "roughly linear manner" with increasing urinary fluoride, this is only apparent in the results for the TT genotype; based on the dose response, the authors concluded that the "safety threshold" was 1.73 mg/L. Ding *et al.* (2011) looked at mean differences for 10 different exposure groups and found notable decreases from the mean above approximately 1 mg/L.

Although there is less confidence in the findings from higher risk-of-bias studies, six studies identified with potential dose-response information demonstrated a similar uncertainty at the lower fluoride concentrations. Surprisingly, three of the studies (Aravind *et al.* 2016, Qin *et al.* 1990 [translated in Qin *et al.* 2008], Xu *et al.* 1994) found that the lowest IQ scores were in areas with the lowest and the highest fluoride concentrations. In these studies, the lowest fluoride concentrations ranged from 0.1–0.2 mg/L fluoride in Qin *et al.* (1990) [translated in Qin *et al.* 2008] to <1.2 mg/L in Aravind *et al.* (2016). Li *et al.* (1995) and Sebastian and Sunitha (2015) only observed lower IQ scores at concentrations above 2 mg/L. Das and Mondal (2016) found a steady decline in IQ with increasing urinary fluoride levels or exposure dose.

To further examine the dose response for lower IQ in the lower exposure region (e.g., <1.5 mg/L fluoride in drinking water or urine), a meta-analysis (**Appendix 5**) using a linear mixed model to analyze mean-effect estimates was performed. Twelve observations from 9 studies were included that reported one or more IQ measurements associated with drinking water fluoride exposures of < 1.5 mg/L and a reference group. This analysis did not show a statistically significant association with the mean SMD in children's IQ scores between exposed and reference groups (SMD = 0.32; 95% CI: −0.57, 1.20). A dose-response meta-analysis including seven observations from four studies with at least one exposure group < 1.5 mg/L urinary fluoride showed a non-statistically significant decrease in mean SMD (SMD = −0.13; 95% CI: −0.29, 0.03). Based on these results, effects of fluoride exposure on children's IQ at levels < 1.5 mg/L remain unclear and more studies at lower exposure levels are needed. A dose-response meta-analysis of studies with individual-level data could not be conducted due to the small number of studies (n = 10), the various types of exposure metrics, and the different types of reported effect estimates. More studies with lower levels of fluoride exposure from drinking water are still needed to fully understand potential effects at exposures in ranges typically found in the United States (i.e., <1.5 mg/L). Of note, the negative association between IQ and fluoride exposure via drinking water was statistically significant when extending the dose-response meta-analysis to include IQ measures from groups exposed to <2 mg/L in drinking water (SMD = −0.27; 95% CI: −0.36, −0.17). A statistically significant decrease in mean SMD in children's IQ scores was not seen in urinary fluoride measures of < 2 mg/L (SMD = −0.09; 95% CI: −0.22, 0.03) (**Table A5-3**).

When generalizing findings from the cited studies to exposures in the United States from fluoride in drinking water, it is important to consider that drinking water only comprises a portion of total exposures to fluoride. Although it can be assumed that children in all the studies cited in this document

are also exposed to fluoride from sources other than drinking water, these other exposures are likely to vary considerably depending on individual circumstances. Fluoride concentrations in drinking water alone do not reflect the magnitude of fluoride exposures to children who consume excessive amounts of fluoridated toothpaste or to formula-fed babies who consume powdered formula that is reconstituted with fluoridated water. A few studies also support the possibility of heightened sensitivities to the detrimental cognitive effects of fluoride exposures in individuals with certain polymorphisms in dopamine receptor D2, or catechol-O-methyltransferase (Cui *et al.* 2018, Zhang *et al.* 2015b), impacting dopamine catabolism and receptor sensitivity. Differential exposures to fluoride and genetic susceptibilities of children to fluoride appear to warrant further research.

Table 8. Human Studies with Lower Fluoride Exposure and Effects on IQ		
Study	**Exposure measures [mean ± SD (range)]**	**Notes**
Lower risk-of-bias studies		
Bashash *et al.* (2017)	Maternal urine during pregnancy (mg/L) 0.90 ± 0.35 (0.23–2.36)	Authors concluded that the model suggested a nonlinear relationship with no clear association between IQ scores and maternal urine below 0.8 mg/L.
	Children's urine (mg/L) 0.82 ± 0.38 (0.18–2.8)	
Green *et al.* (2019)	Maternal urine during pregnancy (mg/L) 0.51 ± 0.36 (0.06–2.44) 0.40 ± 0.27 non-fluoridated areas 0.69 ± 0.42 fluoridated areas	Statistical methods indicated that including quadratic or natural-log effects of maternal urine or intake did not significantly improve the model. In addition, the authors tested separate models with two linear splines to see if the effect of maternal urinary fluoride or maternal fluoride intake significantly differed between lower and higher levels based on knots set at 0.5, 0.8, and 1.0 mg/L for urine and 0.4, 0.8, and 1 mg for intake. There were no differences.
	Maternal intake during pregnancy (mg/day) 0.54 ± 0.44 (0.01–2.65) 0.30 ± 0.26 non-fluoridated areas 0.93 ± 0.43 fluoridated areas	
	Drinking water (mg/L)* 0.31 ± 0.23 (0.04–0.87[1]) 0.13 ± 0.06 non-fluoridated areas 0.59 ± 0.08 fluoridated areas	
Cui *et al.* (2018)	Drinking water (mg/L)* 0.20–1.00 non-endemic 1.52–2.49 endemic	Study authors noted that the IQ decreased in a "roughly linear manner as the log-urine fluoride increased." TT genotypes of the dopamine receptor D2 gene had the strongest negative correlation between log-urine fluoride and IQ scores. The study authors determined a safety threshold of urine fluoride levels in subgroup TT as 1.73 mg/L. Drinking water fluoride levels were used to select children from different areas but were not used in the analysis.
	Children's urine Levels not provided; log-transformed with range of approximately −1.2–2.2	
Ding *et al.* (2011)	Drinking water (mg/L) * 1.31 ± 1.05 (0.24–2.84)	Although there was a significant correlation between urinary fluoride and IQ score, the main drop in IQ occurred at urinary fluoride levels of approximately 0.7–1.2 mg/L. At levels below 0.7 mg/L, data suggest a plateau with no apparent change in IQ compared with the mean. Drinking water fluoride levels were not used in the analysis.
	Children's urine (mg/L) 0.10–3.55	

Table 8. Human Studies with Lower Fluoride Exposure and Effects on IQ

Study	Exposure measures [mean ± SD (range)]	Notes
Sudhir *et al.* (2009)	Drinking water (mg/L)* <0.7 (level 1 villages) 0.7–1.2 (level 2 villages) 1.3–4.0 (level 3 villages) >4.0 (level 4 villages)	The number of intellectually impaired children gradually increased with increasing fluoride concentration in drinking water with an increase in IQ grade (which indicates a decrease in IQ) observed in the 0.7–1.2-mg/L villages. Children were placed in exposure groups based on Water Works Department records. Although children brought in water for verification of strata, it was not collected from all children but only from the first child using a different source of water. Therefore, these are considered group-level data. Note that all groups had a large proportion in the "intellectually impaired" category.
Wang *et al.* (2020b)	Drinking water (mg/L)* 1.39 ± 1.01 (0.20–3.90) Children's urine (mg/L) 1.28 ± 1.30 (0.01–5.54)	There was a significantly lower IQ in quartile 3 (1.00–1.90 mg/L) and quartile 4 (>1.90 mg/L), but not quartile 2 (0.70–1.00 mg/L) of drinking water. Urinary fluoride was only associated with a significantly lower IQ in quartile 4 (>2.28 mg/L).
Xiang *et al.* (2003a)	Drinking water (mg/L)* 0.36 ± 0.15 (0.18–0.76) non-endemic village 2.47 ± 0.79 (0.57–4.50) endemic village Endemic subgroups: group A: 0.75 ± 0.14 group B: 1.53 ± 0.27 group C: 2.46 ± 0.30 group D: 3.28 ± 0.25 group E: 4.16 ± 0.22 Children's urine (mg/L) 1.11 ± 0.39 (0.37–2.50) non-endemic village 3.47 ± 1.95 (0.90–12.50) endemic village	IQ in group A in the endemic village was not significantly lower than the non-endemic village, but IQ in all other groups was significantly lower. Although there were only 9 children in group A, the IQ difference was <1 point. Based on simple regression, there was a steady decline in IQ with increasing urinary fluoride.
Yu *et al.* (2018)	Drinking water (mg/L)* 0.50 ± 0.27 normal 2.00 ± 0.75 high Children's urine (mg/L) 0.41 ± 0.49 normal 1.37 ± 1.08 high	Study authors reported that participants' intelligence presented inverse non-linear dose-response relationships with fluoride content, with obvious decreases at relatively high level of fluoride exposure (drinking water fluoride levels at 3.4–3.90 mg/L and urinary fluoride levels at 1.60–2.50 mg/L). Study authors also note a decreased odds for having IQ ≥ 130 with drinking water fluoride at 0.20–1.40 mg/L (40% decrease with each 0.5-mg/L increase in fluoride), but not at higher concentrations.

Table 8. Human Studies with Lower Fluoride Exposure and Effects on IQ		
Study	**Exposure measures [mean ± SD (range)]**	**Notes**
Zhang *et al.* (2015b)	Drinking water (mg/L)* 0.63 (0.58–0.68) control 1.40 (1.23–1.57) endemic fluorosis	There was a steady decline in IQ with increasing serum or urinary fluoride levels. A simple correlation did not find drinking water fluoride significantly correlated with IQ.
	Children's urine (mg/L) 1.1 ± 0.67 control 2.4 ± 1.01 endemic fluorosis	
	Children's serum (mg/L) 0.06 ± 0.03 control 0.18 ± 0.11 endemic fluorosis	
Higher risk-of-bias studies		
Aravind *et al.* (2016)	Drinking water (mg/L)* <1.2 low fluoride area 1.2–2 medium fluoride area >2 high fluoride area	Mean IQ scores (transformed into percentiles) were highest in the medium fluoride area for both boys and girls.
Das and Mondal (2016)	Groundwater (mg/L)* 2.11 ± 1.64 (0.25–9.40)	Based on simple regression, there was a steady decline in IQ with increasing urinary fluoride and increasing exposure dose. Groundwater fluoride levels were not used in the analysis but were used in calculating the children's exposure dose.
	Children's urine (mg/L) 0.45–17.00	
	Children's exposure dose (mg/kg-day) 0.017–0.203	
Li *et al.* (1995)	Children's urine (mg/L) 1.02 non-fluorosis area 1.81 slight fluorosis area 2.01 medium fluorosis area 2.69 severe fluorosis area	A significantly lower IQ score was observed in the medium and severe fluorosis areas compared to the non-fluorosis area. Children's urine was used as an individual measure of exposure to verify that the areas had different fluoride exposure levels; however, analysis was conducted based on residential area.
Qin *et al.* (1990) [translated in Qin *et al.* 2008]	Drinking water (mg/L)* 0.1–0.2 low fluoride area 0.5–1.0 normal fluoride area 2.1–4.0 high fluoride area	Average IQ scores (transformed into percentages) were significantly lower in both the low and high fluoride areas compared with the normal fluoride area.
Sebastian and Sunitha (2015)	Drinking water (mg/L)* 0.40 (low fluoride village) 1.2 (normal fluoride village) 2.0 (high fluoride village)	A significantly lower mean IQ score of children living in the high fluoride area compared with the low and normal fluoride areas was reported. Binary regression models using the low fluoride village as a reference observed an increased odds ratio (1.74) for increased IQ scores in the normal fluoride village and a decreased odds ratio (0.59) in the high fluoride village.
Xu *et al.* (1994)	Drinking water (mg/L)* 0.8 (control area) 0.38 (low fluoride area) 1.8 (high fluoride area)	Both low and high fluoride areas had IQ levels approximately 3 points below the IQ levels in the control area. There was no difference in IQ between the low and high fluoride areas.

*Data are group-level exposure data; exposure data without the asterisks are individual-level exposure data.

[1]Range data were obtained from Till *et al.* (2018).

The conclusion that fluoride is *presumed to be a cognitive neurodevelopmental hazard in children* is based on the consistency of the data; however, most lower risk-of-bias studies observed effects with drinking water concentrations above 1.5 mg/L. As noted above, describing the effects at 1.5 mg/L or below, which is more relevant to the exposures observed in the U.S. population, including from community water fluoridation, is more difficult. In the reviewed studies, when limiting studies to those that evaluated IQ at fluoride levels across a continuum that included the low dose range, results are less consistent.

Limitations of the Evidence Base

Few limitations exist in the lower risk-of-bias epidemiological studies used for the basis of the hazard conclusion. The main limitations in lower risk-of-bias epidemiological studies include:

- Few studies were available that assessed the association between fluoride exposure and the following:
    - Neurodevelopmental or cognitive effects in subjects from communities served by optimally fluoridated versus non-fluoridated water systems.
    - Neurobehavioral (i.e., cognitive) effects (particularly IQ) in adults.
    - Attention-related disorders including ADHD.
- Studies rarely separated the results by gender or provided information to indicate that gender was not a modifying factor, which limits the ability to evaluate how the association between fluoride exposure and cognitive neurodevelopmental effects in children might differ by gender.

Limitations in the epidemiological studies with higher risk of bias include:

- Many of the original publications were in a foreign language and provided limited details on methodology.
- Some studies lacked information regarding exposure and/or had serious limitations in the exposure assessment. Exposure assessment concerns include limited individual exposure information, a lack of information on fluoride sampling methods and timing of the exposure measurements, a lack of quantitation of levels of fluoride in drinking water, and a lack of individual-level information on fluorosis in areas reported to be endemic for fluorosis.
- The comparison groups in studies conducted in areas endemic for fluorosis may have still been exposed to high levels of fluoride or levels similar to those used in water fluoridation in the United States. This factor may have limited the ability to detect true effects.
- Most studies did not provide sufficient direct information (e.g., participation rates) to evaluate selection bias.
- Failure to address potential confounders was a main issue. Many studies conducted simple statistical analyses without accounting for any potential confounders. In cases where adjustments in analyses were made, often these studies did not account for potential confounders considered critical for that study population and outcome.

- Studies conducted in areas with high, naturally-occurring fluoride levels in drinking water often did not account for potential exposures to arsenic or iodine deficiencies in study subjects.
- Many studies did not account for potential exposures to lead as a residual confounder.
- Many studies lacked information on whether the outcome assessors were blind to the exposure group, including studies that examined children in their schools and subjects from high-fluoride communities.

Limitations in the animal studies include:

- The main limitation in the animal studies was the inability to separate possible learning and memory effects from effects on motor activity/coordination or sensory effects.
- Few learning and memory studies in animals evaluated motor activity or sensory effects. Studies that did evaluate motor activity or sensory effects often lacked discussion on general health of the animals when the endpoints measured could be affected by deficits in motor activity or sensory, such as latency to achieve a desired result.

A key data gap in the human and animal bodies of evidence includes the need for mechanistic insight into fluoride-related neurodevelopmental or cognitive changes.

Limitations of the Systematic Review

There are no major limitations of the systematic review. The human body of evidence included a large database of observational studies. Most of the observational studies were cross-sectional; however, nine of these were considered to be functionally prospective in nature. In addition, the systematic review covered a wide range of study designs, populations, and measures of fluoride exposure. The systematic review was designed to cover reports on all potential mechanistic data including effects on the thyroid. After review of the studies evaluating thyroid effects, studies that only evaluated goiters and other effects on thyroid size were not considered in this review. This is not considered a limitation because changes in thyroid size are not functional changes to the thyroid that could specifically indicate a mechanism for thyroid involvement in neurodevelopment. In addition, review of the mechanistic data was limited to in vivo studies with at least one concentration below 20 ppm. This is not considered a limitation for the systematic review since the mechanistic body of evidence was used to evaluate biological plausibility for the effects observed in humans; therefore, data were limited to concentrations that would be more reflective of human exposures. The decision to not more closely evaluate the in vitro data is not considered a limitation because there were sufficient in vivo data, and no key events were identified where in vitro data would provide additional insight.

CONCLUSION

Because the majority of available studies evaluated cognitive neurodevelopmental effects in children, the focus of the hazard conclusions is on cognitive neurodevelopmental effects, primarily IQ. When focusing on findings from studies with exposures in ranges typically found in drinking water in the

United States (0.7 mg/L for optimally fluoridated community water systems)[9] that can be evaluated for dose response, effects on cognitive neurodevelopment are inconsistent and, therefore, unclear. However, when considering all the evidence, including studies with exposures to fluoride levels higher than 1.5 mg/L in water, NTP concludes that fluoride is *presumed to be a cognitive neurodevelopmental hazard to humans*. This conclusion is based on a moderate level of evidence that shows a consistent and robust pattern of findings in human studies across several different populations demonstrating that higher fluoride exposure (e.g., >1.5 mg/L in drinking water) is associated with lower IQ and other cognitive effects in children. Limited and weaker evidence is considered to provide an inadequate level of evidence that fluoride is associated with cognitive effects in adults. The evidence from animal studies is inadequate to inform conclusions on cognitive effects, and the mechanisms underlying fluoride-associated cognitive neurodevelopmental effects are not well characterized.

[9]As of April 2020, 1.08% of persons living in the United States (~ 3.5 million people) were served by CWS supplying ≥ 1.1 mg/L naturally occurring fluoride. CWS supplying water at ≥ 1.5 mg/L naturally occurring fluoride served 0.59% of the U.S. population(~ 1.9 million people), and systems supplying water at ≥ 2 mg/L naturally occurring fluoride served 0.31% of the U.S. population (~1 million people) (https://www.cdc.gov/fluoridation/data-tools/reporting-system.html).

REFERENCES

Adedara IA, Abolaji AO, Idris UF, Olabiyi BF, Onibiyo EM, Ojuade TD, Farombi EO. 2017a. Neuroprotective influence of taurine on fluoride-induced biochemical and behavioral deficits in rats. *Chem Biol Interact* 261: 1-10.

Adedara IA, Olabiyi BF, Ojuade TD, Idris UF, Onibiyo EM, Farombi EO. 2017b. Taurine reverses sodium fluoride-mediated increase in inflammation, caspase-3 activity, and oxidative damage along the brain-pituitary-gonadal axis in male rats. *Can J Phys Pharmacol* 95: 1019-1029.

Agustina F, Sofro ZM, Partadiredja G. 2018. Subchronic administration of high-dose sodium fluoride causes deficits in cerebellar purkinje cells but not motor coordination of rats. *Biol Trace Elem Res* 188(2): 424-433.

Ahmad KR, Noor S, Jabeen S, Nauroze T, Kanwal MA, Raees K, Abbas T. 2017. Amelioration by jambul fruit extract of fluoride-induced hepato-nephronal histopathologies and impaired neuromotor capacity in mice. *Fluoride* 50: 2-14.

Akinrinade ID, Memudu AE, Ogundele OM. 2015a. Fluoride and aluminium disturb neuronal morphology, transport functions, cholinesterase, lysosomal and cell cycle activities. *Pathophysiology* 22: 105-115.

Akinrinade ID, Memudu AE, Ogundele OM, Ajetunmobi OI. 2015b. Interplay of glia activation and oxidative stress formation in fluoride and aluminium exposure. *Pathophysiology* 22: 39-48.

An J, Mei S, Liu A, Fu Y, Wang C. 1992. [Effect of high level of fluoride on children's intelligence]. *Chin J Control Endem Dis* 7(2): 93-94.

Aravind A, Dhanya RS, Narayan A, Sam G, Adarsh VJ, Kiran M. 2016. Effect of fluoridated water on intelligence in 10-12-year-old school children. *J Int Soc Prev Community Dent* 6(Suppl 3): S237-S242.

Arbuckle TE, Fraser WD, Fisher M, Davis K, Liang CL, Lupien N, Bastien S, Velez MP, von Dadelszen P, Hemmings DG, Wang J, Helewa M, Taback S, Sermer M, Foster W, Ross G, Fredette P, Smith G, Walker M, Shear R, Dodds L, Ettinger AS, Weber JP, D'Amour M, Legrand M, Kumarathasan P, Vincent R, Luo ZC, Platt RW, Mitchell G, Hidiroglou N, Cockell K, Villeneuve M, Rawn DF, Dabeka R, Cao XL, Becalski A, Ratnayake N, Bondy G, Jin X, Wang Z, Tittlemier S, Julien P, Avard D, Weiler H, Leblanc A, Muckle G, Boivin M, Dionne G, Ayotte P, Lanphear B, Séguin JR, Saint-Amour D, Dewailly E, Monnier P, Koren G, Ouellet E. 2013. Cohort profile: The maternal-infant research on environmental chemicals research platform. *Paediatr Perinat Epidemiol* 27(4): 415-425.

Baba NA, Raina R, Verma PK, Sultana M. 2014. Alterations in plasma and tissue acetylcholinesterase activity following repeated oral exposure of chlorpyrifos alone and in conjunction with fluoride in Wistar rats. *Proc Indian Natl Sci Acad B Biol Sci* 84: 969-972.

Bai A, Li Y, Fan Z, Li X, Li P. 2014. [Intelligence and growth development of children in coal-burning-borne arsenism and fluorosis areas: An investigation study]. *Chin J Endemiol* 33(2): 160-163.

Banala R, Nagapuri K, Mohd K, Reddy M, Karnati P. 2018. Carica Papaya leaf extract as a neuroprotective agent against behavioral and neurotransmitter changes in brain of the rat treated with sodium fluoride in pre- and post-natal periods. *Pharmacogn Mag* 14(55): 123-131.

Banala RR, Karnati PR. 2015. Vitamin A deficiency: An oxidative stress marker in sodium fluoride (NaF) induced oxidative damage in developing rat brain. *Int J Dev Neurosci* 47: 298-303.

Barberio AM, Hosein FS, Quinonez C, McLaren L. 2017a. Fluoride exposure and indicators of thyroid functioning in the Canadian population: Implications for community water fluoridation. *J Epidemiol Community Health* 71: 1019-1025.

Barberio AM, Quinonez C, Hosein FS, McLaren L. 2017b. Fluoride exposure and reported learning disability diagnosis among Canadian children: Implications for community water fluoridation. *Can J Public Health* 108: 229-239.

Bartos M, Gumilar F, Gallegos CE, Bras C, Dominguez S, Monaco N, Esandi MDC, Bouzat C, Cancela LM, Minetti A. 2018. Alterations in the memory of rat offspring exposed to low levels of fluoride during gestation and lactation: Involvement of the alpha7 nicotinic receptor and oxidative stress. *Reprod Toxicol* 81: 108-114.

Bashash M, Marchand M, Hu H, Till C, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Green R, Schnaas L, Mercado-Garcia A, Hernandez-Avila M, Tellez-Rojo MM. 2018. Prenatal fluoride exposure and attention deficit hyperactivity disorder (ADHD) symptoms in children at 6-12 years of age in Mexico City. *Environ Int* 121(Pt 1): 658-666.

Bashash M, Thomas D, Hu H, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Ettinger AS, Wright R, Zhang Z, Liu Y, Schnaas L, Mercado-Garcia A, Tellez-Rojo MM, Hernandez-Avila M. 2017. Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6-12 years of age in Mexico. *Environ Health Perspect* 125(9): 1-12.

Begg CB, Mazumdar M. 1994. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 50(4): 1088-1101.

Bhatnagar M, Rao P, Sushma J, Bhatnagar R. 2002. Neurotoxicity of fluoride: Neurodegeneration in hippocampus of female mice. *Indian J Exp Biol* 40: 546-554.

Bhatnagar M, Sukhwal P, Suhalka P, Jain A, Joshi C, Sharma D. 2011. Effects of fluoride in drinking water on NADPH-diaphorase neurons in the forebrain of mice: A possible mechanism of fluoride neurotoxicity. *Fluoride* 44: 195-209.

Broadbent JM, Thomson WM, Moffitt TE, Poulton R. 2015. Community water fluoridation and intelligence response. *Am J Public Health* 105: 3-4.

CDC (Centers for Disease Control and Prevention). 2013. *Community water fluoridation: Fluoridation statistics*. Atlanta, GA. Available: https://www.cdc.gov/fluoridation/statistics/2012stats.htm [accessed 19 August 2019].

Chen J, Niu Q, Xia T, Zhou G, Li P, Zhao Q, Xu C, Dong L, Zhang S, Wang A. 2018. ERK1/2-mediated disruption of BDNF-TrkB signaling causes synaptic impairment contributing to fluoride-induced developmental neurotoxicity. *Toxicology* 410: 222-230.

Chen Y. 2012. Organophosphate-induced brain damage: Mechanisms, neuropsychiatric and neurological consequences, and potential therapeutic strategies. *Neurotox* 33: 391-400.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 1991. [Research on the intellectual development of children in high fluoride areas]. *Chin J Control Endem Dis* 6(Suppl): 99-100.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 2008. Research on the intellectual development of children in high fluoride areas. *Fluoride* 41: 120-124.

Chinoy NJ, Narayana MV. 1992. Studies on fluorosis in Mehsana District of North Gujarat. *Proc Zool Soc* 45: 157-161.

Choi AL, Sun G, Zhang Y, Grandjean P. 2012. Developmental fluoride neurotoxicity: A systematic review and meta-analysis. *Environ Health Perspect* 120: 1362-1368.

Choi AL, Zhang Y, Sun G, Bellinger DC, Wang K, Yang XJ, Li JS, Zheng Q, Fu Y, Grandjean P. 2015. Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study. *Neurotoxicol Teratol* 47: 96-101.

Chouhan S, Flora SJS. 2008. Effects of fluoride on the tissue oxidative stress and apoptosis in rats: Biochemical assays supported by IR spectroscopy data. *Toxicology* 254: 61-67.

Chouhan S, Lomash V, Flora SJ. 2010. Fluoride-induced changes in haem biosynthesis pathway, neurological variables and tissue histopathology of rats. *J Appl Toxicol* 30: 63-73.

Cochran WG. 1954. The combination of estimates from different experiments. *Biometrics* 10(1): 101-129.

Crippa A, Discacciati A, Bottai M, Spiegelman D, Orsini N. 2018. One-stage dose–response meta-analysis for aggregated data. *Stat Methods Med Res* 28(5): 1579-1596.

Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. 2020. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. *Neurosci Lett* 729: 134981.

Cui Y, Zhang B, Ma J, Wang Y, Zhao L, Hou C, Yu J, Zhao Y, Zhang Z, Nie J, Gao T, Zhou G, Liu H. 2018. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. *Ecotoxicol Environ Saf* 165: 270-277.

Das K, Mondal NK. 2016. Dental fluorosis and urinary fluoride concentration as a reflection of fluoride exposure and its impact on IQ level and BMI of children of Laxmisagar, Simlapal Block of Bankura District, W.B., India. *Environ Monit Assess* 188: 218.

Ding Y, Sun H, Han H, Wang W, Ji X, Liu X, Sun D. 2011. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. *J Hazard Mater* 186: 1942-1946.

Dong YT, Wei N, Qi XL, Liu XH, Chen D, Zeng XX, Guan ZZ. 2017. Attenuating effect of vitamin E on the deficit of learning and memory of rats with chronic fluorosis: The mechanism may involve muscarinic acetylcholine receptors. *Fluoride* 50: 354-364.

Du L, Wan C, Cao X, Liu J. 1992. [The effect of fluorine on the developing human brain]. *Chin J Pathol* 21(4): 218-220.

Du L, Wan C, Cao X, Liu J. 2008. The effect of fluorine on the developing human brain. *Fluoride* 41: 327-330.

Duan J, Zhao M, Wang L, Fang D, Wang Y, Wang W. 1995. A comparative analysis of the results of multiple tests in patients with chronic industrial fluorosis. *Guizhou Med J* 18(3): 179-180.

Duan Q, Jiao J, Chen X, Wang X. 2018. Association between water fluoride and the level of children's intelligence: A dose-response meta-analysis. *Public Health* 154: 87-97.

Duval S, Tweedie R. 2000a. A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *J Am Stat Assoc* 95(449): 89-98.

Duval S, Tweedie R. 2000b. Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56(2): 455-463.

Egger M, Smith G, Schneider M, Minder C, eds. 2008. *Systematic reviews in health care: meta-analysis in context*. London, UK: BMJ Publishing Group.

Egger M, Smith GD, Schneider M, Minder C. 1997. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315(7109): 629-634.

Eswar P, Nagesh L, Devaraj CG. 2011. Intelligent quotients of 12-14 year old school children in a high and low fluoride village in India. *Fluoride* 44: 168-172.

Fan Z, Dai H, Bai A, Li P, Li T, Li G. 2007. The effect of high fluoride exposure in children’s intelligence. *J Environ Health* 24(10): 802-803.

Gais S, Schonauer M. 2017. Untangling a cholinergic pathway from wakefulness to memory. *Neuron* 94(4): 696-698.

Gao Q, Liu YJ, Guan ZZ. 2009. Decreased learning and memory ability in rats with fluorosis: Increased oxidative stress and reduced cholinesterase activity in the brain. *Fluoride* 42: 277-285.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008a. [Effects of fluoride on learning and memory and cholinesterase activity in rat brains]. *Chin J Endemiol* 27: 128-130.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008b. [Level of oxidative stress in rat brains and learning and memory function of rats with chronic fluorosis]. *Chin J Endemiol* 27: 371-373.

Garman RH, Li AA, Kaufmann W, Auer RN, Bolon B. 2016. Recommended methods for brain processing and quantitative analysis in rodent developmental neurotoxicity studies. *Toxicol Pathol* 44(1): 14-42.

Ge QD, Tan Y, Luo Y, Wang WJ, Zhang H, Xie C. 2018a. MiR-132, miR-204 and BDNF-TrkB signaling pathway may be involved in spatial learning and memory impairment of the offspring rats caused by fluorine and aluminum exposure during the embryonic stage and into adulthood. *Environ Toxicol Pharmacol* 63: 60-68.

Ge Y, Chen L, Yin Z, Song X, Ruan T, Hua L, Liu J, Wang J, Ning H. 2018b. Fluoride-induced alterations of synapse-related proteins in the cerebral cortex of ICR offspring mouse brain. *Chemosphere* 201: 874-883.

Green R, Lanphear B, Hornung R, Flora D, Martinez-Mier EA, Neufeld R, Ayotte P, Muckle G, Till C. 2019. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. *JAMA Pediatr*: E1-E9.

Guner S, Uyar-Bozkurt S, Haznedaroglu E, Mentes A. 2016. Dental fluorosis and catalase immunoreactivity of the brain tissues in rats exposed to high fluoride pre- and postnatally. *Biol Trace Elem Res* 174: 150-157.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 1991. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. *Chin J Epidemiol* 10(2): 98-100.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 2008a. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. *Fluoride* 41: 125-128.

Guo ZY, He YH, Zhu QX. 2001. [Research on the neurobehavioral function of workers occupationally exposed to fluoride]. *Ind Hlth & Occup Dis* 27(6): 346-348.

Guo ZY, He YH, Zhu QX. 2008b. Research on the neurobehavioral function of workers occupationally exposed to fluoride. *Fluoride* 41: 152-155.

Guyatt GH, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, Debeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schunemann HJ. 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 64(4): 383-394.

He MX, Zhang CN. 2010. [Investigation of children's intelligence quotient and dental fluorosis in drinking water-type of endemic fluorosis area in Pucheng County, Shaanxi Province before and after drinking water change]. *Chin J Endemiol* 29: 547-548.

Health Canada. 2015. *Third report on human biomonitoring of environmental chemicals in Canada - Results of the Canadian Health Measures Survey Cycle 3 (2012–2013)*. Ottawa, Ontario: Canadian Minstry of Health. Available: https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/ewh-semt/alt_formats/pdf/pubs/contaminants/chms-ecms-cycle3/chms-ecms-cycle3-eng.pdf.

Higgins JP, Green S. 2011. *Cochrane handbook for systematic reviews of interventions,* In: The Cochrane Collaboration. Vol 4, New York, NY: John Wiley & Sons.

Higgins JT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, eds. 2019. *Cochrane Handbook for systematic reviews of interventions version 6.0 (updated July 2019)*. Cochrane.

Hong FG, Cao YX, Yang D, Wang H. 2001. [Research on the effects of fluoride on child intellectual development under different environmental conditions]. *Chin Prim Health Care* 15(3): 56-57.

Hong FG, Cao YX, Yang D, Wang H. 2008. Research on the effects of fluoride on child intellectual development under different environmental conditions. *Fluoride* 41: 156-160.

Hosur MB, Puranik RS, Vanaki S, Puranik SR. 2012. Study of thyroid hormones free triiodothyronine (FT3), free thyroxine (FT4) and thyroid stimulating hormone (TSH) in subjects with dental fluorosis. *Eur J Dent* 6: 184-190.

Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, Sedykh A, Thayer K, Merrick BA, Walker V, Rooney A, Shah RR. 2020. SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation. *Environ Int* 138: 105623.

Jacqmin H, Commenges D, Letenneur L, Barberger-Gateau P, Dartigues JF. 1994. Components of drinking water and risk of cognitive impairment in the elderly. *Am J Epidemiol* 139: 48-57.

Jia B, Zong L, Lee JY, Lei J, Zhu Y, Xie H, Clemens JL, Feller MC, Na Q, Dong J, McLane MW, Jones-Beatty K, Burd I. 2019. Maternal supplementation of low dose fluoride alleviates adverse perinatal outcomes following exposure to intrauterine inflammation. *Scientific Reports* 9(1): 2575.

Jiang C, Zhang S, Liu H, Guan Z, Zeng Q, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Zhang X, Cui Y, Yu L, Wang A. 2014. Low glucose utilization and neurodegenerative changes caused by sodium fluoride exposure in rat's developmental brain. *Neuromolecular Med* 16: 94-105.

Jones S, Burt BA, Petersen PE, Lennon MA. 2005. The effective use of fluorides in public health. *Bull World Health Organ* 83: 670-676.

Kang J, Cheng Y, Wu K, Lin S, He G, Jin Y. 2011. [Effect of exposure to fluoride and arsenic in drinking water of Hangjinhouqi on children's intelligence]. *Chin School Health*: 679-681.

Karimzade S, Aghaei M, Mahvi AH. 2014. Investigation of intelligence quotient in 9-12 year-old children exposed to high- and low-drinking water fluoride in West Azerbaijan Province, Iran. *Fluoride* 47: 9-14.

Khan AM, Raina R, Dubey N, Verma PK. 2017. Effect of deltamethrin and fluoride co-exposure on the brain antioxidant status and cholinesterase activity in Wistar rats. *Drug Chem Toxicol* 41(2): 1-5.

Khan SA, Singh RK, Navit S, Chadha D, Johri N, Navit P, Sharma A, Bahuguna R. 2015. Relationship between dental fluorosis and intelligence quotient of school going children in and around Lucknow District: A cross-sectional study. *J Clin Diagn Res* 9(11): 10-15.

Khandare AL, Gourineni SR, Validandi V. 2017. Dental fluorosis, nutritional status, kidney damage, and thyroid function along with bone metabolic indicators in school-going children living in fluoride-affected hilly areas of Doda District, Jammu and Kashmir, India. *Environ Monit Assess* 189: 579.

Khandare AL, Validandi V, Gourineni SR, Gopalan V, Nagalla B. 2018. Dose-dependent effect of fluoride on clinical and subclinical indices of fluorosis in school going children and its mitigation by supply of safe drinking water for 5 years: An Indian study. *Environ Monit Assess* 190: 110.

Kheradpisheh Z, Mahvi AH, Mirzaei M, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018a. Correlation between drinking water fluoride and TSH hormone by ANNs and ANFIS. *J Environ Health Sci Eng* 16(1): 11-18.

Kheradpisheh Z, Mirzaei M, Mahvi AH, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018b. Impact of drinking water fluoride on human thyroid hormones: A case-control study. *Sci Rep* 8: 2674.

Kinawy AA, Al-Eidan AA. 2018. Impact of prenatal and postnatal treatment of sodium fluoride and aluminum chloride on some hormonal and sensorimotor aspects in rats. *Biol Trace Elem Res* 186: 1-8.

Kumar V, Chahar P, Kajjari S, Rahman F, Bansal DK, Kapadia JM. 2018. Fluoride, thyroid hormone derangements and its correlation with tooth eruption pattern among the pediatric population from endemic and non-endemic fluorosis areas. *J Contemp Dent Pract* 19(12): 1512-1516.

Kundu H, Basavaraj P, Singla A, Gupta R, Singh K, Jain S. 2015. Effect of fluoride in drinking water on children's intelligence in high and low fluoride areas of Delhi. *J Indian Assoc Public Health Dent* 13(2): 116-121.

Li F, Chen X, Huang R, Xie Y. 2009. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. *J Environ Health* 26(4): 838-840.

Li J, Yao L, Shao QL, Wu CY. 2004. [Effects of high fluoride level on neonatal neurobehavioral development]. *Chin J Endemiol* 23(5): 463-465.

Li J, Yao L, Shao QL, Wu CY. 2008a. Effects of high fluoride level on neonatal neurobehavioral development. *Fluoride* 41: 165-170.

Li M, Gao Y, Cui J, Li Y, Li B, Liu Y, Sun J, Liu X, Liu H, Zhao L, Sun D. 2016. Cognitive impairment and risk factors in elderly people living in fluorosis areas in China. *Biol Trace Elem Res* 172: 53-60.

Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. 2010. [Investigation and analysis of children's IQ and dental fluorosis in high fluoride area]. *Chin J Pest Control* 26(3): 230-231.

Li X, Zhang J, Niu R, Manthari RK, Yang K, Wang J. 2019. Effect of fluoride exposure on anxiety- and depression-like behavior in mouse. *Chemosphere* 215: 454-460.

Li XS, Zhi JL, Gao RO. 1995. Effect of fluoride exposure on the intelligence of children. *Fluoride* 28: 189-192.

Li Y, Li X, Wei S. 1994. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. *J West Chin University Med Sci* 25(2): 188-191.

Li Y, Li X, Wei S. 2008b. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. *Fluoride* 41: 331-335.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2003. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. *Chin J Public Health Manag* 19(4): 337-338.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2008c. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. *Fluoride* 41: 161-164.

Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. 1991. [High fluoride and low iodine environment and subclinical cretinism in Xinjiang]. *Endem Dis Bull* 6(2): 62-67.

Liu S, Lu Y, Sun Z, Wu L, Wang X, Yan S. 2000. [Report on the intellectual ability of children living in high-fluoride water areas]. *Chin J Control Endem Dis* 15(4): 231-232.

Liu SL, Lu Y, Sun ZR, Wu L, Lu WL, Wang XW, Yan S. 2008. Report on the intellectual ability of children living in high-fluoride water areas. *Fluoride* 41: 144-147.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. Alterations of nAChRs and ERK1/2 in the brains of rats with chronic fluorosis and their connections with the decreased capacity of learning and memory. *Toxicol Lett* 192: 324-329.

Lou DD, Guan ZZ, Liu YJ, Liu YF, Zhang KL, Pan JG, Pei JJ. 2013. The influence of chronic fluorosis on mitochondrial dynamics morphology and distribution in cortical neurons of the rat brain. *Arch Toxicol* 87: 449-457.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S, Mentes A, DenBesten PK. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. *Physiol Behav* 206: 76-83.

Lu Y, Sun ZR, Wu LN, Wang X, Lu W, Liu SS. 2000. Effect of high-fluoride water on intelligence in children. *Fluoride* 33: 74-78.

Ma Q, Huang H, Sun L, Zhou T, Zhu J, Cheng X, Duan L, Li Z, Cui L, Ba Y. 2017. Gene-environment interaction: Does fluoride influence the reproductive hormones in male farmers modified by ERα gene polymorphisms? *Chemosphere* 188: 525-531.

Malin AJ, Riddell J, McCague H, Till C. 2018. Fluoride exposure and thyroid function among adults living in Canada: Effect modification by iodine status. *Environ Int* 121(Pt 1): 667-674.

Malin AJ, Till C. 2015. Exposure to fluoridated water and attention deficit hyperactivity disorder prevalence among children and adolescents in the United States: An ecological association. *Environ Health* 14: 17.

Manusha S, Sudhakar K, Reddy KP. 2019. Protective effects of allium sativum extract against sodium fluoride induced neurotoxicity. *Int J Pharm Sci Res* 10(2): 625-633.

McPherson CA, Zhang G, Gilliam R, Brar SS, Wilson R, Brix A, Picut C, Harry GJ. 2018. An evaluation of neurotoxicity following fluoride exposure from gestational through adult ages in Long-Evans hooded rats. *Neurotoxicol Res*: 1-18.

Mesram N, Nagapuri K, Banala RR, Nalagoni CR, Karnati PR. 2016. Quercetin treatment against NaF induced oxidative stress related neuronal and learning changes in developing rats. *J King Saud Univ Sci* 29: 221-229.

Michael M, Barot VV, Chinoy NJ. 1996. Investigations of soft tissue functions in fluorotic individuals of North Gujarat. *Fluoride* 29: 63-71.

Miller K, Howard B, Phillips J, Shah M, Mav D, Thayer K, Shah R. 2016. SWIFT-Active screener: Reducing literature screening effort through machine learning for systematic reviews, Cochrane Colloquium Seoul, Seoul, Korea.

Moher D, Liberati A, Tetzlaff J, Altman D. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med* 6(7): e1000097.

Mondal D, Dutta G, Gupta S. 2016. Inferring the fluoride hydrogeochemistry and effect of consuming fluoride-contaminated drinking water on human health in some endemic areas of Birbhum District, West Bengal. *Environ Geochem Health* 38: 557-576.

Morgan L, Allred E, Tavares M, Bellinger D, Needleman H. 1998. Investigation of the possible associations between fluorosis, fluoride exposure, and childhood behavior problems. *Pediatr Dent* 20: 244-252.

Müller S, Scealy JL, Welsh AH. 2013. Model selection in linear mixed models. *Statist Sci* 28(2): 135-167.

Mustafa DE, Younis UM, Elhag SAA. 2018. The relationship between the fluoride levels in drinking water and the schooling performance of children in rural areas of Khartoum State, Sudan. *Fluoride* 51(2): 102-113.

Nagarajappa R, Pujara P, Sharda AJ, Asawa K, Tak M, Aapaliya P, Bhanushali N. 2013. Comparative assessment of intelligence quotient among children living in high and low fluoride areas of Kutch, India: A pilot study. *Iran J Public Health* 42: 813-818.

Nageshwar M, Sudhakar K, Reddy KP. 2018. Quercetin ameliorates oxidative stress, neural damage of brain and behavioral impairment of rat with fluoride exposure. *Int J Pharm Sci Res* 9(8): 3247-3256.

Nageshwar M, Sudhakar K, Reddy NCC, Reddy KP. 2017. Neuroprotective effects of curcumin on sodium fluoride induced behavioural and enzymatic changes in brain and muscles of rat. *J Environ Biol* 38: 675-681.

NASEM (National Academies of Sciences, Engineering and Medicine). 2020. *Review of the draft NTP monograph: Systematic review of fluoride exposure and neurodevelopmental and cognitive health effects*. Washington, DC: The National Academies Press. Available: https://doi.org/10.17226/25715.

NIOSH (National Institute for Occupational Safety and Health). 1984. *Fluoride in urine*. In: Manual of Analytical Methods Vol 11. Method 8308. Washington, DC: US Department of Health and Human Services: 1-3.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. *Environ Pollut* 233: 889-899.

Nkpaa KW, Onyeso GI. 2018. Rutin attenuates neurobehavioral deficits, oxidative stress, neuro-inflammation and apoptosis in fluoride treated rats. *Neurosci Lett* 682: 92-99.

NRC (National Research Council). 2006. *Committee on fluoride in drinking water, board on environmental studies and toxicology. Fluoride in drinking water: A scientific review of EPA's standards.* Available: http://www.nap.edu/catalog/11571/fluoride-in-drinking-water-a-scientific-review-of-epas-standards [accessed 19 August 2019].

NTP (National Toxicology Program). 2015. *Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration*. Research Triangle Park, NC. Available: http://ntp.niehs.nih.gov/go/38673 [accessed 19 August 2019].

NTP (National Toxicology Program). 2016. *Systematic literature review on the effects of fluoride on learning and memory in animal studies*. NTP Research Report 1. Research Triangle Park, NC. Available: https://ntp.niehs.nih.gov/ntp/results/pubs/rr/reports/01fluoride_508.pdf [accessed 19 August 2019].

NTP (National Toxicology Program). 2019. *Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration*. Research Triangle Park, NC. Available: https://ntp.niehs.nih.gov/go/systematic_review [accessed 19 August 2019].

OEHHA (California Office of Environmental Health Hazard Assessment). 2011. *Meeting synopsis and slide presentations: carcinogen identification committee meeting held on October 12, 2011*. Available: http://oehha.ca.gov/prop65/public_meetings/cic101211synop.html [accessed 19 August 2019].

Peckham S, Lowery D, Spencer S. 2015. Are fluoride levels in drinking water associated with hypothyroidism prevalence in England? A large observational study of GP practice data and fluoride levels in drinking water. *J Epidemiol Community Health* 69: 619-624.

Podgorski J, Berg M. 2020. Global threat of arsenic in groundwater. *Science* 368(6493): 845-850.

Poureslami HR, Horri A, Garrusi B. 2011. A comparative study of the IQ of children age 7-9 in a high and a low fluoride water city in Iran. *Fluoride* 44: 163-167.

Pulungan ZS, Sofro ZM, Partadiredja G. 2016. Sodium fluoride does not affect the working memory and number of pyramidal cells in rat medial prefrontal cortex. *Anat Sci Int* 93(1): 128-138.

Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. 1990. [Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children]. *J Control Endem Dis* 5(4): 203-204.

Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. 2008. Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children. *Fluoride* 41: 115-119.

Raju S, Sivanesan SK, Gudemalla K. 2019. Cognitive enhancement effect of Ginkgo biloba extract on memory and learning impairments induced by fluoride neurotoxicity. *Int J Res Pharm Sci* 10(1): 129-134.

Razdan P, Patthi B, Kumar JK, Agnihotri N, Chaudhari P, Prasad M. 2017. Effect of fluoride concentration in drinking water on intelligence quotient of 12-14-year-old children in Mathura District: A cross-sectional study. *J Int Soc Prev Community Dent* 7: 252-258.

Reddy YP, Tiwari SK, Shaik AP, Alsaeed A, Sultana A, Reddy PK. 2014. Effect of sodium fluoride on neuroimmunological parameters, oxidative stress and antioxidative defenses. *Toxicol Mech Methods* 24: 31-36.

Ren D, Li K, Liu D. 1989. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. *Chin J Control Endem Dis* 4(4): 251.

Ren D, Li K, Liu D. 2008. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. *Fluoride* 41: 319-320.

Riddell JK, Malin AJ, Flora D, McCague H, Till C. 2019. Association of water fluoride and urinary fluoride concentrations with attention deficit hyperactivity disorder in Canadian youth. *Environ Int* 133: 105190.

Rocha-Amador D, Navarro M, Trejo-Acevedo A, Carrizales L, Perez-Maldonado I, Diaz-Barriga F, Calderon J. 2009. Use of the Rey-Osterrieth Complex Figure Test for neurotoxicity evaluation of mixtures in children. *Neurotox* 30: 1149-1154.

Rocha-Amador D, Navarro ME, Carrizales L, Morales R, Calderon J. 2007. Decreased intelligence in children and exposure to fluoride and arsenic in drinking water. *Cad Saude Publica* 23(Suppl 4): S579-587.

Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. *Environ Health Perspect* 122(7): 711-718.

Rooney AA, Cooper GS, Jahnke GD, Lam J, Morgan RL, Ratcliffe JM, Kraft AD, Schünemann HJ, Schwingl P, Walker TD, Thayer KA, Lunn RM. 2016. How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. *Environ Int* 92-93: 617-629.

Rotton J, Tikofsky RS, Feldman HT. 1982. Behavioral effects of chemicals in drinking water. *J Appl Psychol* 67: 230-238.

Russ TC, Killin LOJ, Hannah J, Batty GD, Deary IJ, Starr JM. 2019. Aluminium and fluoride in drinking water in relation to later dementia risk. *Brit J Psychol* 14: 1-6.

Saxena S, Sahay A, Goel P. 2012. Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India. *J Neurosci Rural Pract* 3: 144-149.

SCHER (Scientific Committee on Health and Environmental Risks). 2011. *Critical review of any new evidence on the hazard profile, health effects, and human exposure to fluoride and the fluoridating agents of drinking water*. European Commission Directorate-General for Health and Consumers Scientific Committees. Available: http://ec.europa.eu/health/scientific_committees/environmental_risks/docs/scher_o_139.pdf [accessed 19 August 2019].

Sebastian ST, Sunitha S. 2015. A cross-sectional study to assess the intelligence quotient (IQ) of school going children aged 10-12 years in villages of Mysore District, India with different fluoride levels. *J Indian Soc Pedod Prev Dent* 33: 307-311.

Seraj B, Shahrabi M, Falahzade M, Falahzade F, Akhondi N. 2006. Effect of high fluoride concentration in drinking water on children's intelligence. *J Dent Med* 19(2): 80-86.

Seraj B, Shahrabi M, Shadfar M, Ahmadi R, Fallahzadeh M, Eslamlu HF, Kharazifard MJ. 2012. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. *J Dent* 9: 221-229.

Shalini B, Sharma JD. 2015. Beneficial effects of emblica officinalis on fluoride-induced toxicity on brain biochemical indexes and learning-memory in rats. *Toxicol Int* 22: 35-39.

Shan KR, Qi XL, Long YG, Nordberg A, Guan ZZ. 2004. Decreased nicotinic receptors in PC12 cells and rat brains influenced by fluoride toxicity: A mechanism relating to a damage at the level in post-transcription of the receptor genes. *Toxicology* 200: 169-177.

Shannon FT, Fergusson DM, Horwood LJ. 1986. Exposure to fluoridated public water supplies and child health and behaviour. *New Zeal Med J* 99: 416-418.

Shao Q. 2003. [Study of cognitive function impairment caused by chronic fluorosis]. *Chin J Endemiol* 22(4): 336-338.

Sharma C, Suhalka P, Bhatnagar M. 2018. Curcumin and resveratrol rescue cortical-hippocampal system from chronic fluoride-induced neurodegeneration and enhance memory retrieval. *Int J Neurosci*: 1-15.

Sharma JD, Sohu D, Jain P. 2009. Prevalence of neurological manifestations in a human population exposed to fluoride in drinking water. *Fluoride* 42: 127-132.

Shivaprakash PK, Ohri K, Noorani H. 2011. Relation between dental fluorosis and intelligence quotient in school children of Bagalkot District. *J Indian Soc Pedod Prev Dent* 29: 117-120.

Singh N, Verma KG, Verma P, Sidhu GK, Sachdeva S. 2014. A comparative study of fluoride ingestion levels, serum thyroid hormone & TSH level derangements, dental fluorosis status among school children from endemic and non-endemic fluorosis areas. *Springerplus* 3: 7.

Singh V, Singh C, Sandeep T, Sandeep K, Vikas G, Mukesh T, Anurag T. 2013. A correlation between serum vitamin, acetylcholinesterase activity and IQ in children with excessive endemic fluoride exposure in Rajasthan, India. *Int Res J Medical Sci* 1(3): 12-16.

Soto-Barreras U, Escalante-Villalobos KY, Holguin-Loya B, Perez-Aguirre B, Nevarez-Rascon A, Martinez-Martinez RE, Loyola-Rodriguez JP. 2019. Effect of fluoride in drinking water on dental caries and IQ in children. *Fluoride* 52: 474-482.

Still CN, Kelley P. 1980. The incidence of primary degenerative dementia vs. water fluoride content in South Carolina, USA. *Neurotox* 1: 125-132.

Sudhakar K, Nageshwar M, Pratap Reddy K. 2017. Seed extract of Abelmoschus moschatus medik reverses NAF-induced behavioral changes through neurodegeneration and oxidative stress in brain of rat. *Asian J Pharm Clin Res* 10: 165-171.

Sudhakar K, Nageshwar M, Reddy KP. 2018a. Abelmoschus moschatus extract reverses altered pain and neurohistology of a rat with developmental exposure of fluoride. *J Appl Pharm Sci* 8(6): 94-104.

Sudhakar K, Nageshwar M, Reddy KP. 2018b. Protective effect of okra, Abelmoschus moschatus seed extract on developing brain of rats during pre- and post-natal fluoride exposure. *Int J Pharm Sci Res* 9: 1519-1528.

Sudhakar K, Reddy KP. 2018. Protective effects of Abelmoschus moschatus seed extract on neurotransmitter system of developing brain of Wistar rats with gestational and post-natal exposure of sodium fluoride. *Int J Green Pharm* 12: S131-S139.

Sudhir KM, Chandu GN, Prashant GM, Subba Reddy VV. 2009. Effect of fluoride exposure on intelligence quotient (IQ) among 13-15 year old school children of known endemic area of fluorosis, Nalgonda District, Andhra Pradesh. *J Indian Assoc Public Health Dent* 2009(13): 88-94.

Sun M, Li S, Wang Y, Li F. 1991. [Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis]. *J Guiyang Med Coll* 16(3): 204-206.

Sun Z, Zhang Y, Xue X, Niu R, Wang J. 2018. Maternal fluoride exposure during gestation and lactation decreased learning and memory ability, and glutamate receptor mRNA expressions of mouse pups. *Hum Exp Toxicol* 37: 87-93.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2000. [Effects of high fluoride drinking water on the cerebral function of mice]. *Chin J Epidemiol* 19: 262-263.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2008. Effects of high fluoride drinking water on the cerebral function of mice. *Fluoride* 41: 148-151.

Susheela AK, Bhatnagar M, Vig K, Mondal NK. 2005. Excess fluoride ingestion and thyroid hormone derangements in children living in Delhi, India. *Fluoride* 38: 98-108.

Till C, Green R, Flora D, Hornung R, Martinez-Mier EA, Blazer M, Farmus L, Ayotte P, Muckle G, Lanphear B. 2020. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. *Environ Int* 134: 105315.

Till C, Green R, Grundy JG, Hornung R, Neufeld R, Martinez-Mier EA, Ayotte P, Muckle G, Lanphear B. 2018. Community water fluoridation and urinary fluoride concentrations in a national sample of pregnant women in Canada. *Environ Health Perspect* 126(10): 107001.

Trivedi M, Sangai N, Patel R, Payak M, Vyas S. 2012. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. *Fluoride* 45(4): 377-383.

Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. 2007. Effect of high fluoride water on intelligence of school children in India. *Fluoride* 40: 178-183.

Tsunoda M, Aizawa Y, Nakano K, Liu Y, Horiuchi T, Itai K, Tsunoda H. 2005. Changes in fluoride levels in the liver, kidney, and brain and in neurotransmitters of mice after subacute administration of fluoride. *Fluoride* 38: 284-292.

US DHHS (U.S. Department of Health and Human Services). 2015. *U.S. Public Health Service recommendation for fluoride concentration in drinking water for the prevention of dental caries*. 318-331. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4547570/ [accessed 19 August 2019].

US EPA (U.S. Environmental Protection Agency). 2010. *Fluoride: Exposure and relative source contribution analysis*. 820-R-10-015. Washington, DC. Available: http://www.epa.gov/dwstandardsregulations/fluoride-risk-assessment-and-relative-source-contribution [accessed 19 August 2019].

Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, Costilla-Salazar R, Calderon Hernandez J, Alcaraz Contreras Y, Rocha-Amador DO. 2017. In utero exposure to fluoride and cognitive development delay in infants. *Neurotox* 59: 65-70.

Villa A, Anabalon M, Zohouri V, Maguire A, Franco AM, Rugg-Gunn A. 2010. Relationships between fluoride intake, urinary fluoride excretion and fluoride retention in children and adults: An analysis of available data. *Caries Res* 44(1): 60-68.

Wang A, Duan L, Huang H, Ma J, Zhang Y, Ma Q, Guo Y, Li Z, Cheng X, Zhu J, Zhou G, Ba Y. 2020a. Association between fluoride exposure and behavioural outcomes of school-age children: A pilot study in China. *Int J Environ Health Res*: 1-10.

Wang G, Gao M, Zhang M, Yang M, Xiang Q. 2012. Correlation between total fluoride intake and children's IQ. *J Southeast Univ Med Ed*: 743-746.

Wang G, Yang D, Jia F, Wang H. 1996. [A study of the IQ levels of four- to seven-year-old children in high fluoride areas]. *Endem Dis Bull* 11(1): 60-62.

Wang G, Yang D, Jia F, Wang H. 2008b. A study of the IQ levels of four- to seven-year-old children in high fluoride areas. *Fluoride* 41: 340-343.

Wang J, Zhang Y, Guo Z, Li R, Xue X, Sun Z, Niu R. 2018. Effects of perinatal fluoride exposure on the expressions of miR-124 and miR-132 in hippocampus of mouse pups. *Chemosphere* 197: 117-122.

Wang M, Liu L, Li H, Li Y, Liu H, Hou C, Zeng Q, Li P, Zhao Q, Dong L, Zhou G, Yu X, Liu L, Guan Q, Zhang S, Wang A. 2020b. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. *Environ Int* 134: 105229.

Wang SX, Wang ZH, Cheng XT, Li J, Sang ZP, Zhang XD, Han LL, Qiao XY, Wu ZM, Wang ZQ. 2005. [Investigation and evaluation on intelligence and growth of children in endemic fluorosis and arsenism areas]. *Chin J Endemiol* 24: 179-182.

Wang SX, Wang ZH, Cheng XT, Li J, Sang ZP, Zhang XD, Han LL, Qiao XY, Wu ZM, Wang ZQ. 2007. Arsenic and fluoride exposure in drinking water: Children's IQ and growth in Shanyin County, Shanxi Province, China. *Environ Health Perspect* 115: 643-647.

Wang SY, Zhang HX, Fan W, Fang SJ, Kang PP, Chen XG, Yu MJ. 2005b. [The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children]. *J Appl Clin Ped* 20(9): 897-899.

Wang SY, Zhang HX, Fan W, Fang SJ, Kang PP, Chen XG, Yu MJ. 2008a. The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children. *Fluoride* 41: 344-348.

Wang X, Wang L, Hu P, Guo X, Luo X. 2001. [Effects of high iodine and high fluorine on children's intelligence and thyroid function]. *Chin J Endemiol* 20(4): 288-290.

Wang Z, Wang S, Zhang X, Li J, Zheng X, Hu C. 2006. Investigation of children's growth and development under long-term fluoride exposure. *Chin J Control Endem Dis* 21(4): 239-241.

Watanabe M, Kono K, Orita Y, Dote T, Usuda K, Takahashi Y, Yoshida Y. 1995. Influence of dietary fluoride intake on urinary fluoride concentration and evaluation of corrected levels in spot urine. *Fluoride* 28(2): 61-70.

Wei N, Li Y, Deng J, Xu S, Guan Z. 2014. [The effects of comprehensive control measures on intelligence of school-age children in coal-burning-borne endemic fluorosis areas]. *Chin J Endemiol* 33(3): 320-322.

WHO (World Health Organization). 2008. *Guidelines for drinking-water quality [electronic resource]: Incorporating 1st and 2nd addenda*. Third Edition. Vol. 1. Geneva, Switzerland. Available: https://apps.who.int/iris/bitstream/handle/10665/204411/9789241547611_eng.pdf?sequence=1&isAllowed=y.

WHO (World Health Organization). 2011. *Guidelines for drinking-water quality*. Fourth edition. Available: https://apps.who.int/iris/bitstream/handle/10665/44584/9789241548151_eng.pdf?sequence=1&isAllowed=y&ua=1.

Xiang Q, Liang Y, Chen B, Chen L. 2011. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. *Fluoride* 44: 191-194.

Xiang Q, Liang Y, Chen L, Wang C, Chen B, Chen X, Zhou M. 2003a. Effect of fluoride in drinking water on children's intelligence. *Fluoride* 36: 84-94.

Xiang Q, Wang Y, Yang M, Zhang M, Xu Y. 2013. Level of fluoride and arsenic in household shallow well water in Wamiao and Xinhuai villages in Jiangsu province, China. *Fluoride* 46: 192-197.

Xiang QY, Liang YX, Zhou MS, Zang HB. 2003b. Blood lead of children in Wamiao-Xinhuai intelligence study. *Fluoride* 36: 198-199.

Xu Y, Lu C, Zhang X. 1994. [The effect of fluorine on the level of intelligence in children]. *Endem Dis Bull* 9(2): 83-84.

Yang L, Jin P, Wang X, Zhou Q, Lin X, Xi S. 2018. Fluoride activates microglia, secretes inflammatory factors and influences synaptic neuron plasticity in the hippocampus of rats. *Neurotox* 69: 108-120.

Yang Y, Wang X, Guo X, Hu P. 1994. [The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine]. *Chin J Epidemiol* 15(4): 296-298.

Yang Y, Wang X, Guo X, Hu P. 2008. The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine. *Fluoride* 41: 336-339.

Yao L, Zhou J, Wang S, Cui K, Lin F. 1996. Analysis on TSH and intelligence level of children with dental Fluorosis in a high fluoride area. *Lit Inf Prev Med* 2(1): 26-27.

Yao Y. 1997. Comparable analysis on the physical and mental development of children in endemic fluorosis area with water improvement and without water improvement. *Lit Inf Prev Med* 3(1): 42-43.

Yasmin S, Ranjan S, D'Souza D. 2013. Effect of excess fluoride ingestion on human thyroid function in Gaya Region, Bihar, India. *Toxicol Environ Chem* 95: 1235-1243.

Yazdi SM, Sharifian A, Dehghani-Beshne M, Momeni VR, Aminian O. 2011. Effects of fluoride on psychomotor performance and memory of aluminum potroom workers. *Fluoride* 44: 158-162.

Yu X, Chen J, Li Y, Liu H, Hou C, Zeng Q, Cui Y, Zhao L, Li P, Zhou Z, Pang S, Tang S, Tian K, Zhao Q, Dong L, Xu C, Zhang X, Zhang S, Liu L, Wang A. 2018. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. *Environ Int* 118: 116-124.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 1996. [Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis]. *Chin J Endemiol* 15(5): 257-259.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 2008. Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis. *Fluoride* 41: 134-138.

Yuan J, Li Q, Niu R, Wang J. 2019. Fluoride exposure decreased learning ability and the expressions of the insulin receptor in male mouse hippocampus and olfactory bulb. *Chemosphere* 224: 71-76.

Zhang J, Yao H, Chen Y. 1998. [The effect of high levels of arsenic and fluoride on the development of children's intelligence]. *Chin J Public Health* 17(2): 119.

Zhang KL, Lou DD, Guan ZZ. 2015a. Activation of the AGE/RAGE system in the brains of rats and in SH-SY5Y cells exposed to high level of fluoride might connect to oxidative stress. *Neurotoxicol Teratol* 48: 49-55.

Zhang S, Zhang X, Liu H, Qu W, Guan Z, Zeng Q, Jiang C, Gao H, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Cui Y, Yu L, Wang A. 2015b. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. *Toxicol Sci* 144: 238-245.

Zhao LB, Liang GH, Zhang DN, Wu XR. 1996. Effect of a high fluoride water supply on children's intelligence. *Fluoride* 29: 190-192.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao Q, Tian Z, Zhou G, Niu Q, Chen J, Li P, Dong L, Xia T, Zhang S, Wang A. 2020. SIRT1-dependent mitochondrial biogenesis supports therapeutic effects of resveratrol against neurodevelopment damage by fluoride. *Theranostics* 10: 4822-4838.

Zheng X, Sun Y, Ke L, Ouyang W, Zhang Z. 2016. Molecular mechanism of brain impairment caused by drinking-acquired fluorosis and selenium intervention. *Environ Toxicol Pharmacol* 43: 134-139.

Zhou G, Tang S, Yang L, Niu Q, Chen J, Xia T, Wang S, Wang M, Zhao Q, Liu L, Li P, Dong L, Yang K, Zhang S, Wang A. 2019. Effects of long-term fluoride exposure on cognitive ability and the underlying mechanisms: Role of autophagy and its association with apoptosis. *Toxicol Appl Pharmacol* 378: 114608.

Zhou T, Duan L-J, Ding Z, Yang R-P, Li S-H, Xi Y, Cheng X-M, Hou J-X, Wen S-B, Chen J, Cui L-X, Ba Y. 2012. Environmental fluoride exposure and reproductive hormones in male living in endemic fluorosis villages in China. *Life Sci J* 9(4): 1-7.

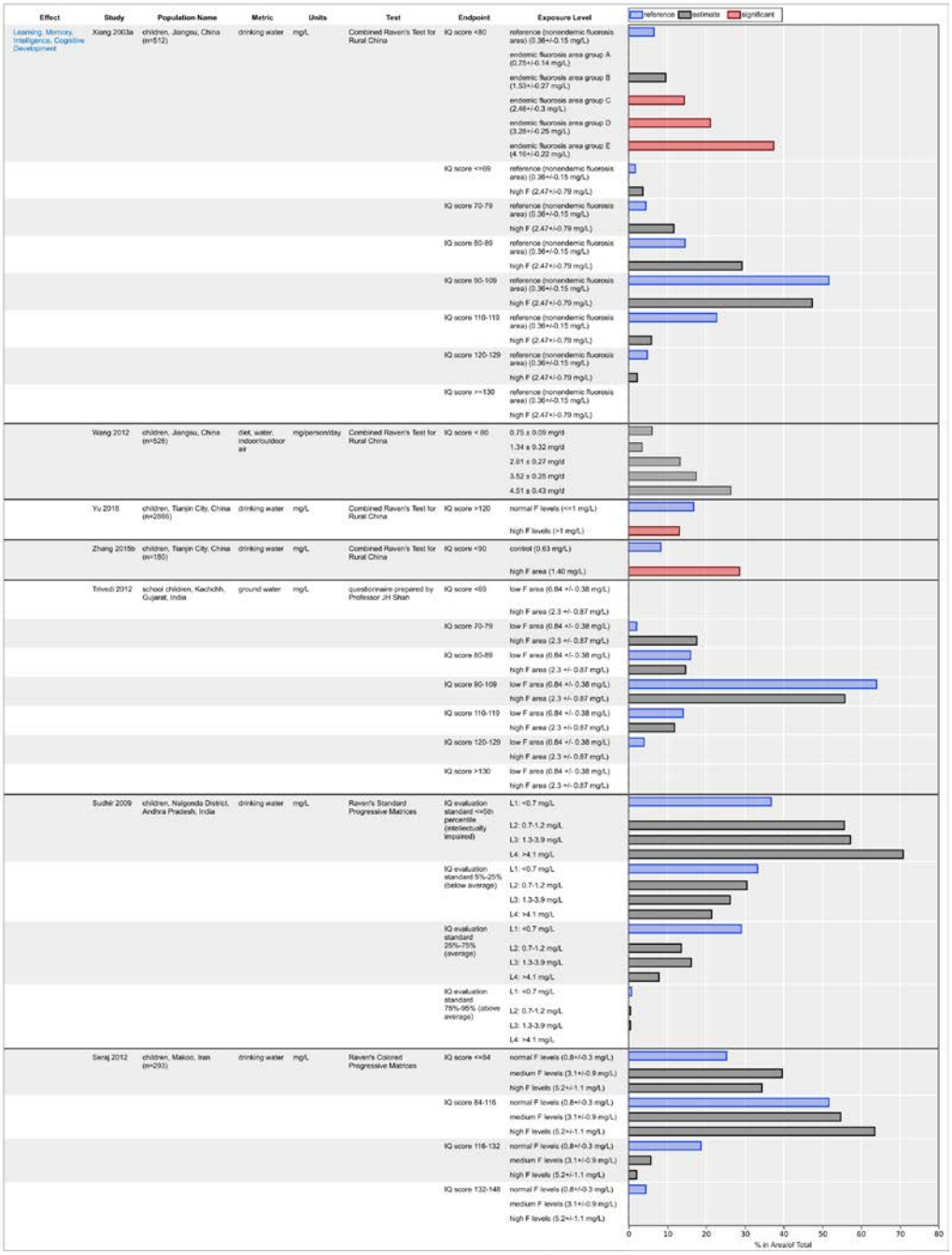
Zhu YP, Xi SH, Li MY, Ding TT, Liu N, Cao FY, Zeng Y, Liu XJ, Tong JW, Jiang SF. 2017. Fluoride and arsenic exposure affects spatial memory and activates the ERK/CREB signaling pathway in offspring rats. *Neurotox* 59: 56-64.

Zohouri FV, Swinbank CM, Maguire A, Moynihan PJ. 2006. Is the fluoride/creatinine ratio of a spot urine sample indicative of 24-h urinary fluoride? *Community Dent Oral Epidemiol* 34(2): 130-138.

DATA FIGURES

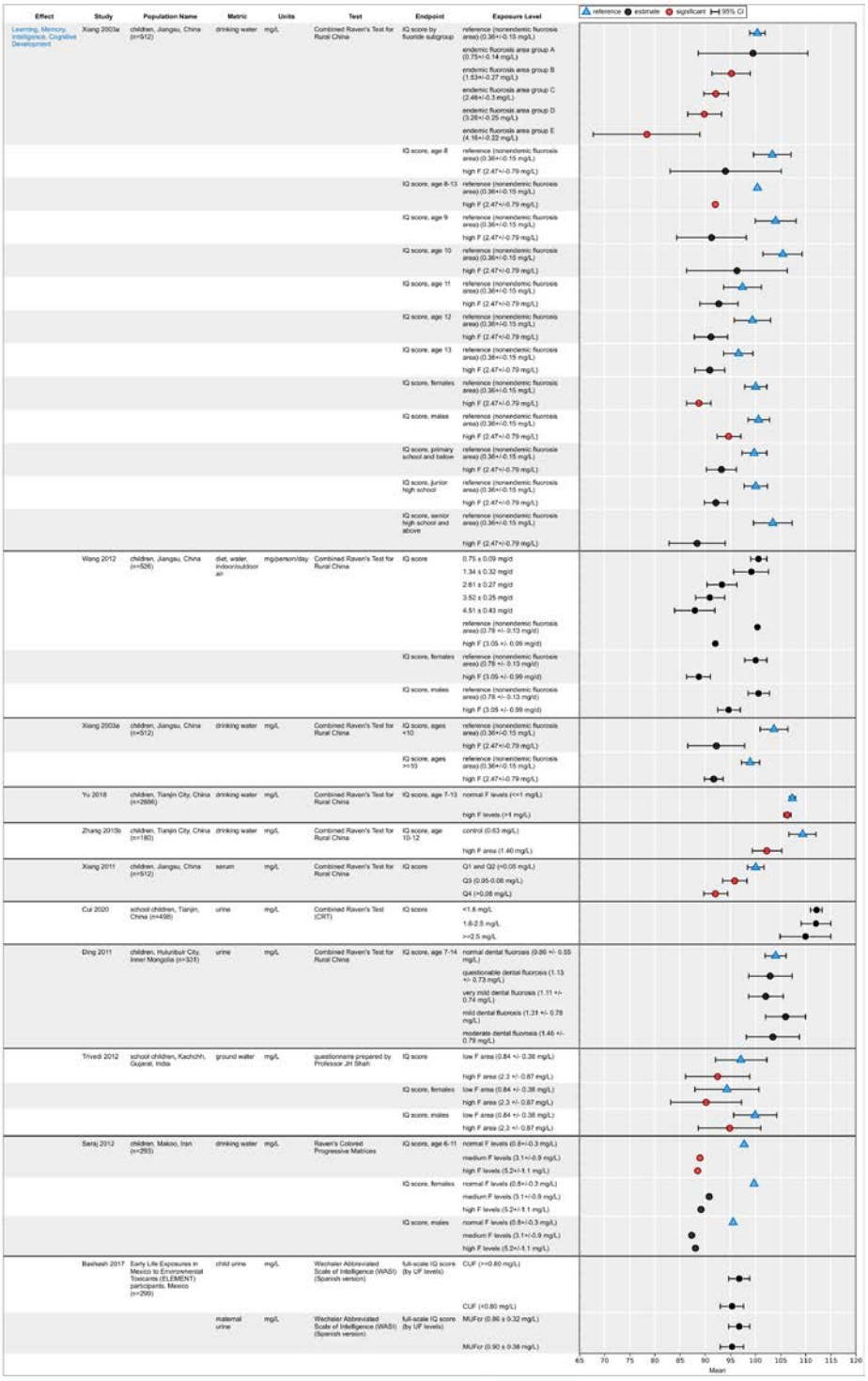
Neurodevelopmental or Cognitive Effects and Outcomes

Figure D1. IQ Distribution in Children by Fluoride Exposure (lower risk-of-bias studies; presented as % in area or % of total group)



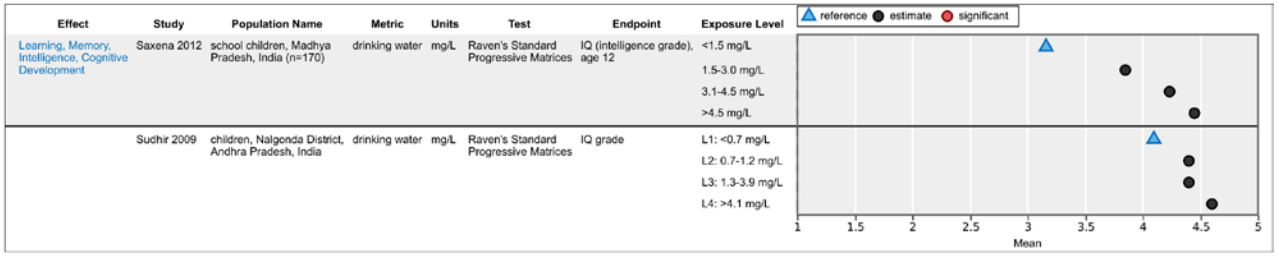
Interactive figure and additional study details in HAWC here. "F" represents fluoride. Differences in intelligence between the reference group and treatment groups were statistically significant although significance was not reported separately for each score level.

Figure D2. Mean IQ in Children by Fluoride Exposure (lower risk-of-bias studies)



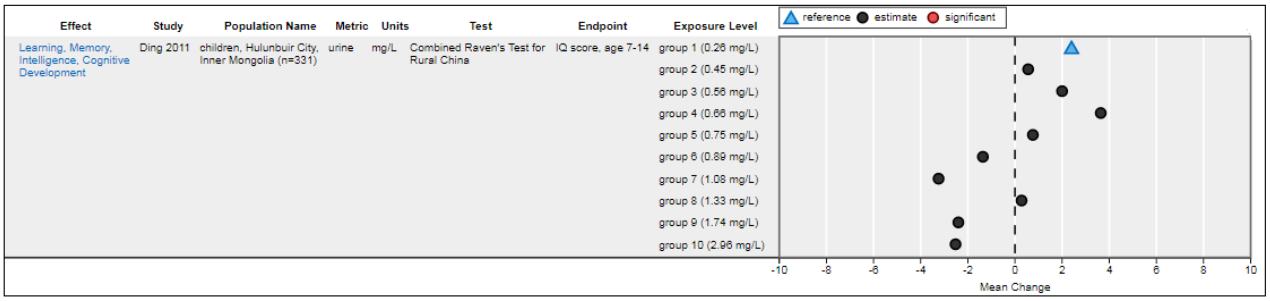
Interactive figure and additional study details in HAWC here. "F" represents fluoride. Three additional publications based on subsample of the larger Yu *et al.* (2018) cohort were identified (Zhao *et al.* 2019, Zhou *et al.* 2019, Zhao *et al.* 2020); however, results from these studies are not presented here. The main study by Yu *et al.* (2018) is considered a better representation of the IQ results. For all studies, SDs are available and can be viewed in HAWC by clicking the data points within the plot area; however, 95% CIs could not be calculated for Seraj *et al.* (2012) because Ns are not available for exposure groups.

Figure D3. Intelligence Grade in Children by Fluoride Exposure (lower risk-of-bias studies; presented as mean)



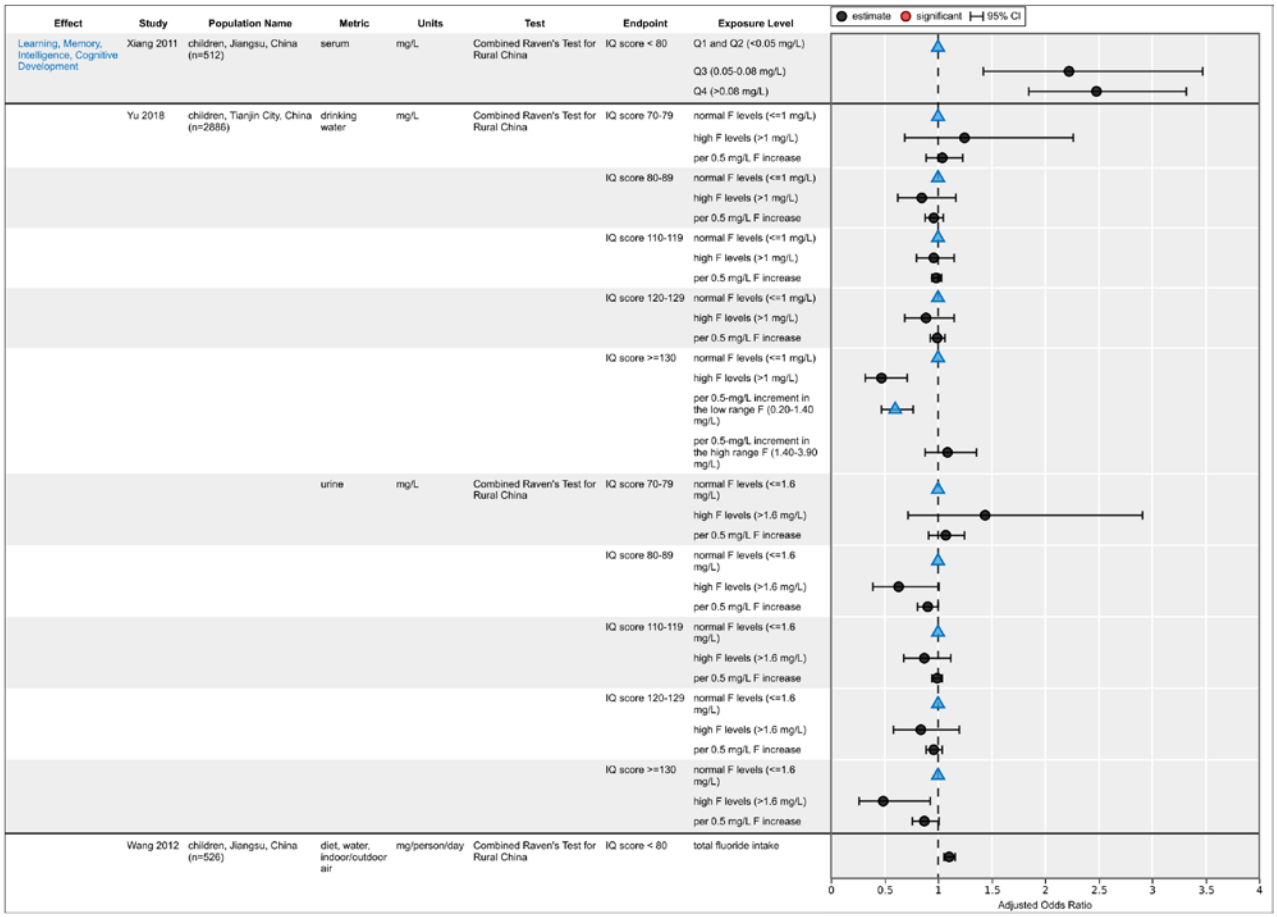
Interactive figure and additional study details in HAWC here. For Saxena *et al.* (2012), children's intelligence was measured using the Raven's Standard Progressive Matrices. Children's scores were converted to percentile and specific grades were allotted based on the percentiles. Grades ranged from intellectually superior (Grade I) to intellectually impaired (Grade V). Results for Soto-Barreras 2019 are not presented here. Outcomes in the study were presented as levels of fluoride exposure associated with each intelligence grade. Results reported were not significant.

Figure D4. Mean Change in IQ in Children by Fluoride Exposure (lower risk-of-bias studies)



Interactive figure and additional study details in HAWC here. For Ding *et al.* (2011), SDs are available and can be viewed in HAWC by clicking the data points within the plot area; however, 95% CIs could not be calculated because Ns for each exposure group are not available.

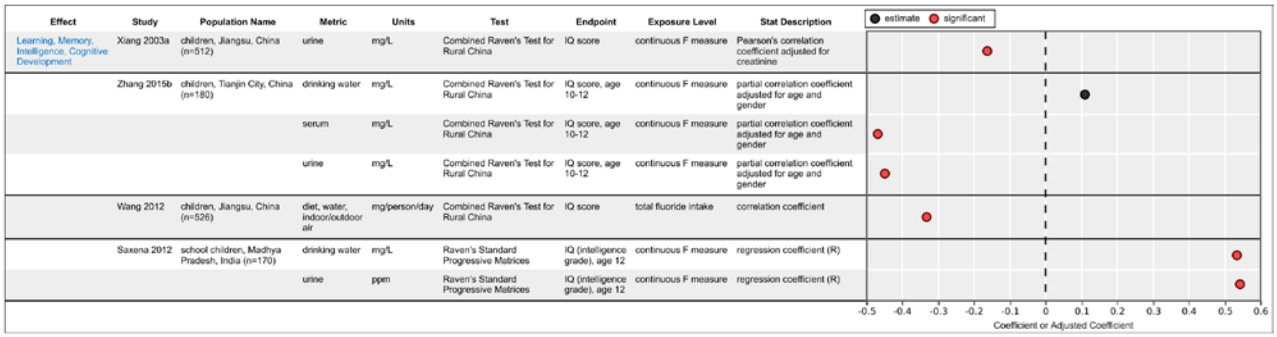
Figure D5. IQ Score in Children by Fluoride Exposure (lower risk-of-bias studies; presented as adjusted OR)



Effect	Study	Population Name	Metric	Units	Test	Endpoint	Exposure Level
Learning, Memory, Intelligence, Cognitive Development	Xiang 2011	children, Jiangsu, China (n=512)	serum	mg/L	Combined Raven's Test for Rural China	IQ score < 80	Q1 and Q2 (<0.05 mg/L)
							Q3 (0.05-0.08 mg/L)
							Q4 (>0.08 mg/L)
	Yu 2018	children, Tianjin City, China (n=2886)	drinking water	mg/L	Combined Raven's Test for Rural China	IQ score 70-79	normal F levels (<=1 mg/L)
							high F levels (>1 mg/L)
							per 0.5 mg/L F increase
						IQ score 80-89	normal F levels (<=1 mg/L)
							high F levels (>1 mg/L)
							per 0.5 mg/L F increase
						IQ score 110-119	normal F levels (<=1 mg/L)
							high F levels (>1 mg/L)
							per 0.5 mg/L F increase
						IQ score 120-129	normal F levels (<=1 mg/L)
							high F levels (>1 mg/L)
							per 0.5 mg/L F increase
						IQ score >=130	normal F levels (<=1 mg/L)
							high F levels (>1 mg/L)
							per 0.5-mg/L increment in the low range F (0.20-1.40 mg/L)
							per 0.5-mg/L increment in the high range F (1.40-3.90 mg/L)
			urine	mg/L	Combined Raven's Test for Rural China	IQ score 70-79	normal F levels (<=1.6 mg/L)
							high F levels (>1.6 mg/L)
							per 0.5 mg/L F increase
						IQ score 80-89	normal F levels (<=1.6 mg/L)
							high F levels (>1.6 mg/L)
							per 0.5 mg/L F increase
						IQ score 110-119	normal F levels (<=1.6 mg/L)
							high F levels (>1.6 mg/L)
							per 0.5 mg/L F increase
						IQ score 120-129	normal F levels (<=1.6 mg/L)
							high F levels (>1.6 mg/L)
							per 0.5 mg/L F increase
						IQ score >=130	normal F levels (<=1.6 mg/L)
							high F levels (>1.6 mg/L)
							per 0.5 mg/L F increase
	Wang 2012	children, Jiangsu, China (n=526)	diet, water, indoor/outdoor air	mg/person/day	Combined Raven's Test for Rural China	IQ score < 80	total fluoride intake

Interactive figure and additional study details in HAWC here. For Xiang *et al.* (2011), there was a significant linear trend across different levels of serum fluoride for IQ score < 80 ($p < 0.001$). For Yu *et al.* (2018), significance levels by IQ score were not reported.

Figure D6. Correlations between IQ Score and Fluoride Exposure in Children (lower risk-of-bias studies; presented as coefficient)

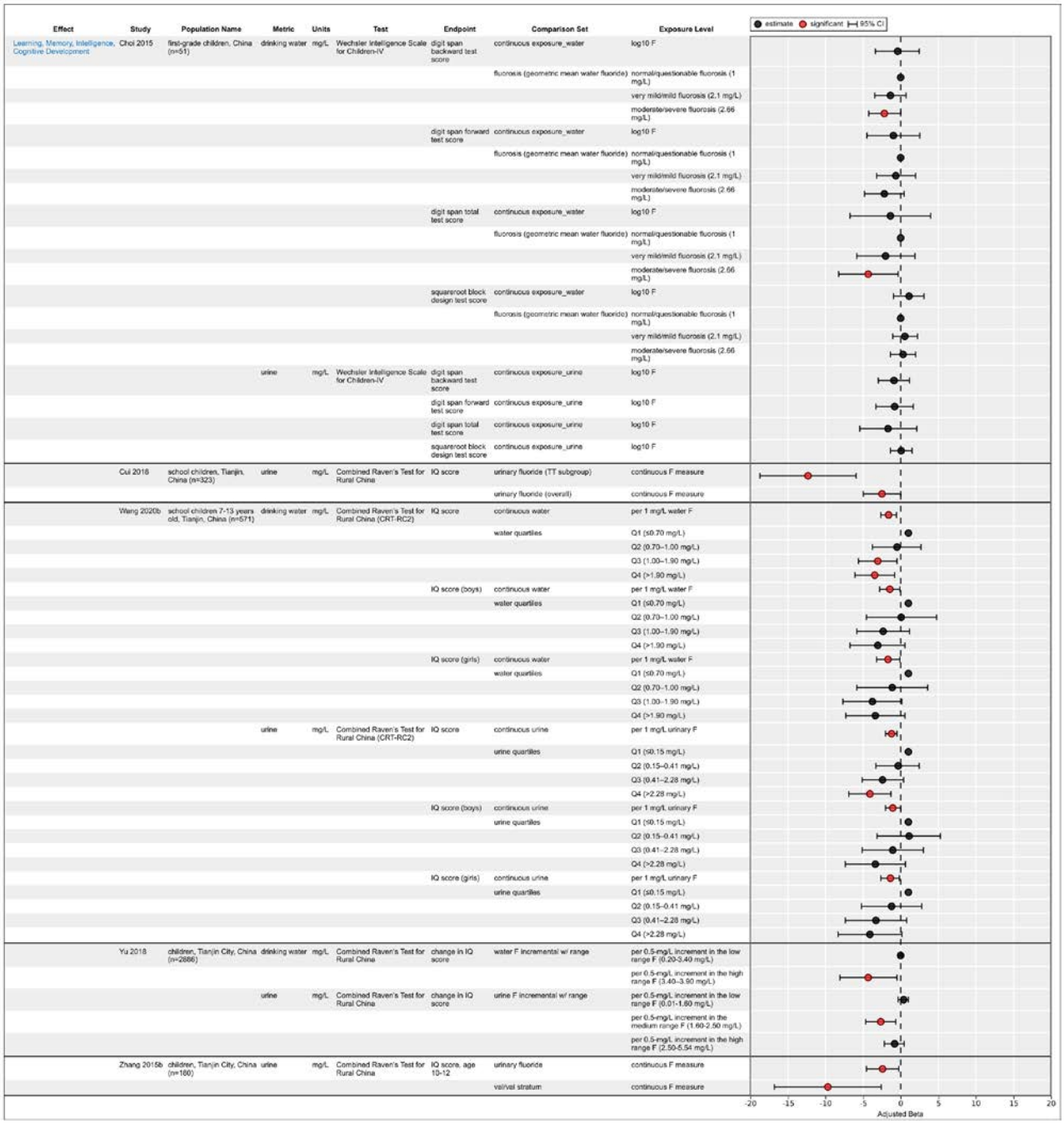


Effect	Study	Population Name	Metric	Units	Test	Endpoint	Exposure Level	Stat Description
Learning, Memory, Intelligence, Cognitive Development	Xiang 2003a	children, Jiangsu, China (n=512)	urine	mg/L	Combined Raven's Test for Rural China	IQ score	continuous F measure	Pearson's correlation coefficient adjusted for creatinine
	Zhang 2015b	children, Tianjin City, China (n=180)	drinking water	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
			serum	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
			urine	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	continuous F measure	partial correlation coefficient adjusted for age and gender
	Wang 2012	children, Jiangsu, China (n=526)	diet, water, indoor/outdoor air	mg/person/day	Combined Raven's Test for Rural China	IQ score	total fluoride intake	correlation coefficient
	Saxena 2012	school children, Madhya Pradesh, India (n=170)	drinking water	mg/L	Raven's Standard Progressive Matrices	IQ (intelligence grade), age 12	continuous F measure	regression coefficient (R)
			urine	ppm	Raven's Standard Progressive Matrices	IQ (intelligence grade), age 12	continuous F measure	regression coefficient (R)

Interactive figure and additional study details in HAWC here. "F" represents fluoride. For Saxena *et al.* (2012), a significant relationship between water fluoride level and intelligence grade was observed. Increasing intelligence grades reflected increasing levels of impairment (reduced intelligence) in children. Zhao *et al.* (2020) and Zhou *et al.* (2019) also had correlations, but these were based on a subsample of the Yu *et al.* (2018) study (which presented betas and provided a better representation of the IQ data).

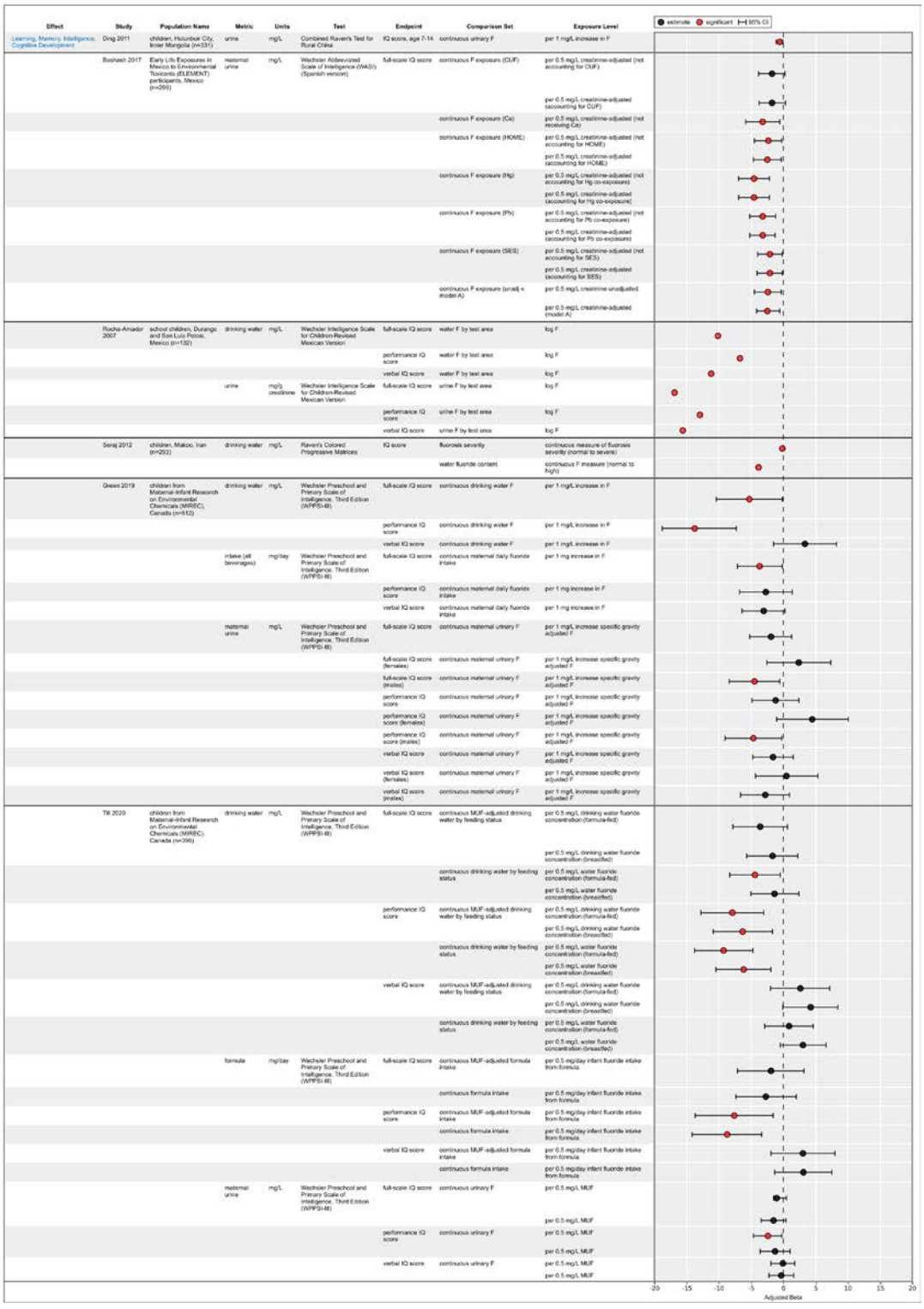
Figure D7. Correlations between IQ Score and Fluoride Exposure in Children (lower risk-of-bias studies; presented as adjusted beta)—(a) China; (b) all other areas

(a)



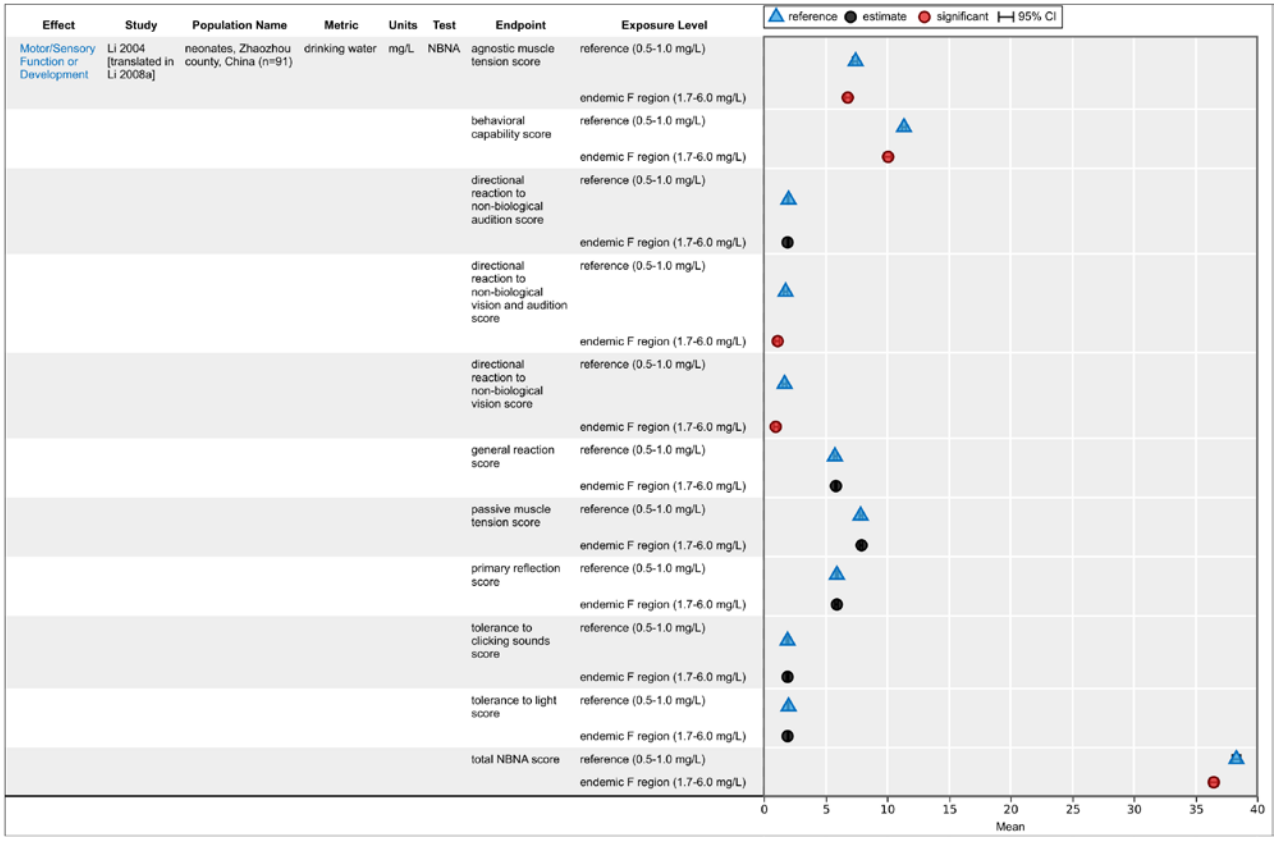
Effect	Study	Population Name	Metric	Units	Test	Endpoint	Comparison Set	Exposure Level
Learning, Memory, Intelligence, Cognitive Development	Choi 2015	first-grade children, China (n=51)	drinking water	mg/L	Wechsler Intelligence Scale for Children-IV	digit span backward test score	continuous exposure_water	log10 F
							fluorosis (geometric mean water fluoride)	normal/questionable fluorosis (1 mg/L)
								very mild/mild fluorosis (2.1 mg/L)
								moderate/severe fluorosis (2.66 mg/L)
						digit span forward test score	continuous exposure_water	log10 F
							fluorosis (geometric mean water fluoride)	normal/questionable fluorosis (1 mg/L)
								very mild/mild fluorosis (2.1 mg/L)
								moderate/severe fluorosis (2.66 mg/L)
						digit span total test score	continuous exposure_water	log10 F
							fluorosis (geometric mean water fluoride)	normal/questionable fluorosis (1 mg/L)
								very mild/mild fluorosis (2.1 mg/L)
								moderate/severe fluorosis (2.66 mg/L)
						squareroot block design test score	continuous exposure_water	log10 F
							fluorosis (geometric mean water fluoride)	normal/questionable fluorosis (1 mg/L)
								very mild/mild fluorosis (2.1 mg/L)
								moderate/severe fluorosis (2.66 mg/L)
			urine	mg/L	Wechsler Intelligence Scale for Children-IV	digit span backward test score	continuous exposure_urine	log10 F
						digit span forward test score	continuous exposure_urine	log10 F
						digit span total test score	continuous exposure_urine	log10 F
						squareroot block design test score	continuous exposure_urine	log10 F
	Cui 2016	school children, Tianjin, China (n=323)	urine	mg/L	Combined Raven's Test for Rural China	IQ score	urinary fluoride (TT subgroup)	continuous F measure
							urinary fluoride (overall)	continuous F measure
	Wang 2020b	school children 7-13 years old, Tianjin, China (n=571)	drinking water	mg/L	Combined Raven's Test for Rural China (CRT-RC2)	IQ score	continuous water	per 1 mg/L water F
							water quartiles	Q1 (≤0.70 mg/L)
								Q2 (0.70–1.00 mg/L)
								Q3 (1.00–1.90 mg/L)
								Q4 (>1.90 mg/L)
						IQ score (boys)	continuous water	per 1 mg/L water F
							water quartiles	Q1 (≤0.70 mg/L)
								Q2 (0.70–1.00 mg/L)
								Q3 (1.00–1.90 mg/L)
								Q4 (>1.90 mg/L)
						IQ score (girls)	continuous water	per 1 mg/L water F
							water quartiles	Q1 (≤0.70 mg/L)
								Q2 (0.70–1.00 mg/L)
								Q3 (1.00–1.90 mg/L)
								Q4 (>1.90 mg/L)
			urine	mg/L	Combined Raven's Test for Rural China (CRT-RC2)	IQ score	continuous urine	per 1 mg/L urinary F
							urine quartiles	Q1 (≤0.15 mg/L)
								Q2 (0.15–0.41 mg/L)
								Q3 (0.41–2.28 mg/L)
								Q4 (>2.28 mg/L)
						IQ score (boys)	continuous urine	per 1 mg/L urinary F
							urine quartiles	Q1 (≤0.15 mg/L)
								Q2 (0.15–0.41 mg/L)
								Q3 (0.41–2.28 mg/L)
								Q4 (>2.28 mg/L)
						IQ score (girls)	continuous urine	per 1 mg/L urinary F
							urine quartiles	Q1 (≤0.15 mg/L)
								Q2 (0.15–0.41 mg/L)
								Q3 (0.41–2.28 mg/L)
								Q4 (>2.28 mg/L)
	Yu 2018	children, Tianjin City, China (n=2686)	drinking water	mg/L	Combined Raven's Test for Rural China	change in IQ score	water F incremental w/ range	per 0.5-mg/L increment in the low range F (0.20-3.40 mg/L)
								per 0.5-mg/L increment in the high range F (3.40–3.90 mg/L)
			urine	mg/L	Combined Raven's Test for Rural China	change in IQ score	urine F incremental w/ range	per 0.5-mg/L increment in the low range F (0.01-1.80 mg/L)
								per 0.5-mg/L increment in the medium range F (1.60-2.50 mg/L)
								per 0.5-mg/L increment in the high range F (2.50-5.54 mg/L)
	Zhang 2015b	children, Tianjin City, China (n=160)	urine	mg/L	Combined Raven's Test for Rural China	IQ score, age 10-12	urinary fluoride	continuous F measure
							val/val stratum	continuous F measure

(b)



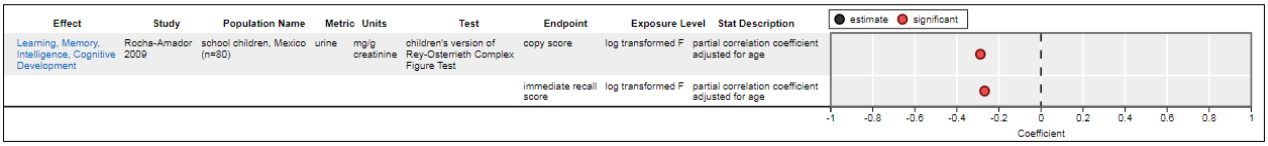
Interactive figure and additional study details in HAWC here for part (a) and here for part (b). "F" represents fluoride. For Yu *et al.* (2018), authors note an obvious decrease in the IQ score at water fluoride exposure levels between 3.40 mg/L and 3.90 mg/L and a similar adverse effect on IQ scores at urinary fluoride exposure levels from 1.60 mg/L to 2.50 mg/L, and so the changes in IQ score are indicated as significant; however, significance levels by change in IQ score were not reported.

Figure D8. Mean Motor/Sensory Scores in Children by Fluoride Exposure (lower risk-of-bias studies)



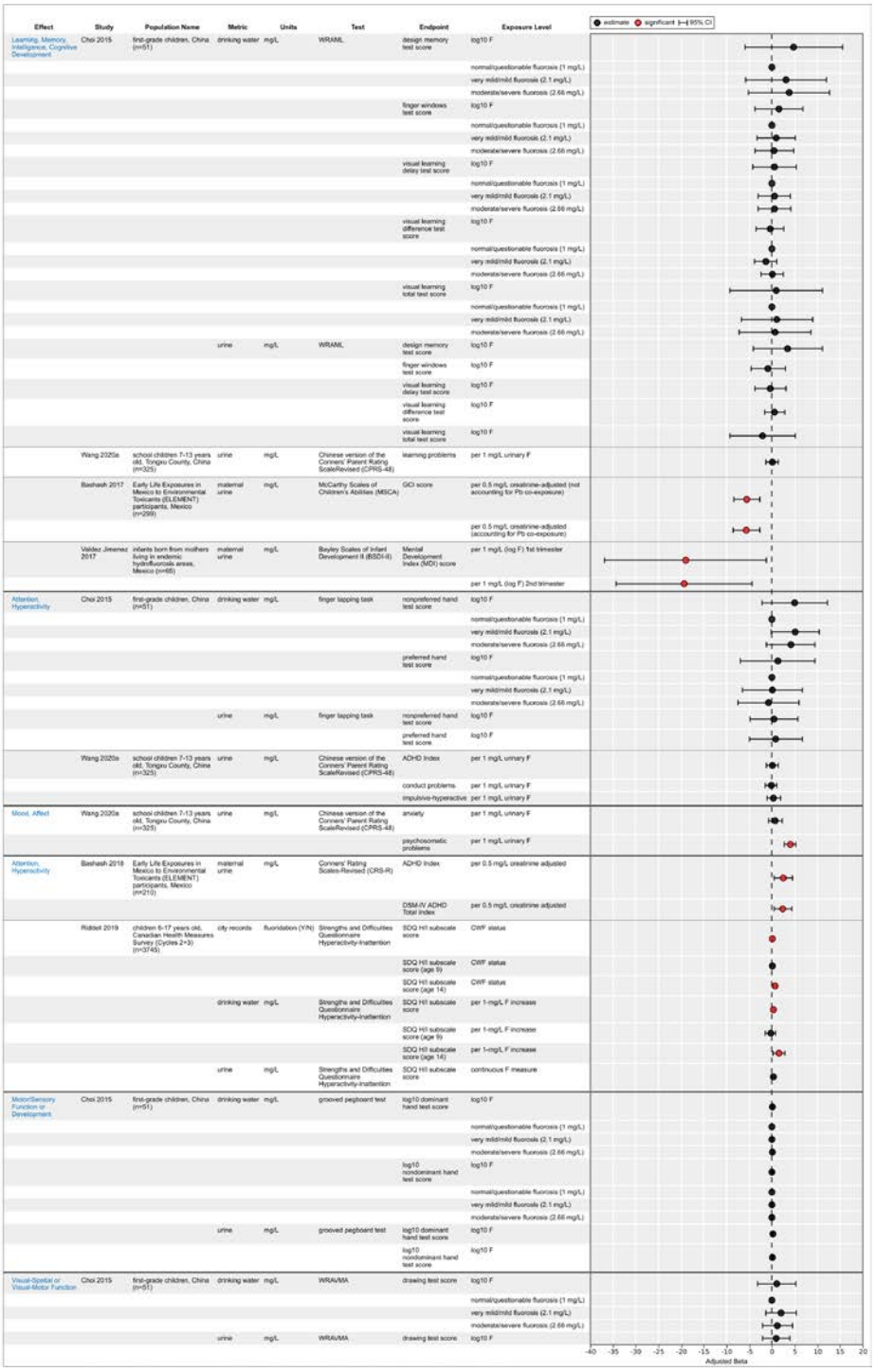
Interactive figure and additional study details in HAWC here. "F" represents fluoride. 95% CIs are small and are within figure symbols and may be difficult to see. Values for SDs and 95% CIs can be viewed in HAWC by clicking the data points within the plot area.

Figure D9. Correlations between Other Neurological Effects and Fluoride Exposure in Children (lower risk-of-bias studies; presented as coefficient)



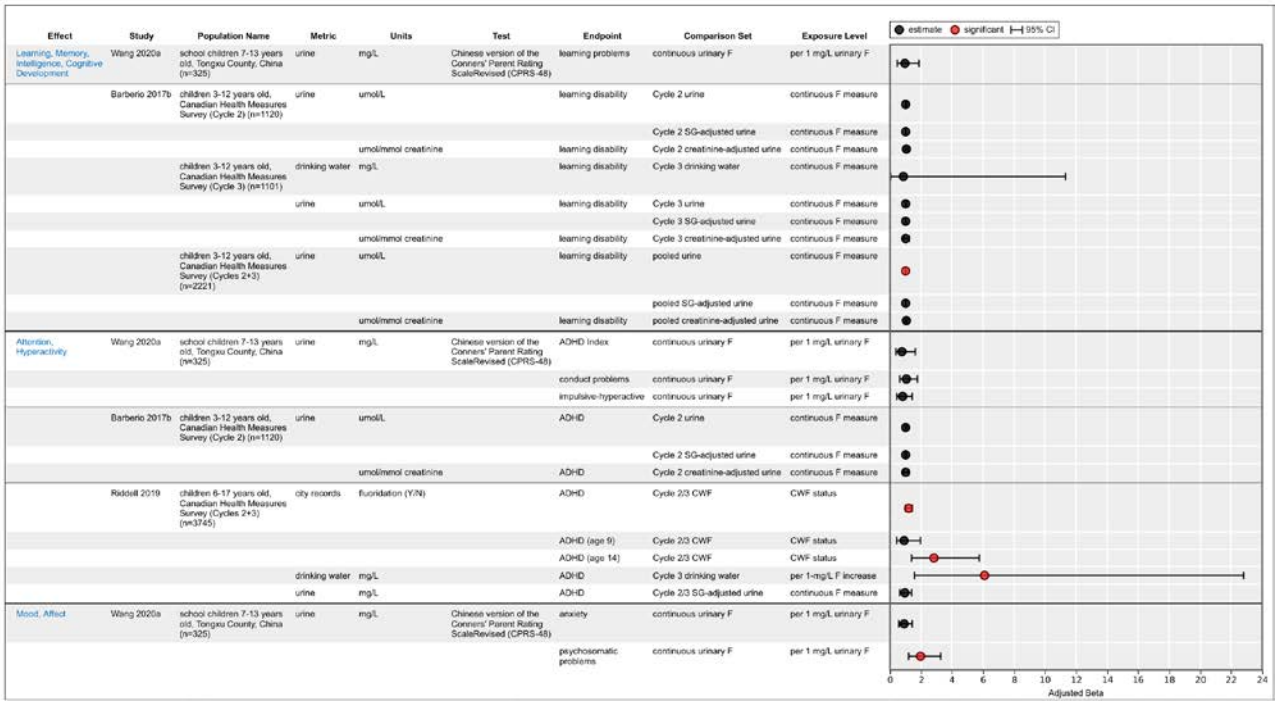
Interactive figure and additional study details in HAWC here. "F" represents fluoride.

Figure D10. Correlations between Other Neurological Effects and Fluoride Exposure in Children (lower risk-of-bias studies; presented as adjusted beta)



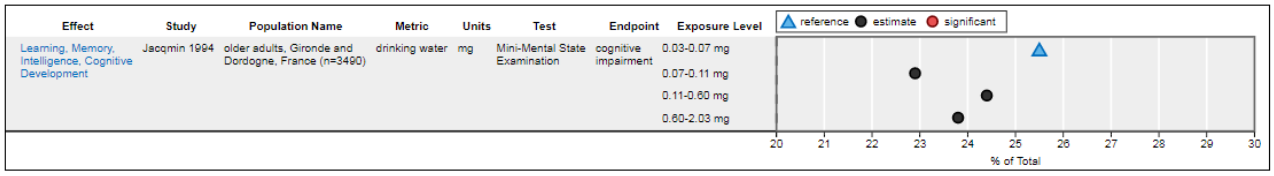
Interactive figure and additional study details in HAWC here. "F" represents fluoride. Bashash *et al.* (2018) observed significant associations between maternal urinary fluoride and ADHD-like symptoms related to inattention (an increase in 0.5 mg/L of maternal urinary fluoride was associated with a 2.84-point increase in the DSM-IV Inattention Index and a 2.54-point increase in Cognitive Problems and Inattention Index). These two scales contributed to the global ADHD Index and the DSM-IV ADHD Total Index shown here.

Figure D11. Correlations between Other Neurological Effects and Fluoride Exposure in Children (lower risk-of-bias studies; presented as adjusted OR)



Interactive figure and additional study details in HAWC here. "F" represents fluoride. Drinking water results for Barberio *et al.* (2017b) have a large confidence interval and are not completely visible in the figure. 95% CIs are 0.068–11.33 and can be viewed in HAWC by clicking the OR within the plot area.

Figure D12. Cognitive Impairment in Adults by Fluoride Exposure (lower risk-of-bias studies; presented as % of total group)



Interactive figure and additional study details in HAWC here. Results from Li *et al.* (2016) suggested that fluoride exposure may be a risk factor for cognitive impairment in elderly subjects; however, results from the study were not conducive to presentation in this visualization.

ABOUT THIS REVIEW

Sources of Support

National Institute of Environmental Health Sciences/Division of the National Toxicology Program

Contributors

Evaluation Team

The evaluation team is composed of federal staff and contractor staff support.

Name	Affiliation
Kyla Taylor, PhD	NIEHS/DNTP, Project Lead
John Bucher, PhD	NIEHS/DNTP
Andrew Rooney, PhD	NIEHS/DNTP
Vickie Walker	NIEHS/DNTP
Cynthia J. Willson, PhD, DVM, DACVP	ILS
Louise Assem, PhD	ICF [f]
Carlye A. Austin, PhD	ICF [d]
Robyn Blain, PhD	ICF [abcdefg]
Natalie Blanton, MPH	ICF [a]
Kristin Bornstein, PhD	ICF [f]
Canden Byrd	ICF [a]
Michelle Cawley	ICF [c]
Jonathan Cohen, PhD	ICF [g]
Ryan Cronk, PhD	ICF [d]
Sorina E. Eftim, PhD	ICF [fg]
Anna Engstrom, PhD	ICF [abcdf]
Jeremy S. Frye, MLS	ICF [a]
Kaitlin Geary	ICF [c]
Susan Goldhaber, MPH	ICF [e]
Ali Goldstone, MPH	ICF [fg]
Pamela Hartman, MEM	ICF [abceg]
Cara Henning, PhD	ICF [a]
Melinda Hoang, MPH	ICF [d]
Tao Hong	ICF [ef]
Jennifer Hsieh, MSPH	ICF [f]
Penelope Kellar	ICF [a]
Courtney Lemeris	ICF [a]
Cynthia J. Lin, PhD MSPH	ICF [g]
Alex Lindahl, MPH	ICF [f]
Kristen Magnuson, MESM	ICF [abcde]
Rachel McGill	ICF [c]
Shannon McGinnis, PhD	ICF [d]

Name	Affiliation
William Mendez, PhD	ICF [b]
Whitney Mitchell	ICF [a]
Revathi Muralidharan	ICF [de]
Johanna Rochester, PhD	ICF [def]
Amanda Ross, MEd	ICF [d]
Pam Ross, MSPH	ICF [df]
Jennifer Seed, PhD	ICF [de]
Karen E. Setty, MESM, PhD	ICF [d]
Codi Sharp	ICF [de]
Robert Shin, MHS	ICF [e]
Kelly Shipkowski, PhD	ICF [df]
Christopher A. Sibrizzi, MPH	ICF [abcdefg]
Raquel A. Silva, PhD	ICF [d]
Samantha J. Snow, PhD, DABT	ICF [ef]
Parnian Soleymani, MS	ICF [d]
Anna Stamatogiannakis	ICF [a]
Nicole Vetter, MLS	ICF [a]
River Williams, BS	ICF [c]

Note: the roles of individual contractors differed: [a] indicates monograph development; [b] indicates review of data, results, and analyses; [c] indicates database and HAWC support; [d] indicates literature screening, [e] indicates data extraction, [f] indicates risk-of-bias assessment, [g] indicates meta-analysis

Peer Reviewers

The peer reviewers were outside experts selected for their experience with fluoride, developmental neurobehavioral toxicity, and systematic review procedures. Peer reviewers were screened for conflict of interest prior to their service and did not report any conflicts of interest. Service as a peer reviewer does not necessarily indicate that the reviewer endorses the final document.

Protocol Reviewers

Name	Affiliation
Joseph Braun, PhD [a]	Brown University
Marie Sutton, PhD [a]	Health Research Board
Thomas Zoeller, PhD [a]	University of Massachusetts, Amherst
Thomas Webster, PhD [a]	Boston University
Gail Wasserman, PhD [a]	Columbia University
Suril Mehta, MPH [b]	NIEHS
Tianjing Li, PhD [b]	University of Colorado Denver

[a] Reviewer of the systematic review protocol; [b] Reviewer of meta-analysis protocol
no conflicts of interest declared

Technical Review of Draft Monograph

Name	Affiliation
Freya Kamal, PhD	NIEHS (retired)

no conflicts of interest declared

Protocol History and Revisions

Date	Activity or revision
December 14, 2016	**Draft evaluation protocol reviewed:** sent to technical advisors for peer review
April 10, 2017	**Draft human risk-of-bias protocol reviewed;** sent to technical advisors for peer review
May 2, 2017	**Draft animal risk-of-bias protocol reviewed;** sent to technical advisors for peer review
June 2017	**Evaluation protocol finalized:** Review protocol finalized for use and posting
May 2019	**Revised protocol:** Revised review protocol posted
September 2020	**Revised protocol:** Revised review protocol posted

APPENDICES

Appendix 1. Literature Search Strategy

The strategy for this search is broad for the consideration of neurodevelopmental or cognitive endpoints and comprehensive for fluoride as an exposure or treatment in order to ensure inclusion of relevant papers. The search terms for PubMed are provided below. The specific search strategies for other databases are available in the protocol (https://ntp.niehs.nih.gov/go/785076).

Database	Search Terms
PUBMED	((Fluorides[mh:noexp] OR fluorides, topical[mh] OR sodium fluoride[mh] OR Fluorosis, Dental[mh] OR fluorosis[tiab] OR fluorid*[tiab] OR flurid*[tiab] OR fluorin*[tiab] OR florin*[tiab]) NOT (18F[tiab] OR f-18[tiab] OR 19F[tiab] OR f-19[tiab] OR f-labeled[tiab] OR "fluorine-18"[tiab] OR "fluorine-19"[tiab] OR pet-scan[tiab] OR radioligand*[tiab])) AND ((Aryl Hydrocarbon Hydroxylases[mh] OR Aryl Hydrocarbon Receptor Nuclear Translocator[mh] OR Behavior and Behavior Mechanisms[mh] OR Gene Expression Regulation[mh] OR Glucuronosyltransferase[mh] OR Intelligence tests[mh] OR Malate Dehydrogenase[mh] OR Mediator Complex Subunit 1[mh] OR Mental disorders[mh] OR Mental processes[mh] OR Monocarboxylic Acid Transporters[mh] OR Myelin Basic Protein[mh] OR nervous system[mh] OR nervous system diseases[mh] OR nervous system physiological phenomena[mh] OR Neurogranin[mh] OR Oligodendroglia[mh] OR Peroxisome Proliferator-Activated Receptors[mh] OR Psychological Phenomena and Processes[mh] OR Receptors, thyroid hormone[mh] OR Receptors, thyrotropin[mh] OR Retinoid X Receptors[mh] OR thyroid diseases[mh] OR thyroid hormones[mh] OR Thyrotropin-releasing hormone[mh] OR Thyroxine-Binding Proteins[mh] OR Pregnane X Receptor[supplementary concept] OR thyroid-hormone-receptor interacting protein[supplementary concept] OR Constitutive androstane receptor[supplementary concept] OR Academic performance[tiab] OR auditory[tiab] OR cortical[tiab] OR delayed development[tiab] OR developmental impairment[tiab] OR developmental-delay*[tiab] OR developmental-disorder*[tiab] OR euthyroid[tiab] OR gait[tiab] OR glia*[tiab] OR gliogenesis[tiab] OR hyperactiv*[tiab] OR impulse-control[tiab] OR iodide peroxidase[tiab] OR IQ[tiab] OR ischemi*[tiab] OR locomotor[tiab] OR mental deficiency[tiab] OR mental development[tiab] OR mental illness[tiab] OR mental-deficit[tiab] OR mobility[tiab] OR mood[tiab] OR morris-maze[tiab] OR morris-water[tiab] OR motor abilit*[tiab] OR Motor activities[tiab] OR motor performance[tiab] OR nerve[tiab] OR neural[tiab] OR neurobehav*[tiab] OR Neurocognitive impairment[tiab] OR neurodegenerat*[tiab] OR Neurodevelopment*[tiab] OR neurodisease*[tiab] OR neurologic*[tiab] OR neuromuscular[tiab] OR neuron*[tiab] OR neuropath*[tiab] OR obsessive compulsive[tiab] OR OCD[tiab] OR olfaction[tiab] OR olfactory[tiab] OR open-field-test[tiab] OR passive avoidance[tiab] OR plasticity[tiab] OR senil*[tiab] OR sociab*[tiab] OR speech*[tiab] OR spelling[tiab] OR stereotypic-movement*[tiab] OR synap*[tiab] OR tauopath*[tiab] OR Thyroglobulin[tiab] OR Thyroid disease*[tiab] OR thyroid gland[tiab] OR thyroid hormone*[tiab] OR thyronine*[tiab] OR visual motor[tiab] OR Visuospatial processing[tiab] OR water maze[tiab]) OR ((active-avoidance[tiab] OR ADHD[tiab] OR alzheimer*[tiab] OR amygdala[tiab] OR antisocial[tiab] OR anxiety[tiab] OR anxious[tiab] OR asperger*[tiab] OR attention deficit[tiab] OR autism[tiab] OR autistic[tiab] OR behavioral[tiab] OR behaviors[tiab] OR behavioural[tiab] OR behaviours[tiab] OR bipolar[tiab] OR cerebellum[tiab] OR cognition[tiab] OR cognitive[tiab] OR communication-disorder*[tiab] OR comprehension[tiab] OR cranial[tiab] OR dementia[tiab] OR dendrit*[tiab] OR dentate-gyrus[tiab] OR depression[tiab] OR dextrothyroxine[tiab] OR diiodothyronine*[tiab] OR diiodotyrosine[tiab] OR down syndrome[tiab] OR dyslexia[tiab] OR entorhinal cortex[tiab] OR epilep*[tiab] OR gangli*[tiab] OR goiter[tiab] OR graves-disease[tiab] OR hearing[tiab] OR hippocamp*[tiab] OR human development[tiab] OR hyperthyroid*[tiab] OR hypothalam*[tiab] OR hypothyroid*[tiab] OR impulsiv*[tiab] OR Intellectual

Database	Search Terms
	disability[tiab] OR intelligence[tiab] OR language[tiab] OR learning[tiab] OR lewy bod*[tiab] OR long-term potentiation[tiab] OR long-term synaptic depression[tiab] OR memory[tiab] OR mental disorder*[tiab] OR mental recall[tiab] OR monoiodotyrosine[tiab] OR Motor activity[tiab] OR motor skill*[tiab] OR multiple sclerosis[tiab] OR myxedema[tiab] OR Nervous system[tiab] OR nervous-system[tiab] OR neurit*[tiab] OR optic[tiab] OR palsy[tiab] OR panic[tiab] OR parahippocamp*[tiab] OR paranoia[tiab] OR paranoid[tiab] OR parkinson*[tiab] OR perception[tiab] OR perforant*[tiab] OR personality[tiab] OR phobia[tiab] OR problem solving[tiab] OR proprioception[tiab] OR psychomotor[tiab] OR reflex[tiab] OR risk taking[tiab] OR schizophrenia[tiab] OR seizure*[tiab] OR sensation*[tiab] OR sleep[tiab] OR smell[tiab] OR spatial behavior[tiab] OR stroke[tiab] OR substantia-nigra[tiab] OR taste[tiab] OR thyroiditis[tiab] OR thyrotoxicosis[tiab] OR Thyrotropin[tiab] OR thyroxine[tiab] OR triiodothyronine[tiab] OR vision[tiab]) NOT medline[sb]))

Appendix 2. List of Included Studies

Studies in Humans

As described in **Figure 4**, 159 human studies were included; however, full data extraction was only conducted on studies with neurological outcomes or thyroid hormone data. Data extraction was completed using HAWC. Data were extracted from a subset of included studies in humans (n = 116) and are available in HAWC based on outcome. The following lists of references are organized as studies that are available in HAWC followed by studies that are not available in HAWC. Specifically, data for primary neurodevelopmental or cognitive outcomes (learning, memory, and intelligence) and secondary neurobehavioral outcomes (anxiety, aggression, motor activity, or biochemical changes), as well as thyroid hormone level data, were extracted from included human studies and are available in HAWC. Data for included studies identified through the 2020 literature search update were only extracted for primary neurodevelopmental or cognitive outcomes; a subset of these studies (n = 5) also included secondary neurobehavioral outcomes and/or thyroid hormone level data that were not extracted because those data would not materially advance the human or mechanistic findings. Included human studies that only evaluated other thyroid-related effects such as goiters or thyroid size (n = 43) were not extracted and are not available in HAWC. The list below presents the 159 human studies that were included in the review. An overview of the screening results is outlined in the study selection diagram (**Figure 4**) that reports numbers of included studies as well as numbers of studies excluded for each reason at the full text review stage.

Studies Available in HAWC

An J, Mei S, Liu A, Fu Y, Wang C. 1992. [Effect of high level of fluoride on children's intelligence]. *Chin J Control Endem Dis* 7(2): 93-94.

Aravind A, Dhanya RS, Narayan A, Sam G, Adarsh VJ, Kiran M. 2016. Effect of fluoridated water on intelligence in 10-12-year-old school children. *J Int Soc Prev Community Dent* 6(Suppl 3): S237-S242.

Bai A, Li Y, Fan Z, Li X, Li P. 2014. [Intelligence and growth development of children in coal-burning-borne arsenism and fluorosis areas: An investigation study]. *Chin J Endemiol* 33(2): 160-163.

Barberio AM, Hosein FS, Quinonez C, McLaren L. 2017. Fluoride exposure and indicators of thyroid functioning in the Canadian population: Implications for community water fluoridation. *J Epidemiol Community Health* 71: 1019-1025.

Barberio AM, Quinonez C, Hosein FS, McLaren L. 2017. Fluoride exposure and reported learning disability diagnosis among Canadian children: Implications for community water fluoridation. *Can J Public Health* 108: 229-239.

Bashash M, Thomas D, Hu H, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Ettinger AS, Wright R, Zhang Z, Liu Y, Schnaas L, Mercado-Garcia A, Tellez-Rojo MM, Hernandez-Avila M. 2017. Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6-12 years of age in Mexico. *Environ Health Perspect* 125(9): 1-12.

Bashash M, Marchand M, Hu H, Till C, Martinez-Mier EA, Sanchez BN, Basu N, Peterson KE, Green R, Schnaas L, Mercado-Garcia A, Hernandez-Avila M, Tellez-Rojo MM. 2018. Prenatal fluoride exposure and attention deficit hyperactivity disorder (ADHD) symptoms in children at 6-12 years of age in Mexico City. *Environ Int* 121(Pt 1): 658-666.

Broadbent JM, Thomson WM, Moffitt TE, Poulton R. 2015. Community water fluoridation and intelligence response. *Am J Public Health* 105: 3-4.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 1991. [Research on the intellectual development of children in high fluoride areas]. *Chin J Control Endem Dis* 6(Suppl): 99-100.

Chen YX, Han FL, Zhoua ZL, Zhang HQ, Jiao XS, Zhang SC, Huang MC, Chang TQ, Dong YF. 2008. Research on the intellectual development of children in high fluoride areas. *Fluoride* 41: 120-124.

Chinoy NJ, Narayana MV. 1992. Studies on fluorosis in Mehsana District of North Gujarat. *Proc Zool Soc* 45: 157-161.

Choi AL, Zhang Y, Sun G, Bellinger DC, Wang K, Yang XJ, Li JS, Zheng Q, Fu Y, Grandjean P. 2015. Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study. *Neurotoxicol Teratol* 47: 96-101.

Cui Y, Zhang B, Ma J, Wang Y, Zhao L, Hou C, Yu J, Zhao Y, Zhang Z, Nie J, Gao T, Zhou G, Liu H. 2018. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. *Ecotoxicol Environ Saf* 165: 270-277.

Cui Y, Yu J, Zhang B, Guo B, Gao T, Liu H. 2020. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. *Neurosci Lett* 729: 134981.

Das K, Mondal NK. 2016. Dental fluorosis and urinary fluoride concentration as a reflection of fluoride exposure and its impact on IQ level and BMI of children of Laxmisagar, Simlapal Block of Bankura District, W.B., India. *Environ Monit Assess* 188: 218.

Ding Y, Sun H, Han H, Wang W, Ji X, Liu X, Sun D. 2011. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. *J Hazard Mater* 186: 1942-1946.

Du L, Wan C, Cao X, Liu J. 1992. [The effect of fluorine on the developing human brain]. *Chin J Pathol* 21(4): 218-220.

Du L, Wan C, Cao X, Liu J. 2008. The effect of fluorine on the developing human brain. *Fluoride* 41: 327-330.

Duan J, Zhao M, Wang L, Fang D, Wang Y, Wang W. 1995. A comparative analysis of the results of multiple tests in patients with chronic industrial fluorosis. Guizhou Med J 18(3): 179-180.
Erickson JD, Hay S. 1976. Water fluoridation and congenital malformations: No association. *J Am Dent Assoc* 93: 981-984.

Erickson JD, Hay S. 1976. Water fluoridation and congenital malformations: No association. *J Am Dent Assoc* 93: 981-984.

Erickson JD. 1980. Down syndrome, water fluoridation, and maternal age. *Teratology* 21: 177-180.

Eswar P, Nagesh L, Devaraj CG. 2011. Intelligent quotients of 12-14 year old school children in a high and low fluoride village in India. *Fluoride* 44: 168-172.

Fan Z, Dai H, Bai A, Li P, Li T, Li G. 2007. Effect of high fluoride exposure in children's intelligence. *J Environ Health* 24(10): 802-803.

Green R, Lanphear B, Hornung R, Flora D, Martinez-Mier EA, Neufeld R, Ayotte P, Muckle G, Till C. 2019. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. *JAMA Pediatr*: E1-E9.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 1991. [A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning]. *Chin J Epidemiol* 10(2): 98-100.

Guo XC, Wang RY, Cheng CF, Wei WS, Tang LM, Wang QS, Tang DX, Liu GW, He GD, Li SL. 2008. A preliminary investigation of the IQs of 7-13 year-old children from an area with coal burning-related fluoride poisoning. *Fluoride* 41: 125-128.

Guo ZY, He YH, Zhu QX. 2001. [Research on the neurobehavioral function of workers occupationally exposed to fluoride]. *Ind Hlth & Occup Dis* 27(6): 346-348.

Guo ZY, He YH, Zhu QX. 2008. Research on the neurobehavioral function of workers occupationally exposed to fluoride. *Fluoride* 41: 152-155.

He H, Cheng ZS, Liu WQ. 1989. [Effects of fluorine on the human fetus]. *J Control Endem Dis* 4(3): 136-138.

He H, Cheng ZS, Liu WQ. 2008. Effects of fluorine on the human fetus. *Fluoride* 41: 321-326.

He MX, Zhang CN. 2010. [Investigation of children's intelligence quotient and dental fluorosis in drinking water-type of endemic fluorosis area in Pucheng County, Shaanxi Province before and after drinking water change]. *Chin J Endemiol* 29: 547-548.

Hong FG, Cao YX, Yang D, Wang H. 2001. [Research on the effects of fluoride on child intellectual development under different environmental conditions]. *Chin Prim Health Care* 15(3): 56-57.

Hong FG, Cao YX, Yang D, Wang H. 2008. Research on the effects of fluoride on child intellectual development under different environmental conditions. *Fluoride* 41: 156-160.

Hosur MB, Puranik RS, Vanaki S, Puranik SR. 2012. Study of thyroid hormones free triiodothyronine (FT3), free thyroxine (FT4) and thyroid stimulating hormone (TSH) in subjects with dental fluorosis. *Eur J Dent* 6: 184-190.

Jacqmin H, Commenges D, Letenneur L, Barberger-Gateau P, Dartigues JF. 1994. Components of drinking water and risk of cognitive impairment in the elderly. *Am J Epidemiol* 139: 48-57.

Kang J, Cheng Y, Wu K, Lin S, He G, Jin Y. 2011. Effect of exposure to fluoride and arsenic in drinking water of Hangjinhouqi on children's intelligence. *Chinese School Health*: 679-681.

Karimzade S, Aghaei M, Mahvi AH. 2014. Investigation of intelligence quotient in 9-12 year-old children exposed to high- and low-drinking water fluoride in West Azerbaijan Province, Iran. *Fluoride* 47: 9-14.

Khan SA, Singh RK, Navit S, Chadha D, Johri N, Navit P, Sharma A, Bahuguna R. 2015. Relationship between dental fluorosis and intelligence quotient of school going children in and around Lucknow District: A cross-sectional study. *J Clin Diagn Res* 9(11): 10-15.

Khandare AL, Gourineni SR, Validandi V. 2017. Dental fluorosis, nutritional status, kidney damage, and thyroid function along with bone metabolic indicators in school-going children living in fluoride-affected hilly areas of Doda District, Jammu and Kashmir, India. *Environ Monit Assess* 189: 579.

Khandare AL, Validandi V, Gourineni SR, Gopalan V, Nagalla B. 2018. Dose-dependent effect of fluoride on clinical and subclinical indices of fluorosis in school going children and its mitigation by supply of safe drinking water for 5 years: An Indian study. *Environ Monit Assess* 190: 110.

Kheradpisheh Z, Mahvi AH, Mirzaei M, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018. Correlation between drinking water fluoride and TSH hormone by ANNs and ANFIS. *J Environ Health Sci Eng* 16(1): 11-18.

Kheradpisheh Z, Mirzaei M, Mahvi AH, Mokhtari M, Azizi R, Fallahzadeh H, Ehrampoush MH. 2018. Impact of drinking water fluoride on human thyroid hormones: A case-control study. *Sci Rep* 8: 2674.

Kumar V, Chahar P, Kajjari S, Rahman F, Bansal DK, Kapadia JM. 2018. Fluoride, thyroid hormone derangements and its correlation with tooth eruption pattern among the pediatric population from endemic and non-endemic fluorosis areas. *J Contemp Dent Pract* 19(12): 1512-1516.

Kundu H, Basavaraj P, Singla A, Gupta R, Singh K, Jain S. 2015. Effect of fluoride in drinking water on children's intelligence in high and low fluoride areas of Delhi. *J Indian Assoc Public Health Dent* 13(2): 116-121.

Lamberg M, Hausen H, Vartiainen T. 1997. Symptoms experienced during periods of actual and supposed water fluoridation. *Community Dent Oral Epidemiol* 25: 291-295.

Li F, Chen X, Huang R, Xie Y. 2009. The impact of endemic fluorosis caused by the burning of coal on the development of intelligence in children. *J Environ Health* 26(4): 838-840.

Li J, Yao L, Shao QL, Wu CY. 2004. [Effects of high fluoride level on neonatal neurobehavioral development]. *Chin J Endemiol* 23(5): 463-465.

Li J, Yao L, Shao QL, Wu CY. 2008. Effects of high fluoride level on neonatal neurobehavioral development. *Fluoride* 41: 165-170.

Li M, Gao Y, Cui J, Li Y, Li B, Liu Y, Sun J, Liu X, Liu H, Zhao L, Sun D. 2016. Cognitive impairment and risk factors in elderly people living in fluorosis areas in China. *Biol Trace Elem Res* 172: 53-60.

Li X, Hou G, Yu B, Yuan C, Liu Y, L Z. 2010. Investigation and analysis of children’s intelligence and dental fluorosis in high fluoride area. *J Med Pest Control* 26(3): 230-231.

Li XS, Zhi JL, Gao RO. 1995. Effect of fluoride exposure on the intelligence of children. *Fluoride* 28: 189-192.

Li Y, Li X, Wei S. 1994. [Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved]. *J West Chin University Med Sci* 25(2): 188-191.

Li Y, Li X, Wei S. 2008. Effects of high fluoride intake on child mental work capacity: Preliminary investigation into the mechanisms involved. *Fluoride* 41: 331-335.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2003. [Effects of endemic fluoride poisoning on the intellectual development of children in Baotou]. *Chin J Public Health Manag* 19(4): 337-338.

Li YP, Jing XY, Chen D, Lin L, Wang ZJ. 2008. Effects of endemic fluoride poisoning on the intellectual development of children in Baotou. *Fluoride* 41: 161-164.

Lin F, Ai H, Zhao H, Lin J, Jhiang J, Maimaiti. 1991. High fluoride and low iodine environment and subclinical cretinism in Xinjiang. *Endem Dis Bull* 6(2): 62-67.

Liu S, Lu Y, Sun Z, Wu L, Wang X, Yan S. 2000. [Report on the intellectual ability of children living in high-fluoride water areas]. *Chin J Control Endem Dis* 15(4): 231-232.

Liu SL, Lu Y, Sun ZR, Wu L, Lu WL, Wang XW, Yan S. 2008. Report on the intellectual ability of children living in high-fluoride water areas. *Fluoride* 41: 144-147.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S, Mentes A, DenBesten PK. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. *Physiol Behav* 206: 76-83.

Lu Y, Sun ZR, Wu LN, Wang X, Lu W, Liu SS. 2000. Effect of high-fluoride water on intelligence in children. *Fluoride* 33: 74-78.

Malin AJ, Till C. 2015. Exposure to fluoridated water and attention deficit hyperactivity disorder prevalence among children and adolescents in the United States: An ecological association. *Environ Health* 14: 17.

Malin AJ, Riddell J, McCague H, Till C. 2018. Fluoride exposure and thyroid function among adults living in Canada: Effect modification by iodine status. *Environ Int* 121(Pt 1): 667-674.

Michael M, Barot VV, Chinoy NJ. 1996. Investigations of soft tissue functions in fluorotic individuals of North Gujarat. *Fluoride* 29: 63-71.

Mondal D, Dutta G, Gupta S. 2016. Inferring the fluoride hydrogeochemistry and effect of consuming fluoride-contaminated drinking water on human health in some endemic areas of Birbhum District, West Bengal. *Environ Geochem Health* 38: 557-576.

Morgan L, Allred E, Tavares M, Bellinger D, Needleman H. 1998. Investigation of the possible associations between fluorosis, fluoride exposure, and childhood behavior problems. *Pediatr Dent* 20: 244-252.

Mustafa DE, Younis UM, Elhag SAA. 2018. The relationship between the fluoride levels in drinking water and the schooling performance of children in rural areas of Khartoum State, Sudan. *Fluoride* 51(2): 102-113.

Nagarajappa R, Pujara P, Sharda AJ, Asawa K, Tak M, Aapaliya P, Bhanushali N. 2013. Comparative assessment of intelligence quotient among children living in high and low fluoride areas of Kutch, India: A pilot study. *Iran J Public Health* 42: 813-818.

Peckham S, Lowery D, Spencer S. 2015. Are fluoride levels in drinking water associated with hypothyroidism prevalence in England? A large observational study of GP practice data and fluoride levels in drinking water. *J Epidemiol Community Health* 69: 619-624.

Poureslami HR, Horri A, Garrusi B. 2011. A comparative study of the IQ of children age 7-9 in a high and a low fluoride water city in Iran. *Fluoride* 44: 163-167.

Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. 1990. [Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children]. *J Control Endem Dis* 5(4): 203-204.

Qin LS, Huo SY, Chen RL, Chang YZ, Zhao MY. 2008. Using the Raven's Standard Progressive Matrices to determine the effects of the level of fluoride in drinking water on the intellectual ability of school-age children. *Fluoride* 41: 115-119.

Ratan SK, Rattan KN, Pandey RM, Singhal S, Kharab S, Bala M, Singh V, Jhanwar A. 2008. Evaluation of the levels of folate, vitamin B12, homocysteine and fluoride in the parents and the affected neonates with neural tube defect and their matched controls. *Pediatr Surg Int* 24: 803-808.

Razdan P, Patthi B, Kumar JK, Agnihotri N, Chaudhari P, Prasad M. 2017. Effect of fluoride concentration in drinking water on intelligence quotient of 12-14-year-old children in Mathura District: A cross-sectional study. *J Int Soc Prev Community Dent* 7: 252-258.

Ren D, Li K, Liu D. 1989. [A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas]. *Chin J Control Endem Dis* 4(4): 251.

Ren D, Li K, Liu D. 2008. A study of the intellectual ability of 8-14 year-old children in high fluoride, low iodine areas. *Fluoride* 41: 319-320.

Riddell JK, Malin AJ, Flora D, McCague H, Till C. 2019. Association of water fluoride and urinary fluoride concentrations with attention deficit hyperactivity disorder in Canadian youth. *Environ Int* 133: 105190.

Rocha-Amador D, Navarro ME, Carrizales L, Morales R, Calderon J. 2007. Decreased intelligence in children and exposure to fluoride and arsenic in drinking water. *Cad Saude Publica* 23 Suppl 4: S579-587.

Rocha-Amador D, Navarro M, Trejo-Acevedo A, Carrizales L, Perez-Maldonado I, Diaz-Barriga F, Calderon J. 2009. Use of the Rey-Osterrieth Complex Figure Test for neurotoxicity evaluation of mixtures in children. *Neurotox* 30: 1149-1154.

Rotton J, Tikofsky RS, Feldman HT. 1982. Behavioral effects of chemicals in drinking water. *J Appl Psychol* 67: 230-238.

Russ TC, Killin LOJ, Hannah J, Batty GD, Deary IJ, Starr JM. 2019. Aluminium and fluoride in drinking water in relation to later dementia risk. *Brit J Psychol* 14: 1-6.

Saxena S, Sahay A, Goel P. 2012. Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India. *J Neurosci Rural Pract* 3: 144-149.

Sebastian ST, Sunitha S. 2015. A cross-sectional study to assess the intelligence quotient (IQ) of school going children aged 10-12 years in villages of Mysore District, India with different fluoride levels. *J Indian Soc Pedod Prev Dent* 33: 307-311.

Seraj B, Shahrabi M, Falahzade M, Falahzade F, Akhondi N. 2006. Effect of high fluoride concentration in drinking water on children's intelligence. *J Dent Med* 19(2): 80-86.

Seraj B, Shahrabi M, Shadfar M, Ahmadi R, Fallahzadeh M, Eslamlu HF, Kharazifard MJ. 2012. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. *J Dent* 9: 221-229.

Shannon FT, Fergusson DM, Horwood LJ. 1986. Exposure to fluoridated public water supplies and child health and behaviour. *New Zeal Med J* 99: 416-418.

Shao Q. 2003. Study of cognitive function impairment caused by chronic fluorosis. *Chin J Endemiol* 22(4): 336-338.

Sharma JD, Sohu D, Jain P. 2009. Prevalence of neurological manifestations in a human population exposed to fluoride in drinking water. *Fluoride* 42: 127-132.

Shivaprakash PK, Ohri K, Noorani H. 2011. Relation between dental fluorosis and intelligence quotient in school children of Bagalkot District. *J Indian Soc Pedod Prev Dent* 29: 117-120.

Singh A, Jolly SS, Devi P, Bansal BC, Singh SS. 1962. Endemic fluorosis: An epidemiological, biochemical and clinical study in the Bhatinda District of Panjab. *Indian J Med Res* 50: 387-398.

Singh N, Verma KG, Verma P, Sidhu GK, Sachdeva S. 2014. A comparative study of fluoride ingestion levels, serum thyroid hormone & TSH level derangements, dental fluorosis status among school children from endemic and non-endemic fluorosis areas. *Springerplus* 3: 7.

Singh V, Singh C, Sandeep T, Sandeep K, Vikas G, Mukesh T, Anurag T. 2013. A correlation between serum vitamin, acetylcholinesterase activity and IQ in children with excessive endemic fluoride exposure in Rajasthan, India. *Int Res J Medical Sci* 1(3): 12-16.

Soto-Barreras U, Escalante-Villalobos KY, Holguin-Loya B, Perez-Aguirre B, Nevarez-Rascon A, Martinez-Martinez RE, Loyola-Rodriguez JP. 2019. Effect of fluoride in drinking water on dental caries and IQ in children. *Fluoride* 52: 474-482.

Still CN, Kelley P. 1980. The incidence of primary degenerative dementia vs. water fluoride content in South Carolina, USA. *Neurotox* 1: 125-132.

Sudhir KM, Chandu GN, Prashant GM, Subba Reddy VV. 2009. Effect of fluoride exposure on intelligence quotient (IQ) among 13-15 year old school children of known endemic area of fluorosis, Nalgonda District, Andhra Pradesh. *J Indian Assoc Public Health Dent* 2009(13): 88-94.

Sun M, Li S, Wang Y, Li F. 1991. Measurement of intelligence by drawing test among the children in the endemic area of Al-F combined toxicosis. *J Guiyang Med Coll* 16(3): 204-206.

Susheela AK, Bhatnagar M, Vig K, Mondal NK. 2005. Excess fluoride ingestion and thyroid hormone derangements in children living in Delhi, India. *Fluoride* 38: 98-108.

Tamboli BL, Mathur GM, Mathur AP, Lalla SK, Goyal OP. 1980. Prevalence of fluorosis in Pratabpura and Surajpura villages, District Ajmer (Rajasthan). *Indian J Med Res* 71: 57-67.

Till C, Green R, Flora D, Hornung R, Martinez-Mier EA, Blazer M, Farmus L, Ayotte P, Muckle G, Lanphear B. 2020. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. *Environ Int* 134: 105315.

Tripathi P, Sultana N. 2007. Fluoride content of groundwater and prevalence of dental, skeletal and neurological stage of fluorosis in Tehsil Purwa of Unnao. *Indian J Environ Prot* 27: 737-739.

Trivedi MH, Verma RJ, Chinoy NJ, Patel RS, Sathawara NG. 2007. Effect of high fluoride water on intelligence of school children in India. *Fluoride* 40: 178-183.

Trivedi M, Sangai N, Patel R, Payak M, Vyas S. 2012. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six villages of the Mundra Region, Kachchh, Gujurat, India. *Fluoride* 45(4): 377-383.

Valdez Jimenez L, Lopez Guzman OD, Cervantes Flores M, Costilla-Salazar R, Calderon Hernandez J, Alcaraz Contreras Y, Rocha-Amador DO. 2017. In utero exposure to fluoride and cognitive development delay in infants. *Neurotox* 59: 65-70.

Wang A, Duan L, Huang H, Ma J, Zhang Y, Ma Q, Guo Y, Li Z, Cheng X, Zhu J, Zhou G, Ba Y. 2020. Association between fluoride exposure and behavioural outcomes of school-age children: A pilot study in China. *Int J Environ Health Res*: 1-10.

Wang G, Yang D, Jia F, Wang H. 1996. [A study of the IQ levels of four- to seven-year-old children in high fluoride areas]. *Endem Dis Bull* 11(1): 60-62.

Wang G, Yang D, Jia F, Wang H. 2008. A study of the IQ levels of four- to seven-year-old children in high fluoride areas. *Fluoride* 41: 340-343.

Wang G, Gao M, Zhang M, Yang M, Xiang Q. 2012. Correlation between total fluoride intake and children's IQ. *J Southeast Univ Med Ed*: 743-746.

Wang M, Liu L, Li H, Li Y, Liu H, Hou C, Zeng Q, Li P, Zhao Q, Dong L, Zhou G, Yu X, Liu L, Guan Q, Zhang S, Wang A. 2020. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. *Environ Int* 134: 105229.

Wang S, Wang L, Hu P, Guo S, Law S. 2001. [Effects of high iodine and high fluorine on children's intelligence and thyroid function]. *Chin J Endemiol* 20(4): 288-290.

Wang SX, Wang ZH, Cheng XT, Li J, Sang ZP, Zhang XD, Han LL, Qiao XY, Wu ZM, Wang ZQ. 2005. [Investigation and evaluation on intelligence and growth of children in endemic fluorosis and arsenism areas]. *Chin J Endemiol* 24: 179-182.

Wang SX, Wang ZH, Cheng XT, Li J, Sang ZP, Zhang XD, Han LL, Qiao XY, Wu ZM, Wang ZQ. 2007. Arsenic and fluoride exposure in drinking water: Children's IQ and growth in Shanyin County, Shanxi Province, China. *Environ Health Perspect* 115: 643-647.

Wang SY, Zhang HX, Fan W, Fang SJ, Kang PP, Chen XG, Yu MJ. 2005. [The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children]. *J Appl Clin Ped* 20(9): 897-899.

Wang SY, Zhang HX, Fan W, Fang SJ, Kang PP, Chen XG, Yu MJ. 2008. The effects of endemic fluoride poisoning caused by coal burning on the physical development and intelligence of children. *Fluoride* 41: 344-348.

Wang Z, Wang S, Zhang X, Li J, Zheng X, Hu C. 2006. Investigation of children's growth and development under long-term fluoride exposure. *Chin J Control Endem Dis* 21(4): 239-241.

Wei N, Li Y, Deng J, Xu S, Guan Z. 2014. [The effects of comprehensive control measures on intelligence of school-age children in coal-burning-borne endemic fluorosis areas]. *Chin J Endemiol* 33(2): 320-322.

Xiang Q, Liang Y, Chen L, Wang C, Chen B, Chen X, Zhou M. 2003. Effect of fluoride in drinking water on children's intelligence. *Fluoride* 36: 84-94.

Xiang Q, Liang Y, Chen B, Chen L. 2011. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. *Fluoride* 44: 191-194.

Xu Y, Lu C, Zhang X. 1994. Effect of fluoride on children's intelligence. *Endem Dis Bull* 2: 83-84.

Yang Y, Wang X, Guo X, Hu P. 1994. [The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine]. *Chin J Epidemiol* 15(4): 296-298.

Yang Y, Wang X, Guo X, Hu P. 2008. The effects of high levels of fluoride and iodine on child intellectual ability and the metabolism of fluoride and iodine. *Fluoride* 41: 336-339.

Yao L, Zhou J, Wang S, Cui K, Lin F. 1996. Analysis of TSH levels and intelligence of children residing in high fluorosis areas. *Lit Inf Prev Med* 2(1): 26-27.

Yao Y. 1997. Comparative assessment of the physical and mental development of children in endemic fluorosis area with water improvement and without water improvement. *Lit Inf Prev Med* 3(1): 42-43.

Yasmin S, Ranjan S, D'Souza D. 2013. Effect of excess fluoride ingestion on human thyroid function in Gaya Region, Bihar, India. *Toxicol Environ Chem* 95: 1235-1243.

Yazdi SM, Sharifian A, Dehghani-Beshne M, Momeni VR, Aminian O. 2011. Effects of fluoride on psychomotor performance and memory of aluminum potroom workers. *Fluoride* 44: 158-162.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 1996. [Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis]. *Chin J Endemiol* 15(5): 257-259.

Yu YN, Yang WX, Dong Z, Wan CW, Zhang JT, Liu JL, Xiao KQ, Huang YS, Lu BF. 2008. Neurotransmitter and receptor changes in the brains of fetuses from areas of endemic fluorosis. *Fluoride* 41: 134-138.

Yu X, Chen J, Li Y, Liu H, Hou C, Zeng Q, Cui Y, Zhao L, Li P, Zhou Z, Pang S, Tang S, Tian K, Zhao Q, Dong L, Xu C, Zhang X, Zhang S, Liu L, Wang A. 2018. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. *Environ Int* 118: 116-124.

Zhang J, Yao H, Chen Y. 1998. [Effect of high level of fluoride and arsenic on children's intelligence]. *Chin J Public Health* 17(2): 57.

Zhang S, Zhang X, Liu H, Qu W, Guan Z, Zeng Q, Jiang C, Gao H, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Cui Y, Yu L, Wang A. 2015. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. *Toxicol Sci* 144: 238-245.

Zhao LB, Liang GH, Zhang DN, Wu XR. 1996. Effect of a high fluoride water supply on children's intelligence. *Fluoride* 29: 190-192.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao Q, Tian Z, Zhou G, Niu Q, Chen J, Li P, Dong L, Xia T, Zhang S, Wang A. 2020. SIRT1-dependent mitochondrial biogenesis supports therapeutic effects of resveratrol against neurodevelopment damage by fluoride. *Theranostics* 10: 4822-4838.

Zhou G, Tang S, Yang L, Niu Q, Chen J, Xia T, Wang S, Wang M, Zhao Q, Liu L, Li P, Dong L, Yang K, Zhang S, Wang A. 2019. Effects of long-term fluoride exposure on cognitive ability and the underlying mechanisms: Role of autophagy and its association with apoptosis. *Toxicol Appl Pharmacol 378*: 114608.

Studies Not Available in HAWC

Bachinskii PP, Gutsalenko OA, Naryzhniuk ND, Sidora VD, Shliakhta AI. 1985. [Action of the body fluorine of healthy persons and thyroidopathy patients on the function of hypophyseal-thyroid the system]. *Probl Endokrinol* 31: 25-29.

Balabolkin MI, Mikhailets ND, Lobovskaia RN, Chernousova NV. 1995. [The interrelationship of the thyroid and immune statuses of workers with long-term fluorine exposure]. *Ter Arkh* 67: 41-42.

Baum K, Boerner W, Reiners C, Moll E. 1981. [Bone density and thyroid function in adolescents in relation to fluoride content of drinking water]. *Fortschr Med* 99: 1470-1472.

Berry WTC, Whittles JH. 1963. Absence of effect of fluoride upon the incidence of thyroid enlargements in Wiltshire schoolgirls. *Mon Bull Minist Health Public Health Lab Serv* 22: 50-52.

Cherkinskii SN, Zaslavskaia RM. 1956. [Significance of fluorides in potable water in the development of endemic goiter]. *Probl Endokrinol Gormonoter* 2: 70-75.

Choubisa SL. 2001. Endemic fluorosis in southern Rajasthan, India. *Fluoride* 34: 61-70.

Chuka A, Zhukovskil V, Mirku I, Postel'Niku D. 1964. Prezhdevremennoe starenie naseleniya v zone rasprostraneniya endemicheskogo zoba. *Vestnik Akad Med Nauk Sssr* 19: 23-27.

Dai HX, Zeng P, Wang KY, Zhang XG, Ma ZJ, Zhou YG, Fan ZX, Guo SH. 2013. [Analysis of a survey results of patients with suspected high iodine goiter in Liuji Town Fuping County of Shaanxi Province]. *Chin J Endemiol* 32: 408-411.

Day T, Powell-Jackson P. 1972. Fluoride, water hardness, and endemic goitre. *Lancet* 299(7761): 1135-1138.

Desai VK, Solanki DM, Bansal RK. 1993. Epidemiological study of goitre in endemic fluorosis district of Gujarat. *Fluoride* 26: 187-190.

Díaz-Cadórniga FJ, Delgado E, Tartón T, Valdés MM, Méndez A, Fernández MT, Rojo C. 2003. Endemic goiter associated with high iodine intake in primary school children in the Saharawi Arab Democratic Republic. *Endocrinol Nutr* 50: 357-362.

Eichner R, Borner W, Henschler D, Kohler W, Moll E. 1981. [Osteoporosis therapy and thyroid function. Influence of 6 months of sodium fluoride treatment on thyroid function and bone density]. *Fortschr Med* 99: 342-348.

Fiorentini S, Galeazzi M, Visintin B. 1947. II fluoro in natura come agente morbigeno II. La fluorosi die Campagnano di Roma. III. Un focolaio di fluorosi umana a Campagnano di Roma. IV. Osservazioni radiologiche sui processi alveolari, sulle ossa mascellari, e sul paradenzio degli abitanti die Campagnano. V. Zona fluorotica intorno a Campagnano di Roma. VI. Frequenza e caratteri clinici della carie dentale in soggetti fluorotici. *Rend Ist Superiore Sanita* 10: 721-804.

Gabovich RD, Verzhikovskaia NV. 1958. [Effect of fluoride compounds on absorption of radioactive iodine by the thyroid gland in humans and in experimental conditions]. *Probl Endokrinol Gormonoter* 4: 49-54.

Galletti P, Held HR, Korrodi H, Wegmann T. 1956. [Investigations on the possible damaging effect of fluorine on the thyroid gland]. *Helv Med Acta* 23: 601-605.

Galletti PM, Joyet G, Jallut O. 1957. [Effect of sodium fluoride on thyroid function in Basedow's Disease]. *Helv Med Acta* 24: 209-215.

Galletti PM, Joyet G. 1958. Effect of fluorine on thyroidal iodine metabolism in hyperthyroidism. *J Clin Endocrinol Metab* 18: 1102-1110.

Gas'kov AI, Savchenkov MF, Iushkov NN. 2005. [The specific features of the development of iodine deficiencies in children living under environmental pollution with fluorine compounds]. *Gig Sanit*: 53-55.

Gedalia I, Brand N. 1963. The relationship of fluoride and iodine in drinking water in the occurrence of goiter. *Arch Int Pharmacodyn Ther* 142: 312-315.

Grimm H. 1973. [The physical development of schoolchildren under the influence of drinking water fluoridation in Karl Marx Stadt]. *Dtsch Gesundheitsw* 28: 2363-2369.

Hasling C, Nielsen HE, Melsen F, Mosekilde L. 1987. Safety of osteoporosis treatment with sodium fluoride, calcium phosphate and vitamin D. *Miner Electrolyte Metab* 13: 96-103.

Hidehiko T. 1958. On the relation between the distribution of endemic goiter and the fluorine content of natural water in Hidaka Province, Hokkaido. *Folia Pharmacol Jpn* 54: 225-229.

Hoffmann-Axthelm W. 1953. [Observations on the influence of fluorine on dental enamel and thyroid gland]. *Dtsch Zahnarztl Z* 8: 757-765.

Jentzer A. 1956. [Effect of fluorine on the iodine content of the human thyroid gland]. *Bull Schweiz Akad Med Wiss* 12: 539-543.

Jooste PL, Weight MJ, Kriek JA, Louw AJ. 1999. Endemic goitre in the absence of iodine deficiency in schoolchildren of the Northern Cape Province of South Africa. *Eur J Clin Nutr* 53: 8-12.

Kolomiitseva MG. 1961. [The content of fluorine in the external environment of the Upper Altai autonomous region and its role in the etiology of endemic goiter]. *Gig Sanit* 26: 101-103.

Korrodi H, Wegmann T, Galleti P, Held HR. 1955. [Caries prophylaxis and the untoward effects of fluor on the thyroid gland]. *Schweiz Med Wochenschr* 85: 1016-1019.

Kutlucan A, Kale Koroglu B, Numan Tamer M, Aydin Y, Baltaci D, Akdogan M, Ozturk M, Vural H, Ermis F. 2013. The investigation of effects of fluorosis on thyroid volume in school-age children. *Med Glas* 10: 93-98.

Latham MC, Grech P. 1967. The effects of excessive fluoride intake. *Am J Public Health* 57: 651-660.

Leone NC, Leatherwood EC, Petrie IM, Lieberman L. 1964. Effect of fluoride on thyroid gland: Clinical study. *J Am Dent Assoc* 69: 179-180.

Levi JE, Silberstein HE. 1955. Lack of effect of fluorine ingestion on uptake of iodine 131 by the thyroid gland. *J Lab Clin Med* 45: 348-351.

McGlashan N, Chelkowska E, Sasananan S. 2010. A survey of goiter morbidity in Ban Mae Toen, northwest Thailand. *Southeast Asian J Trop Med Public Health* 41: 1200-1208.

Rathore S, Meena C, Gonmei Z, Dwivedi S, Toteja GS, Bala K. 2018. Study of excess fluoride ingestion and thyroid hormone derangement in relation with different fluoride levels in drinking water among children of Jodhpur District, Rajasthan, India. *Asian J Microbiol Biotechnol Environ Sci* 20: 327-331.

Reisenauer R, Rezler D, Křemenová J, Preininger Q. 1961. [Fluorization of the waters in Czechoslovakia. IV. Endocrinological control of results of two years' fluorization of drinking-water in school children]. *Cesk Stomatol* 61: 91-97.

Romer TEZ, Kowalczyk B, Kacprzak M, Wiktorowski M. 1976. [Incidence of goiter in pubertal girls of the Piotrkow Region and iodide content in drinking water]. *Endokrynol Pol* 27: 373-380.

Savchenkov MF, Efimova NV, Manueva RS, Nikolaeva LA, Shin NS. 2016. [Thyroid gland pathology in children population exposed to the combination of iodine deficiency and fluoride pollution of environment]. *Gig Sanit* 95: 1201-1205.

Shtifanova AK. 1962. [The fluorine content in water, soil and vegetal products of the Alma-Atinsk District areas and its role in the etiology of dental caries and endemic goiter]. *Zdravookhranenie Kazakhstana*: 60-63.

Siddiqui AH. 1969. Incidence of simple goiter in areas of endemic fluorosis in Nalgonda District, Andhra Pradesh, India. *Fluoride* 2: 200-205.

Sidora VD, Shliakhta AI, Iugov VK, Kas'ianenko AS, Piatenko VG. 1983. [Indices of the pituitary-thyroid system in residents of cities with various fluorine concentrations in drinking water]. *Probl Endokrinol* 29: 32-35.

Sung FC, Chen KP, Chen CY, Tai PW, Yang CF. 1973. Studies of the effect of salt iodization on endemic goiter in Taiwan. IV. A survey of drinking water in relation to endemic goiter. *J Fomosan Med Assoc* 72: 96-103.

Tokar VI, Voroshnin VV, Sherbakov SV. 1989. [Chronic effects of fluorides on the pituitary-thyroid system in industrial workers]. *Gig Tr Prof Zabol*: 19-22.

Wespi HJ. 1954. [Iodized-fluoridized salt for the prevention of goiter and caries]. *Schweiz Med Wochenschr* 84: 885-890.

Yu YN. 1985. [Effects of chronic fluorosis on the thyroid gland]. *Chin Med J* 65: 747-7479.

Studies in Non-human Animals

As described in **Figure 4**, 339 non-human mammal studies were included; however, full data extraction was only conducted on studies with primary neurological outcomes and/or secondary functional neurological outcomes (e.g., motor activity). Data extraction was completed using HAWC. Data were extracted from a subset of included studies in animals (n = 123) and are available in HAWC based on outcome. The following lists of references are organized as studies that are available in HAWC followed by studies that are not available in HAWC. Specifically, all primary outcomes and functional neurological secondary outcomes (e.g., motor activity) were extracted from animal studies and are available in HAWC, including studies from the NTP (2016) assessment. Studies are also available in HAWC that evaluated mechanistic effects related to oral fluoride exposure at or below 20 ppm fluoride drinking water equivalents for categories of mechanistic endpoints with the largest amount of available data (i.e., biochemistry of the brain or neurons, neurotransmission, oxidative stress, and histopathology [n = 70]); however, these mechanistic data were generally not extracted. Several animal studies assessed primary neurological outcomes and/or functional neurological secondary outcomes and mechanistic effects in the four mechanistic categories listed above (n = 56). In total, 140 animal studies are available in HAWC (70 with primary neurological outcomes and/or secondary functional neurological outcomes without relevant mechanistic data; 15 with relevant mechanistic data only; and 55 with primary/or secondary functional neurological outcomes with relevant mechanistic data). Studies that evaluated other mechanistic endpoints, as well as studies that only assessed mechanistic effects at fluoride levels above 20 ppm fluoride drinking water equivalents, are not available in HAWC (n = 199). The list below presents the 339 non-human animal studies that were included in the review. An overview of the screening results is outlined in the study selection diagram (**Figure 4**) that reports numbers of included studies as well as numbers of studies excluded for each reason at the full text review stage.

Studies Available in HAWC

Adedara IA, Abolaji AO, Idris UF, Olabiyi BF, Onibiyo EM, Ojuade TD, Farombi EO. 2017. Neuroprotective influence of taurine on fluoride-induced biochemical and behavioral deficits in rats. *Chem Biol Interact* 261: 1-10.

Adedara IA, Olabiyi BF, Ojuade TD, Idris UF, Onibiyo EM, Farombi EO. 2017. Taurine reverses sodium fluoride-mediated increase in inflammation, caspase-3 activity, and oxidative damage along the brain-pituitary-gonadal axis in male rats. *Can J Phys Pharmacol* 95: 1019-1029.

Agustina F, Sofro ZM, Partadiredja G. 2018. Subchronic administration of high-dose sodium fluoride causes deficits in cerebellar purkinje cells but not motor coordination of rats. *Biol Trace Elem Res* 188(2): 424-433.

Ahmad KR, Noor S, Jabeen S, Nauroze T, Kanwal MA, Raees K, Abbas T. 2017. Amelioration by jambul fruit extract of fluoride-induced hepato-nephronal histopathologies and impaired neuromotor capacity in mice. *Fluoride* 50: 2-14.

Akinrinade ID, Memudu AE, Ogundele OM. 2015. Fluoride and aluminium disturb neuronal morphology, transport functions, cholinesterase, lysosomal and cell cycle activities. *Pathophysiology* 22: 105-115.

Akinrinade ID, Memudu AE, Ogundele OM, Ajetunmobi OI. 2015. Interplay of glia activation and oxidative stress formation in fluoride and aluminium exposure. *Pathophysiology* 22: 39-48.

Baba NA, Raina R, Verma PK, Sultana M. 2014. Alterations in plasma and tissue acetylcholinesterase activity following repeated oral exposure of chlorpyrifos alone and in conjunction with fluoride in Wistar rats. *Proc Indian Natl Sci Acad B Biol Sci* 84: 969-972.

Bagmut I, Kolisnyk I, Titkova A, Babiy L, Filipchenko S. 2018. The antioxidant system enzymes' activity in rats' brain, intoxicated with sodium fluoride in subtoxic doses. *Arch Balkan Med Union* 53(4): 506-511.

Balaji B, Kumar EP, Kumar A. 2015. Evaluation of standardized bacopa monniera extract in sodium fluoride-induced behavioural, biochemical, and histopathological alterations in mice. *Toxicol Ind Health* 31: 18-30.

Balayssac D, Richard D, Authier N, Nicolay A, Jourdan D, Eschalier A, Coudore F. 2002. Absence of painful neuropathy after chronic oral fluoride intake in Sprague-Dawley and Lou/C rats. *Neurosci Lett* 327: 169-172.

Banala RR, Karnati PR. 2015. Vitamin A deficiency: An oxidative stress marker in sodium fluoride (NaF) induced oxidative damage in developing rat brain. *Int J Dev Neurosci* 47: 298-303.

Banala R, Nagapuri K, Mohd K, Reddy M, Karnati P. 2018. Carica Papaya leaf extract as a neuroprotective agent against behavioral and neurotransmitter changes in brain of the rat treated with sodium fluoride in pre- and post-natal periods. *Pharmacogn Mag* 14(55): 123-131.

Banji D, Banji OJ, Pratusha NG, Annamalai AR. 2013. Investigation on the role of spirulina platensis in ameliorating behavioural changes, thyroid dysfunction and oxidative stress in offspring of pregnant rats exposed to fluoride. *Food Chem* 140: 321-331.

Baran-Poesina V, Negres S, Dobrescu D, Dimcevici-Poesina N, Dimcevici-Poesina A, Feghiu A, Soare T, Militaru M. 2013. Experimental pharmacological researches regarding the influence of sodium fluoride in allopathic and homeopathic doses on central nervous system's performances: A correlation between behavioral response in classic maze test and morphological aspects of cerebral cortex. *Farmacia* 61: 781-799.

Bartos M, Gumilar F, Bras C, Gallegos CE, Giannuzzi L, Cancela LM, Minetti A. 2015. Neurobehavioural effects of exposure to fluoride in the earliest stages of rat development. *Physiol Behav* 147: 205-212.

Bartos M, Gumilar F, Gallegos CE, Bras C, Dominguez S, Monaco N, Esandi MDC, Bouzat C, Cancela LM, Minetti A. 2018. Alterations in the memory of rat offspring exposed to low levels of fluoride during gestation and lactation: Involvement of the alpha7 nicotinic receptor and oxidative stress. *Reprod Toxicol* 81: 108-114.

Basha PM, Rai P, Begum S. 2011. Fluoride toxicity and status of serum thyroid hormones, brain histopathology, and learning memory in rats: A multigenerational assessment. *Biol Trace Elem Res* 144: 1083-1094.

Basha PM, Sujitha NS. 2012. Combined impact of exercise and temperature in learning and memory performance of fluoride toxicated rats. *Biol Trace Elem Res* 150: 306-313.

Bataineh HN, Nusier MK. 2006. Impact of 12-week ingestion of sodium fluoride on aggression, sexual behavior, and fertility in adult male rats. *Fluoride* 39: 293-301.

Bera I, Sabatini R, Auteri P, Flace P, Sisto G, Montagnani M, Potenza MA, Marasciulo FL, Carratu MR, Coluccia A, Borracci P, Tarullo A, Cagiano R. 2007. Neurofunctional effects of developmental sodium fluoride exposure in rats. *Eur Rev Med Pharmacol Sci* 11: 211-224.

Bhatnagar M, Rao P, Sushma J, Bhatnagar R. 2002. Neurotoxicity of fluoride: Neurodegeneration in hippocampus of female mice. *Indian J Exp Biol* 40: 546-554.

Bhatnagar M, Sukhwal P, Suhalka P, Jain A, Joshi C, Sharma D. 2011. Effects of fluoride in drinking water on NADPH-diaphorase neurons in the forebrain of mice: A possible mechanism of fluoride neurotoxicity. *Fluoride* 44: 195-209.

Chen H, Geng D. 2011. [The change of cognition induced by chronic fluoride in rats]. *Acta Academiae Medicinae Xuzhou* 31(5): 319-322.

Chen J, Niu Q, Xia T, Zhou G, Li P, Zhao Q, Xu C, Dong L, Zhang S, Wang A. 2018. ERK1/2-mediated disruption of BDNF-TrkB signaling causes synaptic impairment contributing to fluoride-induced developmental neurotoxicity. *Toxicology* 410: 222-230.

Chinoy NJ, Shah SD. 2004. Biochemical effects of sodium fluoride and arsenic trioxide toxicity and their reversal in the brain of mice. *Fluoride* 37: 80-87.

Chioca LR, Raupp IM, Da Cunha C, Losso EM, Andreatini R. 2008. Subchronic fluoride intake induces impairment in habituation and active avoidance tasks in rats. *Eur J Pharmacol* 579: 196-201.

Chouhan S, Flora SJS. 2008. Effects of fluoride on the tissue oxidative stress and apoptosis in rats: Biochemical assays supported by IR spectroscopy data. *Toxicology* 254: 61-67.

Chouhan S, Lomash V, Flora SJ. 2010. Fluoride-induced changes in haem biosynthesis pathway, neurological variables and tissue histopathology of rats. *J Appl Toxicol* 30: 63-73.

Cui YS, Zhong Q, Li WF, Liu ZH, Wang Y, Hou CC. 2017. [Effects of fluoride exposure on thyroid hormone level and intelligence in rats]. *Chin J Ind Hyg Occup Dis* 35: 888-892.

Dabrowska E. 1997. Effect of different fluorine doses on the supraoptic nucleus of the rat. *Folia Histochem Cytobiol* 35: 115-116.

Dong Y, Wang Y, Wei N, Guan Z. 2015. [Expression levels of brain muscarinic acetylcholine receptor in offspring rats of drinking-water borne fluorosis]. *Chin J Endemiol* 34: 326-330.

Dong YT, Wang Y, Wei N, Zhang QF, Guan ZZ. 2015. Deficit in learning and memory of rats with chronic fluorosis correlates with the decreased expressions of M1 and M3 muscarinic acetylcholine receptors. *Arch Toxicol* 89: 1981-1991.

Dong YT, Wei N, Qi XL, Liu XH, Chen D, Zeng XX, Guan ZZ. 2017. Attenuating effect of vitamin E on the deficit of learning and memory of rats with chronic fluorosis: The mechanism may involve muscarinic acetylcholine receptors. *Fluoride* 50: 354-364.

Dong YW, Y. Wei, N. Guan, Z. 2015. [Expression of muscarinic acetylcholine receptors in the brain of rats with chronic fluorosis]. *Chin J Endemiol* 34(2): 84-88.

Ekambaram P, Paul V. 2001. Calcium preventing locomotor behavioral and dental toxicities of fluoride by decreasing serum fluoride level in rats. *Environ Toxicol Pharmacol* 9: 141-146.

Ekambaram P, Paul V. 2002. Modulation of fluoride toxicity in rats by calcium carbonate and by withdrawal of fluoride exposure. *Pharmacol Toxicol* 90: 53-58.

Ekambaram P, Paul V. 2003. Effect of vitamin D on chronic behavioral and dental toxicities of sodium fluoride in rats. *Fluoride* 36: 189-197.

El-lethey HS, Kamel MM, Shaheed IB. 2010. Neurobehavioral toxicity produced by sodium fluoride in drinking water of laboratory rats. *J Am Sci* 6(5): 54-63.

El-lethey HS, Kamel MM. 2011. Effects of black tea in mitigation of sodium fluoride potency to suppress motor activity and coordination in laboratory rats. *J Am Sci* 7(4): 243-254.

El-lethey HS, Shaheed IB. 2011. Potential health impact of black tea against Na-F-induced alterations in territorial aggression, sexual behaviour and fertility of male rats. *Life Sci J* 8: 828-839.

Elliott L. 1967. Lack of effect of administration of fluoride on the central nervous system of rats. *Acta Pharmacol Toxicol (Copenh)* 25: 323-328.

Flace P, Benagiano V, Vermesan D, Sabatini R, Inchingolo AM, Auteri P, Ambrosi G, Tarullo A, Cagiano R. 2010. Effects of developmental fluoride exposure on rat ultrasonic vocalization, acoustic startle reflex and pre-pulse inhibition. *Eur Rev Med Pharmacol Sci* 14: 507-512.

Gabovich RD. 1962. [On the problem of the effect of fluorine in drinking water on the functional state of the central nervous system]. *Gig Sanit* 27: 10-12.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008. [Level of oxidative stress in rat brains and learning and memory function of rats with chronic fluorosis]. *Chin J Endemiol* 27: 371-373.

Gao Q, Liu YJ, Wu CX, Long YG, Guan ZZ. 2008. [Effects of fluoride on learning and memory and cholinesterase activity in rat brains]. *Chin J Endemiol* 27: 128-130.

Gao Q, Liu YJ, Guan ZZ. 2009. Decreased learning and memory ability in rats with fluorosis: Increased oxidative stress and reduced cholinesterase activity in the brain. *Fluoride* 42: 277-285.

Gao Y, Liu L, Young L, Huan L, Jin H. 2009. Effects of learning and memory of fluoride and the antagonism of selenium in rats. *Studies of Trace Elements and Health* 26(2): 1-3.

Ge QD, Tan Y, Luo Y, Wang WJ, Zhang H, Xie C. 2018. MiR-132, miR-204 and BDNF-TrkB signaling pathway may be involved in spatial learning and memory impairment of the offspring rats caused by fluorine and aluminum exposure during the embryonic stage and into adulthood. *Environ Toxicol Pharmacol* 63: 60-68.

Ge Y, Chen L, Yin Z, Song X, Ruan T, Hua L, Liu J, Wang J, Ning H. 2018. Fluoride-induced alterations of synapse-related proteins in the cerebral cortex of ICR offspring mouse brain. *Chemosphere* 201: 874-883.

Gopal K, Saxena R, Gupta GSD, Rana MD, Agrawal D. 2006. Fluoride induced alterations in neurobehavioural and cardiovascular responses in rats. *J Adv Zool* 27: 1-7.

Gui CZ, Ran LY, Wu CX, Long YG, He J, Zhang H, Guan ZZ. 2009. [Changes in learning and memory ability and brain cholinesterase activity in the rats with coal burning fluorosis]. *Chin J Endemiol* 28: 497-500.

Gui CZ, Ran LY, Li JP, Guan ZZ. 2010. Changes of learning and memory ability and brain nicotinic receptors of rat offspring with coal burning fluorosis. *Neurotoxicol Teratol* 32: 536-541.

Gui CZ, Ran LY, Guan ZZ. 2011. [Expression levels of brain nicotinic acetylcholine receptor mRNA and protein in coal-burning type of fluorosis rats]. *Chin J Endemiol* 30: 239-242.

Guner S, Uyar-Bozkurt S, Haznedaroglu E, Mentes A. 2016. Dental fluorosis and catalase immunoreactivity of the brain tissues in rats exposed to high fluoride pre- and postnatally. *Biol Trace Elem Res* 174: 150-157.

Han H, Du W, Zhou B, Zhang W, Xu G, Niu R, Sun Z. 2014. Effects of chronic fluoride exposure on object recognition memory and mRNA expression of SNARE complex in hippocampus of male mice. *Biol Trace Elem Res* 158: 58-64.

Hong JH, Ge YM, Ning HM, Wang JD. 2005. [Effects of High Fluoride and Low Iodine on Learning-Memory and TchE of Brain in Offspring Rats]. *Chin Prev Med* 6: 489-491.

Inkielewicz I, Krechniak J. 2004. Fluoride effects on glutathione peroxidase and lipid peroxidation in rats. *Fluoride* 37: 7-12.

Jain A, Mehta VK, Chittora RA, Mahdi A, Bhatnagar M. 2015. Melatonin ameliorates fluoride induced neurotoxicity in young rats: An in vivo evidence. *Asian J Pharm Clin Res* 8: 164-167.

Jetti R, Raghuveer CV, Mallikarjuna RC. 2016. Protective effect of ascorbic acid and Ginkgo biloba against learning and memory deficits caused by fluoride. *Toxicol Ind Health* 32: 183-187.

Jia B, Zong L, Lee JY, Lei J, Zhu Y, Xie H, Clemens JL, Feller MC, Na Q, Dong J, McLane MW, Jones-Beatty K, Burd I. 2019. Maternal supplementation of low dose fluoride alleviates adverse perinatal outcomes following exposure to intrauterine inflammation. *Sci Rep* 9(1): 2575.

Jiang C, Zhang S, Liu H, Guan Z, Zeng Q, Zhang C, Lei R, Xia T, Wang Z, Yang L, Chen Y, Wu X, Zhang X, Cui Y, Yu L, Wang A. 2014. Low glucose utilization and neurodegenerative changes caused by sodium fluoride exposure in rat's developmental brain. *Neuromolecular Med* 16: 94-105.

Jiang S, Su J, Yao S, Zhang Y, Cao F, Wang F, Wang H, Li J, Xi S. 2014. Fluoride and arsenic exposure impairs learning and memory and decreases mGluR5 expression in the hippocampus and cortex in rats. *PLoS One* 9(4): e96041.

Khan AM, Raina R, Dubey N, Verma PK. 2017. Effect of deltamethrin and fluoride co-exposure on the brain antioxidant status and cholinesterase activity in Wistar rats. *Drug Chem Toxicol* 41(2): 1-5.

Kinawy AA, Al-Eidan AA. 2018. Impact of prenatal and postnatal treatment of sodium fluoride and aluminum chloride on some hormonal and sensorimotor aspects in rats. *Biol Trace Elem Res*: 1-8.

Kivrak Y. 2012. Effects of fluoride on anxiety and depression in mice. *Fluoride* 45: 302-306.

Li M, Cui J, Gao YH, Zhang W, Sun LY, Liu XN, Liu Y, Sun DJ. 2015. Pathological changes and effect on the learning and memory ability in rats exposed to fluoride and aluminum. *Toxicol Res* 4: 1366-1373.

Li X, Zhang J, Niu R, Manthari RK, Yang K, Wang J. 2019. Effect of fluoride exposure on anxiety- and depression-like behavior in mouse. *Chemosphere* 215: 454-460.

Liu F, Ma J, Zhang H, Liu P, Liu YP, Xing B, Dang YH. 2014. Fluoride exposure during development affects both cognition and emotion in mice. *Physiol Behav* 124: 1-7.

Liu WX. 1989. [Experimental study of behavior and cerebral morphology of rat pups generated by fluorotic female rat]. *Chin J Pathol* 18: 290-292.

Liu YJ, Gao Q, Wu CX, Long YG, Guan ZZ. 2009. [Modified expression of extracellular signal-regulated protein kinase signal transduction in rat brains and changed capacity of learning and memory of rats with chronic fluorosis]. *Chin J Endemiol* 28: 32-35.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. Alterations of nAChRs and ERK1/2 in the brains of rats with chronic fluorosis and their connections with the decreased capacity of learning and memory. *Toxicol Lett* 192: 324-329.

Liu YJ, Gao Q, Long YG, Yu YN, Guan ZZ. 2011. [Influence of chronic fluorosis on expression of phospho-Elk-1 in rat brains]. *Chin J Endemiol* 30: 251-255.

Lou DD, Guan ZZ, Liu YJ, Liu YF, Zhang KL, Pan JG, Pei JJ. 2013. The influence of chronic fluorosis on mitochondrial dynamics morphology and distribution in cortical neurons of the rat brain. *Arch Toxicol* 87: 449-457.

Lu F, Zhang Y, Trivedi A, Jiang X, Chandra D, Zheng J, Nakano Y, Uyghurturk DA, Jalai R, Guner S, Mentes A, DenBesten PK. 2019. Fluoride related changes in behavioral outcomes may relate to increased serotonin. *Physiol Behav* 206: 76-83.

Ma J, Liu F, Liu P, Dong YY, Chu Z, Hou TZ, Dang YH. 2015. Impact of early developmental fluoride exposure on the peripheral pain sensitivity in mice. *Int J Dev Neurosci* 47: 165-171.

Manusha S, Sudhakar K, Reddy KP. 2019. Protective effects of allium sativum extract against sodium fluoride induced neurotoxicity. *Int J Pharm Sci Res* 10(2): 625-633.

McPherson CA, Zhang G, Gilliam R, Brar SS, Wilson R, Brix A, Picut C, Harry GJ. 2018. An evaluation of neurotoxicity following fluoride exposure from gestational through adult ages in Long-Evans hooded rats. *Neurotoxicol Res*: 1-18.

Mesram N, Nagapuri K, Banala RR, Nalagoni CR, Karnati PR. 2016. Quercetin treatment against NaF induced oxidative stress related neuronal and learning changes in developing rats. *J King Saud Univ Sci* 29: 221-229.

Mullenix PJ, Denbesten PK, Schunior A, Kernan WJ. 1995. Neurotoxicity of sodium fluoride in rats. *Neurotoxicol Teratol* 17: 169-177.

Nageshwar M, Sudhakar K, Reddy NCC, Reddy KP. 2017. Neuroprotective effects of curcumin on sodium fluoride induced behavioural and enzymatic changes in brain and muscles of rat. *J Environ Biol* 38: 675-681.

Nageshwar M, Sudhakar K, Reddy KP. 2018. Quercetin ameliorates oxidative stress, neural damage of brain and behavioral impairment of rat with fluoride exposure. *Int J Pharm Sci Res* 9(8): 3247-3256.

Nian W, Wang X, Shao D, Yu Q, Ouyang W, Zhang Z, Ruan Q. 2018. Effects of subchronic exposure to fluorine on hippocampal injury in mice and its molecular mechanism. *Acta Sci Circumst* 38(11): 4512-4519.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. *Environ Pollut* 233: 889-899.

Niu R, Sun Z, Wang J, Cheng Z. 2008. Effects of fluoride and lead on locomotor behavior and expression of nissl body in brain of adult rats. *Fluoride* 41: 276-282.

Niu R, Sun Z, Cheng Z, Li Z, Wang J. 2009. Decreased learning ability and low hippocampus glutamate in offspring rats exposed to fluoride and lead. *Environ Toxicol Pharmacol* 28: 254-258.

Niu R, Liu S, Wang J, Zhang J, Sun Z. 2014. Proteomic analysis of hippocampus in offspring male mice exposed to fluoride and lead. *Biol Trace Elem Res* 162: 227-233.

Niu R, Xue X, Zhao Y, Sun Z, Yan X, Li X, Feng C, Wang J. 2015. Effects of fluoride on microtubule ultrastructure and expression of Tubalpha1a and Tubbeta2a in mouse hippocampus. *Chemosphere* 139: 422-427.

Niu R, Chen H, Manthari RK, Sun Z, Wang J, Zhang J, Wang J. 2018. Effects of fluoride on synapse morphology and myelin damage in mouse hippocampus. *Chemosphere* 194: 628-633.

Nkpaa KW, Onyeso GI. 2018. Rutin attenuates neurobehavioral deficits, oxidative stress, neuro-inflammation and apoptosis in fluoride treated rats. *Neurosci Lett* 682: 92-99.

Paul V, Ekambaram P, Jayakumar AR. 1998. Effects of sodium fluoride on locomotor behavior and a few biochemical parameters in rats. *Environ Toxicol Pharmacol* 6: 187-191.

Pereira M, Dombrowski PA, Losso EM, Chioca LR, Da Cunha C, Andreatini R. 2011. Memory impairment induced by sodium fluoride is associated with changes in brain monoamine levels. *Neurotoxicol Res* 19: 55-62.

Pulungan ZS, Sofro ZM, Partadiredja G. 2016. Sodium fluoride does not affect the working memory and number of pyramidal cells in rat medial prefrontal cortex. *Anat Sci Int* 93(1): 128-138.

Raghu J, Raghuveer VC, Rao MC, Somayaji NS, Babu PB. 2013. The ameliorative effect of ascorbic acid and Ginkgo biloba on learning and memory deficits associated with fluoride exposure. *Interdiscip Toxicol* 6: 217-221.

Raju S, Sivanesan SK, Gudemalla K. 2019. Cognitive enhancement effect of Ginkgo biloba extract on memory and learning impairments induced by fluoride neurotoxicity. *Int J Res Pharm Sci* 10(1): 129-134.

Reddy MM, Karnati PR. 2015. Protective effects of aqueous extract of fruit pulp of tamarindus indica on motor activity and metabolism of the gastrocnemius muscle of rats treated with fluoride. *Int J Toxicol Pharmacol Res* 7: 241-246.

Reddy YP, Tiwari SK, Shaik AP, Alsaeed A, Sultana A, Reddy PK. 2014. Effect of sodium fluoride on neuroimmunological parameters, oxidative stress and antioxidative defenses. *Toxicol Mech Methods* 24: 31-36.

Rumiantsev GI, Novikov SM, Mel'nikova NN, Levchenko NI, Kozeeva EE. 1988. [Experimental study of the biological effect of salts of hydrofluosilicic acid]. *Gig Sanit*: 80-82.

Sarkozi K, Horvath E, Vezer T, Papp A, Paulik E. 2015. Behavioral and general effects of subacute oral arsenic exposure in rats with and without fluoride. *Int J Environ Health Res* 25: 418-431.

Shah SD, Chinoy NJ. 2004. Adverse effects of fluoride and/or arsenic on the cerebral hemisphere of mice and recovery by some antidotes. *Fluoride* 37: 162-171.

Shalini B, Sharma JD. 2015. Beneficial effects of emblica officinalis on fluoride-induced toxicity on brain biochemical indexes and learning-memory in rats. *Toxicol Int* 22: 35-39.

Shan KR, Qi XL, Long YG, Nordberg A, Guan ZZ. 2004. Decreased nicotinic receptors in PC12 cells and rat brains influenced by fluoride toxicity: A mechanism relating to a damage at the level in post-transcription of the receptor genes. *Toxicology* 200: 169-177.

Sharma C, Suhalka P, Bhatnagar M. 2018. Curcumin and resveratrol rescue cortical-hippocampal system from chronic fluoride-induced neurodegeneration and enhance memory retrieval. *Int J Neurosci*: 1-15.

Shen X, Zhang Z, Xu X. 2004. [Effect of iodine and selenium on learning memory impairment induced by fluorosis and blood biochemical criterion of rats]. *Occupation and Health* 20(1): 6-8.

Sudhakar K, Nageshwar M, Pratap Reddy K. 2017. Seed extract of Abelmoschus moschatus medik reverses NAF-induced behavioral changes through neurodegeneration and oxidative stress in brain of rat. *Asian J Pharm Clin Res* 10: 165-171.

Sudhakar K, Nageshwar M, Reddy KP. 2018. Protective effect of okra, Abelmoschus moschatus seed extract on developing brain of rats during pre- and post-natal fluoride exposure. *Int J Pharm Sci Res* 9: 1519-1528.

Sudhakar K, Nageshwar M, Reddy KP. 2018. Abelmoschus moschatus extract reverses altered pain and neurohistology of a rat with developmental exposure of fluoride. *J Appl Pharm Sci* 8(6): 94-104.

Sudhakar K, Reddy KP. 2018. Protective effects of Abelmoschus moschatus seed extract on neurotransmitter system of developing brain of Wistar rats with gestational and post-natal exposure of sodium fluoride. *Int J Green Pharm* 12: S131-S139.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2000. [Effects of high fluoride drinking water on the cerebral function of mice]. *Chin J Epidemiol* 19: 262-263.

Sun ZR, Liu FZ, Wu LN, Lu Y, Yu DK. 2008. Effects of high fluoride drinking water on the cerebral function of mice. *Fluoride* 41: 148-151.

Sun Z, Zhang Y, Xue X, Niu R, Wang J. 2018. Maternal fluoride exposure during gestation and lactation decreased learning and memory ability, and glutamate receptor mRNA expressions of mouse pups. *Hum Exp Toxicol* 37: 87-93.

Trivedi MH, Verma RJ, Chinoy NJ. 2007. Amelioration by black tea of sodium fluoride-induced changes in protein content of cerebral hemisphere, cerebellum and medulla oblongata in brain region of mice. *Acta Poloniae Pharm* 64: 221-225.

Trivedi MH, Verma RJ, Chinoy NJ. 2009. Mitigation of sodium fluoride induced toxicity in mice brain by black tea infusion. *Fluoride* 42: 29-33.

Trivedi MH, Verma RJ, Sangai NP, Chinoy NJ. 2011. Black tea extract mitigation of NaF-induced lipid peroxidation in different regions of mice brains. *Fluoride* 44: 243-254.

Trivedi MH, Verma RJ, Sangai NP, Chinoy NJ. 2012. Mitigation by black tea extract of sodium fluoride induced histopathological changes in brain of mice. *Fluoride* 45: 13-26.

Tsunoda M, Aizawa Y, Nakano K, Liu Y, Horiuchi T, Itai K, Tsunoda H. 2005. Changes in fluoride levels in the liver, kidney, and brain and in neurotransmitters of mice after subacute administration of fluoride. *Fluoride* 38: 284-292.

Varner JA, Jensen KF, Horvath W, Isaacson RL. 1998. Chronic administration of aluminum-fluoride or sodium-fluoride to rats in drinking water: Alterations in neuronal and cerebrovascular integrity. *Brain Res* 784(1-2): 284-298.

Verma RJ, Trivedi MH, Chinoy NJ. 2007. Black tea amelioration of sodium fluoride-induced alterations of DNA, RNA, and protein contents in the cerebral hemisphere, cerebellum, and medulla oblongata regions of mouse brain. *Fluoride* 40: 7-12.

Wang G, Li J, Zhu H, Zhu J. 2006. Effect of different doses of chronic exposure of fluoride on rat learning and memory behavior. *Studies of Trace Elements and Health* 23(2): 1-2.

Wang JD, Ge YM, Ning HM, Wang SL. 2004. Effects of high fluoride and low iodine on biochemical indexes of the brain and learning-memory of offspring rats. *Fluoride* 37: 201-208.

Wang J, Zhang Y, Guo Z, Li R, Xue X, Sun Z, Niu R. 2018. Effects of perinatal fluoride exposure on the expressions of miR-124 and miR-132 in hippocampus of mouse pups. *Chemosphere* 197: 117-122.

Wei N, Dong Y, Wang Y, Guan Z. 2014. [Effects of chronic fluorosis on neurobehavioral development in offspring of rats and antagonistic effect of vitamin E]. *Chin J Endemiol* 33: 125-128.

Whitford GM, Whitford JL, Hobbs SH. 2009. Appetitive-based learning in rats: Lack of effect of chronic exposure to fluoride. *Neurotoxicol Teratol* 31: 210-215.

Wu CX, Gu XL, Ge YM, Zhang JH, Wang JD. 2006. Effects of high fluoride and arsenic on brain biochemical indexes and learning-memory in rats. *Fluoride* 39: 274-279.

Wu NP, Zhao ZL, Gao WH, Li XQ. 1995. [Behavioral teratology in rats exposed to fluoride.] *Chin J Endemiol* 12(5): 271-273.

Wu NP, Zhao ZL, Gao WH, Li XQ. 2008. Behavioral teratology in rats exposed to fluoride. *Fluoride* 41: 129-133.

Xu X, Shen X, Zhang Z. 2001. Effect of fluorosis on mice learning and memory behaviors and brain SOD activity and MDA content *China Public Health* 17(1): 8-10.

Yang L, Jin P, Wang X, Zhou Q, Lin X, Xi S. 2018. Fluoride activates microglia, secretes inflammatory factors and influences synaptic neuron plasticity in the hippocampus of rats. *Neurotox* 69: 108-120.

Yu Q, Shao D, Zhang R, Ouyang W, Zhang Z. 2019. Effects of drinking water fluorosis on L-type calcium channel of hippocampal neurons in mice. *Chemosphere* 220: 169-175.

Yuan J, Li Q, Niu R, Wang J. 2019. Fluoride exposure decreased learning ability and the expressions of the insulin receptor in male mouse hippocampus and olfactory bulb. *Chemosphere* 224: 71-76.

Zhang C, Ren C, Chen H, Geng R, Fan H, Zhao H, Guo K, Geng D. 2013. The analog of Ginkgo biloba extract 761 is a protective factor of cognitive impairment induced by chronic fluorosis. *Biol Trace Elem Res* 153: 229-236.

Zhang KL, Lou DD, Guan ZZ. 2015. Activation of the AGE/RAGE system in the brains of rats and in SH-SY5Y cells exposed to high level of fluoride might connect to oxidative stress. *Neurotoxicol Teratol* 48: 49-55.

Zhang J, Zhu W, Zhang Z. 2009. [The effect of fluorine exposure of pregnant rats on the learning and memory capabilities of baby rats]. *Chinese Journal of Public Health* 25(11): 1347-1348.

Zhang J, Zhu WJ, Xu XH, Zhang ZG. 2011. Effect of fluoride on calcium ion concentration and expression of nuclear transcription factor kappa-B rho65 in rat hippocampus. *Exp Toxicol Pathol* 63: 407-411.

Zhang J, Zhang Z. 2013. Effects of chronic fluorosis on camkiiα, c-FOS, BAX, and BCL-2 channel signaling in the hippocampus of rats. *Fluoride* 46: 135-141.

Zhang Z, Shen X, Xu X. 2001. [Effects of selenium on the damage of learning-memory ability of mice induced by fluoride]. *J Hyg Res* 30: 144-146.

Zhang Z, Xu X, Shen X, Xua XH. 1999. [Effect of fluoride exposure on synaptic structure of brain areas related to learning-memory in mice]. *J Hyg Res* 28(4): 210-212.

Zhang Z, Xu X, Shen X, Xua XH. 2008. Effect of fluoride exposure on synaptic structure of brain areas related to learning-memory in mice. *Fluoride* 41: 139-143.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao XL, Wu JH. 1998. Actions of sodium fluoride on acetylcholinesterase activities in rats. *Biomed Environ Sci* 11: 1-6.

Zheng X, Sun Y, Ke L, Ouyang W, Zhang Z. 2016. Molecular mechanism of brain impairment caused by drinking-acquired fluorosis and selenium intervention. *Environ Toxicol Pharmacol* 43: 134-139.

Zhu W, Zhang J, Zhang Z. 2011. Effects of fluoride on synaptic membrane fluidity and PSD-95 expression level in rat hippocampus. *Biol Trace Elem Res* 139: 197-203.

Zhu YL, Zheng YJ, LV XM, Ma Y, Zhang J. 2012. Effects of fluoride exposure on performance in water labyrinth and monoamine neurotransmitters of rats. *Journal of Xinjiang Medical University* 3: 014.

Zhu YP, Xi SH, Li MY, Ding TT, Liu N, Cao FY, Zeng Y, Liu XJ, Tong JW, Jiang SF. 2017. Fluoride and arsenic exposure affects spatial memory and activates the ERK/CREB signaling pathway in offspring rats. *Neurotox* 59: 56-64.

Studies Not Available in HAWC

Abdelaleem MM, El-Tahawy NFG, Abozaid SMM, Abdel-Hakim SA. 2018. Possible protective effect of curcumin on the thyroid gland changes induced by sodium fluoride in albino rats: Light and electron microscopic study. *Endocr Regul* 52: 59-68.

Abd-Elhakim YM, Mohammed AT, Ali HA. 2018. Impact of subchronic exposure to triclosan and/or fluoride on estrogenic activity in immature female rats: The expression pattern of calbindin-D9k and estrogen receptor alpha genes. *J Biochem Mol Toxicol* 32(2): 22027.

Abdumajidov OR. 2004. [Sex differences in lipid peroxidation and antioxidant defense of the brain tissue in intoxication with low doses of inorganic compounds]. *Uzbekiston Tibbiet Zhurnali*: 58-60.

Adebayo OL, Shallie PD, Salau BA, Ajani EO, Adenuga GA. 2013. Comparative study on the influence of fluoride on lipid peroxidation and antioxidants levels in the different brain regions of well-fed and protein undernourished rats. *J Trace Elem Med Biol* 27: 370-374.

Adedara IA, Ojuade TJD, Olabiyi BF, Idris UF, Onibiyo EM, Ajeigbe OF, Farombi EO. 2016. Taurine ameliorates renal oxidative damage and thyroid dysfunction in rats chronically exposed to fluoride. *Biol Trace Elem Res*: 1-8.

Ahmed SK, Kalleny NK, Attia AAEM, Elkateb LA. 2015. The possible protective role of chromium chloride against sodium fluoride-induced changes in the structure of the cerebellar cortex of the adult male albino rat. *Egypt J Histol* 38: 402-414.

Al Badawi MH, Mahmoud OM, Salem NA. 2016. Therapeutic potential of omega-3 against sodium fluoride toxicity on the cerebellar cortex of adult male albino rats: Histological and immunohistochemical study. *Egypt J Histol* 39: 170-178.

Alhayani A, Elshal EB, Aal IHA, Al-Shammeri E, Kabra H. 2013. Does vitamin E protect against sodium fluoride toxicity on the cerebellar cortex of albino rats? *Middle East J Sci Res* 16: 1019-1026.

Ameeramja J, Raghunath A, Perumal E. 2018. Tamarind seed coat extract restores fluoride-induced hematological and biochemical alterations in rats. *Environ Sci Pollut Res Int* 25(26): 26157-26166.

Antonyan OA. 1980. [Lipid per oxidation in fluorosis and the protective role of dietary factors]. *Zh Eksp Klin Med* 20: 381-388.

Ardelean I, Racoveanu N, Manescu S, Lupulescu A, Diaconescu M, Ghelerter L. 1964. Experimental investigations concerning the effect of fluorine on the thyroid gland. *Rum Med Rev* 8: 17-20.

Atmaca N, Atmaca HT, Kanici A, Anteplioglu T. 2014. Protective effect of resveratrol on sodium fluoride-induced oxidative stress, hepatotoxicity and neurotoxicity in rats. *Food Chem Toxicol* 70: 191-197.

Auskaps AM, Shaw JH. 1955. Hemoglobin concentration, thyroid weight and growth rate in rats during minimum fluoride ingestion. *J Nutr* 55: 611-621.

Bagmut I, Kolisnyk I, Titkova A, Petrenko T, Filipchenko S. 2018. Content of catecholamines in blood serum of rats under fluoride intoxication. *Georgian Med News* (280-281): 125-129.

Bakalyan PH, Antonyan OA. 1981. [Effect of fluorosis on glutathione peroxidase and glutathione reductase activities and sulfhydryl groups]. *Zh Eksp Klin Med* 21: 10-14.

Basha PM, Madhusudhan N. 2010. Pre and post natal exposure of fluoride induced oxidative macromolecular alterations in developing central nervous system of rat and amelioration by antioxidants. *Neurochem Res* 35: 1017-1028.

Basha PM, Madhusudhan N. 2011. Effect of maternal exposure of fluoride on oxidative stress markers and amelioration by selected antioxidants in developing central nervous system of rats. *Biologia* 66: 187-193.

Basha PM, Rai P, Begum S. 2011. Evaluation of fluoride-induced oxidative stress in rat brain: A multigeneration study. *Biol Trace Elem Res* 142: 623-637.

Basha PM, Sujitha NS. 2012. Combined influence of intermittent exercise and temperature stress on the modulation of fluoride toxicity. *Biol Trace Elem Res* 148: 69-75.

Basha PM, Saumya SM. 2013. Suppression of mitochondrial oxidative phosphorylation and TCA enzymes in discrete brain regions of mice exposed to high fluoride: Amelioration by panax ginseng (ginseng) and lagerstroemia speciosa (banaba) extracts. *Cell Mol Neurobiol* 33: 453-464.

Basha MP, Begum S, Madhusudhan N. 2014. Antioxidants in the management of fluoride induced neural oxidative stress in developing rats. *Int J Pharm Sci Res* 5: 201-206.

Benetato G, Giuran AM, Cirmaciu R, Cirje M, Petrescu A, Vacariu A. 1970. [Effect of fluorine in drinking water on the metabolism of Ca and Mg and on neuromuscular excitability: Experimental studies and clinical observations]. *Rev Roum Physiol* 7: 335-352.

Bharti VK, Srivastava RS. 2009. Fluoride-induced oxidative stress in rat's brain and its amelioration by buffalo (Bubalus bubalis) pineal proteins and melatonin. *Biol Trace Elem Res* 130: 131-140.

Bhatnagar M, Rao P, Saxena A, Bhatnagar R, Meena P, Barbar S, Chouhan A, Vimal S. 2006. Biochemical changes in brain and other tissues of young adult female mice from fluoride in their drinking water. *Fluoride* 39: 280-284.

Bilgili A, Akdogan M, Yildiz M, Eraslan G, Cetin N. 2004. The effects of fluoride on thyroid hormones in rabbits. *Indian Vet J* 81: 986-988.

Bobek S, Kahl S, Ewy Z. 1976. Effect of long-term fluoride administration on thyroid hormones level blood in rats. *Endocrinol Exp* 10: 289-295.

Bouaziz H, Ammar E, Ghorbel H, Ketata S, Jamoussi K, Ayadi F, Guermazi F, Zeghal N. 2004. Effect of fluoride ingested by lactating mice on the thyroid function and bone maturation of their suckling pups. *Fluoride* 37: 133-142.

Bouaziz H, Soussia L, Guermazi F, Zeghal N. 2005. Fluoride-induced thyroid proliferative changes and their reversal in female mice and their pups. *Fluoride* 38: 185-192.

Bouaziz HB, Amara I, Essefi M, Croute F, Zeghal N. 2010. Fluoride-induced brain damages in suckling mice. *Pestic Biochem Physiol* 96: 24-29.

Chauhan SS, Ojha S, Mahmood A. 2013. Effects of fluoride and ethanol administration on lipid peroxidation systems in rat brain. *Indian J Exp Biol* 51: 249-255.

Chen J, Chen X, Yang K, Xia T, Xie H. 2002. [Studies on DNA damage and apoptosis in rat brain induced by fluoride]. *Chin J Prev Med* 36: 222-224.

Chirumari K, Reddy PK. 2007. Dose-dependent effects of fluoride on neurochemical milieu in the hippocampus and neocortex of rat brain. *Fluoride* 40: 101-110.

Chouhan S, Yadav A, Kushwah P, Kaul RK, Flora SJS. 2011. Silymarin and quercetin abrogates fluoride induced oxidative stress and toxic effects in rats. *Mol Cell Toxicol* 7: 25-32.

Clay AB, Suttie JW. 1987. Effect of dietary fluoride on dairy cattle: Growth of young heifers. *J Dairy Sci* 70: 1241-1251.

Czechowicz K, Osada A, Slesak B. 1974. Histochemical studies on the effect of sodium fluoride on metabolism in Purkinje's cells. *Folia Histochem Cytochem* 12: 37-44.

Demole V, Lerch P. 1956. [Normality of fixation of radioactive iodine in the thyroid of rats during experimental fluorosis]. *Helv Physiol Pharmacol Acta* 14(4): 62-63.

Dhurvey V, Patil V, Thakare M. 2017. Effect of sodium fluoride on the structure and function of the thyroid and ovary in albino rats (rattus norvegicus). *Fluoride* 50: 235-246.

Domzalska E. 1966. [Influence of sodium fluoride on hypophysis, thyroid gland, parathyroid, and adrenal gland in the white rat]. *Czas Stomatol* 19: 839-844.

El-lethey HS, Kamel MM, Shaheed IB. 2011. Perinatal exposure to sodium fluoride with emphasis on territorial aggression, sexual behaviour and fertility in male rats. *Life Sci J* 8: 686-694.

Flora SJS, Mittal M, Mishra D. 2009. Co-exposure to arsenic and fluoride on oxidative stress, glutathione linked enzymes, biogenic amines and DNA damage in mouse brain. *J Neurol Sci* 285: 198-205.

Flora SJS, Mittal M, Pachauri V, Dwivedi N. 2012. A possible mechanism for combined arsenic and fluoride induced cellular and DNA damage in mice. *Metallomics* 4: 78-90.

Gabovich RD, Verzhikovskaia NV. 1958. [Effect of fluoride compounds on absorption of radioactive iodine by the thyroid gland in humans and in experimental conditions]. *Probl Endokrinol Gormonoter* 4: 49-54.

Galamini-Ligori M, Di Blasi F. 1961. [Action of sodium fluoride on the thyroid of hypophysectomized rats]. *Boll Soc Ital Biol Sper* 37: 1503-1506.

Galletti P, Held HR, Korrodi H, Wegmann T. 1956. [Investigations on the possible damaging effect of fluorine on the thyroid gland]. *Helv Med Acta* 23: 601-605.

Ge Y, Ning H, Feng C, Wang H, Yan X, Wang S, Wang J. 2006. Apoptosis in brain cells of offspring rats exposed to high fluoride and low iodine. *Fluoride* 39: 173-178.

Ge Y, Niu R, Zhang J, Wang J. 2011. Proteomic analysis of brain proteins of rats exposed to high fluoride and low iodine. *Arch Toxicol* 85: 27-33.

Ge YM, Ning HM, Wang SL, Wang JD. 2005. DNA damage in thyroid gland cells of rats exposed to long-term intake of high fluoride and low iodine. *Fluoride* 38: 318-323.

Ge YM, Ning HM, Wang SL, Wang JD. 2005. Comet assay of DNA damage in brain cells of adult rats exposed to high fluoride and low iodine. *Fluoride* 38: 209-214.

Ge YM, Ning HM, Wang SL, Wang JD. 2005. Effects of high fluoride and low iodine on brain histopathology in offspring rats. *Fluoride* 38: 127-132.

Ge YM, Ning HM, Gu XL, Yin M, Yang XF, Qi YH, Wang JD. 2013. Effects of high fluoride and low iodine on thyroid function in offspring rats. *J Integr Agric* 12: 502-508.

Guan ZZ. 1986. [Morphology of the brain of the offspring of rats with chronic fluorosis]. *Chin J Pathol* 15: 297-299.

Guan Z, Wang Y, Xiao K. 1997. [Influence of experimental fluorosis on phospholipid content and fatty acid composition in rat brain]. *Chin Med J* 77: 592-596.

Guan Z-Z, Wang Y-N, Xiao K-Q, Dai D-Y, Chen Y-H, Liu J-L, Sindelar P, Dallner G. 1998. Influence of chronic fluorosis on membrane lipids in rat brain. *Neurotoxicol Teratol* 20: 537-542.

Guan ZZ, Shan KR, Wang YN, Dallner G. 2006. [Changes of lipids and nicotinic receptors in rat brains and pheochromocytoma with fluorosis]. *Chin J Endemiol* 25: 121-124.

Gushchin SK. 1951. [Effect of sodium fluoride on iodine metabolism in rabbit tissue organs; on the etiology of endemic goiter]. *Gig Sanit* 2: 45-48.

Hamza RZ, Al-Harbi MS. 2014. Sodium fluoride induced neurotoxicity and possible antioxidant role of selenium and curcumin in male mice. *Biosci Biotechnol Res Asia* 11: 81-87.

Hamza RZ, El-Shenawy NS, Ismail HAA. 2015. Protective effects of blackberry and quercetin on sodium fluoride-induced oxidative stress and histological changes in the hepatic, renal, testis and brain tissue of male rat. *J Basic Clin Physiol Pharmacol* 26: 237-251.

Haojun Z, Yaoling W, Ke Z, Jin L, Junling W. 2012. Effects of NaF on the expression of intracellular Ca2+ fluxes and apoptosis and the antagonism of taurine in murine neuron. *Toxicol Mech Methods* 22: 305-308.

Hara K. 1980. Studies on fluorosis especially effects of fluoride on thyroid metabolism. *J Dent Health* 30: 42-57.

Harris NO, Hayes RL. 1955. A tracer study of the effect of acute and chronic exposure to sodium fluoride on the thyroid iodine metabolism of rats. *J Dent Res* 34: 470-477.

Hassan HA, Abdel-Aziz AF. 2010. Evaluation of free radical-scavenging and anti-oxidant properties of black berry against fluoride toxicity in rats. *Food Chem Toxicol* 48: 1999-2004.

Hoogstratten B, Leone NCLG, Shupe J, Greenwood DA, Lieberman J. 1965. Effect of fluorides on hematopoietic system, liver, and thyroid gland in cattle. *J Amer Med Assoc* 192: 26-32.

Inkielewicz I, Rogowska M, Krechniak J. 2006. Lipid peroxidation and antioxidant enzyme activity in rats exposed to fluoride and ethanol. *Fluoride* 39: 53-59.

Inkielewicz I, Czarnowski W. 2008. Oxidative stress parameters in rats exposed to fluoride and aspirin. *Fluoride* 41: 76-82.

Inkielewicz-Stepniak I, Czarnowski W. 2010. Oxidative stress parameters in rats exposed to fluoride and caffeine. *Food Chem Toxicol* 48: 1607-1611.

Jiang P, Li G, Zhou X, Wang C, Qiao Y, Liao D, Shi D. 2018. Chronic fluoride exposure induces neuronal apoptosis and impairs neurogenesis and synaptic plasticity: Role of GSK-3beta/beta-catenin pathway. *Chemosphere* 214: 430-435.

Jiang SF, Xi SH, Yao SQ, Tong JW, Zhang YS, Wang Q, Su J, Li MY. 2013. [Effects of fluoride, arsenic and co-exposure on expression of Bcl-2 and Bax in hippocampus and cerebral cortex of rats]. *Chin J Endemiol* 32: 365-369.

Jiang Y, Guo X, Sun Q, Shan Z, Teng W. 2016. Effects of excess fluoride and iodide on thyroid function and morphology. *Biol Trace Elem Res* 170: 382-389.

Jin TX, Guan ZZ, Zhang H. 2011. [The effect of fluoride on α subunit of calcium/calmodulin-dependent protein kinase-II mRNA and protein expression in central nervous system]. *Chin J Endemiol* 30: 247-250.

Jonderko G, Kita K, Pietrzak J, Primus-Slowinska B, Ruranska B, Zylka-Wloszczyk M, Straszecka J. 1983. [Effect of subchronic sodium fluoride poisoning on the thyroid gland of rabbits with normal and increased supply of iodine]. *Endokrynol Pol* 34: 195-203.

Kahl S, Bobek S. 1975. [Effect of fluoride administration on radiothyroxine turnower in rats]. *Endokrynol Pol* 26: 391-396.

Kahl S, Ewy Z. 1975. Effect of single and long term sodium fluoride administration on biosynthesis of the thyroid hormone in rats. *Fluoride* 8: 191-198.

Kapoor V, Prasad T, Paliwal VK. 2001. Blood biochemical constituents in calves following subclinical levels of fluoride toxicosis. *Fluoride* 34: 126-131.

Karawya FS, Zahran NM, Azzam EZ. 2015. Is water fluoridation a hidden cause of obesity? Histological study on thyroid follicular cells of albino rats. *Egypt J Histol* 38: 547-557.

Kaur T, Bijarnia RK, Nehru B. 2009. Effect of concurrent chronic exposure of fluoride and aluminum on rat brain. *Drug Chem Toxicol* 32: 215-221.

Kelimu A, Liu KT, Lian J, Hu HH, Zheng YJ, Wang TM. 2008. [Effects of vitamin C and E on the ultrastructure in liver, kidney and brain of fluorosis rats]. *Chin J Endemiol* 27: 378-381.

Kinawy AA. 2019. Synergistic oxidative impact of aluminum chloride and sodium fluoride exposure during early stages of brain development in the rat. *Environ Sci Pollut Res Int* 26(11): 10951-10960.

Knizhnikov VA. 1959. [Effect of potable water with high fluoride concentration on thyroid function]. *Gig Sanit* 24: 20-25.

Knizhnikov VA, Tsypin AB, Shcherbova EN, Bugryshev PF. 1963. [The effect of drinking water with an increased fluorine content on the bioelectrical activity of the brain and heart under experimental conditions]. *Gig Sanit* 28: 16-19.

Kondo T, Yoshida M, Kasahara K. 1976. [Acute fluorosis in female rats: Time of inhibition and recovery of cholinesterase in serum and salivary glands]. *Jpn J Dent Health* 26: 187-192.

Kowalewska M. 1974. [Biopotentials of the organ of hearing in chronic sodium fluoride poisoning]. *J Pol Otolaryngol* 28: 417-424.

Krechniak J, Inkielewicz I. 2005. Correlations between fluoride concentrations and free radical parameters in soft tissues of rats. *Fluoride* 38: 293-296.

Leonard BE. 1972. Effect of phentolamine on the increase in brain glycolysis following the intraventricular administration of dibutyryl-3,5-cyclic adenosine monophosphate and sodium fluoride to mice. *Biochem Pharmacol* 21: 115-117.

Liu G, Zhang W, Jiang P, Li X, Liu C, Chai C. 2012. Role of nitric oxide and vascular endothelial growth factor in fluoride-induced goitrogenesis in rats. *Environ Toxicol Pharmacol* 34: 209-217.

Li H, Cai Q, Wang D. 2012. [Effect of fluoride on the expression of rat thyroid peroxidase mRNA]. *Chin J Endemiol* 31: 515-517.

Li H, Cai Q, Wang D. 2012. [Effects of fluoride on rat thyroid morphology, thyroid peroxidase activity and the expression of thyroid peroxidase protein]. *Chin J Endemiol* 31: 271-274.

Liu H, Hou C, Zeng Q, Zhao L, Cui Y, Yu L, Wang L, Zhao Y, Nie J, Zhang B, Wang A. 2016. Role of endoplasmic reticulum stress-induced apoptosis in rat thyroid toxicity caused by excess fluoride and/or iodide. *Environ Toxicol Pharmacol* 46: 277-285.

Liu YJ, Gao Q, Wu CX, Guan ZZ. 2010. [Changes of the c-Jun N-terminal kinase in the brains of rats with chronic fluorosis]. *Chin J Endemiol* 29: 608-612.

Liu YJ, Guan ZZ, Gao Q, Pei JJ. 2011. Increased level of apoptosis in rat brains and SH-SY5Y cells exposed to excessive fluoride-A mechanism connected with activating JNK phosphorylation. *Toxicol Lett* 204: 183-189.

Lohakare J, Pattanaik AK. 2013. Effects of addition of fluorine in diets differing in protein content on urinary fluoride excretion, clinical chemistry and thyroid hormones in calves. *Brazilian J Anim Sci* 42: 751-758.

Long YG, Wang YN, Chen J, Jiang SF, Nordberg A, Guan ZZ. 2002. Chronic fluoride toxicity decreases the number of nicotinic acetylcholine receptors in rat brain. *Neurotoxicol Teratol* 24: 751-757.

Lou DD, Liu YF, Zhang KL, Yu YN, Guan ZZ. 2011. [Changes of reactive oxygen species level and mitochondria fission-fusion in cortical neurons of rats with chronic fluorosis]. *Chin J Endemiol* 30: 256-260.

Lou DD, Liu YF, Qin SL, Zhang KL, Yu YN, Guan ZZ. 2012. [Changed transcription level of mitochondrial fission and fusion gene loci in cortical neurons of rats with chronic fluorosis]. *Chin J Endemiol* 31: 125-129.

Lou DD, Zhang KL, Qin SL, Liu YF, Yu YN, Guan ZZ. 2012. [Alteration of mitochondrial distribution and gene expression of fission 1 protein in cortical neurons of rats with chronic fluorosis]. *Chin J Pathol* 41: 243-247.

Lou DD, Pan JG, Zhang KL, Qin SL, Liu YF, Yu YN, Guan ZZ. 2013. [Changed expression of mito-fusion 1 and mitochondrial fragmentation in the cortical neurons of rats with chronic fluorosis]. *Chin J Prev Med* 47: 170-174.

Lou DD, Zhang KL, Pan JG, Qin SL, Liu YF, Yu YN, Guan ZZ. 2013. [Influence of chronic fluorosis on the expression of mitochondrial fission protein dynamin-related 1 in the cortical neurons of rats]. *Chin J Prev Med* 47: 561-564.

Lou DD, Zhang KL, Qin SL, Liu YF, Liu YJ, Guan ZZ. 2013. [Effects of chronic fluorosis on 4.8 kb mitochondrial DNA in liver, kidney and brain of rats]. *Chin J Endemiol* 32: 121-124.

Lou DD, Guan ZZ, Pei JJ. 2014. Alterations of apoptosis and expressions of Bax and Bcl-2 in the cerebral cortices of rats with chronic fluorosis. *Fluoride* 47: 199-207.

Luo GY, Niu RY, Sun ZL, Zhang JH, Wang JM, Wang C, Wang JD. 2011. Reduction of CaMKII expression in the hippocampus of rats from ingestion of fluoride and/or lead. *Fluoride* 44: 63-69.

Ma T, Liu D, Song K. 1999. Cytochemical study of neuron enzyme at anterior horn of spinal cord in rats with experimental fluorosis. *J Chin Med Univ* 28: 81-82.

Ma TX, Yu HT, Song KQ. 2008. [Expression of c-fos and Caspase 8 in cerebral cortex of rats with experimental fluorosis]. *Chin J Endemiol* 27: 131-133.

Mach Z, Zygulska-Machowa H. 1959. O wplywie fluoru na przemiane J131 [Russian and English summ.]. *Endokrynol Pol* 10: 157-162.

Machida H. 1989. [A study on the rabbit thermoregulatory system effects of high dose sodium fluoride]. *Dent Sci Rep* 89: 607-626.

Madan J, Puri JP, Singh JK. 2009. Growth, feed efficiency and blood profile of buffalo calves consuming high levels of fluoride. *Trop Anim Health Prod* 41: 295-298.

Madhusudhan N, Basha PM, Begum S, Ahmed F. 2009. Fluoride-induced neuronal oxidative stress and its amelioration by antioxidants in developing rats. *Fluoride* 42: 179-187.

Madhusudhan N, Basha PM, Rai P, Ahmed F, Prasad GR. 2010. Effect of maternal fluoride exposure on developing CNS of rats: Protective role of Aloe vera, Curcuma longa and Ocimum sanctum. *Indian J Exp Biol* 48: 830-836.

Manocha SL, Warner H, Olkowski ZL. 1975. Cytochemical response of kidney, liver and nervous system of fluoride ions in drinking water. *Histochem J* 7: 343-355.

Mansour HH, Tawfik SS. 2012. Efficacy of lycopene against fluoride toxicity in rats. *Pharm Biol* 50: 707-711.

Mietkiewski K, Walczak M, Trojanowicz R. 1966. [Effect of sodium fluoride on the neurosecretory system in guinea pigs]. *Endokrynol Pol* 17: 121-131.

Mohamed NE. 2016. The role of calcium in ameliorating the oxidative stress of fluoride in rats. *Biol Trace Elem Res* 170: 128-144.

Muhlemann HR, Schneider R. 1956. [Mitotic activity of rat thyroid epithelium after administration of fluoridated drinking water]. *Schweiz Med Wochenschr* 86: 625-627.

Nabavi SF, Eslami S, Moghaddam AH, Nabavi SM. 2011. Protective effects of curcumin against fluoride-induced oxidative stress in the rat brain. *Neurophysiology* 43: 287-291.

Nabavi SF, Moghaddam AH, Nabavi SM, Eslami S. 2011. Protective effect of curcumin and quercetin on thyroid function in sodium fluoride intoxicated rats. *Fluoride* 44: 147-152.

Nabavi SF, Habtemariam S, Jafari M, Sureda A, Nabavi SM. 2012. Protective role of gallic acid on sodium fluoride induced oxidative stress in rat brain. *Bull Environ Contam Toxicol* 89: 73-77.

Nabavi SF, Nabavi SM, Latifi AM, Mirzaei M, Habtemariam S, Moghaddam AH. 2012. Mitigating role of quercetin against sodium fluoride-induced oxidative stress in the rat brain. *Pharm Biol* 50: 1380-1383.

Nabavi SF, Nabavi SM, Habtemariam S, Moghaddam AH, Sureda A, Mirzaei M. 2013. Neuroprotective effects of methyl-3-O-methyl gallate against sodium fluoride-induced oxidative stress in the brain of rats. *Cell Mol Neurobiol* 33: 261-267.

Nabavi SM, Sureda A, Nabavi SF, Latifi AM, Moghaddam AH, Hellio C. 2012. Neuroprotective effects of silymarin on sodium fluoride-induced oxidative stress. *J Fluor Chem* 142: 79-82.

Narayanaswamy M, Piler MB. 2010. Effect of maternal exposure of fluoride on biometals and oxidative stress parameters in developing CNS of rat. *Biol Trace Elem Res* 133: 71-82.

Narbutt B, Romer TE, Grabski J, Szymik N. 1971. [Influence of natrium fluoride on the structure of the rat thyroid]. *Endokrynol Pol* 22: 445-451.

Narbutt B, Romer TE, Grabski J, Szymik N. 1971. [The influence of sodium fluoride on the morphology of the thyroid gland in rats]. *Endokrynol Pol* 22: 361-365.

Niu RY, Sun ZL, Cheng ZT, Liu HT, Chen HC, Wang JD. 2008. Effects of fluoride and lead on N-methyl-D-aspartate receptor 1 expression in the hippocampus of offspring rat pups. *Fluoride* 41: 101-110.

Niu R, Wang J, Sun Z, Xue X, Yan X, Zhang J. 2015. Transcriptional regulatory dynamics of the hypothalamic-pituitary-testicular axis in male mice exposed to fluoride. *Environ Toxicol Pharmacol* 40: 557-562.

Niu R, Zhang Y, Liu S, Liu F, Sun Z, Wang J. 2015. Proteome alterations in cortex of mice exposed to fluoride and lead. *Biol Trace Elem Res* 164: 99-105.

Ogilvie AL. 1952. Histological findings in the kidney, liver, pancreas, adrenal and thyroid gland of the rat following sodium fluoride administration. *J Dent Res* 31: 598-598.

Okayasu I, Tsuchida M, Yanagisawa F. 1985. Hyperplastic nodules of thyroid parafollicular cells (C cells) in rats induced by prolonged low dose ingestion of NaF. *Fluoride* 18: 111-117.

Pal S, Sarkar C. 2014. Protective effect of resveratrol on fluoride induced alteration in protein and nucleic acid metabolism, DNA damage and biogenic amines in rat brain. *Environ Toxicol Pharmacol* 38: 684-699.

Pan Y, Lu P, Yin L, Chen K, He Y. 2015. Effect of fluoride on the proteomic profile of the hippocampus in rats. *Z Naturforsch C* 70: 151-157.

Phillips PH, Lamb AR. 1934. Histology of certain organs and teeth in chronic toxicosis due to fluorin. *Arch Path* 17: 169-176.

Portela ML. 1972. [Biochemical effects in the prolonged ingestion of fluorides in rats]. *Arch Latinoam Nutr* 22: 291-308.

Prestes DS, Filappi A, Schossler DR, Duarte FA, Dressler VL, Flores EMM, Cecim M. 2009. Functional and histological evaluations of thyroid of sheep submitted to sodium fluoride administration. *Arq Bras Med Vet Zootec* 61: 293-298.

Puentes F, Cremer HD. 1966. Experiments on fluoride-iodine antagonism in the thyroid gland. *Adv Fluorine Res* 4: 213-220.

Qian W, Miao K, Li T, Zhang Z. 2013. Effect of selenium on fluoride-induced changes in synaptic plasticity in rat hippocampus. *Biol Trace Elem Res* 155: 253-260.

Qing-Feng S, Ying-Peng X, Tian-Tong X. 2019. Matrix metalloproteinase-9 and p53 involved in chronic fluorosis induced blood-brain barrier damage and neurocyte changes. *Arch Med Sci* 15(2): 457-466.

Qiu YH, Kong DM, Yang Q, Zhao N. 2010. [Influence of high-fluoride on thyroid function and brain damage in rats]. *Chin J Endemiol* 29: 146-149.

Raghavendra M, Ravindra RK, Raghuveer YP, Narasimha JK, Uma MRV, Navakishor P. 2016. Alleviatory effects of hydroalcoholic extract of cauliflower (brassica oleracea var. botrytis) on thyroid function in fluoride intoxicated rats. *Fluoride* 49: 84-90.

Rakhov GM. 1964. [Effect of calcium and fluorine in drinking water on the iodine metabolism and status of the thyroid gland in iodine insufficiency in food]. *Gig Sanit* 29: 12-17.

Ranpariya VL, Parmar SK, Sheth NR, Chandrashekhar VM. 2011. Neuroprotective activity of matricaria recutita against fluoride-induced stress in rats. *Pharm Biol* 49: 696-701.

Reddy KP, Sailaja G, Krishnaiah C. 2009. Protective effects of selenium on fluoride induced alterations in certain enzymes in brain of mice. *J Environ Biol* 30: 859-864.

Rogalska A, Kuter K, Zelazko A, Glogowska-Gruszka A, Swietochowska E, Nowak P. 2017. Fluoride alteration of [3H]glucose uptake in Wistar rat brain and peripheral tissues. *Neurotoxicol Res* 31: 436-443.

Saka O, Hallac P, Urgancioğlu I. 1965. The effect of fluoride on the thyroid of the rat. *New Istanbul Contrib Clin Sci* 8: 87-90.

Samanta A, Chanda S, Bandyopadhyay B, Das N. 2016. Establishment of drug delivery system nanocapsulated with an antioxidant (+)-catechin hydrate and sodium meta borate chelator against sodium fluoride induced oxidative stress in rats. *J Trace Elem Med Biol* 33: 54-67.

Sarkar C, Das N, Pal S, Dinda B. 2014. Oxidative stress induced alteration of protein and nucleic acid metabolism in fluoride-intoxicated rat brain: Protection by 3α-hydroxy olean-12-en-27-oic acid isolated from neanotis wightiana. *Int J Pharm Sci Res* 5: 3047-3066.

Sarkar C, Pal S. 2014. Ameliorative effect of resveratrol against fluoride-induced alteration of thyroid function in male Wistar rats. *Biol Trace Elem Res* 162: 278-287.

Sarkar C, Pal S, Das N, Dinda B. 2014. Ameliorative effects of oleanolic acid on fluoride induced metabolic and oxidative dysfunctions in rat brain: Experimental and biochemical studies. *Food Chem Toxicol* 66: 224-236.

Seffner W, Teubener W, Runde H, Wiedner H, Vogt J, Otto G, Zschau E, Geinitz D, Franke J. 1990. Boron as an antidote to fluorosis? II. Studies on various organs of pigs. *Fluoride* 23: 68-79.

Selim AOA, El-Haleem MR, Ibrahim IH. 2012. Effect of sodium fluoride on the thyroid gland of growing male albino rats: Histological and biochemical study. *Egypt J Histol* 35: 470-482.

Shao Q, Wang Yn, Guan Z. 2000. [Influence of free radical inducer on the level of oxidative stress in brain of rats with fluorosis]. *Chin J Prev Med* 34: 330-332.

Sharma C, Suhalka P, Sukhwal P, Jaiswal N, Bhatnagar M. 2014. Curcumin attenuates neurotoxicity induced by fluoride: An in vivo evidence. *Pharmacogn Mag* 10: 61-65.

Shashi A. 1992. Studies on alterations in brain lipid metabolism following experimental fluorosis. *Fluoride* 25: 77-84.

Shashi A. 1993. Nucleic acid levels in thyroid gland in acute and chronic fluoride intoxication. *Fluoride* 26: 191-196.

Shashi A, Singh JP, Thapar SP. 1994. Effect of long-term administration of fluoride on levels of protein, free amino acids and RNA in rabbit brain. *Fluoride* 27: 155-159.

Shashi A. 2003. Histopathological investigation of fluoride-induced neurotoxicity in rabbits. *Fluoride* 36: 95-105.

Shashi A, Neetika S, Bhardwaj M. 2009. Neuronal DNA damage and apoptosis in brain of rat exposed to fluoride. *Asian J Microbiol Biotechnol Environ Sci* 11: 629-632.

Shayiq RM, Raza H, Kidwai AM. 1986. Fluoride and lipid peroxidation a comparative study in different rat tissues. *Bull Environ Contam Toxicol* 37: 70-76.

Shen QF, Li HN, Xu TT, Xia YP. 2012. [Damage of blood brain barrier of spinal cord in rats with chronic fluorosis]. *Chin Med J* 92: 2357-2361.

Shen Q, Tian R, Li H, Xu T, Xia Y. 2014. [White matter injury of spinal cord in rats with chronic fluorosis and recovery after defluoridation]. *Chin Med J* 94: 1189-1192.

Shen X, Zhang Z, Xu X. 2004. [Influence of combined iodine and fluoride on phospholipid and fatty acid composition in brain cells of rats]. *J Hyg Res* 33: 158-161.

Shivarajashankara YM, Shivashankara AR, Bhat PG, Rao SH. 2001. Effect of fluoride intoxication on lipid peroxidation and antioxidant systems in rats. *Fluoride* 34: 108-113.

Shivarajashankara YM, Shivashankara AR, Bhat PG, Rao SH. 2002. Brain lipid peroxidation and antioxidant systems of young rats in chronic fluoride intoxication. *Fluoride* 35: 197-203.

Shivarajashankara YM, Shivashankara AR, Bhat PG, Rao SM, Rao SH. 2002. Histological changes in the brain of young fluoride-intoxicated rats. *Fluoride* 35: 12-21.

Siebenhuner L, Miloni E, Burgi H. 1984. [Effects of fluoride on thyroid hormone biosynthesis: Studies in a highly sensitive test system]. *Klin Wochenschr* 62: 859-861.

Singh R, Srivastava AK, Gangwar NK. 2017. Clinico-pathological studies on the co-exposure of cypermethrin and fluoride in experimental rats with ameliorative action of Vitamin E. *Vet Pract* 18(2): 207-210.

Soni KK, Shrivastava VK. 2007. Sodium fluoride induced histopathological changes in thyroid gland of male mus musculus. *Biochem Cell Arch* 7: 317-320.

Stee EW. 1968. *Effect of sodium fluoride and AMOX (NF3O) on growth and thyroid function in the rat*. No. AMRL-TR-67-189. Wright-Patterson Air Force Base, OH: pp. 67.

Štolc V, Podoba J. 1960. Effect of fluoride on the biogenesis of thyroid hormones. *Nature* 188: 855-856.

Sugivama Y. 1967. [The effect of sodium fluoride administration on the parathyroid glands]. *Hirosaki Med J* 19: 520-529.

Sun Y, Ke L, Zheng X, Li T, Ouyang W, Zhang Z. 2016. Effects of different levels of calcium intake on brain cell apoptosis in fluorosis rat offspring and its molecular mechanism. *Biol Trace Elem Res*: 1-12.

Takata H. 1958. The effect of fluorine upon the uptake of I131 by the thyroid glands. *Folia Pharmacol Jpn* 54: 230-236.

Teng Y, Zhang J, Zhang Z, Feng J. 2017. The effect of chronic fluorosis on calcium ions and CaMKIIα, and c-fos expression in the rat hippocampus. *Biol Trace Elem Res*: 295-302.

Trabelsi M, Guermazi F, Zeghal N. 2001. Effect of fluoride on thyroid function and cerebellar development in mice. *Fluoride* 34: 165-173.

Tsuchida M, Okayasu I, Kohyama Y, Kurihara H, Tanaka H, Yanagisawa F, Date C, Hayashi M, Mui K, Asada M. 1986. Effects of long term, low dose ingestion of fluoride on the thyroid gland in rats. *Stud Environ Sci* 27: 307-312.

Vani ML, Reddy KP. 2000. Effects of fluoride accumulation on some enzymes of brain and gastrocnemius muscle of mice. *Fluoride* 33: 17-26.

Wang C, Liang C, Ma J, Manthari RK, Niu R, Wang J, Wang J, Zhang J. 2018. Co-exposure to fluoride and sulfur dioxide on histological alteration and DNA damage in rat brain. *J Biochem Mol Toxicol* 32.

Wang H, Yang Z, Zhou B, Gao H, Yan X, Wang J. 2009. Fluoride-induced thyroid dysfunction in rats: Roles of dietary protein and calcium level. *Toxicol Ind Health* 25: 49-57.

Wang J, Niu R, Sun Z, Lv L, Smith GW. 2008. Effects of protein and calcium supplementation on bone metabolism and thyroid function in protein and calcium deficient rabbits exposed to fluoride. *Fluoride* 41: 283-291.

Wang JD, Ge YM, Ning HM, Wang SL. 2004. Effects of high fluoride and low iodine on oxidative stress and antioxidant defense of the brain in offspring rats. *Fluoride* 37: 264-270.

Wang JL. 2007. [Effect of fluoride on the intracellular Ca2+ in neurons of mice]. *Chin J Endemiol* 26: 505-507.

Wang Y, Guan Z, Xiao K. 1997. [Changes of coenzyme Q content in brain tissues of rats with fluorosis]. *Chin J Prev Med* 31: 330-333.

Wang Y, Dong Y, Wei N, Guan Z. 2015. [Influence of chronic fluorosis on expression of quinone oxidoreductase-1 and heme oxygenase-1 in rat brains]. *Chin J Endemiol* 34: 250-253.

Wedzisz A, Cieciura J. 1988. Effect of small sodium fluoride feed supplements on the serum thyroid hormone content of rats. *Bromatol Chem Toksykol* 21: 174-175.

Wei N, Dong YT, Deng J, Wang Y, Qi XL, Yu WF, Xiao Y, Zhou JJ, Guan ZZ. 2018. Changed expressions of N-methyl-d-aspartate receptors in the brains of rats and primary neurons exposed to high level of fluoride. *J Trace Elem Med Biol* 45: 31-40.

Yan N, Liu Y, Liu S, Cao S, Wang F, Wang Z, Xi S. 2016. Fluoride-induced neuron apoptosis and expressions of inflammatory factors by activating microglia in rat brain. *Mol Neurobiol* 53: 4449-4460.

Yang H, Xing R, Liu S, Yu H, Li P. 2016. Gamma-Aminobutyric acid ameliorates fluoride-induced hypothyroidism in male Kunming mice. *Life Sci* 146: 1-7.

Yang H, Xing R, Liu S, Yu H, Li P. 2019. Analysis of the protective effects of gamma-aminobutyric acid during fluoride-induced hypothyroidism in male Kunming mice. *Pharm Biol* 57(1): 29-37.

Yang M, Ren Z, Zhou B, Guan Z, Yu W. 2017. [Expression of endonuclease G in the brain tissue of rats with chronic fluorosis]. *Chin J Endemiol* 36: 327-332.

Yuan SD, Xie QW, Lu FY. 1993. Changes of serotonin content and turnover rate in hypothalamus of female rat during fluorosis. *Fluoride* 26: 57-60.

Zhai JX, Guo ZY, Hu CL, Wang QN, Zhu QX. 2003. [Studies on fluoride concentration and cholinesterase activity in rat hippocampus]. *Chin J Ind Hyg Occup Dis* 21: 102-104.

Zhan CW, Huo DJ. 1988. Ultrastructural findings in liver, kidneys, thyroid-gland and cardiac-muscle of rabbits following sodium-fluoride administration. *Fluoride* 21: 32-38.

Zhan XA, Xu ZR, Li JX, Wang M. 2005. Effects of fluorosis on lipid peroxidation and antioxidant systems in young pigs. *Fluoride* 38: 157-161.

Zhan XA, Li JX, Wang M, Xu ZR. 2006. Effects of fluoride on growth and thyroid function in young pigs. *Fluoride* 39: 95-100.

Zhang KL, Lou DD, Liu YF, Qin SL, Guan ZZ. 2012. [Changes of P-glycoprotein and nuclear factor κB in the cerebral cortex of rat with chronic fluorosis]. *Chin J Endemiol* 31: 613-616.

Zhang KL, Lou DD, Guan ZZ. 2013. [Expression of receptor for advanced glycation endproducts and nuclear factor κB in brain hippocampus of rat with chronic fluorosis]. *Chin J Endemiol* 32: 625-628.

Zhang WD, Zhang Y, Liu GY, Jiang P, Chai CY. 2008. [Effects of fluoride on ultrastructure of thyroids in rats]. *Chin J Endemiol* 27: 622-624.

Zhang ZG, Wang XY, Nian WW, Liao QX, Zhang R, Ouyang W. 2017. Effects of calcium on drinking fluorosis-induced hippocampal syntaptic plasticity impairment in the offspring of rats. *Transl Neurosci* 8: 191-200.

Zhao W, Zhu H, Yu Z, Aoki K, Misumi J, Zhang X. 1998. Long-term effects of various iodine and fluorine doses on the thyroid and fluorosis in mice. *Endocr Regul* 32: 63-70.

Zhao WY. 1988. [A preliminary study of the interaction of iodine and fluoride in experimental iodine goiter and fluorosis]. *Chin J Prev Med* 22: 146-148.

Zhao XL, Gao WH, Zhao ZL. 1994. [Effects of sodium fluoride on the activity of Ca2+Mg(2+)-ATPase in synaptic membrane in rat brain]. *Chin J Prev Med* 28: 264-266.

Zhavoronkov AA, Polyakova GA. 1973. Morphological and functional state of the hypothalamo-hypophyseal neurosecretory system in experimental fluorosis. *Bull Exp Biol Med* 75: 194-196.

Zhou B, Luo G, Wang C, Niu R, Wang J. 2014. Effects of fluoride on expression of cytokines in the hippocampus of adult rats. *Fluoride* 47: 191-198.

In Vitro Experimental Studies

As described in **Figure 4**, 60 in vitro experimental studies were included; however, data extraction was not conducted on in vitro studies. Therefore, in vitro experimental studies are not available in HAWC with the exception of in vitro studies that also reported in vivo non-human animal data that meet the relevant criteria for being made available in HAWC. The following lists of references are organized as studies that are available in HAWC (n = 6) followed by studies that are not available in HAWC (n = 54).

Studies Available in HAWC

Chen J, Niu Q, Xia T, Zhou G, Li P, Zhao Q, Xu C, Dong L, Zhang S, Wang A. 2018. ERK1/2-mediated disruption of BDNF-TrkB signaling causes synaptic impairment contributing to fluoride-induced developmental neurotoxicity. *Toxicology* 410: 222-230.

Niu Q, Chen J, Xia T, Li P, Zhou G, Xu C, Zhao Q, Dong L, Zhang S, Wang A. 2018. Excessive ER stress and the resulting autophagic flux dysfunction contribute to fluoride-induced neurotoxicity. *Environ Pollut* 233: 889-899.

Shan KR, Qi XL, Long YG, Nordberg A, Guan ZZ. 2004. Decreased nicotinic receptors in PC12 cells and rat brains influenced by fluoride toxicity: A mechanism relating to a damage at the level in post-transcription of the receptor genes. *Toxicology* 200: 169-177.

Zhang KL, Lou DD, Guan ZZ. 2015. Activation of the AGE/RAGE system in the brains of rats and in SH-SY5Y cells exposed to high level of fluoride might connect to oxidative stress. *Neurotoxicol Teratol* 48: 49-55.

Zhao Q, Niu Q, Chen J, Xia T, Zhou G, Li P, Dong L, Xu C, Tian Z, Luo C, Liu L, Zhang S, Wang A. 2019. Roles of mitochondrial fission inhibition in developmental fluoride neurotoxicity: Mechanisms of action in vitro and associations with cognition in rats and children. *Arch Toxicol* 93(3): 709-726.

Zhao XL, Wu JH. 1998. Actions of sodium fluoride on acetylcholinesterase activities in rats. *Biomed Environ Sci* 11: 1-6.

Studies Not Available in HAWC

Ardelean I, Racoveanu N, Manescu S, Lupulescu A, Diaconescu M, Ghelerter L. 1964. Experimental investigations concerning the effect of fluorine on the thyroid gland. *Rum Med Rev* 8: 17-20.

Chen J, Chen X, Yang K. 2000. [Effects of selenium and zinc on the DNA damage caused by fluoride in pallium neural cells of rats]. *J Hyg Res* 29: 216-217.

Chen L, Ning H, Yin Z, Song X, Feng Y, Qin H, Li Y, Wang J, Ge Y, Wang W. 2017. The effects of fluoride on neuronal function occurs via cytoskeleton damage and decreased signal transmission. *Chemosphere* 185: 589-594.

Chen R, Zhao LD, Liu H, Li HH, Ren C, Zhang P, Guo KT, Zhang HX, Geng DQ, Zhang CY. 2017. Fluoride induces neuroinflammation and alters Wnt signaling pathway in BV2 microglial cells. *Inflammation* 40: 1123-1130.

Cheng TJ, Chen TM, Chen CH, Lai YK. 1998. Induction of stress response and differential expression of 70 kDa stress proteins by sodium fluoride in HeLa and rat brain tumor 9L cells. *J Cell Biochem* 69: 221-231.

Deng MF, Zhu D, Liu YP, He WW, Gui CZ, Guan ZZ. 2018. Attenuation by 7-nitroindazole of fluoride-induced toxicity in SH-SY5Y cells exposed to high fluoride: Effects on nitric oxide, nitric oxide synthetase activity, nNOS, and apoptosis. *Fluoride* 51(4): 328-339.

Flores-Mendez M, Ramirez D, Alamillo N, Hernandez-Kelly LC, Del Razo LM, Ortega A. 2014. Fluoride exposure regulates the elongation phase of protein synthesis in cultured Bergmann glia cells. *Toxicol Lett* 229: 126-133.

Gao Q, Liu YH, Guan ZZ. 2008. Oxidative stress might be a mechanism connected with the decreased alpha 7 nicotinic receptor influenced by high-concentration of fluoride in SH-SY5Y neuroblastoma cells. *Toxicol In Vitro* 22: 837-843.

Goschorska M, Gutowska I, Baranowska-Bosiacka I, Piotrowska K, Metryka E, Safranow K, Chlubek D. 2018. Influence of acetylcholinesterase inhibitors used in Alzheimer's Disease treatment on the activity of antioxidant enzymes and the concentration of glutathione in THP-1 macrophages under fluoride-induced oxidative stress. *Int J Environ Res Pub Health* 16(1).

Guan ZZ, Shan KR, Xiu J, Long YG. 2005. [Fluorosis on expression of nicotinic acetylcholine receptors in protein and gene levels in human SH-SY5Y neuroblastoma cells]. *Chin J Prev Med* 39: 26-29.

Guan ZZ, Shan KR, Wang YN, Dallner G. 2006. [Changes of lipids and nicotinic receptors in rat brains and pheochromocytoma with fluorosis]. *Chin J Endemiol* 25: 121-124.

Haojun Z, Yaoling W, Ke Z, Jin L, Junling W. 2012. Effects of NaF on the expression of intracellular Ca2+ fluxes and apoptosis and the antagonism of taurine in murine neuron. *Toxicol Mech Methods* 22: 305-308.

Hong-Liang L, Qiang Z, Yu-Shan C, Lei Z, Gang F, Chang-Chun H, Liang Z, Aiguo W. 2014. Fluoride-induced thyroid cell apoptosis. *Fluoride* 47: 161-169.

Inkielewicz-Stepniak I, Radomski MW, Wozniak M. 2012. Fisetin prevents fluoride- and dexamethasone-induced oxidative damage in osteoblast and hippocampal cells. *Food Chem Toxicol* 50: 583-589.

Jin TX, Guan ZZ, Zhang H. 2011. [The effect of fluoride on α subunit of calcium/calmodulin-dependent protein kinase-II mRNA and protein expression in central nervous system]. *Chin J Endemiol* 30: 247-250.

Kariya T, Kotani M, Field JB. 1974. Effects of sodium fluoride and other metabolic inhibitors on basal and TSH stimulated cyclic AMP and thyroid metabolism. *Metab Clin Exper* 23: 967-973.

Ke L, Zheng X, Sun Y, Ouyang W, Zhang Z. 2016. Effects of sodium fluoride on lipid peroxidation and PARP, XBP-1 expression in PC12 cell. *Biol Trace Elem Res* 173: 161-167.

Lee J, Han YE, Favorov O, Tommerdahl M, Whitsel B, Lee CJ. 2016. Fluoride induces a volume reduction in CA1 hippocampal slices via MAP kinase pathway through volume regulated anion channels. *Exp Neurobiol* 25: 72-78.

Levesque L, Mizzen CA, McLachlan DR, Fraser PE. 2000. Ligand specific effects on aluminum incorporation and toxicity in neurons and astrocytes. *Brain Res* 877: 191-202.

Li H, Gao MT, Xu KY, Wang CY. 2007. Effect of sodium fluoride on the primary porcine thyroid cells and thyroid peroxidase activity. *J Clin Rehabil Tissue Eng Res* 11: 7425-7428.

Li H, Gao MT, Xu KY, Cui MY, Dai X. 2008. [Effect of fluoride on thyroid functioning in primary porcine thyrocyte]. *Chin J Endemiol* 27: 38-40.

Li H, Huang H, Xu Y, Gao Y, Liu Z. 2010. [Toxic effects of fluoride on rat cerebral cortex astrocytes in vitro]. *J Hyg Res* 39: 86-88.

Liu H, Zeng Q, Cui Y, Yu L, Zhao L, Hou C, Zhang S, Zhang L, Fu G, Liu Y, Jiang C, Chen X, Wang A. 2014. The effects and underlying mechanism of excessive iodide on excessive fluoride-induced thyroid cytotoxicity. *Environ Toxicol Pharmacol* 38: 332-340.

Liu HL, Zeng Q, Cui YS, Zhao L, Zhang L, Fu G, Hou CC, Zhang S, Yu LY, Jiang CY, Wang ZL, Chen XM, Wang AG. 2014. The role of the IRE1 pathway in excessive iodide- and/or fluoride-induced apoptosis in Nthy-ori 3-1 cells in vitro. *Toxicol Lett* 224: 341-348.

Liu YJ, Guan ZZ, Gao Q, Pei JJ. 2011. Increased level of apoptosis in rat brains and SH-SY5Y cells exposed to excessive fluoride-A mechanism connected with activating JNK phosphorylation. *Toxicol Lett* 204: 183-189.

Liu Y, Gao Q, Tang Z, Zhang X, Guan Z. 2015. [The expression and correlation between neural nicotinic acetylcholine receptor subunit α3 and mitogen-activated protein kinase cell signaling transduction pathway in human neuroblastoma cell line SH-SY5Y overexposed to fluoride]. *Chin J Endemiol* 34: 553-558.

Madaoui S, Rappaport L, Nunez J. 1974. Prostaglandins and in vitro TSH-dependent iodide binding by rat thyroid glands. *Biochimie* 56: 109-113.

Nakagawa-Yagi Y, Saito Y, Kitoh N, Ogane N, Fujisawa E, Nakamura H. 1993. Fluoride causes suppression of neurite outgrowth in human neuroblastoma via an influx of extracellular calcium. *Biochem Biophys Res Commun* 191: 727-736.

Ong J, Kerr DIB. 1995. Interactions of N-ethylmaleimide and aluminium fluoride with GABA(B) receptor function in rat neocortical slices. *Eur J Pharmacol* 287: 197-200.

Pastan I, Macchia V, Katzen R. 1968. Effect of fluoride on the metabolic activity of thyroid slices. *Endocrinology* 83: 157-160.

Rubakhova VM. 1977. [Effect of serotonin and sodium fluoride on visceral nerve conductors]. *Vyestsi Akademii Navuk BSSR Syeryya Biyalahichnykh Navuk* 1: 117-119.

Shayiq RM, Raza H, Kidwai AM. 1986. Fluoride and lipid peroxidation a comparative study in different rat tissues. *Bull Environ Contam Toxicol* 37: 70-76.

Shuhua X, Ziyou L, Ling Y, Fei W, Sun G. 2012. A role of fluoride on free radical generation and oxidative stress in BV-2 microglia cells. *Mediators Inflamm* 2012: 1-8.

Singh P, Das TK. 2019. Ultrastructural localization of 4-hydroxynonenal adducts in fluoride-exposed cells: Protective role of dietary antioxidants. *Fluoride* 52(1): 49-58.

Taylor P. 1972. Comparison of the effects of various agents on thyroidal adenyl cyclase activity with their effects on thyroid hormone release. *J Endocrinol* 54: 137-145.

Tu W, Zhang Q, Liu Y, Han LY, Wang Q, Chen PP, Zhang S, Wang AG, Zhou X. 2018. Fluoride induces apoptosis via inhibiting SIRT1 activity to activate mitochondrial p53 pathway in human neuroblastoma SH-SY5Y cells. *Toxicol Appl Pharmacol* 347: 60-69.

van der Voet GB, Schijns O, de Wolff FA. 1999. Fluoride enhances the effect of aluminium chloride on interconnections between aggregates of hippocampal neurons. *Arch Physiol Biochem* 107: 15-21.

Wang JL. 2007. [Effect of fluoride on the intracellular Ca2+ in neurons of mice]. *Chin J Endemiol* 26: 505-507.

Wang J, Gao Y, Cheng X, Yang J, Zhao Y, Xu H, Zhu Y, Yan Z, Manthari RK, Mehdi OM, Wang J. 2019. GSTO1 acts as a mediator in sodium fluoride-induced alterations of learning and memory related factors expressions in the hippocampus cell line. *Chemosphere* 226: 201-209.

Wei N, Dong YT, Deng J, Wang Y, Qi XL, Yu WF, Xiao Y, Zhou JJ, Guan ZZ. 2018. Changed expressions of N-methyl-d-aspartate receptors in the brains of rats and primary neurons exposed to high level of fluoride. *J Trace Elem Med Biol* 45: 31-40.

Willems CB-V, Sande J, Dumont JE. 1972. Inhibition of thyroid secretion by sodium fluoride in vitro. *Biochim Biophys Acta* 264: 197-204.

Woodward JJ, Harms J. 1992. Potentiation of N-methyl-D-aspartate-stimulated dopamine release from rat brain slices by aluminum fluoride and carbachol. *J Neurochem* 58: 1547-1554.

Wu J, Cheng M, Liu Q, Yang J, Wu S, Lu X, Jin C, Ma H, Cai Y. 2015. Protective role of tert-butylhydroquinone against sodium fluoride-induced oxidative stress and apoptosis in PC12 cells. *Cell Mol Neurobiol* 35: 1017-1025.

Xia T, Zhang M, He WH, He P, Wang AG. 2007. [Effects of fluoride on neural cell adhesion molecules mRNA and protein expression levels in primary rat hippocampal neurons]. *Chin J Prev Med* 41: 475-478.

Xu B, Xu Z, Xia T, He P, Gao P, He W, Zhang M, Guo L, Niu Q, Wang A. 2011. Effects of the Fas/Fas-L pathway on fluoride-induced apoptosis in SH-SY5Y cells. *Environ Toxicol* 26: 86-92.

Xu Z, Xu B, Xia T, He W, Gao P, Guo L, Wang Z, Niu Q, Wang A. 2013. Relationship between intracellular Ca2+ and ROS during fluoride-induced injury in SH-SY5Y cells. *Environ Toxicol* 28: 307-312.

Yamashita K, Field JB. 1972. Elevation of cyclic guanosine 3,5; monophosphate levels in dog thyroid slices caused by acetylcholine and sodium fluoride. *J Biol Chem* 247: 7062-7066.

Yan L, Liu S, Wang C, Wang F, Song Y, Yan N, Xi S, Liu Z, Sun G. 2013. JNK and NADPH oxidase involved in fluoride-induced oxidative stress in BV-2 microglia cells. *Mediators Inflamm* 2013: 895-975.

Zhang CY, Chen R, Wang F, Ren C, Zhang P, Li Q, Li HH, Guo KT, Geng DQ, Liu CF. 2016. EGb-761 attenuates the anti-proliferative activity of fluoride via DDK1 in PC-12 cells. *Neurochem Res* 42(2): 606-614.

Zhang M, Wang A, He W, He P, Xu B, Xia T, Chen X, Yang K. 2007. Effects of fluoride on the expression of NCAM, oxidative stress, and apoptosis in primary cultured hippocampal neurons. *Toxicology* 236: 208-216.

Zhang M, Wang A, Xia T, He P. 2008. Effects of fluoride on DNA damage, S-phase cell-cycle arrest and the expression of NF-kappaB in primary cultured rat hippocampal neurons. *Toxicol Lett* 179: 1-5.

Zhang S, Zheng X, Sun Y, Wang Y, Zhang Z. 2015. Alterations in oxidative stress and apoptosis in cultured PC12 cells exposed to fluoride. *Fluoride* 48: 213-222.

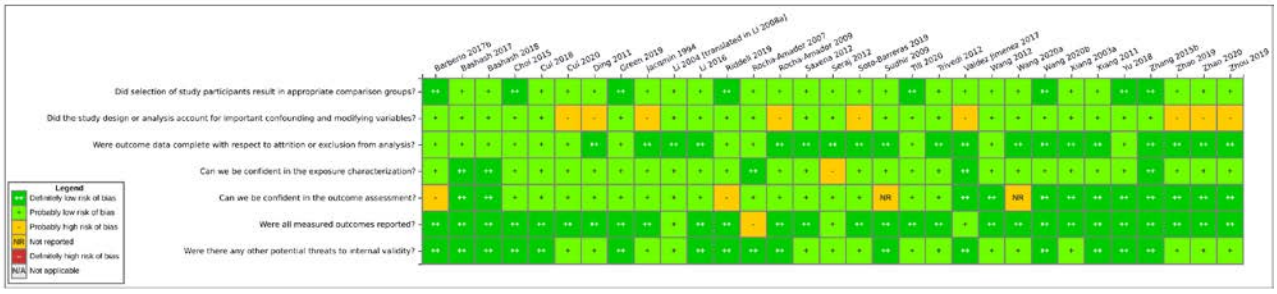
Zhao L, Xiao Y, Deng CM, Tan LC, Guan ZZ. 2016. Protective effect of lovastatin on neurotoxicity of excessive fluoride in primary hippocampal neurons. *Fluoride* 49: 36-46.

Zhao XL, Gao WH, Zhao ZL. 1994. [Effects of sodium fluoride on the activity of Ca2+Mg(2+)-ATPase in synaptic membrane in rat brain]. *Chin J Prev Med* 28: 264-266.

Appendix 3. Risk-of-bias Figures

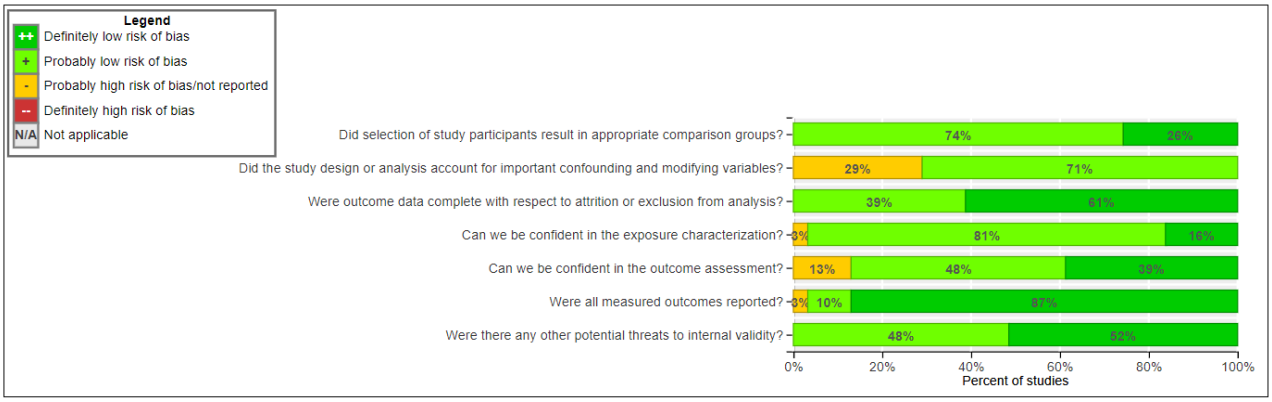
Studies in Humans

Figure A3-1. Risk-of-bias Heatmap for Lower Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure



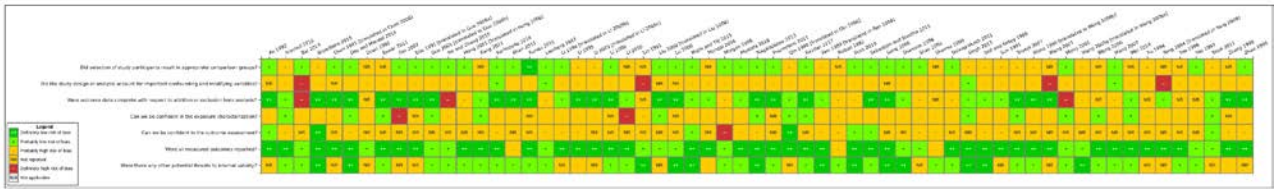
Interactive figure and additional study details in HAWC here.

Figure A3-2. Risk-of-bias Bar Chart for Lower Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure



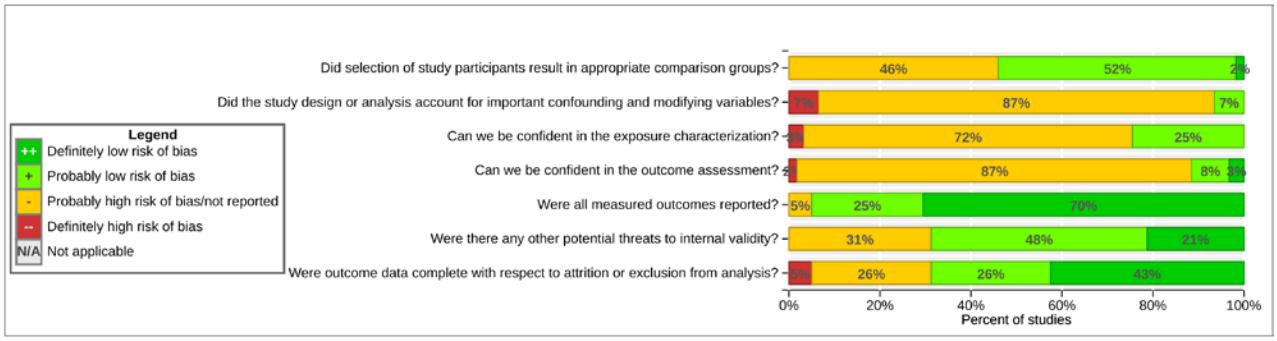
Interactive figure and additional study details in HAWC here.

Figure A3-3. Risk-of-bias Heatmap for Higher Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure



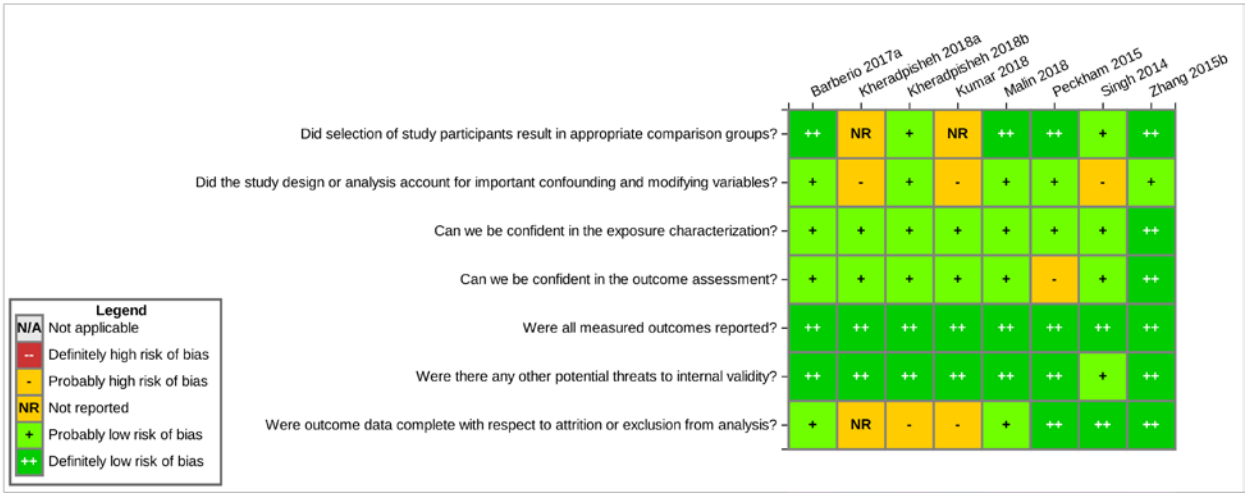
Interactive figure and additional study details in HAWC here.

Figure A3-4. Risk-of-bias Bar Chart for Higher Risk-of-bias Human Neurodevelopmental or Cognitive Studies Following Fluoride Exposure



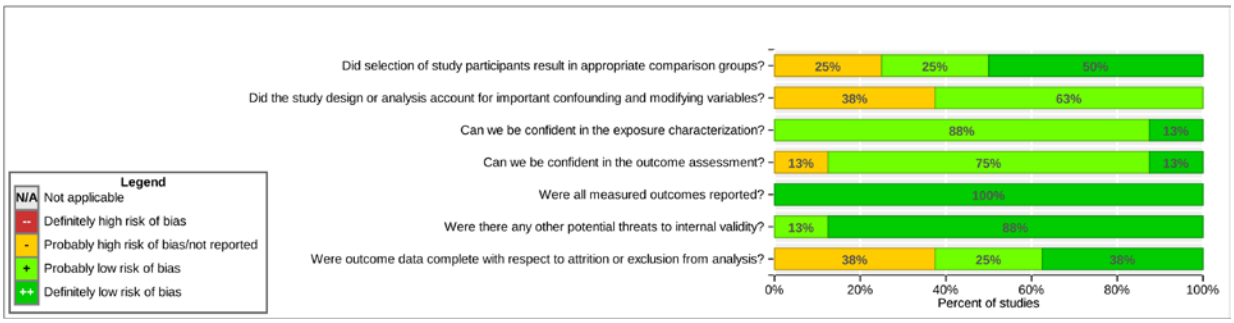
Interactive figure and additional study details in HAWC here.

Figure A3-5. Risk-of-bias Heatmap for Lower Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure



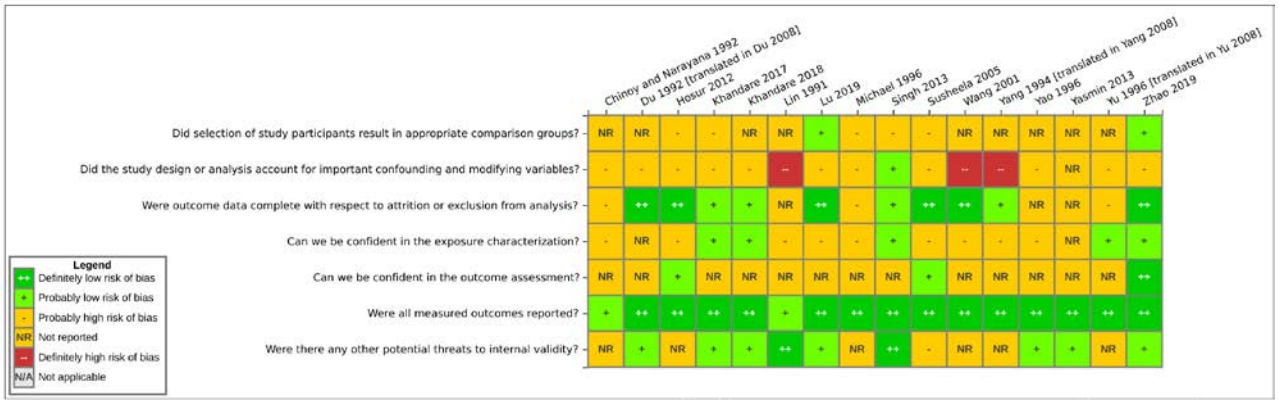
Interactive figure and additional study details in HAWC here.

Figure A3-6. Risk-of-bias Bar Chart for Lower Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure



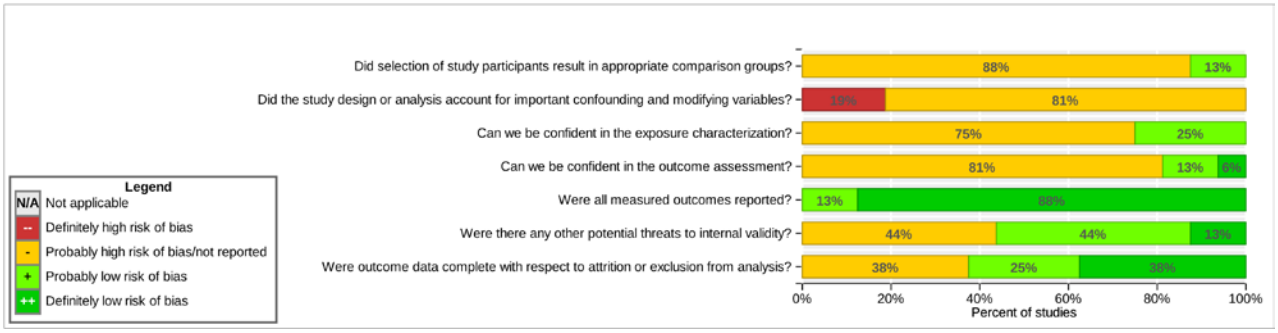
Interactive figure and additional study details in HAWC here.

Figure A3-7. Risk-of-bias Heatmap for Higher Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure



Interactive figure and additional study details in HAWC here.

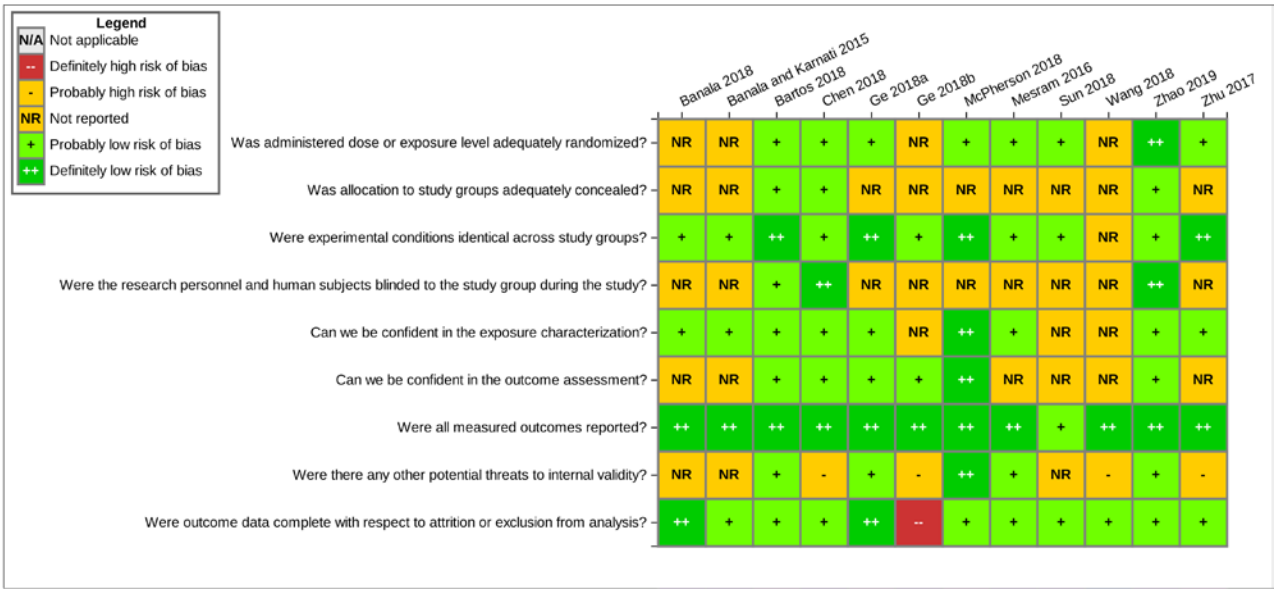
Figure A3-8. Risk-of-bias Bar Chart for Higher Risk-of-bias Human Mechanistic Studies Following Fluoride Exposure



Interactive figure and additional study details in HAWC here.

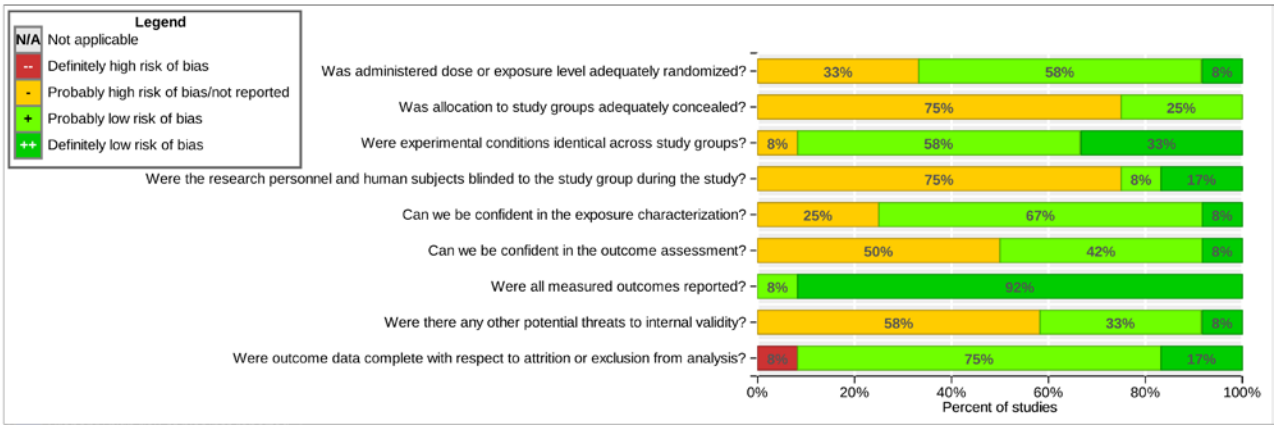
Studies in Non-human Animals

Figure A3-9. Risk-of-bias Heatmap for New Developmental Animal Learning and Memory Studies Following Fluoride Exposure



Interactive figure and additional study details in HAWC here.

Figure A3-10. Risk-of-bias Bar Chart for New Developmental Animal Learning and Memory Studies Following Fluoride Exposure

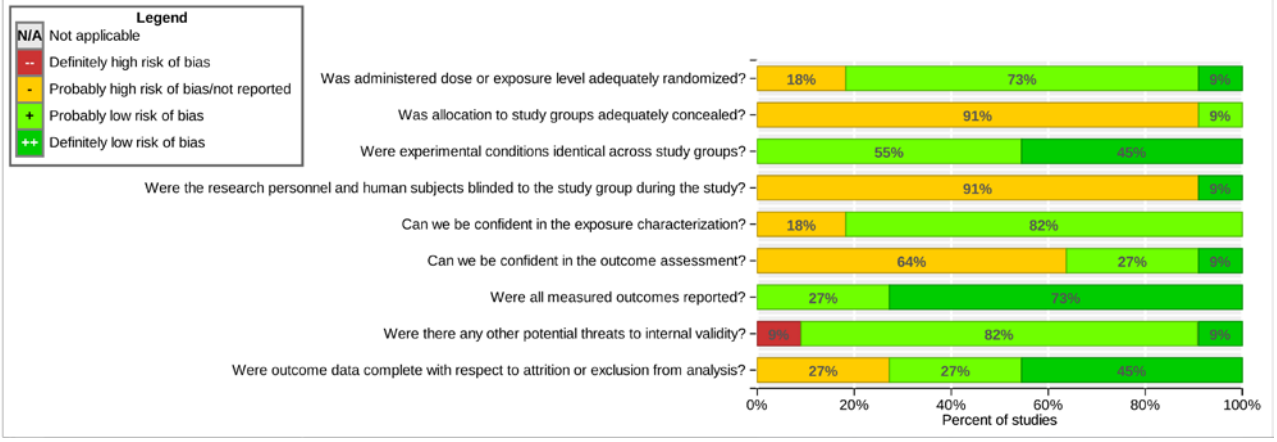


Interactive figure and additional study details in HAWC here.

Figure A3-11. Risk-of-bias Heatmap for New Adult Animal Learning and Memory Studies Following Fluoride Exposure

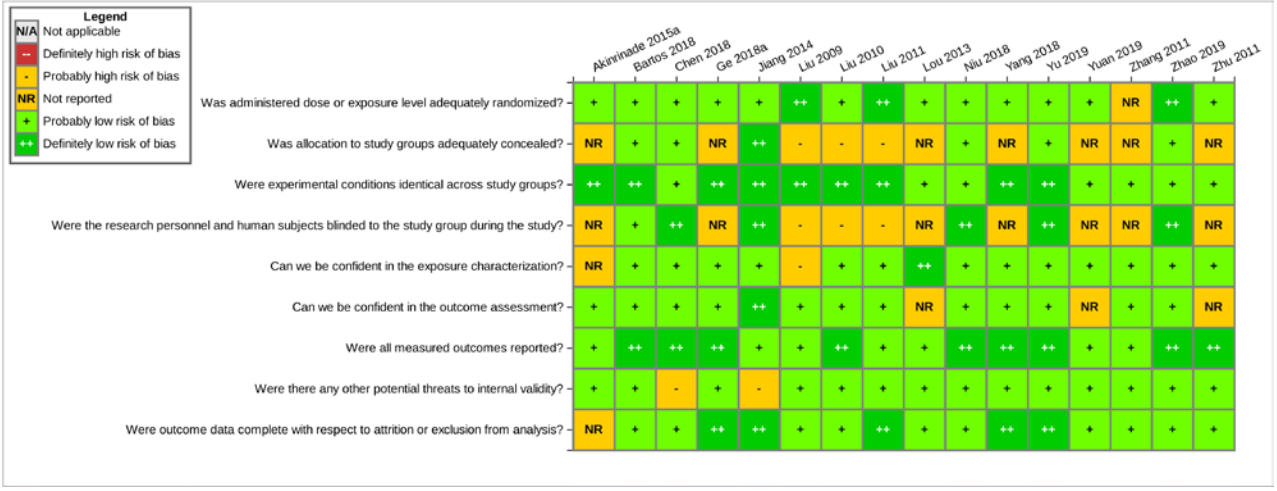
Interactive figure and additional study details in HAWC here.

Figure A3-12. Risk-of-bias Bar Chart for New Adult Animal Learning and Memory Studies Following Fluoride Exposure



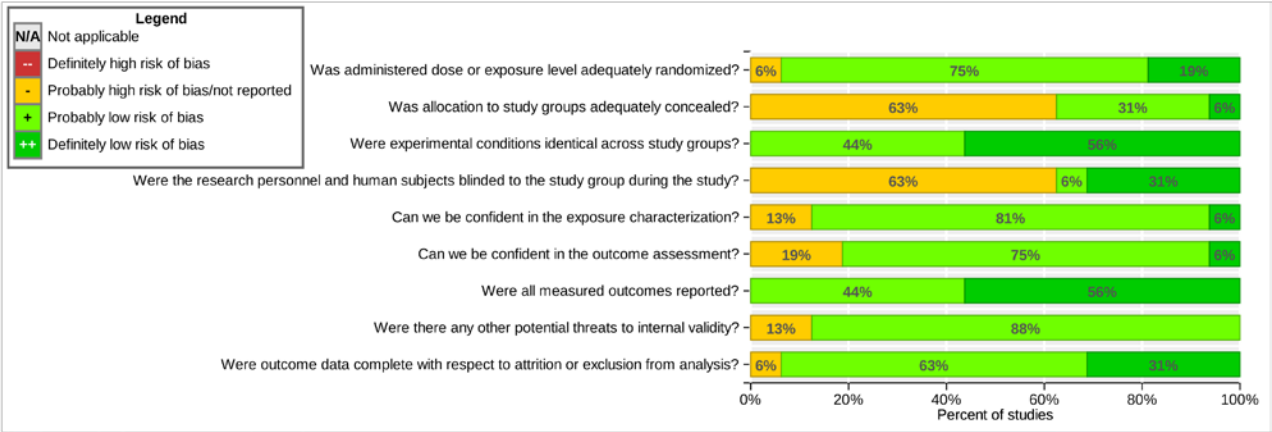
Interactive figure and additional study details in HAWC here.

Figure A3-13. Risk-of-bias Heatmap for Lower Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure



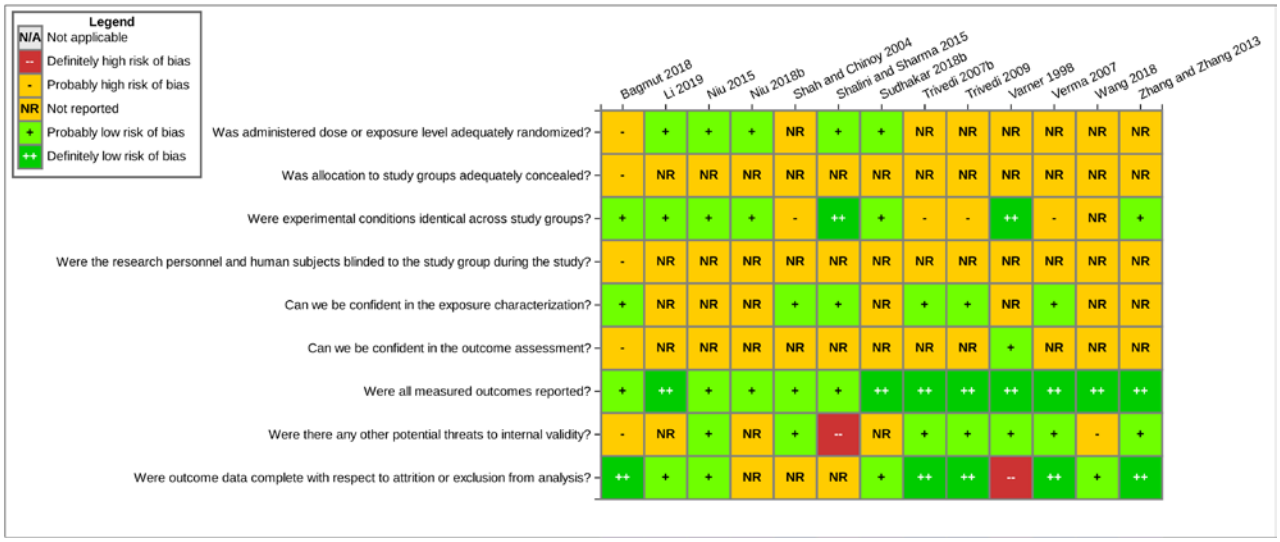
Interactive figure and additional study details in HAWC here.

Figure A3-14. Risk-of-bias Bar Chart for Lower Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure



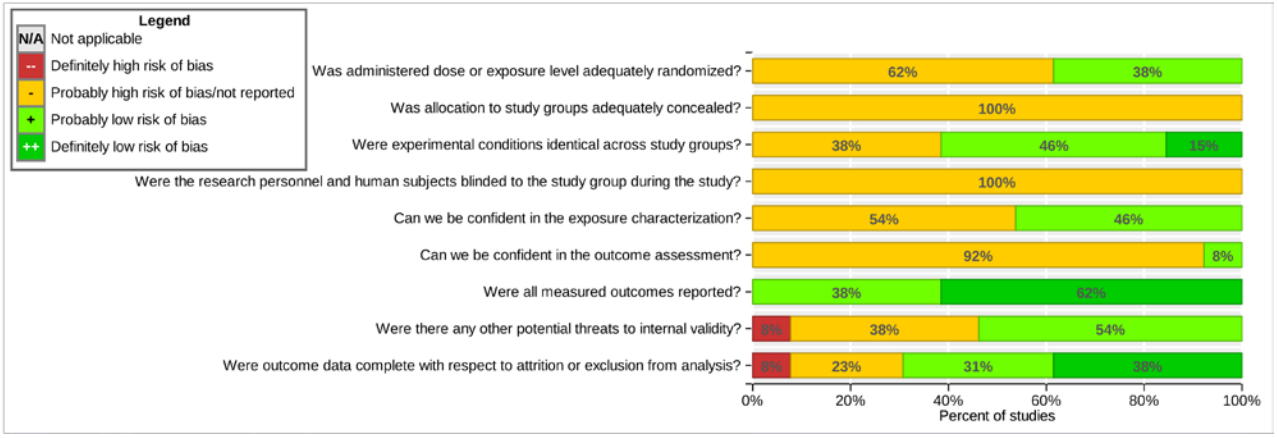
Interactive figure and additional study details in HAWC here.

Figure A3-15. Risk-of-bias Heatmap for Higher Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure



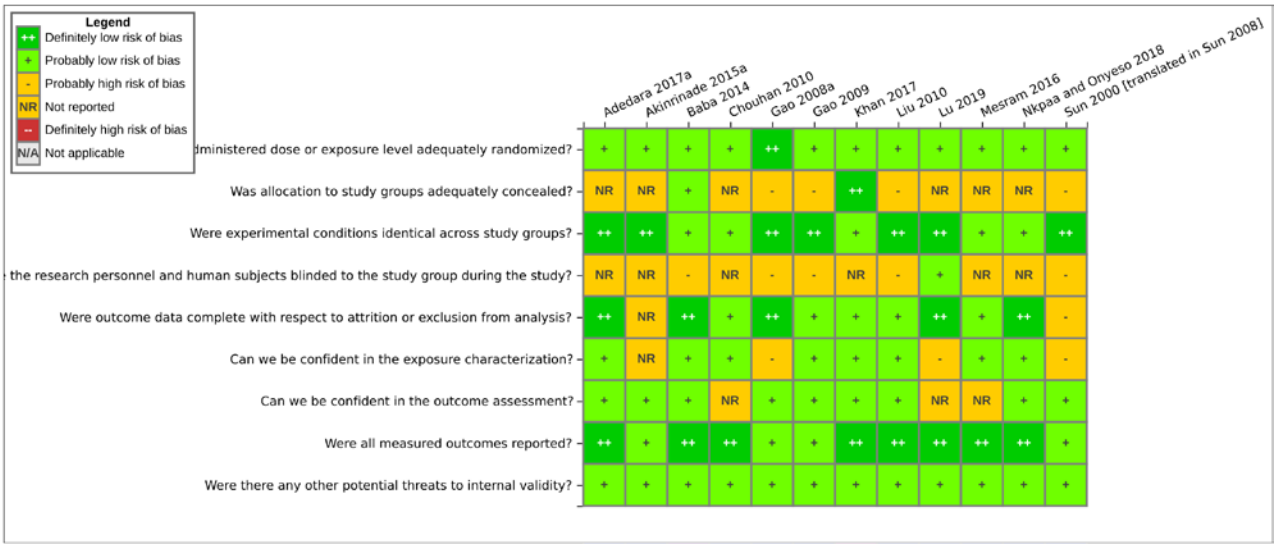
Interactive figure and additional study details in HAWC here.

Figure A3-16. Risk-of-bias Bar Chart for Higher Risk-of-bias Animal Biochemical Studies Following Fluoride Exposure



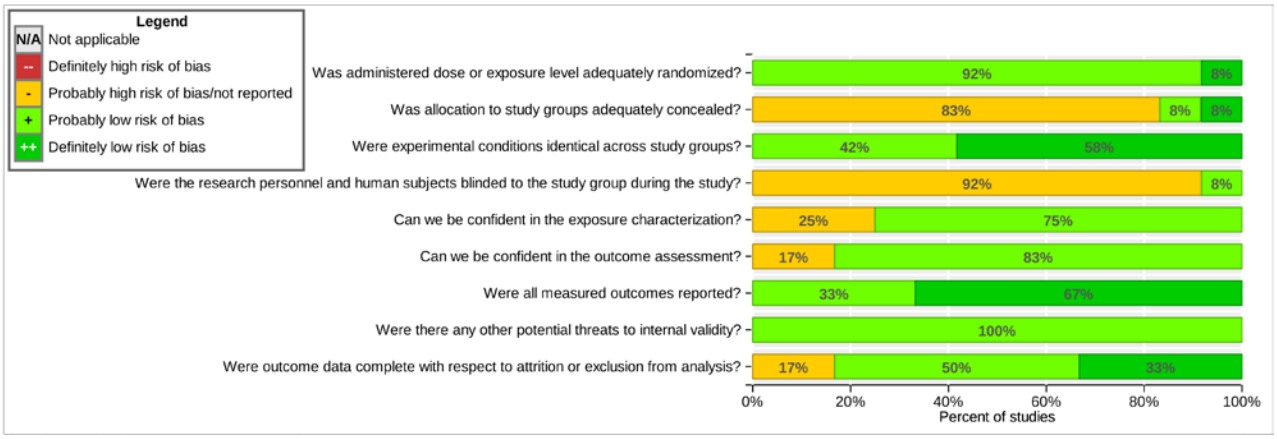
Interactive figure and additional study details in HAWC here.

Figure A3-17. Risk-of-bias Heatmap for Lower Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure



Interactive figure and additional study details in HAWC here.

Figure A3-18. Risk-of-bias Bar Chart for Lower Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure

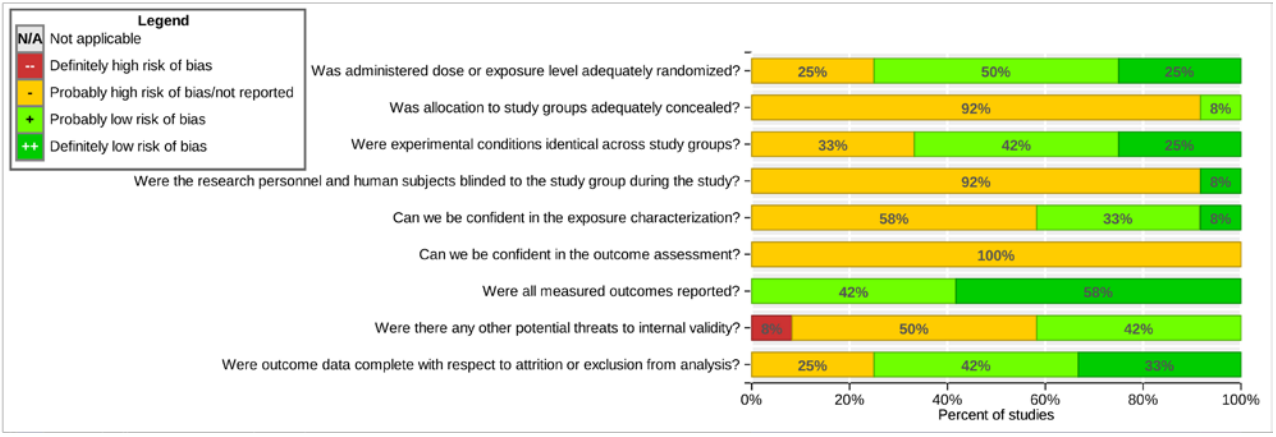


Interactive figure and additional study details in HAWC here.

Figure A3-19. Risk-of-bias Heatmap for Higher Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure

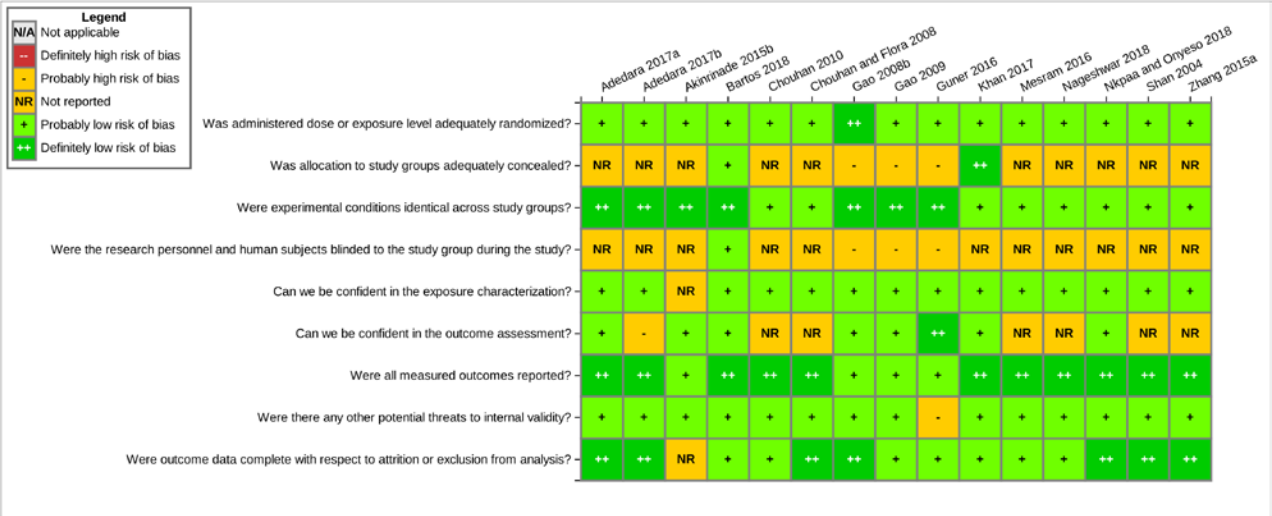
Interactive figure and additional study details in HAWC here.

Figure A3-20. Risk-of-bias Bar Chart for Higher Risk-of-bias Animal Neurotransmission Studies Following Fluoride Exposure



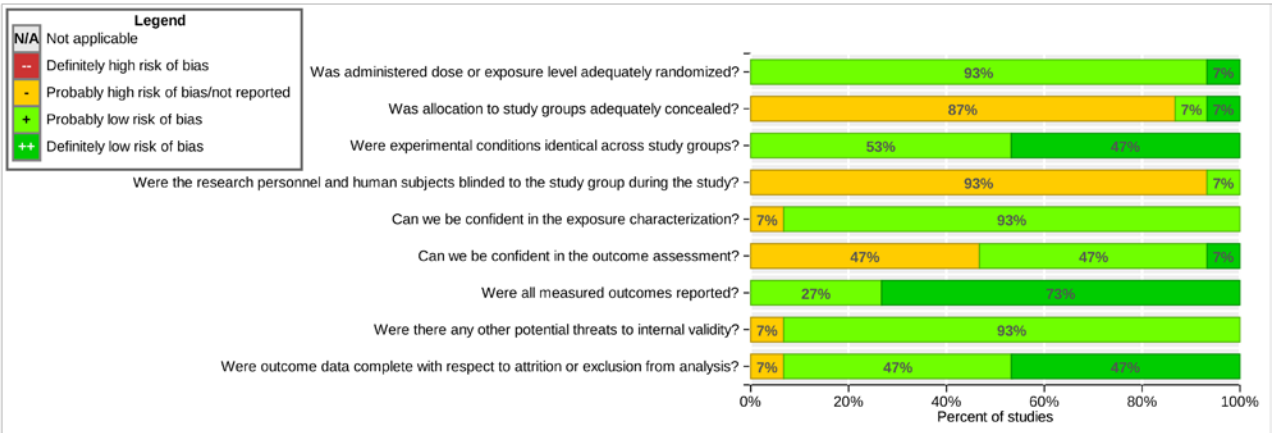
Interactive figure and additional study details in HAWC here.

Figure A3-21. Risk-of-bias Heatmap for Lower Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure



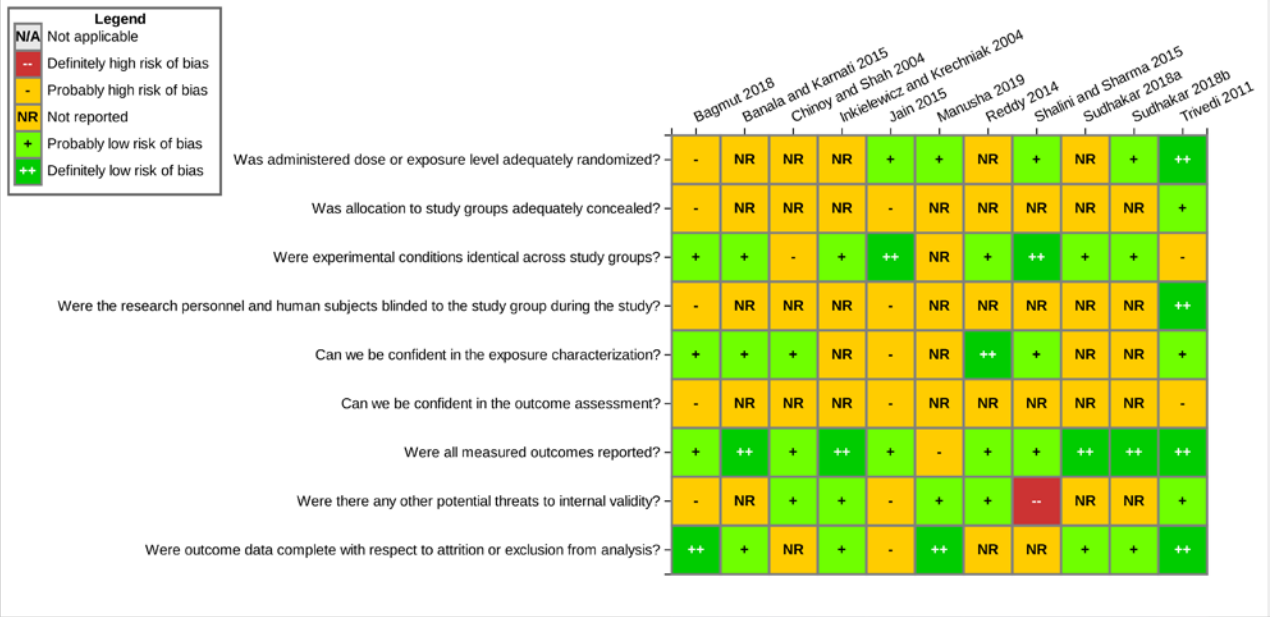
Interactive figure and additional study details in HAWC here.

Figure A3-22. Risk-of-bias Bar Chart for Lower Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure



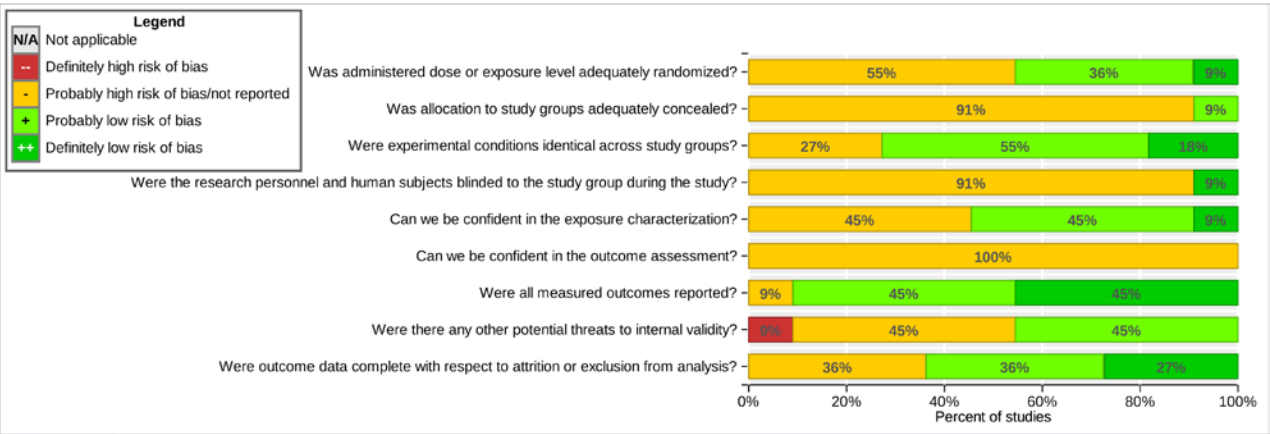
Interactive figure and additional study details in HAWC here.

Figure A3-23. Risk-of-bias Heatmap for Higher Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure



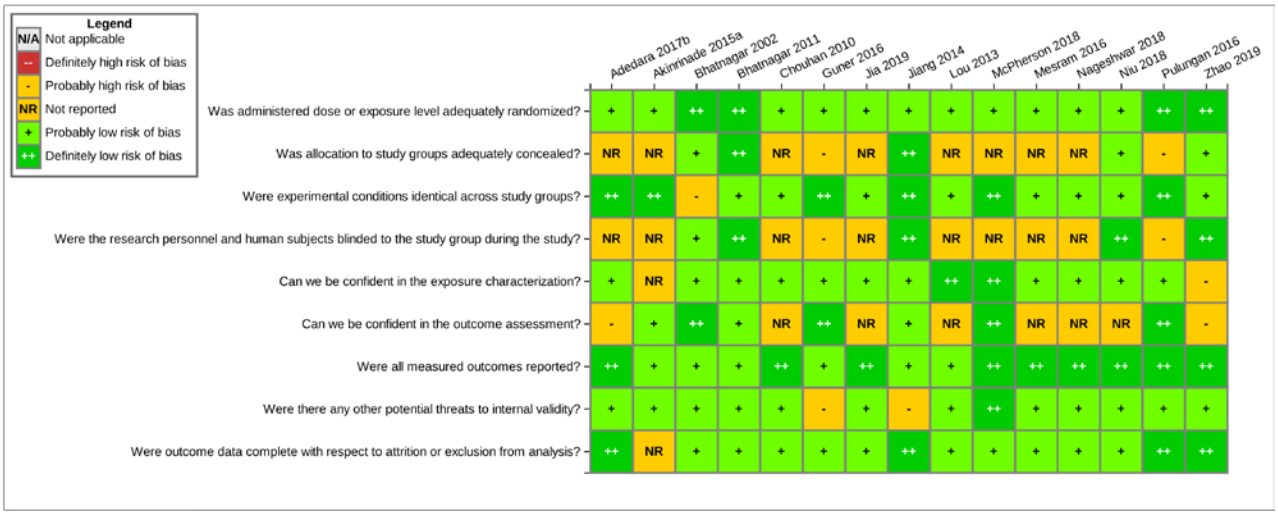
Interactive figure and additional study details in HAWC here.

Figure A3-24. Risk-of-bias Bar Chart for Higher Risk-of-bias Animal Oxidative Stress Studies Following Fluoride Exposure



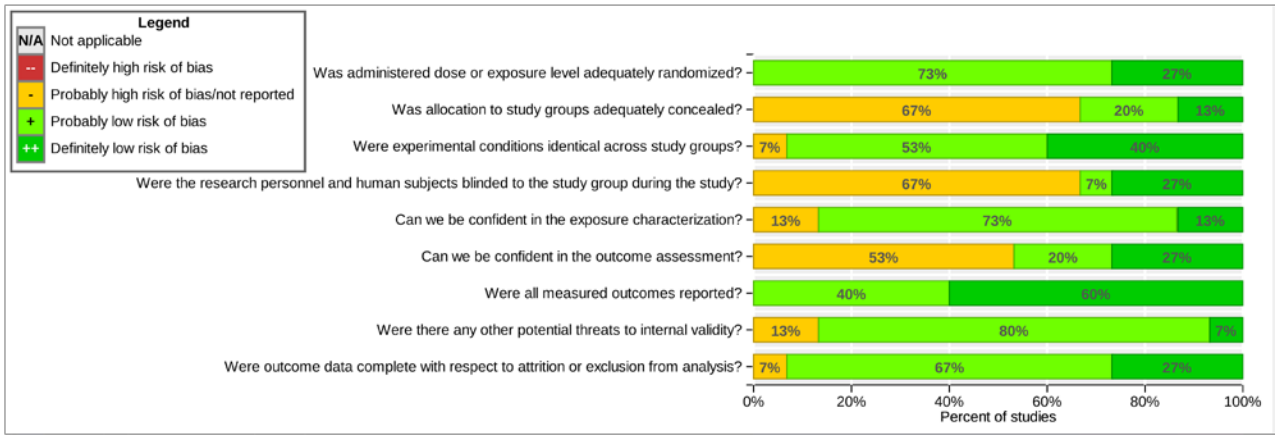
Interactive figure and additional study details in HAWC here.

Figure A3-25. Risk-of-bias Heatmap for Lower Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure



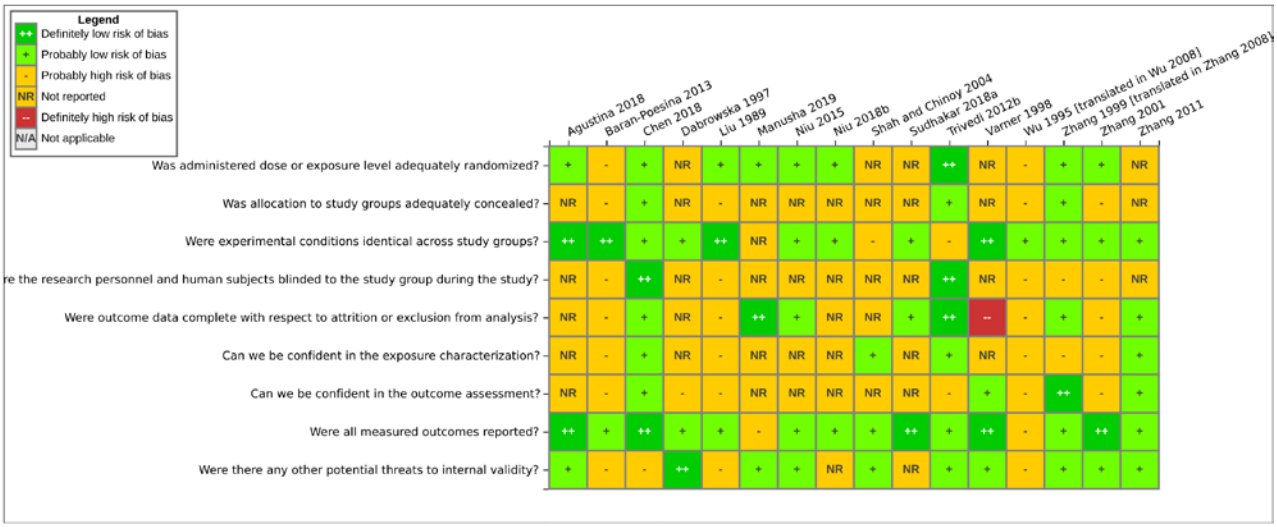
Interactive figure and additional study details in HAWC here.

Figure A3-26. Risk-of-bias Bar Chart for Lower Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure



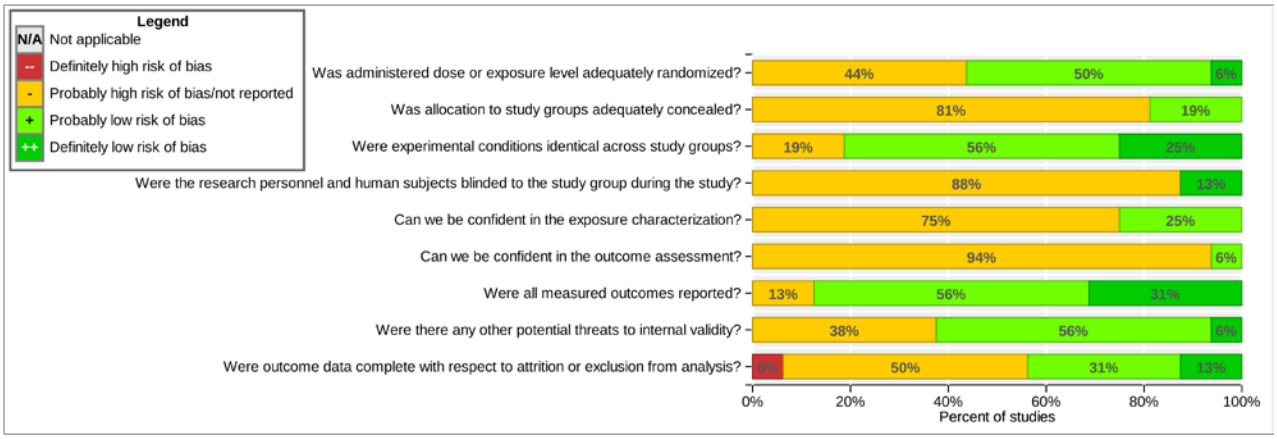
Interactive figure and additional study details in HAWC here.

Figure A3-27. Risk-of-bias Heatmap for Higher Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure



Interactive figure and additional study details in HAWC here.

Figure A3-28. Risk-of-bias Bar Chart for Higher Risk-of-bias Animal Histopathology Studies Following Fluoride Exposure



Interactive figure and additional study details in HAWC here.

Appendix 4. Details for Lower Risk-of-bias Studies

Barberio *et al.* (2017b)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Canadian Health Measures Survey (cycles 2 and 3) participants (children aged 3−12 years)
- ***Study area:*** general population of Canada
- ***Sample size:*** 2,221 children (1,120 from Cycle 2, 1,101 from Cycle 3)
- ***Data relevant to the review:*** Associations between learning disability or ADHD (Cycle 2 only) assessed by parent or child self-report and urinary fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant increase in adjusted OR for learning disability with unadjusted urinary fluoride (1.02; 95% CI: 1.00, 1.03) when Cycle 2 and 3 were combined. No significant associations with urinary fluoride when adjusted for creatinine and/or specific gravity. No significant association between urinary fluoride and ADHD.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* The comparison groups were selected from Cycles 2 and 3 of the Canadian Health Measures Survey. This is a nationally representative sample of residents living in 10 provinces, with clear exclusion criteria provided. Exclusion only represented about 4% of the target population (all Canadian residents 3−79 years old living in 10 provinces). A table of characteristics of the study population is provided.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the subjects were recruited from the same population using the same methods during the same time frame and exposure groups were similar.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study adjusted for sex, age (3–12 years old), household education, and household income adequacy. Variables to discern fluoride source, including drinking water and dental products, were also considered. Cycle 2 data also included adjustments for: 1) children for whom tap water (vs. bottled or other) was the primary source of drinking water at home or away from home and 2) children who had lived in his or her current home for 3 or more years. Confounders such as parental behavioral and mental health disorders, smoking, and nutrition were not discussed. Co-exposure to lead and arsenic are less likely an issue in this population and the lack of information is not considered to appreciably bias the results.
  - *Potentially important study-specific confounders:* All key confounders were considered in this study.
    - *Direction/magnitude of effect*: Not applicable.

- *Basis for rating:* Probably low risk of bias based on direct evidence that key confounders were addressed and indirect evidence that the methods used to collect the information were valid and reliable and that co-exposures were not an issue.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariate data were missing for less than 5% of all analyses, apart from household income; household income was reported for only 71–77% of participants and was imputed for the remainder.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Estimates of urinary fluoride (µmol/L) from spot urine were available for a subsample of respondents. Analysis was performed under standardized operating procedures at the Human Toxicology Laboratory of the Institut National de Santé Publique du Québec (accredited under ISO 17025). Fluoride content of urine samples was analyzed using an Orion pH meter with a fluoride ion-selective electrode with limits of detection of 20 µg/L (Cycle 2) and 10 µg/L (Cycle 3). Urinary dilution was addressed by using creatinine-adjusted levels as well as specific gravity-adjusted levels. In Cycle 3 only, estimates of the fluoride concentration of tap water samples collected from randomly selected households were available. The subsample of households selected for tap water sample collection corresponded to the person-level urine fluoride subsample. Analysis of the fluoride concentration of tap water was performed using a basic anion exchange chromatography procedure, with a limit of detection of 0.006 mg/L. QC methods were not addressed.
    - *Direction/magnitude of effect***:** There is not any specific direction or magnitude of bias expected. Urinary fluoride levels are reflective of a recent exposure. Having a single concurrent measurement may not be reflective of the exposure associated with the outcome, but if subjects lived in the same area throughout life the exposure may be an adequate representation. Although there is possible exposure misclassification it would be non-differential.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* The primary outcome variable, diagnosis of a learning disability by a health professional, was based on a single item from a household survey asked to all respondents: "Do you have a learning disability?". Answer options were: "yes", "no", "don't know", or the participant refused to answer. For Cycle 2, those who indicated having a learning disability were also asked what kind, with the answer options of: "ADD", "ADHD", "dyslexia", or "other". This question was omitted in Cycle 3 and the reason for omission is not described. Parents or guardians answered all questions for children aged 3–11 years, while children 12 years and older answered questions

themselves. The self-reporting of a learning disability did not appear to have been confirmed by medical records or a health professional. (- for methods based on self-report of diagnosis by a health care professional also in Cycle 3 no specific disabilities were described). Blinding was not a concern as spot urine samples were sent to a separate lab and self-reports would not have knowledge of their urine or tap water exposure level (+ for blinding). Overall rating = -.

- *Basis for rating:* Probably high risk of bias based on indirect evidence that the outcome was measured using an insensitive method in the study population.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods sections were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses were appropriate for the study.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include individual exposure measurements and the addressing of potential key confounders but was limited by the cross-sectional study design and insensitive outcome measures.

Bashash *et al.* (2017)

Study Details:

- ***Study design*:** Prospective cohort
- ***Population*:** Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) participants (pregnant mothers and their children aged 4 or 6–12 years).
- ***Study area:*** Mexico City, Mexico
- ***Sample size:*** 299 mother–child pairs, of whom 287 and 211 had data for the general cognitive index (GCI) and IQ analyses, respectively.
- ***Data relevant to the review:*** Adjusted and unadjusted associations between GCI or IQ scores and maternal or child's urinary fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant effect between maternal urinary fluoride and IQ score (adjusted β = −2.50; 95% CI: −4.12, −0.59) and GCI score (adjusted β = −3.15; 95% CI: −5.42, −0.87). No significant effects associated with children's urine.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Study participants were selected from two different cohorts from three hospitals in Mexico City that serve low-to-moderate income populations. One cohort was from an observational study of prenatal lead exposure and neurodevelopment outcomes and the other was from a randomized trial of the effect of calcium on maternal blood lead levels. The authors state that participants had no history of psychiatric disorders, high-risk pregnancies, gestational diabetes, illegal drug use, or continuous prescription drugs, but they do not include any information on smoking habits. Study populations appear to be similar, but there may be some differences because subjects were selected from two different cohorts that were recruited from slightly different time periods.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar despite the subjects coming from different original study populations where different methods were used for recruitment.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Data were collected via questionnaire on maternal age, education, marital status at first prenatal visit, birth order, birth weight, gestational age at delivery, maternal smoking, maternal IQ, and HOME scores. All models were adjusted for gestational age at birth, child's sex, birth weight, birth order, child's age at testing, maternal marital status, smoking history, age at delivery, maternal IQ, education, and cohort, with additional testing for children's urinary fluoride, mercury, lead, and calcium. Sensitivity analyses additionally adjusted for HOME score. Confounders not considered included BMI, iodine deficiency, arsenic, and maternal mental health and nutrition. Arsenic is assumed not to be a potential co-exposure in this population as the study authors did not discuss it as an issue but did discuss other co-exposures. Arsenic may have been included in the water quality control program in Mexico City.
  - *Potentially important study-specific confounders:* All key confounders were addressed.
    - *Direction/magnitude of effect***:** Not applicable.
  - *Basis for rating:* Probably low risk of bias based on direct evidence that key confounders including other potential co-exposures were addressed and indirect evidence that the methods used to collect the information were valid and reliable and that arsenic is not likely to be an issue in this study population.
- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although there was a large amount of attrition, the study authors clearly describe all reasons for attrition and also provide characteristics to compare those participants included to those excluded. There were some slight differences between those included and those excluded, but there is nothing to indicate that the attrition would potentially bias the results.

  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Urinary fluoride concentrations were determined in spot urine samples (2nd morning void) collected from mothers (during at least one trimester) and children ages 6–12 years. Fluoride content was measured using ion-selective electrode-based assays. QC methods were described including between laboratory correlations. All samples were measured in duplicate. Extreme outliers were excluded. Urinary dilution was addressed by using creatinine-adjusted levels.
    - *Direction/magnitude of effect***:** Not applicable.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Outcome was assessed using the McCarthy Scales of Children's Abilities (MSCA) in 4-year-old children (translated into Spanish) and the Wechsler Abbreviated Scale of Intelligence (WASI) in 6–12-year-olds. The WASI is a well-established test and the validity of both tests is well documented by the authors. Inter-examiner reliability was evaluated and reported with a correlation of 0.99 (++ for methods). The study report stated that psychologists were blind to the children's fluoride exposure (++ for blinding). Overall rating for methods and blinding = ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, outcome blindly assessed, and the prospective cohort study design.

Bashash *et al.* (2018)

Study Details:

- ***Study design*:** Prospective cohort
- ***Population*:** ELEMENT participants (pregnant mothers and their children aged 6–12 years)
- ***Study area:*** Mexico City, Mexico
- ***Sample size:*** 210 mother–child pairs
- ***Data relevant to the review:*** Associations between ADHD and other attention/impulsivity scores and maternal urinary fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant associations between maternal urinary fluoride and Conners' Rating Scales-Revised (CRS-R) scores, including Cognitive Problems + Inattention Index (adjusted β = 2.54; 95% CI: 0.44, 4.63), DSM-IV Inattention Index (adjusted β = 2.84; 95% CI: 0.84, 4.84), DSM-IV ADHD Total Index (adjusted β = 2.38; 95% CI: 0.42, 4.34), and ADHD Index (adjusted β = 2.47; 95% CI: 0.43, 4.50).

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Participants were a subset of mother-child dyads enrolled in various longitudinal birth cohort studies of the Early Life Exposure in Mexico to Environmental Toxicants (ELEMENT) project. Subjects were included from two of the four cohorts for which maternal urinary samples were available. Participants in cohort 2A were recruited between 1997 and 1999, and participants in cohort 3 were recruited from 2001 to 2003. Inclusion and exclusion criteria were applied consistently across the two cohorts. A table of subject characteristics was provided in the study and any differences were considered in the analysis. Study populations appear to be similar, but there may be some differences because subjects were selected from two different cohorts: one from an observational study on prenatal lead exposure and the other from a randomized trial on the effects of calcium on blood lead levels. In addition, they were recruited from slightly different time periods.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposed groups were similar, and any differences were taken into account in the analysis.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Questionnaires were used to collect information on maternal age, maternal education, history of smoking, and marital status during the first pregnancy visit. Child information at birth included birth weight, sex, birth order, and gestational age as calculated by the nurse. Mothers also responded to an SES questionnaire during the visit when the psychometric tests were administered. The Home Observation for Measurement of the Environment (HOME) score was evaluated in a subset of participants. Covariates were selected a priori. Models adjusted for maternal age at

delivery, years of education, marital status, smoking history, gestational age at birth, age at outcome assessment, child's sex, birth order, SES, cohort, and calcium intervention.

- *Potentially important study-specific confounders:* None identified, although this study did not specifically address arsenic or other co-exposures. Bashash *et al.* (2017) addressed potential co-exposure to lead and mercury but did not address arsenic. Arsenic was potentially addressed as part of the water quality program in Mexico City.
  - *Direction/magnitude of effect*: Not applicable.
- *Basis for rating:* Probably low risk of bias based on direct evidence that key confounders were addressed and indirect evidence that the methods used to collect the information were valid and reliable and that arsenic and other potential co-exposures are not likely to be an issue in this study population.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although there was a large amount of attrition from the original cohorts, it was unlikely related to outcome or exposure and there were very little missing data from those included in the study. Of the 231 mothers with a minimum of one maternal urine fluoride measurement and matching outcome identified for the project, only 17 were excluded based on incomplete demographic and outcome information.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Mothers provided at least one spot urine sample during pregnancy. As described in Bashash *et al.* (2017), urinary concentrations were determined on second morning void. Fluoride content was measured using ion-selective electrode-based assay. Bashash *et al.* (2017) describes QC methods. All samples were measured in duplicate and extreme outliers were excluded. Urinary dilution was addressed by using creatinine-adjusted levels.
    - *Direction/magnitude of effect*: N/A
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Behaviors associated with ADHD were assessed using the Spanish version of the Conners' Rating Scales-Revised, which has been validated for the evaluation of ADHD. Mothers completed the CRS-R at the same follow-up visit that the child completed the CPT-II tests. All tests were applied under the supervision of an experienced psychologist (++ for methods); however, a limitation of the study noted by the authors was only using parent reports and not teacher reports as they can vary from one another. Blinding was not reported, but it is unlikely that the mothers were aware of their urinary fluoride levels. Although mothers may have had knowledge that they were receiving fluoride through fluoridated salt or naturally occurring fluoride in their water, they would not have knowledge that this was relevant to the study purpose as

the ADHD tests were conducted for the original cohort (as was acknowledged by the study authors in the discussion). (++ for blinding). Overall rating = ++.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:*
        - *Statistical analyses:* Statistical analyses used were appropriate for the study.
        - *Other potential concerns:* None identified.
    - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, outcome blindly assessed, and the prospective cohort study design.

Choi *et al.* (2015)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** First grade children (ages 6–8 years)
- ***Study area:*** Mianning County in southern Sichuan, China
- ***Sample size:*** 51 first grade children
- ***Data relevant to the review:*** Associations between learning, memory, IQ (digit span for auditory span and working memory and block design for visual organization and reasoning components of WISC-IV only), visual motor ability, motor ability, and manual dexterity with continuous urine or drinking water fluoride levels. Study also had information based on dental fluorosis score.
- ***Reported association with fluoride exposure:*** Yes: Compared to the normal/questionable dental fluorosis, the moderate/severe dental fluorosis group was associated with significantly lower total (adjusted β = −4.28; 95% CI: −8.22, −0.33) and backward digit span scores (adjusted β = −2.13; 95% CI: −4.24, −0.02). Linear correlation between fluoride in urine (adjusted β = −1.67; 95% CI: −5.46, 2.12) and in drinking water (adjusted β = −1.39; 95% CI: −6.76, 3.98) with total digit span was observed but not significant. Other outcomes not significantly associated with fluoride exposure.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were selected during the same time frame using the same methods. Fifty-one first-grade children residing in Mianning County in southern Sichuan, China were included in this pilot study. It is not specified if the 51 children represented all the first-grade children from this area or if some refused to participate. Children who did not speak Chinese, were not students at the Primary School of Sunshui Village in Mianning County, or those with chronic or acute disease that might affect neurobehavioral function tests were excluded. Demographic characteristics are presented in Table 1 of the study, which indicates that subjects were similar. Potential confounders are adjusted for in the statistical analyses.
  - *Basis for Rating:* Definitely low risk of bias based on direct evidence that the exposure groups were similar and were recruited within the same time frame using the same methods with no evidence of differences in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The parents or guardians completed a questionnaire on demographic and personal characteristics of the children (sex, age at testing, parity, illnesses before age 3, and past medical history) and caretakers (age, parity, education and occupational histories, residential history, and household income). A 20-µL capillary blood sample was collected at the school by a Mianning County Center for Disease Control (CDC) health practitioner and tested for possible iron deficiency which could be used as a covariate of neurodevelopmental performance. Confounders that were not assessed include: maternal BMI, parental mental health, maternal smoking status, maternal reproductive factors, parental IQ, and HOME score. However, the study authors noted that confounding bias appeared to be limited due to the minimal diversity in the social characteristics of the subjects. The study authors indicated that CDC records documented that levels of other contaminants including arsenic and lead were very low in the area. Iodine differences were not specifically addressed, but there is no indication from the information provided that this might be a concern.
  - *Potentially important study-specific confounders:* All key confounders were considered in this study.
    - *Direction/magnitude of effect***:** Not applicable.
  - *Basis for rating:* Probably low risk of bias because there is direct evidence that the key confounders are taken into account and indirect evidence that co-exposure to arsenic is likely not an issue in this area and that methods used for collecting the information were valid and reliable.
- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The majority of results were reported for the 51 children stated to be included in the pilot study. In Table 5 of the study, the N for each dental fluorosis

category only totals 43, but the text indicates 8 children did not have a Dean Index because teeth had not erupted.

- o *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***
  - o *Rating:* Probably low risk of bias (+)
  - o *Summary:* The study used three different measurements of fluoride exposure: well water fluoride concentrations from the residence during pregnancy and onwards, fluoride concentrations from children's first morning urine samples, and degree of children's dental fluorosis. Fluoride concentrations in community well water were measured and recorded by Mianning County CDC; specific methods were not reported, but they likely used standard methods as they were conducted by the CDC and were likely the same as those used to measure the fluoride in urine. Migration of subjects was noted to be limited. Well water fluoride concentrations of the mother's residence during pregnancy and onward were used to characterize a child's lifetime exposure. To provide a measure of the accumulated body burden, each child was given a 330-mL (11.2-oz) bottle of Robust© distilled water (free from fluoride and other contaminants) to drink the night before the clinical examinations, after emptying the bladder and before bedtime. The first urine sample the following morning was collected at home, and the fluoride concentration was determined on a 5-mL sample using an ion-specific electrode at the Mianning CDC. There is no indication that urinary fluoride levels accounted for dilution nor was it clear that the method of administering water to the children and collection methods sufficiently controlled for differences in dilution. One of the investigators, a dentist, performed a blinded dental examination on each child's permanent teeth to rate the degree of dental fluorosis using the Dean Index. The Dean Index is a commonly used index in epidemiological studies and remains the gold standard in the dentistry armamentarium. The Index has the following classifications: normal, questionable, very mild, mild, moderate, and severe. Quality control (QC) procedures are not reported but were likely appropriate.
    - ▪ *Direction/magnitude of effect***:** Current levels were used to assess lifetime exposure. This is likely to be a non-differential exposure misclassification and direction of bias is unknown. Because subject migration appears to be limited, it is likely that the current fluoride levels are adequate reflections of past exposure. Dental fluorosis would be an indicator that exposure occurred in the past and there was a fair correlation between degree of dental fluorosis and current urine and water fluoride levels, with both increasing with increasing levels of dental fluorosis.
  - o *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.
- ***Outcome:***
  - o *Rating:* Probably low risk of bias (+)
  - o *Summary:* The study authors adopted culture-independent tests considered feasible for children aged 6 to 8 years. The Wide Range Assessment of Memory and Learning

(WRAML) was used for the assessment of memory and learning. Three subtests were also used. The Finger Windows subtest assesses sequential visual memory. The Design Memory subtest assesses the ability to reproduce designs from memory following a brief exposure. The Visual Learning subtest assesses the ability to learn the locations of pictured objects over repeated exposures. The Wechsler Intelligence Scale for Children-Revised (WISC-IV) included digit span for auditory span and working memory and block design for visual organization and reasoning. The grooved pegboard test assesses manual dexterity. The tests used have been validated on a western population. Although there is no information provided to indicate that they were validated on the study population, the study authors indicated that the tests were culture-independent (+ for methods). Blinding of the outcome assessors or steps to minimize potential bias was not reported. However, it is unlikely that the assessors had knowledge of the individual exposure as children all came from the same area, and water and urine levels were tested at the CDC. (+ for blinding). Overall = +.

    - *Basis for rating:* Probably low risk of bias based on indirect evidence that all outcomes were assessed blindly using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient details.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:*
        - *Statistical analyses:* Statistical analyses were appropriate. Data were log-transformed when necessary.
        - *Other potential concerns:* It should be noted that this study was a pilot study and, therefore, had a relatively small sample size (i.e., 51 children).
    - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in the confounding, exposure, and outcome risk-of-bias domains. Study strengths include individual fluoride measurements with blinding at outcome assessment likely. All key confounders and many other confounders were taken into account in the study design or analysis.

Cui *et al.* (2018)

Study Details:

- ***Study design*:** Cross-sectional

- ***Population*:** School children aged 7–12 years from four schools in two districts in China with different fluoride levels
- ***Study area:*** Jinghai and Dagang in Tianjin City, China
- ***Sample size:*** 323 school children
- ***Data relevant to the review:*** IQ scores by urine fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant correlation between IQ score and urinary fluoride (adjusted β = −2.47).

Risk of Bias:

- **Author contacts:**
  - Authors were contacted in June 2019 to obtain additional information for risk-of-bias evaluation.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Four schools were selected from the same district in China. The schools were selected based on levels of fluoride in the local drinking water and the degree of school cooperation. No details were provided on the number of schools in given areas or the difficulty in getting school cooperation. It was noted that the residents in the four areas had similar living habits, economic situations, and educational standards. Although authors do not provide the specific data to support this, fluoride levels and IQ scores were provided by different subject characteristics. The areas were classified as historically endemic fluorosis and non-fluorosis. Cluster sampling was used to select the grades in each school according to previously set child ages, and classroom was randomly selected with all students within a selected classroom included. Reasons for exclusion do not appear to be related to exposure or outcome.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and recruited within the same time frame using the same methods, with no evidence of differences in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The measurements of all covariates were obtained by structured questionnaires that were completed by children with the help of their parents. Confounders that were assessed include: child's gender, child's ethnicity, child's age, child's BMI, birth (normal vs abnormal), mother's age at delivery, mother's education, income per family member, mother's smoking/alcohol during pregnancy, family member smoking, environmental noise, iodine region (non-endemic vs iodine-excess-endemic area), factory within 30 m of residence, iodine salt, diet supplements, seafood/pickled food/tea consumption, surface water consumption, physical activity, stress, anger, anxiety/depression, trauma, having a cold 5 times a year, thyroid disease in relatives, mental retardation in relatives, and cancer in relatives. Covariates not considered include parity, maternal and paternal IQ, and quantity and quality of caregiving environment (e.g., HOME score). The authors report that there are no other environmentally toxic substances that may affect intelligence, such as high arsenic or iodine deficiency according to the Tianjin Centers for Disease Prevention and Control.

- *Potentially important study-specific confounders:* All key confounders were considered in this study.
  - *Direction/magnitude of effect*: Not applicable.
- *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key confounders are considered, methods for collecting the information are valid and reliable, and co-exposure to arsenic is likely not an issue in this area.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of the 400 children enrolled, 35 were excluded because they did not have informed consent signed by a guardian or they moved out of the area. Forty-two children were excluded because they did not have a DRD2 genotyping measurement. It is unclear if these children were from the same schools or if they were evenly distributed throughout the study area. It was also unclear if the excluded subjects were similar to those included in the study. In the study, some analyses had fewer than the 323 subjects, but this seems reasonable given the subgroups that were being evaluated.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although children were selected based on area fluoride levels, fluoride in the urine was used in the analysis. Urine was collected from each child the morning of enrollment and analyzed within a week. Fluoride levels were measured using an ion-selective electrode according to the China standard. A brief description of the method was provided, but no QC methods were reported. The study authors did not account for urinary dilution in the spot samples.
    - *Direction/magnitude of effect*: Not accounting for dilution could cause there to be some exposure misclassification. The direction and magnitude would depend on where the differences occurred.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* IQ was measured by professionals using the Combined Raven's Test-The Rural in China method, which is the appropriate test for the study population (++ for methods). Blinding or other methods to reduce bias were not reported. Although it is unlikely that the outcome assessor would have knowledge of the child's urine fluoride levels, there is potential that they would know if the child was from an endemic or non-endemic area if the IQ tests were conducted at the child's school, and there was no information provided on how the IQ tests were administered. Correspondence with the study author noted the cross-sectional nature of the study with outcome and exposure assessed at the same time making the outcome assessors blind to the exposure. However, there is still potential for knowledge of the area (+ for blinding).

    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient details.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:*
        - *Statistical analyses:* Statistical analyses were appropriate. Data were log-transformed when necessary.
        - *Other potential concerns:* None identified.
    - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key confounders were accounted for in the study design or analysis.

Cui *et al.* (2020)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** School children aged 7–12 years
- ***Study area:*** Tianjin City, China (one randomly selected school from each district based on iodine levels in the water), presumably was an expansion of the Cui *et al.* (2018)
- ***Sample size:*** 498 school children
- ***Data relevant to the review:*** IQ scores by urine fluoride levels.
- ***Reported association with fluoride exposure:*** No: No significant difference in IQ score based on a one-way ANOVA in the three different urinary fluoride categories.

Risk of Bias:

- **Author contacts:**
    - Authors were not contacted for the 2020 publication. Authors were contacted in June 2019 for additional information on the Cui *et al.* (2018) publication. Information obtained from that correspondence may have been used for additional information in the 2020 publication.
- ***Population selection:***
    - *Rating:* Probably low risk of bias (+)

- *Summary:* Subjects were recruited from 2014 to 2018. One school was selected from each district where water concentrations of water iodine were <10, 10–100, 100–150, 150–300 and >300 µg/L. In each school, classes were randomly sampled for the appropriate age group of 7–12 years old. A table of subject characteristics was provided by IQ. A total of 620 children were recruited, and 122 children who did not have complete information or enough blood sample were excluded. Reasons for exclusion do not appear to be related to exposure or outcome. The characteristics of the 498 included children are presented.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited within the same time frame using the same methods, with no evidence of differences in participation/response rates.

- ***Confounding:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* It was noted by the study authors that there were no other environmental poisons except water fluoride. Other studies also conducted in this area of China noted specifically that arsenic was not a concern. Iodine was addressed as that was one of the main points of the study. Twenty-one factors (provided in Table 1 of the study) were selected as confounders, and a homemade questionnaire of unspecified validity was used for obtaining the information. It was noted that child age, stress, and anger were significantly associated with IQ although it is unclear if these varied by fluoride level. However, Cui *et al.* (2018) indicates that stress and anger were not significantly associated with fluoride, and it is assumed that results would be similar for this study even though more children were included in the current study.
  - *Potentially important study-specific confounders:* Age (children 7–12 years old)
    - *Direction/magnitude of effect*: Age is a potential confounder for IQ, even in the narrow age range evaluated in this study. The direction of effects may depend on the number of children in each age group within the different urinary fluoride categories; however, these data were not provided. In general, there were fewer subjects ≤ 9 years of age (i.e., 111) compared to > 9 years of age (i.e., 387) with a significantly higher IQ in the ≤9-year-old age group. Therefore, if exposure were higher in the older subjects, this could bias away from the null.
  - *Basis for rating:* Probably high risk of bias because there is indirect evidence that age was not addressed as a confounder and it may be related to both IQ and exposure.
- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of the 620 (20%) children recruited, 122 were excluded due to incomplete information or inadequate blood sample. No information was provided to indicate if there were similarities or differences in the children included versus the children excluded, but exclusion is unlikely to be related to either outcome or exposure.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)

- *Summary:* Children's morning urine was collected with a clean polyethylene tube and fluoride was measured using a fluoride ion-selective electrode following Chinese standard WS/T 89-2015. A brief description was provided, but no QC methods were reported. The study authors do not account for urinary dilution in the spot samples.
  - *Direction/magnitude of effect*: Not accounting for dilution could cause there to be some exposure misclassification. The direction and magnitude would depend on where the differences occurred.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* IQ was measured using the Combined Raven's Test, which is an appropriate test for the study population (++ for methods). Blinding was not mentioned; however, the outcome assessors would not likely have knowledge of the child's urinary fluoride. Subjects appear to have been recruited based on iodine levels and it is, therefore, unlikely that there would be any knowledge of potential fluoride exposure. Correspondence with the study authors for the Cui *et al.* (2018) study also indicated that the outcome assessors would have been blind.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* The IQ scores are stated to be normally distributed, but there is no evidence that this was in fact tested. A t-test or one-way ANOVA was used to make comparisons between IQ and fluoride. The primary focus of the study was to evaluate associations between thyroid hormones or dopamine levels on IQ (not between fluoride and IQ). It should also be noted that regardless of the analysis conducted, there is no adjustment for school and no accounting for the clustering of children from the same school.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements, but the

study is limited by the cross-sectional study design, lack of accounting for urine dilution, and by not addressing age as a potential confounder.

Ding *et al.* (2011)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Elementary school children aged 7–14 years old
- ***Study area:*** Hulunbuir City, Inner Mongolia, China
- ***Sample size:*** 331 school children
- ***Data relevant to the review:*** IQ mean difference based on 10 categories of urine fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant association between urinary fluoride and IQ score (each increase in urinary fluoride of 1 mg/L was associated with an IQ score 0.59 points lower; 95% CI: −1.09, −0.08).

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study randomly selected 340 7–14-year-olds from four nearby elementary schools in Hulunbuir. Authors stated that the four elementary schools appeared to be very similar in teaching quality. The study authors noted that they followed the principles of matching social and natural factors like economic situation, educational standards, and geological environments as much as possible; however, how this was done is unclear and no table of study subject characteristics by group was provided.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited within the same time frame using the same methods, with no evidence of differences in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* It was noted that none of the four sites had other potential neurotoxins including arsenic in their drinking water. While they did not provide the specifics, they did provide a reference. In addition, iodine deficiency was noted as not being issue in any of the four areas. Age was the only confounder adjusted in the model. While dental fluorosis severity by % female was reported, not enough data were provided to determine if it was a confounder that should have been considered in the regression. The study authors note that future studies will include covariates such as parents' educational attainment, mother's age at delivery, and household income.
  - *Potentially important study-specific confounders:* Gender
    - *Direction/magnitude of effect*: There is not enough information to determine if there is an effect from gender. There were some differences in dental fluorosis

level by gender, but it is unclear how this might impact the results or if the distribution of gender differed by age.

- *Basis for rating:* Probably high risk of bias based on indirect evidence that there were differences in gender that were not considered in the study design or analyses.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data were relatively complete (i.e., <5% loss). Of the 340 subjects selected for inclusion, 5 were excluded because they lived in the area for less than a year with an additional 4 not consenting to participate.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analysis was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Spot urine samples were collected and measured using China CDC standards. All samples were analyzed twice using a fluoride ion-selective electrode. Recovery rates were specified as 95–105% with an LOD of 0.05 mg/L. Water samples were collected from small-scale central water supply systems and tube wells with handy pumps and were processed using standard methods similar to the urine samples. Quality assurance validation was reported. A blind professional examiner evaluated the children for dental fluorosis using the Dean's Index. All urine and water samples were above the LOD. Urine levels were the primary exposure measure used in the analysis. The study authors did not account for urinary dilution in the spot samples. The mean urine fluoride concentration was correlated with the dental fluorosis levels.
    - *Direction/magnitude of effect***:** Spot urine samples that did not account for dilution could have exposure misclassification. The misclassification is likely non-differential and potential direction of bias is unknown.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* IQ was determined using the Combined Raven's Test-The Rural in China (CRT-RC3) (++ for methods). Although blinding was not reported, it is unlikely that the IQ assessors had knowledge of the children's urine levels or even of the water levels from the four sites as these were sent to a separate lab for testing (+ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient details.

  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses were reasonable (ANOVA and multiple linear regression), but consideration of homogeneity of variance was not reported. The NASEM (2020) review pointed out a potential concern for the lack of accounting for clustering at the school-level since children were selected from four elementary schools. However, as pointed out in the ***Selection*** domain, the authors stated that they followed the principles of matching social and natural factors like economic situation, educational standards, and geological environments as much as possible and that the four elementary schools appeared to be very similar in teaching quality.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and that there were no other potential threats to risk of bias.
- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements, but the study is limited by the cross-sectional study design, lack of accounting for urine dilution, and by not addressing gender as a potential confounder.

Green *et al.* (2019)

Study Details:

- ***Study design*:** Prospective cohort
- ***Population*:** Maternal-Infant Research on Environmental Chemicals (MIREC) participants (pregnant mothers and their children aged 3–4 years)
- ***Study area:*** 10 cities, Canada
- ***Sample size:*** 512 mother−child pairs (238 from non-fluoridated areas, 162 from fluoridated areas; 264 females, 248 males)
- ***Data relevant to the review:*** Adjusted linear regression models evaluating associations between IQ in both genders together and separate with maternal urinary fluoride across all three trimesters or estimated maternal fluoride intake.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower full-scale IQ with 1-mg/L increases in maternal urinary fluoride in boys (adjusted β = −4.49), but not girls (adjusted β = 2.40) and not in sexes combine (adjusted β = −1.95); significantly lower full-scale IQ with 1-mg increases in maternal intake in sexes combined (adjusted β = −3.66 [no sex interaction]); significantly lower full-scale IQ with 1-mg/L increases in drinking water fluoride in sexes combined (adjusted β = −5.29 [no sex interaction]).

Risk of Bias:

- **Author contacts:**

- Authors were contacted in June 2019 for additional information for the risk if bias evaluation.

- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Pregnant women were recruited from the same population, during the same timeframe, and using the same methods as the MIREC program. Methods were reported in detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were similar and were recruited with the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study considered several possible covariates including maternal age, pre-pregnancy BMI, marriage status, birth country, race, maternal education, employment, income, HOME score, smoking during pregnancy, secondhand smoke in the home, alcohol consumption during pregnancy, parity, child's gender, child's age at testing, gestational age, birth weight, time of void, and time since last void. The study also conducted secondary analyses to test for lead, mercury, arsenic, and PFOA. There is no indication of any other potential co-exposures in this study population. Iodine deficiency or excess could not be assessed but is not expected to differentially occur. The study was not able to assess parental IQ or mental health disorders. Methods used to obtain the information included questionnaires and laboratory tests.
  - *Potentially important study-specific confounders:* All key confounders were addressed.
    - *Direction/magnitude of effect:* Not applicable.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and direct evidence that key confounders including potential co-exposures were addressed.
- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of the 610 recruited children, 601 (98.5%) completed testing. Of the 601 mother-child pairs, 512 (85.2%) had all three maternal urine samples and complete covariate data, and 400 (66.6%) had data available to estimate fluoride intake.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Spot urine samples from all three trimesters of pregnancy were evaluated using appropriate methods, and results were adjusted for creatinine and specific gravity. Fluoride intake was estimated based on fluoride water levels and information on consumption of tap water and other water-based beverages (e.g., tea, coffee) was obtained via questionnaire.
    - *Direction/magnitude of effect***:** There is not any specific direction or magnitude of bias expected. Urinary fluoride levels are reflective of a recent exposure.

Having measurements from all three trimesters of pregnancy provides a better representation of actual exposure than a single measurement although the potential for missed high exposure is possible. However, the possibility of the occurrence of missed high exposure would be similar in all females and would be non-differential. For the fluoride intake, exposure was based on the fluoride levels in the water at the residence. If women worked outside the home and the majority of intake occurred from areas outside the home (and were different from levels in the home), there is potential to bias toward the null.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The Wechsler Preschool and Primary Scale of Intelligence was normalized for ages 2.5–<4.0 and child sex using the U.S population-based norms. Blinding was not reported, but it is unlikely that the outcome assessors had knowledge of the maternal fluoride level or were aware if the city had fluoridated water.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes were reported.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Linear regression was performed. Regression diagnostics were used to test assumptions for linearity, normality, and homogeneity. There were two potential influential observations (based on Cook's distance), and sensitivity analyses re-estimated the models without these two variables. Further sensitivity analyses were also conducted. Although city was accounted for as a covariate in the regression models, the city effect should have been a random effect rather than a fixed effect to account for potential clustering of results within each city. Although the analysis used individual-level exposure rather than city-level exposure, if the exposure levels within a city are highly correlated (which might be expected given that some cities were fully on fluoridated water and others were not), the fixed-effect model could still produce biased estimates. However, correspondence with the study authors indicated that a supplemental analysis using a random effects multi-level model showed similar results to the main model.
    - *Other potential concerns:* None identified.

  - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, prospective cohort design, and addressing potential key confounders.

Li *et al.* (2004) [translated in Li *et al.* 2008a]

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Full term, normal neonates 24–72 hours old from healthy mothers
- ***Study area:*** Zhaozhou County, Heilongiang Province, China
- ***Sample size:*** 91 neonates (46 males and 45 females)
- ***Data relevant to the review:*** Comparison of neurobehavioral capacity between children in the high-fluoride area compared to the control area.
- ***Reported association with fluoride exposure:*** Yes: Significant differences in neurobehavioral assessment total scores between high-fluoride (36.48 ± 1.09) and control groups (38.28 ± 1.10); significant differences in total neurobehavioral capacity scores as measured by non-biological visual orientation reaction and biological visual and auditory orientation reaction between the two groups (11.34 ± 0.56 in controls compared to 10.05 ± 0.94 in high-fluoride group).

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There is indirect evidence that the exposure groups were similar. They were recruited during the same time frame using the same methods. From 2002 to 2003, 273 neonates were born in a hospital in Zhaozhou County, China. Ninety-one of 273 full-term neonates (46 males, 45 females) were randomly selected. Mothers ranged in age from 20 to 31 years, met multiple health criteria, and had not changed residence during pregnancy. Authors report that the two study groups are located in the same area with similar climate, living habits, economic and nutritional conditions, and cultural backgrounds, but do not provide these data in the manuscript. There is no statistically significant difference in the mode of delivery, birth weight, infant length, or sex. Subjects were separated into exposure groups after random selection.
  - *Basis for Rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited within the same time frame using the same methods with no evidence of differences in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)

- *Summary:* No confounders were specifically controlled in the analysis. The study authors note similarities in characteristics in the two populations (i.e., living habits, economic and nutritional conditions, and cultural backgrounds), but do not provide these data nor do they indicate what specific characteristics were considered. There were no significant differences in infant gender, birth method, gestational age, or infant weight and length. All tests were conducted when children were 1–3 days old. No potential co-exposures were discussed. Although arsenic is considered a potential issue in China, water quality maps indicate that there is a 25–50% probability that the drinking water in that area exceeds the WHO guideline for arsenic of 10 µg/L.
- *Potentially important study-specific confounders:* Key confounders, including child's age, child's gender, and measures of socioeconomic status (SES), were similar between exposure groups; however, arsenic was not taken into account. Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on water quality maps, arsenic does not appear to be an issue in Zhaozhou County of the Heilongjiang Province. Iodine deficiencies are not mentioned.
  - *Direction/magnitude of effect*: The presence of arsenic would potentially bias away from the null if it were present with fluoride. Deficiencies in iodine would potentially bias away from the null if it were present in areas of higher fluoride, but toward the null if it were present in areas of lower fluoride.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the key confounders are taken into account, co-exposure to arsenic is likely not an issue in this area, and methods used for collecting the information are valid and reliable.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Although authors did not discuss why they only randomly selected 91 of the 273 neonates available, results were available for all 91 subjects.
  - *Basis for rating:* Definitely low risk of bias based on results being available for all subjects.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were split into control and high-fluoride groups based on fluoride levels in their places of residence. Although the levels were provided (1.7–6.0 mg/L for the high-fluoride group compared to 0.5–1.0 mg/L for the control group), it was not reported how or when these levels were measured. Urine was collected when women were hospitalized, but before labor began. Urine samples were sent to a specific lab for measurement using fluoride ion-selective electrode. It was noted that this procedure strictly followed the internal controls of the laboratory indicating quality control. Level of detection (LOD) was not provided. Urinary fluoride levels were significantly higher in the high-fluoride mothers (3.58 ±1.47 mg/L) compared to the control-group mothers (1.74±0.96 mg/L). There was indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure. Although results were mainly based on exposure area, they were supported by urine data making exposure misclassification less of a concern.

        - *Direction/magnitude of effect*: There is high variability in both water fluoride and urine fluoride in the subjects from the high-exposure area. Although there is no overlap in the water fluoride levels in the exposure areas, there is some overlap in the urine concentrations in the mothers from the two areas. This may reflect the single measurement and pose no specific bias, or it could indicate that some mothers in the high-fluoride area have lower fluoride exposure, which could bias the results toward the null.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measure exposure.
- ***Outcome:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* A standard neonatal behavioral neurological assessment method was carried out by professionals in the pediatric department working in neonatal section trained specifically for these programs and passing the training exams. (+ for methods). The examinations were carried out 1 to 3 days after delivery. Because urine samples were collected on the day of delivery and sent to a separate laboratory, it is likely that the outcome assessors were blind. Although the subjects were separated by fluoride exposure area, it is not likely that the professionals were aware of the exposure as the tests were conducted in the hospital (+ for blinding).
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was assessed blindly using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* The study authors reported numerous endpoints in sufficient detail; however, because they did not provide a list of endpoints tested there is no direct evidence that all were reported.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that all the study's measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:*
        - *Statistical analyses:* Statistical analyses were reasonable (t-test), but consideration of homogeneity of variance was not reported. This was a translated study.
        - *Other potential concerns:* It should be noted that, although the study states that subjects were randomly selected, it is unclear why only 91 subjects were included and if they were randomly selected to obtain equal groups in the high-fluoride and control groups.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that statistical analyses were appropriate and that there were no other potential threats to risk of bias.
- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in the confounding, exposure, and outcome risk-of-bias domains. Study strengths include individual fluoride measurements to support the differences in the two areas. Tests were noted

to be conducted at the hospital providing indirect evidence that blinding was not a concern during the outcome evaluation. Although there was some potential for bias due to the lack of accounting for arsenic or iodine deficiencies, co-exposure to arsenic is likely not a major concern according to groundwater quality maps.

Riddell *et al.* (2019)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Canadian Health Measures Survey (cycles 2 and 3) participants (children aged 6−17 years)
- ***Study area:*** general population, Canada
- ***Sample size:*** 3,745 children
- ***Data relevant to the review:*** Adjusted odds ratios for ADHD and attention symptoms per 1 unit increase in urinary fluoride, by water fluoride in the tap water, or community fluoridation status.
- ***Reported association with fluoride exposure:*** Yes: Significantly increased odds of ADHD diagnosis (adjusted OR = 6.10; 95% CI: 1.60, 22.8) or hyperactivity/inattentive symptoms (adjusted beta = 0.31; 95% CI: 0.04, 0.58) per 1-mg/L increase in tap water fluoride. Also, a significant association between ADHD diagnosis (adjusted OR = 1.21; 95% CI: 1.03, 1.42) or hyperactivity/inattentive symptoms (adjusted beta = 0.11; 95% CI: 0.02, 0.58) and community water fluoridation status. No significant associations with urinary fluoride levels.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were part of Cycles 2 and 3 of the Canadian Health Measures Survey. This is a nationally representative sample of residents living in 10 provinces. Specific inclusion criteria were provided. This study was restricted to children 6–17 years of age with different fluoride measurements that consisted of three participant samples. One of the samples was only available in Cycle 3.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were similar and were recruited with the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariates included in all models included child's age at test, child's sex, ethnicity, BMI, parents' education, total household income, exposure to cigarette smoke inside the home, and log-transformed concurrent blood lead levels. Confounders such as parental behavioral and mental health disorders, quantity and quality of caregiving environment, and co-exposure to arsenic were not discussed. Rationale for selection of covariates was based on relationship to ADHD diagnosis and to fluoride metabolism based on literature review and consultation with an ADHD expert. There is no

information of the source if data for covariates, but this is likely the questionnaires from the Canadian Health Measures Survey, which are considered standardized and validated.

- o *Potentially important study-specific confounders:* All key confounders were considered in this study.
  - *Direction/magnitude of effect***:** Not applicable.
- o *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key confounders are taken into account, co-exposure to arsenic is likely not an issue in this area, and methods used for collecting the information were valid and reliable.

- ***Attrition:***
  - o *Rating:* Probably low risk of bias (+)
  - o *Summary:* There is no information indicating that there were any data excluded due to missing covariates. All exclusions of children were described and reasonable (i.e., drinking bottled water when considered city fluoridation as a measure of fluoride exposure). Outliers were stated to be excluded, but methods for determining this were provided and it was noted that the outliers were 0.27% of the values.
  - o *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - o *Rating:* Probably low risk of bias (+)
  - o *Summary:* **Urinary Fluoride:** Spot urine samples were collected under normal non-fasting conditions and analyzed using an Orion pH meter with a fluoride ion-selective electrode after being diluted with an ionic adjustment buffer. Analysis was performed at the Human Toxicology Laboratory of the Institut National de Sante Publique du Quebec. The precision and accuracy of the fluoride analyses, including quality control and quality assurance, were described by Health Canada (2015). The limits of detection were 20 µg/L for Cycle 2 and 10 µg/L for Cycle 3 with no values below detection. Fluoride levels were adjusted for specific gravity.

    Water Fluoride in Tap water: Tap water was collected at the subjects' homes in Cycle 3 only. Samples were analyzed for fluoride concentrations using anion exchange chromatography procedure with a LOD of 0.006 mg/L. Values below the LOD were imputed with LOD/square root 2. Of the 980 samples, 150 (16%) were below detection.
    Chlorinate Water Fluoride status: This was determined by viewing reports on each city's website or contacting the water treatment plant (provided in supplemental material). Children were excluded if they drank bottled water, had a well, had a home filtration system, lived in the current residence for 2 years or less, or lived in an area with mixed city fluoridation.
    - *Direction/magnitude of effect***:** There is not any specific direction or magnitude of bias expected. Urinary fluoride levels are reflective of a recent exposure, but the study authors adjusted to account for dilution. The possibility of exposure misclassification would be similar in all subjects and would be non-differential. There is less potential for exposure misclassification in regard to tap water or chlorinated water fluoride status as children who drank bottled water were

excluded and children who had a home filtration system were excluded from the chlorinated water status.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:*
    Strengths and Difficulties Questionnaire (SDQ): The questionnaire was administered to youths under 18 years. Children aged 6–11 years had SDQ ratings provided by parents and guardians, but youths aged 12−17 years completed the questionnaire themselves. Tests consist of 25 items with a 3-point scale. Items were divided into five subscales: emotional problems, conduct problems, hyperactivity-inattention, peer problems, and prosocial behavior. The current study only used the hyperactivity-inattention subscale. Validation of this method was not reported (- for methods).
    ADHD: Ninety percent of youths with ADHD are diagnosed after age 6 years. For children aged 6–11 years, ADHD diagnosis was provided by parents, but youths age 12−17 years completed the questionnaire themselves. Cycle 2 asked "Do you have a learning disability?" and if yes asked to specify the type (4 options available and described). In Cycle 3, parents were asked directly whether they had ADHD, and children 12 years and older were asked if they had a physician diagnosis of ADHD and, if so, what subtype. (- for methods because different methods were used and only the children 12 years and older in cycle 3 were asked specifically about doctor diagnosis). Both were measured in both cycles. Blinding is not likely an issue as subjects would not have knowledge of the urine or tap water fluoride levels. However, they would likely have knowledge of the city.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that the outcome was assessed using insensitive methods that varied based subject age.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods sections were reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Logistic regression was used for ADHD results. Box-Tidewell tests were used to check the linearity of the relationship with the continuous predictors. Linear regression was used for the SDQ scores using Huber-White standard errors. All regressions were tested for interactions between age and fluoride and sex and fluoride. Sensitivity analyses were conducted to test the different cycles.
    - *Other potential concerns:* None identified.

- *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.

- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include individual exposure measurements and the addressing of potential key confounders but was limited by the cross-sectional study design and insensitive outcome measures.

Rocha-Amador *et al.* (2007)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 6–10 years
- ***Study area:*** Moctezuma (low fluoride, low arsenic) and Salitral (high fluoride, high arsenic) of San Luis Potosí State and 5 de Febrero (high fluoride, high arsenic) of Durango State, Mexico
- ***Sample size:*** 132 children
- ***Data relevant to the review:*** Associations between full-scale IQ, performance IQ, verbal IQ and child's urine or water fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant associations between fluoride and IQ scores (full-scale IQ adjusted βs of –10.2 with water and −16.9 with urine; CIs not reported); arsenic also present, but the effect was smaller (full-scale IQ adjusted βs of –6.15 with water and –5.72 with urine; CIs not reported).

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* All children in 1st through 3rd grades in three rural areas in Mexico (n = 480) were screened for study eligibility including age, time at residence, and address. Authors report that the three selected communities were similar in population and general demographic characteristics. Children who had lived in the area since birth and were 6–10 years old were eligible to participate (n = 308). Of the 308 children, 155 were randomly selected and the response rate was 85%, but participation was not reported by area. It was noted, however, that no significant differences in age, gender, or time of residence were observed between participants and non-participants. Timeframe for selection was not mentioned but appears to be similar. Sociodemographic characteristics of subjects was provided in Table 1 of the study. There was a significant difference in SES and transferrin saturation, but these were taken into account in the analysis.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the populations were similar and differences were noted and addressed in the analysis.
- ***Confounding:***

- *Rating:* Probably low risk of bias (+)
- *Summary:* The study design or analysis accounted for child's age, sex, SES, transferrin saturation, weight, height, blood lead levels, and mother's education. Arsenic levels were highly correlated with fluoride levels and it was stated that each was tested alone, and arsenic was found to have less of an effect. The authors noted in the methods that they tested for an interaction between arsenic and fluoride. Smoking was not addressed and methods for measuring many of the confounders were not reported.
- *Potentially important study-specific confounders:* Arsenic
  - *Direction/magnitude of effect*: Presence of arsenic, which also demonstrated an association, would bias away from the null.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and direct evidence that key confounders were addressed.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Of 155 children randomly selected for study participation, 85% responded to enroll. According to the authors, there were no significant differences in age, gender, or time of residence between responders and non-responders. However, no data are provided to support this, and no breakdown of responders/non-responders by region is provided. Data were provided for the 132 children agreeing to participate.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Urine samples were collected on the same day as psychological evaluations and were analyzed for fluoride according to NIOSH Method 8308 (Fluoride in Urine). For QC, a reference standard was also used (NIST SRM 2671a). Urine samples were also analyzed for arsenic by using the Atomic Absorption Spectrophotometer with hydride system (Perkin-Elmer, model AAnalyst 100, Wellesley, United States) and used a reference standard for QC. Levels were adjusted for urinary creatinine levels to account for dilution in the spot samples. Tap water samples were collected from each child's home on the day of biological monitoring. Fluoride was measured with a sensitive, specific ion electrode. Detailed methods are provided including internal quality controls. It was noted that in the high fluoride group it was common to drink bottled water low in fluoride and to only use the tap water for cooking; therefore, urine was considered the most appropriate measure of exposure. Only children who had lived at the same residence since birth were included.
    - *Direction/magnitude of effect*: Not applicable.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)

  - *Summary:* Neuropsychological profiles were assessed through the WISC-RM (revised for Mexico). This is a well-established test appropriately adjusted for the study population. However, no additional validation is provided (+ for methods). The study report stated that the test assessors were masked to both arsenic and fluoride water levels (++ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* It was reported that an interaction between fluoride and arsenic was measured, but it was only noted in the discussion that the study design precluded testing statistical interaction between fluoride and arsenic. This provides indirect evidence of selective reporting.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that there was selective reporting.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
  - *Statistical analyses:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and outcomes blindly assessed, but it is limited by the cross-sectional study design and not being able to completely rule out the influence of arsenic in the results.

Rocha-Amador *et al.* (2009)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 6–11 years
- ***Study area:*** Durango, Mexico
- ***Sample size:*** 80 children
- ***Data relevant to the review:*** Associations between visuospatial organization and visual memory (using the Rey-Osterrieth Complex Figure Test, children's version) and urinary fluoride levels in the children.
- ***Reported association with fluoride exposure:*** Yes: Significant correlation between urinary fluoride and visuospatial organization (r = −0.29) and visual memory (r = −0.27) scores. No significant correlations with arsenic.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were from the same population and were recruited during the same time frame using the same methods. Although this study compared three sites with antecedents of environmental pollution to mixtures of either F–As, Pb–As, or DDT–PCBs, authors evaluated each contaminant separately. The only area of interest is the area with F and As contamination. The area in Durango state (5 de Febrero) where drinking water is polluted naturally with F and As at levels exceeding 6 and 19 times, respectively, the World Health Organization (WHO) limits (WHO 2008). Children attending public schools were screened through personal interviews for study eligibility. Inclusion criteria were children between 6 and 11 years old, living in the study area since birth, and whose parents signed the agreement to participate. Children with a neurological disease diagnosed by a physician and reported by the mother were excluded from the study. The final sample for the F–As was 80. Participation rates were not reported. Selected demographic characteristics are presented in Table 1 of the study.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the populations were similar and recruited during the same time frame using the same methods.
- ***Confounding:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* Confounding factors in children tested in the analysis included blood lead (PbB), age, gender, and height-for-age z-scores; only age had significant associations and was included in the final analysis. Arsenic was also assessed and analyzed separately from fluoride. Arsenic in urine was analyzed by atomic absorption spectrophotometer coupled to a hydride system (Perkin-Elmer model AAnalyst 100). Although the model did not adjust for arsenic, arsenic in the F–As group was not associated with either endpoint. PbB was analyzed with a Perkin-Elmer 3110 atomic absorption spectrophotometer using a graphite furnace. Authors note that the mean blood lead level in the F–As study area was 5.2 µg/dL and 8% of the children had values above the reference value of 10 µg/dL. PbB was stated not to affect results and was not included in the final analysis. Other confounding data were obtained during the study interview. Father's education was provided and, in the F–As group, was stated to range from 0–16 years, but this was not considered. Maternal education, smoking, and SES were also not considered. The authors provide an SES score of 5.9 ± 1.4 for the 5 de Febrero region (the fluoride region). It is not clear if this would vary by fluoride or arsenic levels.
  - *Potentially important study-specific confounders:* SES.
    - *Direction/magnitude of effect***:** There are insufficient data to determine the magnitude or direction of effect. If there is an association between fluoride exposure and SES, the direction of effect would depend on the association.

- *Basis for rating:* Probably high risk of bias based on indirect evidence that the SES was not accounted for in the study design or analysis and may have varied by fluoride levels.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are complete. All 80 participants stated to be the final sample for the site of interest (F–As) were included in all analyses.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Fluoride in urine (FU) was analyzed according to method 8308 (''fluoride in urine'') from the National Institute of Occupational Safety and Health (NIOSH 1984) with a sensitive specific ion electrode. As a quality control check, reference standard ''fluoride in freeze dried urine'' (NIST SRM 2671a) was analyzed. The accuracy was 97.0 +/- 6.0%. Levels of FU and AsU were adjusted for urinary creatinine, which was analyzed by a colorimetric method (Bayer Diagnostic Kit, Sera-Pak1 Plus). However, details on the collection methods were not reported.
    - *Direction/magnitude of effect***:** Spot urine samples in a small sample size (i.e., 80 children) may have some exposure misclassification. Adjusting for dilution reduces the potential for misclassification based on differences in dilution. Exposure misclassification would be non-differential.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* IQ is assessed through the Rey-Osterrieth Complex Figure Test (ROCF). This is a less well-established method, although the authors provide citations suggesting it has been validated and standardized for the Mexican population (+ for methods). According to the study report, the neuropsychologist who administered the test was blinded to all exposure types and levels. (++ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study.
    - *Other potential concerns:* None identified.

  - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements and outcomes blindly assessed, but it is limited by the cross-sectional study design, lack of addressing SES in the study population, co-exposure with arsenic, and use of spot samples in a small population.

Saxena *et al.* (2012)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 12 years
- ***Study area:*** Madhya Pradesh, India
- ***Sample size:*** 170 children
- ***Data relevant to the review:*** Mean IQ grade (not standard scores) by water fluoride quartiles or continuous or by continuous urinary fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant correlation between water (r = 0.534; p = 0.000) and urinary (r = 0.542; p = 0.000) fluoride levels and IQ score.

Risk of Bias:

- **Author contacts:**
  - Authors were contacted in August of 2017 to obtain additional information for risk-of-bias evaluation.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There was indirect evidence that subjects were similar and were recruited using the same methods during the same time frame. The study participants were selected from a stratified cluster of geographic areas based on fluoride concentration in groundwater. According to the authors, the selected villages were similar in population and demographic characteristics. Data are provided to show the breakdown in SES, parental education, height/age, and weight/height and no significant differences were noted. Participation was stated to be voluntary, but participation rates were not provided. It is unclear if the 170 subjects were selected with 100% participation or if the 170 subjects were all that were asked to participate, but it appears that all subjects participated. Timing of the recruitment was not provided but is assumed to occur during the same time frame.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and recruited using the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)

- *Summary:* There was indirect evidence that key confounders including potential co-exposures were addressed using reasonable methods. A questionnaire, completed with the assistance of parents, was used to collect information on child characteristics (age, sex, height, weight), residential history, medical history (including illness affecting nervous system and head trauma), educational level of the head of the family (in years), and SES of the family. The SES was recorded according to the Pareek and Trivedi classification. The nutritional status of the children was calculated using the Waterlow's classification, which defines two groups for malnutrition using height for age ratio (chronic condition) and weight for height ratio (acute condition). Within both groups, it categorizes the malnutrition as normal, mildly impaired, moderately impaired, or severely impaired. Urinary lead and arsenic were analyzed using the atomic absorption spectrophotometer (Perkin-Elmer, Wellesley, United States). Urinary iodine was measured using the Dunn method. Authors do not report which covariates were included in the multivariate regression models; however, there was no difference in reported demographic characteristics. All subjects were the same age, and there was no difference in iodine, lead, or arsenic between the groups. Mean urinary arsenic levels did increase with increasing fluoride even though there was no significant difference by group.
- *Potentially important study-specific confounders:* All key confounders were considered in this study.
  - *Direction/magnitude of effect*: Not applicable.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and that key confounders including potential co-exposures were addressed.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results were provided for all 170 children stated to be included in the study.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence of no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* A sample of 200 mL of drinking water was collected at each child's home. The fluoride levels were analyzed by a fluoride ion-selective electrode, Orion 9609BN (Thermo Fisher Scientific Inc., West Palm Beach, United States). Each subject was also asked to collect a sample of their first morning urine. The fluoride content in the urine was determined using a fluoride ion-selective electrode, Orion 9609BN (Thermo Fisher Scientific Inc., West Palm Beach, United States). QA/QC and LOD were not reported and urinary dilution was not assessed. Although only current levels were measured, children who had changed water source since birth were excluded.
    - *Direction/magnitude of effect*: Spot urine samples that did not account for dilution could have exposure misclassification. The misclassification is likely non-differential and not likely to bias in any specific direction. Children who had changed water since birth were excluded, but it was not specifically noted that the fluoride in the water source was stable over the years.

    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Intelligence is assessed using the Raven's Standard Progressive Matrices and categorized into five grade levels. Although it was not noted that the test was validated to the study population, the test is visual and would be applicable to most populations (+ for methods). There is no mention of blinding by test administrators or evaluators and the exposure groups come from different geographic areas. It was also not reported who measured the levels of fluoride from the home or urine samples. Correspondence with the study authors indicated that the outcome assessors were blind to the children's fluoride status (++ for blinding). Overall rating for methods and blinding = +.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:*
        - *Statistical analyses:* Statistical analyses were reasonable (ANOVA), but consideration of homogeneity of variance was not reported.
        - *Other potential concerns:* None identified.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate, and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and the addressing of potential key confounders, but it was limited by the cross-sectional study design and lack of addressing dilution in the urine samples.

Seraj *et al.* (2012)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 6–11 years
- ***Study area:*** five villages, Makoo, Iran
- ***Sample size:*** 293 children

- ***Data relevant to the review:*** IQ (mean and distribution) assessed by Raven's Colored Progressive Matrices and presented by fluoride area, beta was also provided for water fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant correlation between water fluoride and IQ score (adjusted β = −3.865; CIs not reported); significantly higher IQ score in normal area (97.77 ± 18.91) compared with medium (89.03 ± 12.99) and high (88.58 ± 16.01) areas.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were selected from five villages in Makoo. The villages were stated to all be rural with similar general demographic and geographic characteristics and were comparable in terms of SES and parental occupations. Children were 6–11 years old. Age, gender, and education were taken into account in the analysis. No other characteristics were provided or discussed. Participation rates were not reported. There is indirect evidence that the populations were similar, and some possible differences were addressed.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and recruited using the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Age, gender, dental fluorosis intensity, and educational levels (child's and parents') were evaluated as potential confounders. Other potential confounders such as smoking were not discussed. Information was obtained from a detailed questionnaire. Lead was measured, but only found in low levels in the drinking water throughout the study regions. Iodine in the water was also stated to be measured and residents were receiving iodine-enriched salt. Arsenic was not addressed, but there is no evidence that arsenic levels would vary across villages in this area.
  - *Potentially important study-specific confounders:* All key confounders were considered in this study.
    - *Direction/magnitude of effect***:** Not applicable.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and that key confounders including potential co-exposures were addressed or were not likely to be an issue in the study area.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Attrition was low if it occurred. It was noted that 293 out of 314 children living in the villages were recruited. It is not clear if 21 children were excluded based on exclusion criteria or if they refused to participate; however, this accounts for less than 10% of the population and results were available for all 293 subjects.

- *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* Exposure was primarily based on area of residence. Fluoride in the groundwater was analyzed by the SPADNS (Sulfophenylazo dihydroxynaphthalene-disulfonate) method, utilizing 4000 UV-Vis spectrophotometer (Hach Company, Germany) in the environmental health engineering laboratory of the Public Health School of Tehran University of Medical Sciences. Specific details were not provided on methods of collection, samples locations, or if these locations represented the primary sources of drinking water for the subjects. Villages were categorized into normal (0.5–1 ppm), moderate (3.1±0.9 ppm), and high (5.2±1.1 ppm) fluoride based on the mean fluoride content of all seasons presumably for the stated 12-year time period. Subjects were stated to be long-life residents of the village. Dental fluorosis was also measured and increased in severity with fluoride levels; however, all areas had some degree of dental fluorosis. Although authors used an average fluoride level in varying seasons over presumably 12 years, they used a less-established method without reporting reliability or validity, nor did they provide data to indicate that the mean was truly representative of the fluoride levels over time and throughout the village. Although dental fluorosis severity increased with increasing fluoride levels, the data could also indicate potential exposure misclassification.
    - *Direction/magnitude of effect***:** The presence of dental fluorosis in all groups indicates that there may have been different exposure in some children at a younger age. Although there were only about 20 children in the "normal" fluoride group with very mild to mild dental fluorosis, this could bias the results toward the null because those children may have experienced a higher level of fluoride at some point. The other two fluoride groups were exposed to fluoride levels that likely exceeded those in the "normal" fluoride group.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that exposure was assessed using insensitive methods.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Intelligence was evaluated using the Raven's Color Progressive Matrices. This is a well-established method. Although the study authors did not provide data to indicate that the methods were valid in this study population, the test is designed to be culturally diverse. (+ for methods). The study report stated that test administrators were blinded. (++ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that outcomes were blindly assessed using instruments that were valid and reliable in the study population.

- ***Selective Reporting:***
  - *Rating:* Probably low risk of bias (+)

  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported. However, because they did not report the method for obtaining the betas in Table 4 of the study, it is not clear if these were adjusted or unadjusted betas.
  - *Basis for rating:* Probably low risk of bias based on direct evidence that all the study's measured outcomes were reported, but the results were not sufficiently reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Statistical comparisons between groups were reasonable (ANOVA), but consideration of homogeneity of variance was not reported. In addition, the methods for obtaining the betas were not reported.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and outcome. Study strengths include addressing potential key confounders, but it was limited by the cross-sectional study design and the group-level exposure data.

Soto-Barreras *et al.* (2019)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 9–10 years
- ***Study area:*** Chihuahua, Mexico
- ***Sample size:*** 161 children
- ***Data relevant to the review:*** Water fluoride, urinary fluoride, exposure dose, and dental fluorosis index by IQ grade.
- ***Reported association with fluoride exposure:*** No: No significant associations between fluoride exposure and IQ grades.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were selected using a multistage cluster sampling. During the first stage, 13 public elementary schools were randomly selected from a pool of 73 using a cluster sample design. Secondly, only fourth grade students were included. Authors stated that they wanted to keep the same grade level, but they were not specific as to why fourth graders were selected as opposed to any other grade. Lastly, only children whose parents or guardians attended and responded to the survey were included. There is no information provided on how the 13 schools selected may be similar or different from the 60 schools not selected. There is no information provided on the number of

children in the fourth grade to know participant rates. It was only noted that 245 children were examined, but 161 were included after the exclusion rules were applied. Inclusion and exclusion criteria are presented. Reasons for exclusion do not appear to be related to exposure or outcome. Characteristics of participants and non-participants are not compared; however, characteristics of the 161 included children were provided and any differences were taken into account in the analysis.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposed groups were similar and were recruited using similar methods during the same time frame.

- ***Confounding:***
  - *Rating:* Probably high risk of bias (-)
  - *Summary:* No confounders was considered when evaluating fluoride associations with intelligence; they were only applied when evaluating fluoride levels and dental caries. Based on Table 4 of the study, there was no significant association between IQ grade and child's age, sex, parental education, or SES status. No other information was reported or considered. There is no information on potential co-exposures. Based on water quality maps, the arsenic prediction indicates a greater than 50% probability of exceeding the WHO guidelines for arsenic of 10 µg/L in areas of Chihuahua, Mexico.
  - *Potentially important study-specific confounders:* Arsenic.
    - *Direction/magnitude of effect***:** The direction and magnitude of effects is unknown. There is potential for arsenic to occur in the study area, but it is not known how it relates to fluoride exposure. If they occur together in the water, it will bias away from the null; however, if they occurred in different areas, there is potential to bias toward the null.
  - *Basis for rating:* Probably high risk of bias based on indirect evidence that there is potential for exposure to arsenic that was not sufficiently addressed.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* A total of 161 of 245 children were included in the study. Exclusion criteria are presented and are unrelated to outcome or exposure. For the 161 children, there are no missing outcome data.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+);Probably high risk of bias (-)
  - *Summary:* **Urinary Fluoride (probably low risk of bias):** First morning void urine samples were collected based on NIOSH methods. Water samples were also stated to be collected, but it does not appear that methods followed any particular standard, and there is no indication that subjects were provided with collection containers. Analysis was based on a calibration curve using fluoride ion selective electrode. QC methods were mentioned. Based on results, there were values below detection limits, but LODs or % below LOD were not reported.

Daily fluoride exposure (probably high risk of bias): Daily fluoride exposure was based on the water fluoride level, drinking water consumption (based on parental report of how many glasses of water consumed), and body weight.

- *Direction/magnitude of effect***:** Spot urine samples that did not account for dilution could have exposure misclassification. The misclassification is likely non-differential and is not likely to bias in any specific direction. Daily exposure was based partially on parental report of water consumption. The direction and magnitude of effect is unknown.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure. The daily fluoride exposure is probably high risk of bias because there is indirect evidence that the exposure was assessed using methods of unknown validity.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Intellectual ability was evaluated using Raven's Colored Progressive Matrices by an independent examiner. Some details were provided, but it was not stated that the tests were assessed blind; however, there is no indication that subjects were from high fluoride areas and the assessor would not have knowledge of the urine or water fluoride levels. Results for children were converted into a percentile according to age (details not provided) and overall scores were assigned an intellectual grade of I to V as described in the report.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* The main analysis was for dental caries. Although they make conclusions on fluoride and IQ, they do not use the same analytical methods for both outcomes. Table 4 of the study provides a p-value although it is not clear what the p-value represents; it is presumed to be the Kruskal Wallis p-value. It appears that a Kolmogorov-Smirnov test was used to determine variable distribution and a Kruskal Wallis test was used to compare among the groups with a Dunn's test if significant.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.

- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements and outcomes blindly assessed, but is limited by the cross-sectional study design, lack of accounting for urine dilution, and by not addressing potential exposures to arsenic in the study area.

Sudhir *et al.* (2009)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 13–15 years
- ***Study area:*** Nalgonda district (Andhra Pradesh), India
- ***Sample size:*** 1,000 children
- ***Data relevant to the review:*** Mean IQ grade (not standard scores) or IQ distribution by water fluoride strata (<0.7, 0.7-1.2, 1.3-4.0, and >4.0 ppm).
- ***Reported association with fluoride exposure:*** Yes: Significantly increased number of intellectually impaired children with increasing drinking water fluoride levels.

Risk of Bias:

- **Author contacts:**
  - Authors were contacted in September of 2017 for additional information related to risk-of-bias evaluation, but no response was received.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children were selected from the same general population during the same time frame and were then broken down into nearly equal exposure groups. A cross-sectional study was conducted among 13–15-year-old school children of Nalgonda district, Andhra Pradesh between August and October 2006. Data were collected from the school children who were life-long residents of Nalgonda district, Andhra Pradesh and who consumed drinking water from the same source during the first 10 years of life. A stratified random sampling technique was used. The entire geographical area of Nalgonda district was divided into four strata based on different levels of naturally occurring fluoride in the drinking water supply. Children were randomly selected from schools in the different strata. It was noted that the 1,000 selected children were equally divided among all four strata, however, each group did not have 250 children (but instead 243−267 in each group). Participation rates are not reported. Exclusion criteria included: children who had a history of brain disease and head injuries, children whose intelligence had been affected by congenital or acquired disease, children who had migrated or were not permanent residents, children with orthodontic brackets, and children with severe extrinsic stains on their teeth. Age and gender data are presented in Table 1 of the study, but this information is not presented by the different fluoride groups.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and were recruited using the same methods during the same time frame.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Data were collected using a self-administered questionnaire and clinical examination. The self-administered questionnaire requested information on demographic data (appears to cover age and sex), permanent residential address, staple food consumed, liquids routinely consumed, and aids used for oral hygiene maintenance (fluoridated or nonfluoridated). SES was measured using the Kakkar socio-economic status scale (KSESS) with eight closed-ended questions related to parental education, family income, father's occupation, and other factors. All children were asked to fill out the form, and the answers obtained were scored using Kakkar socio-economic status scoring keys. Based on this scoring, children were divided into three groups—lower class, middle class, or upper class. Age, sex, and SES were not found to be significantly associated with IQ. Other confounders including smoking were not addressed. Co-exposures such as arsenic and lead were not addressed; however, there is no indication that lead is a co-exposure in this population and arsenic is not likely a major concern in this area based on water quality maps.
  - *Potentially important study-specific confounders:* Key confounders age, gender, and measures of SES were similar between exposure groups; however, arsenic was not taken into account. Arsenic often occurs in the drinking water along with fluoride in some Indian populations; however, based on water quality maps, this does not appear to be an issue in the Nalgonda district of Andhra Pradesh. Iodine deficiencies are not mentioned.
    - *Direction/magnitude of effect***:** The presence of arsenic would potentially bias away from the null if present with fluoride. Deficiencies in iodine would bias away from the null if present in areas of high fluoride, but toward the null if present in areas of non-high fluoride.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the key confounders are considered, co-exposure to arsenic is likely not an issue in this area, and methods used for collecting the information were valid and reliable.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results were available for the 1,000 children selected to participate.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence of no attrition.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children were placed into one of four strata based on the level of fluoride in drinking water. Collection of water samples was done in the districts. The placement into strata was based on fluoride levels obtained from documented records of District Rural Water Works Department. Once the children were assigned to strata, it was confirmed that the fluoride level of their drinking water was within the strata assigned. This was done using the methodology followed in National Oral Health Survey and Fluoride Mapping 2002–2003. During the initial visits to the schools, the children were interviewed regarding their history of residence and source of drinking water from birth to 10 years. The first child meeting criteria was given a bottle for water collection and

the next child was only given a bottle for collection if the water source was different than that of a previous child. Children were asked to collect the sample of water from the source that was used in the initial 10 years of their life and was collected the next day. It was not reported if all bottles were returned. The water samples collected were subjected to water fluoride analysis using an ion-specific electrode, Orion 720A fluoride meter at District Water Works, Nalgonda to confirm the fluoride levels in the water before commencement of clinical examination. LOD and QA/QC details were not reported.

        - *Direction/magnitude of effect*: There is some potential for exposure misclassification based on recall of the children on the source of water used in their first 10 years of life. The misclassification is likely non-differential and not likely to bias in any specific direction. Children who had changed water since birth were excluded, but it was not specifically noted that the fluoride in the water source was stable over the years.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
    - *Rating:* Probably high risk of bias (NR)
    - *Summary:* The Raven's standard progressive matrices (1992 edition) was used to assess IQ. Exams were carried out by a single examiner. Calibration of the examiner was done before the study and in the middle of the study, but it was not clear if this applied to the IQ evaluation or only to the clinical examination. This Raven's test is a standard test and although there is no information provided to indicate that the methods were reliable and valid in the study population, this test was created to be culturally fair (+ for methods). Blinding or other methods to reduce potential bias were not reported (NR for blinding). No response was received to an e-mail request for clarification in September 2017. Overall rating for methods and blinding = NR.
    - *Basis for rating:* Probably high risk of bias based on indirect evidence that the outcome was not assessed blind and could bias the results.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:*
        - *Statistical analyses:* Statistical analyses were appropriate and no other threats to internal validity were identified.
        - *Other potential concerns:* None identified.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.

- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include verification of exposure measurements and the addressing of potential key confounders, but it was limited by the cross-sectional study design and lack of information on blinding during outcome assessment.

Till *et al.* (2020)

Study Details:

- ***Study design*:** Prospective cohort
- ***Population*:** MIREC participants (pregnant mothers and their children aged 3–4 years)
- ***Study area:*** 10 cities, Canada
- ***Sample size:*** 398 mother−child pairs (247 from non-fluoridated areas, 151 from fluoridated areas; 200 breastfed as infants, 198 formula-fed as infants)
- ***Data relevant to the review:*** Adjusted linear regression models evaluating associations between IQ with water fluoride concentration (with or without adjusting for maternal urine) in formula-fed or breast-fed infants or by fluoride intake from formula.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower performance IQ with water fluoride (adjusted βs = −9.26 formula-fed, −6.19 breastfed) and fluoride intake from formula (adjusted β = −8.76); significantly lower full-scale IQ with water fluoride in formula-fed (adjusted β = −4.40); no significant changes in full-scale IQ for water fluoride in breastfed children or fluoride intake from formula-fed children; no significant changes in verbal IQ scores with fluoride exposure.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Pregnant women were recruited between 2008 and 2011 by the MIREC program from 10 cities across Canada. Inclusion and exclusion criteria were provided. Additional details were stated to be available in Arbuckle *et al.* (2013). A total of 610 children were recruited to participate in the developmental follow-up with 601 children completing all testing. The demographic characteristics of women included in the current analyses (n = 398) were not substantially different from the original MIREC cohort (N = 1945) or the subset without complete water fluoride and covariate data (n = 203). A table of characteristics of the study population is provided. Approximately half of the children lived in nonfluoridated cities and half lived in fluoridated cities.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were similar and were recruited with the same methods during the same time frame.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)

- *Summary:* Covariates were selected a priori that have been associated with fluoride, breast feeding, and children's intellectual ability. Final covariates included child's sex and age at testing, maternal education, maternal race, second-hand smoke in the home, and HOME score. City was considered but was excluded from the models. Confounders that were not assessed include: parental mental health, iodine deficiency/excess, parental IQ, and co-exposure to arsenic and lead. Co-exposure to arsenic is less likely an issue in this Canadian population and the lack of information is not considered to appreciably bias the results.
- *Potentially important study-specific confounders:* All key confounders were considered in this study.
    - *Direction/magnitude of effect***:** Not applicable.
- *Basis for rating:* Probably low risk of bias based on direct evidence that key confounders were addressed and indirect evidence that the methods used to collect the information were valid and reliable and co-exposures were not an issue.

- ***Attrition:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Of 610 children, 601 (98.5%) in the MIREC developmental study who were ages 3–4 years completed the neurodevelopment testing. Of the 601 children who completed the neurodevelopmental testing, 591 (99%) completed the infant feeding questionnaire and 398 (67.3%) reported drinking tap water. It was noted that the demographic characteristics were not substantially different from the original MIREC cohort or the 203 subjects without complete water fluoride or covariate data.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Information on breastfeeding was obtained via questionnaire at 30–48 months. Fluoride concentration in the drinking water was assessed by daily or monthly reports provided by water treatment plants. Water reports were first linked with mothers' postal codes and the daily or weekly amounts were averaged over the first 6 months of each child's life. Additional details can be found in Till *et al.* (2018). Maternal urinary exposure was used to assess fetal fluoride exposure. Procedures can be found in Green *et al.* (2019).
        - *Direction/magnitude of effect***:** There is not any specific direction or magnitude of bias expected. Urinary fluoride levels are reflective of recent exposure. The possibility of the exposure misclassification would be similar in all subjects and would be non-differential. For the fluoride intake from formula, exposure was based on the fluoride levels in the water at the residence and the proportion of time that the infant was not exclusively breastfed. This exposure misclassification would also be non-differential.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***

- *Rating:* Probably low risk of bias (+)
- *Summary:* Intelligence was tested using the Wechsler Preschool and Primary Scale of Intelligence III. This is appropriate for both the study population and age group. This is considered a gold standard test. It was not reported whether the evaluators were blind to the child's fluoride exposure status during the assessment. Although it is unlikely that the assessors had knowledge of the specific drinking water levels or maternal urine levels, there is potential that the outcome assessors had knowledge of the city the child lived in and if the city was fluoridated or non-fluoridated. Correspondence with the study authors on the outcome assessment for Green *et al.* (2019) indicated that it was unlikely that the testers had knowledge of the city's fluoridation. The same is assumed here. Specific measurements included were identified.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Regression diagnostics were used to test assumptions for linearity, normality, and homogeneity. There were two potential influential observations (based on Cook's distance), and sensitivity analyses re-estimated the models without these two variables. Effect modification by breastfeeding status was evaluated. Interestingly, all regression coefficients were divided by 2 to represent change in IQ per 0.5-mg/L change in fluoride. One concern is posed by the lack of accounting for city in the regression models, ideally as a random effect. The authors explored including city as a covariate in the models; however, city was not included either because it was strongly multi-collinear with water fluoride concentration (VIF > 20) (model 1, with water fluoride concentration) or because fluoride intake from formula is a function of water fluoride concentration (assessed at the city level) and was therefore deemed redundant (model 2). However, the models use city-level water fluoride concentrations (and in sensitivity analyses, adjust for maternal urinary fluoride) which warrant exploration of city as a random effect rather than a fixed effect (as would be by just having it included as a covariate). Even including individual-level maternal urinary fluoride might not fully account for lack of city, given that the subjects were from six different cities, with half of them fully on fluoridated water. Hence, even individual-level exposures are likely to be correlated at the city level. Based on a previous analysis (Green *et al.* 2019), it is unlikely that

exclusion of city from models (as a fixed or random effect) would impact the effect estimates.
        - *Other potential concerns:* None identified.
    - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, prospective cohort design, and the addressing of potential key confounders.

Trivedi *et al.* (2012)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 12−13 years
- ***Study area:*** Kachchh, Gujarat, India
- ***Sample size:*** 84 children
- ***Data relevant to the review:*** Mean IQ scores and distribution by low and high fluoride villages.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower IQ score in the high fluoride (92.53 ± 3.13) compared to the low fluoride (97.17 ± 2.54) areas in boys and girls combined (as well as separately). Villages with higher fluoride levels had a larger percentage of subjects with IQ scores of 70–79, while the lower fluoride villages had a greater percentage of IQ scores > 109.

Risk of Bias:

- **Author contacts:**
    - Authors were contacted in September of 2017 to obtain additional information for risk-of-bias evaluation.
- ***Population selection:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* There is insufficient information provided on the sampling methods to determine if the populations were similar. Although it was noted that samples were obtained for groundwater quality from March to May of 2011, there is no indication that the children were selected at the same time or during a similar time frame. Correspondence with the author indicates that children were selected within a week of the water collection based on random selection of a school in the village. Study participants were selected from six different villages of the Mundra region of Gujarat, India. Subjects were grouped into high and low villages based on the level of fluoride in the drinking water of those villages. The number of subjects per village were not reported, but it was noted that there were 50 children in the low fluoride group and 34 children in the high fluoride group. It is not clear if the differences in numbers were based on different participation rates or if there were fewer children in the high fluoride villages. Recruitment methods including any exclusion criteria and participation rates

were not provided. SES was stated to be low and equal based on questionnaire information, but the results were not provided. It should also be noted that only regular students (having attendance more than 80%) of standard 6th and 7th grades were selected, but it was not noted if attendance varied by village. Correspondence with the study author indicated that there was an average of 20 students per class with an average of 40 students per village. It appears that keeping the requirement for 80% attendance was a limiting factor that caused different numbers of children by area; however, this was applied similarly to both groups.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that subjects were similar and recruited using the same methods during the same time frame.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children were stated to be students of the 6th and 7th standard grades. Age was not addressed, but the children would all be of similar age based on the grades included. Results were reported for males and females separately as well as combined. SES and iodine consumption were stated to be analyzed via a questionnaire and were standardized on the basis of the 2011 census of India. Although it was noted in the abstract that the SES was equal (no data provided), the study report did not mention the iodine results. Although the study authors did not address arsenic or lead, they did provide physicochemical analyses for the water samples from the six different villages. Information on arsenic in the water is not provided, but based on water quality maps, arsenic is not expected to be a major concern in this study area.
  - *Potentially important study-specific confounders:* Key confounders age, gender, and measures of SES were similar between exposure groups; however, arsenic was not taken into account. Arsenic often occurs in the drinking water along with fluoride in some Indian populations; however, based on water quality maps, arsenic does not appear to be an issue in the study area.
    - *Direction/magnitude of effect***:** Presence of arsenic would potentially bias away from the null if present with fluoride or toward the null if present in the reference group.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable, that potential co-exposures were not an issue, and that key confounders were addressed.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results were provided for 84 children, but the methods do not indicate how many children were initially selected to participate nor were any exclusion criteria provided. It was noted in the results that 84 children had their groundwater and urine tested, but it was not noted if analyses were restricted to these children or if exposures were assessed in all the children who had IQ measurements. Correspondence with the study author indicated that the main reason for exclusion was a <80% attendance rate, with fluoride and IQ measured on all 84 children who met the criteria.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence of no attrition.

- ***Exposure:***

- *Rating:* Probably low risk of bias (+)
- *Summary:* Children in villages were grouped based on fluoride levels that were assessed in groundwater (low F villages versus high F villages). The average concentration of these levels was considered to be the levels in the drinking water with confirmation using urinary fluoride levels. The groundwater samples were selected to cover major parts of the taluka and represent overall groundwater quality. Ten samples were obtained from each village. Fluoride was measured in the groundwater using ion exchange chromatography. Although urine levels were also significantly higher in the high fluoride village, no information was provided on how or when the urinary samples were obtained or how they were measured. However, correspondence with the study author indicated that the groundwater and urine fluoride levels were available for all 84 children indicating that the urine measures were available for the children that had IQ measures. The urine samples were stated to be collected at the same time that the second water sample was collected.
  - *Direction/magnitude of effect*: Fluoride levels were measured in both the drinking water and urine. Although there is some variability in the measurements, there is no overlap between the two groups and the urine and drinking water levels in the children support each other. Any potential exposure misclassification would be non-differential and direction and magnitude are unknown.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Outcome methods were only noted to be reported in Trivedi *et al.* (2007), which was scored as follows: IQ was measured in the children of both areas using a questionnaire prepared by Professor JH Shah, copyrighted by Akash Manomapan Kendra, Ahmedabad, India, and standardized on the Gujarati population with 97% reliability rate in relation to the Stanford-Binet Intelligence Scale (+ for methods). Blinding or other methods to reduce bias are not reported, but correspondence with the study author indicated that the teachers were blind to the status of fluoride. The teachers administered the tests in the presence of a research fellow. It is not completely clear who scored the tests, but it is assumed the teachers. (+ for blinding). Overall rating for methods and blinding = +.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the outcomes were blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***

- *Rating:* Probably low risk of bias (+)
- *Summary:*
  - *Statistical analyses:* Statistical analyses were reasonable (paired sample T-test), but consideration of homogeneity of variance was not reported.
  - *Other potential concerns:* No other threats to internal validity were identified.
- *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.

- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and the addressing of potential key confounders but was limited by the cross-sectional study design.

Valdez Jimenez *et al.* (2017)

Study Details:

- ***Study design*:** Prospective cohort
- ***Population*:** Infants aged 3–15 months
- ***Study area:*** Durango City and Lagos de Moreno, Jalisco, Mexico
- ***Sample size:*** 65 infants
- ***Data relevant to the review:*** The Bayley Scales of Infant Development II was used to assess Mental Development Index Scale and the Psychomotor Development Index scale in children 3 to 15 month and evaluated for associations with first and second trimester maternal urine fluoride.
- ***Reported association with fluoride exposure:*** Yes: Significant correlation between maternal urinary fluoride and MDI score during first trimester (adjusted β = −19.05; SE = 8.9) and second trimester (adjusted β = −19.34; SE = 7.46).

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were recruited from two endemic areas in Mexico. The study authors do not provide information on the similarities or differences between the two areas nor do they indicate if there were different participation rates. However, recruitment methods were the same. Women receiving prenatal care in health centers located in Durango City and Lagos de Moreno, Jalisco, Mexico were recruited in 2013–2014. Participation rates are not likely to be an issue as characteristics were similar between those who participated and those who did not. Although they did not provide characteristics by area, the characteristics provided do not indicate any differences that may be biased by the selection. Considering the age range for the non-participants, the mean age for non-participants appears to be incorrect (or the age range is incorrect);

however, there does not appear to be a difference that would potentially indicate selection bias.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited with the same methods in the same time frame, with no evidence of differences or issues with participation/response rates.
- ***Confounding:***
    - *Rating:* Probably high risk of bias (-)
    - *Summary:* Questionnaires were used to obtain information about sociodemographic factors, prenatal history, mother's health status before pregnancy (e.g., use of drugs, vaccines, diseases) and the type of water for drinking and cooking. The marginalization index (MI) was obtained from the National Population Council (CONAPO). Two additional surveys were conducted during the 2nd and 3rd trimester of pregnancy to get information about the mother's health, pregnancy evolution, and sources of water consumption. A survey was also conducted to get information about childbirth (type of birth, week of birth, weight and length of the baby at birth, Apgar and health conditions of the baby during the first month of life). This information was corroborated with the birth certificate. Linear regression models included gestational age, children's age, marginality index, and type of drinking water. Bivariate analysis was conducted on the other factors including child's gender prior to conducting multivariable regression models. Some important confounders were not considered, including parental mental health, IQ, smoking, and potential co-exposures.
    - *Potentially important study-specific confounders:* Arsenic is a potential co-exposure in this area of Mexico.
        - *Direction/magnitude of effect***:** If arsenic were present as a co-exposure it would bias the results away from the null.
    - *Basis for rating:* Probably high risk of bias based on indirect evidence that there is a potential for co-exposure with arsenic that was not addressed.
- ***Attrition:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* Out of the 90 women selected for inclusion in the study, 65 approved the participation of their infants. The authors provide a table of characteristics between women who consented to their children's cognitive evaluation and those that only participated in biological monitoring. There were no significant differences between the groups. There were fewer women who provided urine during the second and third trimesters. All specified children are included in the relevant analyses.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* Fluoride exposure is assessed through morning urine samples and water fluoride levels collected from the children's homes. Sampling methodology is appropriately documented, and water levels were quantified through specific

ion-sensitive electrode assays. QC was described and accuracy was >90%. Urinary fluoride was corrected by specific gravity.

- *Direction/magnitude of effect*: Not applicable.

- *Basis for rating:* Definitely low risk of bias based on direct evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Neurodevelopment was assessed with the Bayley Scales of Infant Development II (BSDI-II) that was noted to be reliable and valid for evaluating children from 3 months to 5 years of age. The average age of children assessed was 8 months, with a range of 3–15 months) (++ for methods). The study report stated that a trained psychologist who was blinded about the mother's fluoride exposure evaluated the infants at home (++ for blinding). Overall rating for methods and blinding = ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods were reported. Table 4 of the study only displays data for trimesters 1 and 2. Although 3rd trimester data were collected, they were not reported, likely because data were only available for 29 subjects. No discussion of this was provided.
  - *Basis for rating:* Probably low risk of bias because, although it appears some data were not reported, it is likely because there were insufficient data and not because the authors were selectively reporting the results.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study.
    - *Other potential concerns:* No other potential concerns were identified. In the peer-review report, NASEM (2020) cited the following as potential concerns: "the large difference in numbers of males and females in the offspring (20 males, 45 females), and apparently incorrect probabilities were reported for age differences between participants and nonparticipants, high rates of cesarean deliveries and premature births among participants (degree of overlap not reported), and incorrect comparisons of observed prematurity rates with national expected rates." However, these concerns were taken in consideration in other domains (***Selection, Confounding)***.
  - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Definitely low risk-of-bias ratings in exposure and outcome. Study strengths include individual exposure measurements and

outcome blindly assessed, but it is limited by the cross-sectional study design and lack of accounting for potential co-exposures to arsenic.

Wang *et al.* (2012)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 8–13 years (possibly the same population as Xiang *et al.* (2003a))
- ***Study area:*** Wamiao and Xinhuai villages located in Sihong County, Jiangsu Providence, China
- ***Sample size:*** 526 school children
- ***Data relevant to the review:*** Mean IQ and % low IQ (< 80) by total fluoride intake.
- ***Reported association with fluoride exposure:*** Yes: Significantly lower mean IQ in the high fluoride village (92.02 ± 13.00) compared to the control village (100.41 ± 13.21); when high exposure group was broken into 4 exposure groups, a dose-dependent decrease in IQ and increase in % with low IQ observed; significant correlation between total fluoride intake and IQ (r = −0.332); OR for IQ<80 per increase in total fluoride intake=1.106; 95% CI 1.052–1.163).

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study appears to be the same population as Xiang *et al.* (2003a) and Xiang *et al.* (2011); however, the study does not cite these studies as providing additional information and numbers of children differ; therefore, it may be a separate analysis on the same villages. The years of testing were not provided so it cannot be determined if study subjects are the same. Two villages, Wamiao and Xinhuai, located 64 km apart in Sihong County, Jiangsu Province were selected for the study. Wamiao is a village in a region with severe endemic fluorosis and Xinhuai is a village in a non-endemic fluorosis region. Neither village has fluoride pollution from coal or industrial sources. Villages were stated to be similar in terms of annual per capita income, transportation, education, medical conditions, the natural environment, and lifestyle. All primary students ages 8–13 years currently in school in either village were surveyed with exclusions noted. Of 243 children from Wamiao, 236 (97.12%) were included, and of 305 children from Xinhuai, 290 (95.08%) were included. No table of subject characteristics was provided.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited using the same methods within the same time frame, with direct evidence that there was no difference in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Logistic regression of low IQ rate and total fluoride intake adjusted for age and sex. Both villages had hand-pumped well water for drinking water, but the authors

do not mention if arsenic was also present in the drinking water. However, a publication by Xiang *et al.* (2013) on this study area indicates that Xinhuai (the low fluoride area) had significantly higher arsenic levels compared to Wamiao (the endemic fluorosis area), which would bias toward the null. Areas were stated to be similar in annual per capita income, transportation, education, medical conditions, the natural environment, and lifestyle; however, no details were provided. This study did not address other co-exposures, but other studies on populations in these villages (Xiang *et al.* 2011, Xiang *et al.* 2003a) indicate that iodine and lead are not concerns.

- *Potentially important study-specific confounders:* Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on information provided in Xiang *et al.* (2013), arsenic concentrations were higher in the low fluoride area compared to the high fluoride area.
  - *Direction/magnitude of effect*: Presence of arsenic in this study population would potentially bias toward the null.
- *Basis for rating:* Probably low of risk bias because there is indirect evidence that the key confounders are take into account, methods used for collecting the information were valid and reliable, and co-exposures to arsenic and lead and iodine deficiency are not attributing to the effect observed in this area.

- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Data are reported for all 526 children noted to be included in the study. There is a slight discrepancy in the reported total number of children from the high-fluoride village and the number of participants from the high-fluoride village between this paper (236 participated of 243 total children) and the 2003 and 2011 publications on the same study population (222 of 238). This discrepancy is not explained but is not expected to appreciably bias the results.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+); Probably high risk of bias (-)
  - *Summary:* **Water fluoride (+ probably low risk of bias):** Exposure was based on drinking water levels and fluoride intake. Residents in the Wamiao village were divided into five groups based on fluoride levels in the drinking water. Clean, dry polyethylene bottles were used to collect 50 mL of drinking water from each student's household and fluoride content was measured.

    Total fluoride intake (- probably high risk of bias): Six families from each of the five Wamiao groups were randomly selected as dietary survey households. Intakes of various foods by each person at each meal and intakes of unboiled water, boiled water, and tea were surveyed for four consecutive days. Methods for food collection were described. Five representative households from each village were selected based on geographic location, population distribution, housing structure, and other conditions. Indoor air samples were collected once daily for five consecutive days; outdoor air was sampled at two points once daily for five days. Methods for determining fluoride

content in samples were noted to follow specific guidelines. Calculation of total fluoride intake was stated to follow Appendix A of the People's Republic of China Health Industry Standard with some details provided. Although it is assumed the method is valid, it was not detailed how each fluoride determination was made for each subject, and it appears that total fluoride intake was determined based on data from select subjects and not all subjects.

- *Direction/magnitude of effect*: There is potential for exposure misclassification based on calculating fluoride intake based on measurements from a few select subjects rather than all subjects. The direction and magnitude of effect cannot be assessed based on the information provided.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure. The intake is probably high risk of bias because there is indirect evidence that the exposure was assessed using methods of unknown validity.

- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* IQ of each child was measured with the Combined Raven's Test for Rural China (CRT-RC) (++ for methods). The test was stated to be administered to the children independently in a school classroom under the supervision of three exam proctors. Testing methods, testing language, and testing conditions were all in strict accordance with the CRT-RC guidebook. Major testing personnel received necessary training by the Psychology Department of East China Normal University. The children undergoing IQ testing and the test scorers were kept double-blinded throughout the testing process. (++ for blinding). Overall rating= ++.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Although it was noted that a logistic regression was used to determine the odds of having low IQ with increasing fluoride intake, no details were provided on any of the other tests conducted. Because this is the same population evaluated in Xiang *et al.* (2003a) and Xiang *et al.* (2011), it is assumed that the same methods were used even if this study population consisted of different children.
    - *Other potential concerns:* None identified.

  - *Basis for rating:* Probably low risk if bias based on indirect evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements with blinding at outcome assessment but is limited by the cross-sectional study design and not using individual measurements to calculate fluoride intake. All key confounders were accounted for in the study design or analysis, but there is potential for the presence of arsenic to bias toward the null.

Wang *et al.* (2020a)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** School children aged 7–13 years
- ***Study area:*** Tongxu County, China
- ***Sample size:*** 325 school children
- ***Data relevant to the review:*** Associations between ADHD and other measures of learning disability with urine fluoride concentrations.
- ***Reported association with fluoride exposure:*** Yes: Significant correlation between psychosomatic problems and urinary fluoride (adjusted β = 4.01 [95% CI: 2.74, 5.28]) and increased risk of a T-score > 70 with increasing urinary fluoride (adjusted OR = 1.97 [95% CI: 1.19, 3.27]). No significant associations with ADHD or other measures of learning disability.

Risk of Bias:

- **Author contacts:**
  - Authors were contacted in July of 2020 to obtain additional information for risk-of-bias evaluation. No response was received.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Subjects were recruited in 2017 from Tongxu County, China. Children were selected from four randomly selected primary schools in the area. Selection was based on specified inclusion rules. It was noted that the living habits and diets of the participants from the four schools were well matched, but details were not provided. The area did not have industrial pollution within 1 km of the living environment of the children, and it was noted that the children were not exposed to other neurodevelopmental toxicants (lead, cadmium, arsenic, or mercury). A table of subject characteristics was provided in the study, but not by school or exposure. This is a pilot study, and it is not explicitly stated if all eligible subjects participated in the study. There is no information on participation rates or if they varied by school.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposed groups were recruited using similar methods during the same time frame and that any differences between the exposed groups were considered in the statistical analyses.

- ***Confounding:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* It was noted that subjects were well matched in terms of living habits and diets, but there were no specifics provided. It was noted that there was no industrial exposure or exposure to other neurotoxins such as lead, cadmium, arsenic, or mercury. Covariates were collected using a standardized and structured questionnaire completed by the children and their guardians under the direction of investigators, but reliability or validity of the questionnaire was not reported. Information collected included age, gender, weight, height, parental education level, and parental migration (or work as migrant workers). IQ scores evaluated by the Combined Raven's Test-the Rural in China were used to represent basic cognitive function. Models were adjusted for age, BMI, gender, mother and father migration, and urinary creatinine. Adjustments were not made for parental education, race/ethnicity, maternal demographics (e.g., maternal age, BMI), parental behavioral and mental health disorders (e.g., ADHD, depression), smoking (e.g., maternal smoking status, secondhand tobacco smoke exposure), reproductive factors (e.g., parity), iodine deficiency/excess, maternal (and paternal) IQ, quantity and quality of caregiving environment (e.g., HOME score), or SES other than parental migration. There is no evidence to suggest that SES would differ substantially among the four rural schools in the same area of China that were randomly selected.
    - *Potentially important study-specific confounders:* SES.
        - *Direction/magnitude of effect***:** Direction and magnitude is unknown. It was noted that the subjects were matched in terms of living habits and diet and this could be an indication that SES was not different among the groups, but details were not provided.
    - *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key confounders are considered, that the methods for collecting the information were valid and reliable, and that co-exposure to arsenic is not an issue in this area.
- ***Attrition:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* Data are complete. It was noted that there were 325 subjects included and results were available on all subjects.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- ***Exposure:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Spot urine samples were collected from each child in the early morning into cleaned polyethylene tubes. Fluoride concentrations were measured using fluoride ion-selective electrode (with reference to Ma *et al.* (2017); however, that reference cites Zhou *et al.* (2012). Therefore, no QC methods or LODs were available. Fluoride concentrations were creatinine-adjusted.
        - *Direction/magnitude of effect***:** Spot urine samples only account for recent exposure. Although this could cause there to be some exposure misclassification, the number of subjects should help dilute any issues with the non-differential misclassification.

    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.
- ***Outcome:***
    - *Rating:* Probably high risk of bias (NR)
    - *Summary:* Children's behavior was assessed by the Chinese version of the Conners' Parent Rating Scale-Revised (CPRS-48). The homogeneity reliability of Cronbach α in the Chinese version of CPRS-48 was 0.932; the correlation of Spearman-brown split-half was 0.900; and the retest reliability of total score was 0.594. Raw scores for each subscale are converted into sex- and age-adjusted T-scores within a mean ± standard deviation (SD) of 50 ± 10. The guardians independently completed the CPRS-48 according to the instruction manual under the direction of trained investigators (++ for methods). Blinding is not reported. Although it is unlikely that the outcome assessors were aware of the fluoride levels in the urine, it is unclear if subjects were selected based on areas with endemic fluoride or if parents were aware of fluoride concentrations in the areas. (NR for blinding). Overall rating for methods and blinding = NR.
    - *Basis for rating:* Probably high risk of bias based on no information provided to indicate that the outcome was blindly assessed.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient details.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:*
        - *Statistical analyses:* Multiple linear regression models were used to assess the fluoride association with each behavioral outcome. Logistic regression was used to assess the risk of behavioral problems due to fluoride exposure, but what they used to delineate a behavioral problem was not specified. Sensitivity analyses were performed.
        - *Other potential concerns:* None identified.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that the statistical analyses were appropriate and no other potential threats to risk of bias were identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Probably low risk-of-bias ratings in confounding and exposure. Study strengths include individual exposure measurements, but it is limited by the cross-sectional study design and lack of details on blinding of the outcome assessment. All key confounders were considered in the study design or analysis.

Wang *et al.* (2020b)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** School children aged 7–13 years
- ***Study area:*** Tianjin City, China (possibly a subset of the children from Yu *et al.* (2018))
- ***Sample size:*** 571 school children
- ***Data relevant to the review:*** IQ scores by urine and water fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant correlations between IQ score and water fluoride (adjusted β = −1.587 per 1-mg/L increase) and urinary fluoride (adjusted β = −1.214 per 1-mg/L increase).

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Subjects were from a cross-sectional study conducted in 2015, but no citation was provided on this cohort (presumably the Yu *et al.* (2018) cohort). It was noted that the subjects in that cohort were from districts with historically high or normal fluoride levels. Subjects for this study were selected by using a stratified and multistage random sampling approach. Brief description was provided. The study area consisted of three historically high fluoride areas and four nonendemic areas. A flow diagram was provided for inclusion and exclusion, but this detail was given for all children and not by area. Therefore, it cannot be determined if the participation differed by area. However, there was a 93% recruitment rate, and the 13 excluded due to missing data are not likely excluded due to exposure. Detailed characteristics of the study population are provided. Exclusion criteria included: "children who had congenital or acquired diseases affecting intelligence, or a history of cerebral trauma and neurological disorders, or those with a positive screening test history (like hepatitis B virus infection, Treponema palladium infection and Down's syndrome) and adverse exposures (smoking and drinking) during maternal pregnancy, prior diagnosis of thyroid disease, and children who had had missing values of significant factors (2.2%) were also excluded."
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were recruited using similar methods during the same time frame and that any differences between the exposed groups were accounted for in the statistical analyses.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Study authors noted that the study areas are not exposed to other neurotoxins such as lead, arsenic, or mercury nor were they iodine-deficient. Final models included child's age, child's gender, child's BMI, maternal and paternal education, household income, and low birth weight. Other potential confounders that were considered is unclear as they only noted that the confounders were selected based on current literature. Reasons for exclusion included history of disease affecting intelligence, history of trauma or neurological disorders, positive screening test history,

or exposures such as smoking or drinking during pregnancy. Information was obtained by questionnaire or measurements. Variables such as parental BMI, behavioral and mental health disorders, IQ, and quantity and quality of the caregiving environment were not addressed.

- *Potentially important study-specific confounders:* All key confounders were considered in this study.
  - *Direction/magnitude of effect*: Not applicable.
- *Basis for rating:* Probably low risk of bias because there is direct evidence that the key confounders are taken into account, indirect evidence that the methods for collecting the information were valid and reliable, and co-exposure to arsenic is not an issue in this area.

- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* A detailed chart of the recruitment process is presented. The study had a 93% recruitment rate and only 2.2% of subjects with missing data for certain covariates were excluded.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Children provided spot urine samples, presumably at the time of examination. Water samples were randomly collected from public water supplies in each village. Fluoride concentrations were analyzed using fluoride ion-selective electrode according to the national standardized method in China. There is no indication if the urine samples accounted for dilution.
    - *Direction/magnitude of effect*: Not accounting for dilution could cause there to be some exposure misclassification. The direction and magnitude would depend on where the differences occurred.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using acceptable methods that provide individual levels of exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Assessments of IQ scores were conducted by graduate students at the School of Public Health, Tongji Medical College at the Huazhong University of Science and Technology. Each team member was assigned a single task, meaning that only one person would have conducted the IQ tests. A Combined Raven's Test for Rural China was used. Therefore, the test was appropriate for the study population (++ for method). It was note that the examiner was trained and blind to the exposure (++ for blinding). Overall = ++
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes in the abstract, introduction, and methods are reported in sufficient details.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses were appropriate and no other threats to internal validity were identified. Logistic and multivariate regression models accounting for potential confounders were used. Results are presented as betas or odds ratios and 95% confidence intervals. Regression diagnostics were conducted for all models, including examination of multicollinearity, heteroscedasticity, and influential observations. Mediation and interaction analyses were appropriate. The stratified and multistage random sampling approach for subject selection and the fact that selected villages are similar in population and general demographic characteristics helps to ensure that there was no need to account for village-level effects even when the analysis used water samples from the village.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the statistical analyses were appropriate and no other potential threats to risk of bias were identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key confounders were considered in the study design or analysis.

Xiang *et al.* (2003a)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 8–13 years
- ***Study area:*** Wamiao and Xinhuai villages located in Sihong County, Jiangsu Providence, China
- ***Sample size:*** 512 school children
- ***Data relevant to the review:*** Comparison of IQ (mean and distribution) between Wamiao County (a severe endemic fluorosis area) and Xinhuai County (non-endemic fluorosis area); additional breakdown of the Wamiao area into 5 water fluoride exposure groups.
- ***Reported association with fluoride exposure:*** Yes: Significant dose-related effect of drinking water fluoride on IQ score based on quintile levels with significantly lower IQ scores observed with water fluoride levels of 1.53 mg/L or higher. Pearson correlation coefficient of −0.164 with urinary fluoride. IQ scores for children in the non-endemic region (100.41 ± 13.21) were

significantly higher than the endemic region (92.02 ± 13.00). The lower-bound confidence limit benchmark concentration (BMCL) of 1.85 mg/L was calculated.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Two villages, Wamiao and Xinhuai, located 64 km apart in Sihong County, Jiangsu Province were selected for this study, which was conducted between September and December 2002. Wamiao is located in a severe fluorosis endemic area, and Xinhuai is located in a non-endemic fluorosis area. Neither village has fluoride pollution from burning coal or other industrial sources. All eligible children in each village were included; children who had been absent from either village for 2 years or longer or who had a history of brain disease or head injury were excluded. In Wamiao, 93% of the children (222 out of 238) were included for the study, while in Xinhuai, 95% were included (290 out of 305). The children in Wamiao were divided into five subgroups according to the level of fluoride in their drinking water: <1.0 mg/L (group A), 1.0–1.9 mg/L (group B), 2.0–2.9 mg/L (group C), 3.0–3.9 mg/L (group D), and >3.9 mg/L (group E). Children in Xinhuai (0.18–0.76 mg F/L in the drinking water) served as a control group (group F). Demographic characteristics are not presented, and statistical analyses are not adjusted, but mean IQ scores are stratified by child's age, child's gender, family income, and parental education.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited using the same methods within the same time frame, with direct evidence that there was no difference in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Although information was stated to be collected on personal characteristics, medical history, education levels of the children and parents, family SES, and lifestyle, only child's gender, child's age, family income, and parental education were addressed. Other potential co-exposures, such as arsenic, were not addressed. A separate publication in 2003 [(Xiang *et al.* 2003b), letter to the editor], indicated that blood lead levels were not significantly different between the two areas. Although arsenic was not addressed specifically in this publication, Xiang *et al.* (2013) measured both fluoride and arsenic in the Wamiao and Xinhuai areas. Xinhuai (the low fluoride area) had significantly higher arsenic levels compared to Wamiao (the endemic fluorosis area). This is likely to bias toward the null; however, the study observed a significantly lower IQ score in the endemic fluorosis area. Iodine was tested in a subset of the children and found not to be significantly different between the two groups.
  - *Potentially important study-specific confounders:* Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on information provided in Xiang *et al.* (2013), arsenic concentrations were higher in the low fluoride area compared to the high fluoride area.

        - *Direction/magnitude of effect*: Presence of arsenic in this study population would potentially bias towards the null.
    - *Basis for rating:* Probably low risk of bias because there is indirect evidence that the key confounders are taken into account, methods used for collecting the information were valid and reliable, and co-exposures to arsenic and lead and iodine deficiency are not attributing to the effect observed in this area.
- ***Attrition:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* Data are complete. IQ results were reported for all 512 children included in the study (222 in the endemic area and 290 in the nonendemic area).
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- ***Exposure:***
    - *Rating:* Probably low risk of bias (+)
    - *Summary:* Exposure was based on drinking water and urinary levels of fluoride. The two study areas were selected to reflect a severe endemic area and a nonendemic area. Drinking water was collected from wells and early-morning spot urine samples were collected from a randomly-selected subsample of children. Both water and urine samples were measured using fluoride ion-selective electrode, but no quality control was discussed. Both absolute and creatinine-adjusted urine results were reported.
        - *Direction/magnitude of effect*: There is potential for exposure misclassification because only current levels were assessed. Migration of subjects in or out of the area was not assessed, but the study authors noted that, if the children had been absent from the village for 2 or more years, they were excluded. Misclassification would likely be non-differential, which could bias the results in either direction.
    - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* The IQ of each child was measured with the Combined Raven's Test for Rural China (CRT-RC) (++ for methods). The test was stated to be administered to the children independently in a school classroom, in a double-blind manner, under the supervision of an examiner and two assistants, and in accordance with the directions of the CRT-RC manual regarding test administration conditions, instructions to be given, and test environment. (++ for blinding). Overall rating= ++
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.

  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:*
    - *Statistical analyses:* Data were stated to be analyzed using SAS without reporting the tests conducted. Results provided in the tables indicate that a t-test was conducted, but it was not reported that homogeneity of variance was tested or confirmed. In addition, correlations were tested with Pearson's correlation.
    - A potential concern raised by the NASEM (2020) peer review was the lack of accounting for relationships in exposure between persons from the same village. However, only two villages were included, and the analyses consisted of village-level comparisons; hence, accounting for clustering was not possible. Without controlling for village effects and given the large differences in fluoride concentrations and IQ between villages, the apparent dose-response relationship could be due to a village effect in addition to a fluoride effect. However, the dose-response relationship is still present within the "exposed" village, diminishing the concern for a village-only effect.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that statistical analyses were appropriate and that there were no other threats to risk of bias.
- ***Basis for classification as lower risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements and outcomes blindly assessed but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key confounders were considered in the study design or analysis, but there is potential for the presence of arsenic to bias toward the null.

Xiang *et al.* (2011)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 8–13 years (same population as Xiang *et al.* (2003a))
- ***Study area:*** Wamiao and Xinhuai villages located in Sihong County, Jiangsu Providence, China
- ***Sample size:*** 512 school children
- ***Data relevant to the review:*** Mean IQ scores and odds ratio for having an IQ < 80 presented by serum fluoride quartiles.
- ***Reported association with fluoride exposure:*** Yes: Significant trend on association between quartiles of serum fluoride and children's IQ score < 80 (adjusted ORs for Q1 and Q2; Q1 and Q3; and Q1 and Q4, respectively: 1; 2.22 [95% CI: 1.42, 3.47]; and 2.48 [95% CI: 1.85, 3.32]); significant effects at ≥ 0.05 ppm fluoride.

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* The study population is the same as that was used in the Xiang *et al.* (2003a) study, but a few more measurements were available and different analyses were conducted. The comparison population is considered the same as previously based on the study populations being recruited from similar populations, using similar methods, during the same time frame. Demographic characteristics were not provided.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the exposure groups were similar and were recruited using the same methods within the same time frame, with direct evidence that there was no difference in participation/response rates.
- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* As was noted in the 2003 publication, information was collected on personal characteristics, medical history, education levels in the children and parents, family SES, and lifestyle. In the logistic regression model, age and gender were adjusted in the analysis. In the previous report, no significant associations were observed between groups for family income and parents' education. Urinary iodine and blood lead levels were also stated to be measured and were noted not to be significantly different between the groups. Although the iodine levels were reported in the previous publication, the lead levels were not reported nor were the methods. Lead information is reported in a letter to the editor (Xiang *et al.* (2003b)) and was not significantly different between the areas. Although arsenic was not addressed specifically in this publication, Xiang *et al.* (2013) measured both fluoride and arsenic in the Wamiao and Xinhuai areas. Xinhuai (the low fluoride area) had significantly higher arsenic levels compared to Wamiao (the endemic fluorosis area). This is likely to bias toward the null; however, the study observed a significantly lower IQ score in the endemic fluorosis area and with increasing serum fluoride.
  - *Potentially important study-specific confounders:* Arsenic often occurs in the drinking water along with fluoride in some Chinese populations; however, based on information provided in Xiang *et al.* (2013), arsenic concentrations were higher in the low fluoride area compared to the high fluoride area.
    - *Direction/magnitude of effect***:** Presence of arsenic in this study population would potentially bias toward the null.
  - *Basis for rating:* Probably low of risk bias because there is indirect evidence that the key confounders are taken into account, methods used for collecting the information were valid and reliable, and co-exposures to arsenic and lead and iodine deficiency are not attributing to the effects observed in this area.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Data are reported for all 512 children noted to be included in the study.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.

- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Fluoride levels were measured in serum with a fluoride ion-selective electrode. A fasting venous blood sample was used. No details are provided on validation (including correlation with drinking water levels) or QA. Children who did not reside in their village for at least 2 years were excluded. Results were provided in quartiles, but they combined the lower two quartiles with results ranging from <0.05 mg/L to >0.08 mg/L.
    - *Direction/magnitude of effect***:** Serum fluoride may not be the best estimate for exposure. There is potential for exposure misclassification because only current levels were assessed. Migration of subjects in or out of the area was not assessed, but the study authors noted that, if the children had been absent from the village for 2 or more years, they were excluded. Misclassification would likely be non-differential, which could bias results in either direction.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* IQ was assessed as part of the 2003 evaluation. IQ was measured with the Combined Raven's Test for Rural China which is appropriate for this population (++ for methods). Although this study does not provide details, the original study article from 2003 provides specific details. The study authors indicate in the 2003 publication that the tests were conducted in a double-blind manner and these are the same results and population (++ for methods). Overall rating=++
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses conducted were appropriate for the study. Chi square tests were used to compare categorical variables, and logistic regression was used to evaluate the association between serum fluoride levels and risk of low IQ. A potential concern raised by the NASEM (2020) peer review was the lack of accounting for relationships in exposure between persons from the same village. However, only two villages were included, and the analyses consisted of village-level comparisons; hence, accounting for clustering was not possible. Without controlling for village effects and given the large differences in

fluoride concentrations and IQ between villages, the apparent dose-response relationship could be due to a village effect in addition than a fluoride effect. However, the dose-response relationship is still present within the "exposed" village, diminishing the concern for a village-only effect.

- *Other potential concerns:* None identified.

- *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.

- ***Basis for classification as lower risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements with blinding at outcome assessment but is limited by the cross-sectional study design and use of serum concentrations. All key confounders were considered in the study design or analysis, but there is potential for the presence of arsenic to bias toward the null.

Yu *et al.* (2018)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 7–13 years
- ***Study area:*** Tianjin City, China
- ***Sample size:*** 2,886 school children
- ***Data relevant to the review:*** IQ for normal (≤ 1 mg/L) versus high (> 1 mg/L) water fluoride; betas for IQ score by water and urine fluoride groupings; ORs by IQ category using water and urine fluoride levels.
- ***Reported association with fluoride exposure:*** Yes: Significant difference (p = 0.036) in mean IQ scores in high (106.4 ± 12.3) versus normal (107.4 ± 13.0) water fluoride areas. Distribution of IQ scores was also significantly different (p = 0.003); every 0.5-mg/L increase in water fluoride (between 3.40 and 3.90 mg/L) was associated with an IQ score 4.29 points lower (95% CI: −8.09, −0.48).

Risk of Bias:

- **Author contacts:**
  - Authors were contacted in September 2018 to obtain additional information for the risk-of-bias evaluation.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* School children (2,886), aged 7–13 years, were recruited from the rural areas of Tianjin City, China. After exclusion, 1,636 children were assigned to the "normal-fluoride" exposure group and 1,250 were assigned to the "high-fluoride" exposure group based on a cut-off water fluoride level of 1.0 mg/L. A multi-stage random sampling technique, stratified by area, was performed to select representative samples among local children who were permanent residents since birth. Detailed characteristics of the study population are provided. Exclusion criteria included: 1) children who had

congenital or acquired diseases affecting intelligence, 2) children with a history of cerebral trauma and neurological disorders, 3) children with a positive screening test history (like hepatitis B virus infection, Treponema palladium infection and Down's syndrome), and 4) children with adverse exposures (smoking and drinking) during maternal pregnancy. A table of characteristics was provided by fluoride level with differences adjusted in the analysis.

- *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposed groups were recruited using similar methods during the same time frame and that any differences between the exposed groups were considered in the statistical analyses.

- ***Confounding:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Demographic data were collected by trained investigators during a face-to-face interview with the recruited children and their parents. Questionnaires were not stated to be validated. The developmental status of the children was further assessed by calculation of BMI, and all measurements were conducted by nurses based on recommended standard methods. Variables that presented differential distribution between the normal-fluoride and high-fluoride exposure groups were adjusted in the linear regression analysis of IQ data and included age, sex, paternal and maternal education levels, and low birth weight. Children exposed to smoking in utero were excluded from the study. Sensitivity analyses were conducted by modifying covariates adjusted in multivariable models among demographics (age and sex); development (BMI); socioeconomics (maternal education, paternal education, and household income); history of maternal disease during pregnancy (gestational diabetes, malnutrition, and anemia); and delivery conditions (hypoxia, dystocia, premature birth, post-term birth, and low birth weight). None of the study sites selected were in areas endemic for iodine deficiency disorders nor were other potential neurotoxins like lead, arsenic, and mercury present. Variables such as parental BMI and behavioral and mental health disorders were not addressed.
  - *Potentially important study-specific confounders:* All key confounders were considered in this study.
    - *Direction/magnitude of effect***:** Not applicable.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that methods of obtaining the information were valid and reliable and direct evidence that all key confounders and co-exposures were addressed.
- ***Attrition:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* There were 1,636 children assigned to the "normal-fluoride" exposure group based on water fluoride, and 1,250 children were assigned to the "high-fluoride" exposure group. Exclusion from the original group of 2,886 children was adequately described. A total of 2,380 children provided urine samples. There is no indication that the data presented excludes any additional children or urine samples, but results do not indicate a sample size for all results.

- *Basis for rating:* Probably low risk of bias based on indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

- ***Exposure:***
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* According to the annual surveillance data from the CDC, the drinking water sources and water fluoride concentrations in each village had remained at stable levels over the past decade. During the investigation, water samples were collected randomly from the public water supplies in each village. Spot (early-morning) urine samples from every child and water samples from each village were collected in pre-cleaned, labeled polythene tubes and transported to the lab within 24 hours while frozen. Samples were stored at −80°C until analysis. Concentrations of fluoride ions (mg/L) were analyzed using the national standardized ion-selective electrode method in China; the detection limit was 0.01 mg/L. Samples were diluted with an equal volume of total ionic strength adjusted buffer (TISAB) of pH 5–5.5 for optimal analysis. Double-distilled deionized water was used throughout the experiment. There is no reporting of any QC methods.
    - *Direction/magnitude of effect*: Spot urine samples may lead to non-differential exposure misclassification. The large population size likely dilutes any potential effects of occasional misclassification. Because the drinking water sources of fluoride had been noted to be stable for the past decade and the children were 13 years or younger, there would only be exposure misclassification if there was a lot of migration between areas.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that exposure was consistently assessed using well-established methods that directly measured exposure.

- ***Outcome:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* IQ scores were measured using the second edition of Combined Raven's Test-The Rural in China (CRT-RC2) for children aged 7–13 years (++ for methods). The test was completed by each participant within 40 minutes according to the instruction manual. For each test, 40 children were randomly allocated to one classroom to take the test independently under the supervision of four trained professionals. There is no mention of whether the evaluators were blinded to the fluoride group of each child (normal vs. high fluoride) or whether there were steps taken to ensure consistency in scoring across the evaluators. It is also not clear if the 40 children randomly assigned to the classroom were specific to the village or if a local center was used. Correspondence with the study authors indicated that the four professionals worked together throughout the examination without knowledge of the child's fluoride exposure (++ for blinding).
  - *Basis for rating:* Definitely low risk of bias based on the direct evidence that the outcome was blindly assessed using instruments that were valid and reliable.

- ***Selective Reporting:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* All outcomes outlined in the abstract, introduction, and methods are reported in sufficient detail.

  - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:*
    - *Statistical analyses:* Statistical analyses used were appropriate for the study.
    - *Other potential concerns:* None identified.
  - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements with blinding at outcome assessment but is limited by the cross-sectional study design and lack of accounting for urine dilution. All key confounders including potential co-exposures were considered in the study design or analysis.

Zhang *et al.* (2015b)

Study Details:

- ***Study design*:** Cross-sectional
- ***Population*:** Children aged 10–12 years
- ***Study area:*** Tianjin City, China
- ***Sample size:*** 180 children
- ***Data relevant to the review:*** IQ by control and high fluoride groups; IQ correlations with water, serum, or urinary fluoride levels; betas for IQ with urinary fluoride levels (by genotypes)
- ***Reported association with fluoride exposure:*** Yes: Significant correlation between IQ score and serum fluoride (r = –0.47) and urinary fluoride (r = –0.45); significant difference in IQ score for high-fluoride area (>1 mg/L; 102.33 ± 13.46) compared with control area (<1 mg/L; 109.42 ± 13.30).

Risk of Bias:

- **Author contacts:**
  - Authors were not contacted for additional information because it was not necessary.
- ***Population selection:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary*: Subjects were similar and recruited during the same time frame using the same methods. Authors recruited schoolchildren from a high fluoride area (1.40 mg/L) and a control area (0.63 mg/L) in Tianjin City, China. In accordance with the principles of matching social and natural factors such as educational standard, economic situation, geological environments as much as possible, two areas with different fluoride concentrations in the groundwater were selected by a stratified cluster random sampling of this region. A total of 180 5th grade children aged 10 to 12 years from two primary schools located 18 km apart in the Jinnan District were recruited—Gegu Second

Primary School (from an endemic fluorosis area) and Shuanggang Experimental Primary School (from a non-endemic fluorosis area). The areas are not affected by other drinking water contaminants, such as arsenic or iodine. All subjects were unrelated ethnic Han Chinese and residents in Tianjin with similar physical and mental health status. The authors excluded subjects with known neurological conditions including pervasive developmental disorders and epilepsy. Descriptive statistics of the study population are presented by exposure group in Table 1 of the study. A number of potential differences are taken into account in the statistical analyses.

- *Basis for rating*: Definitely low risk of bias based on direct evidence that the exposure groups were similar and recruited using similar methods during the same time frame.

- **Confounding:**
  - *Rating:* Probably low risk of bias (+)
  - *Summary:* Covariates included in the statistical models were child's age, child's gender, educational levels of parents, drinking water fluoride (mg/L), and levels of thyroid hormones (T3, T4, and TSH). Authors report that the study areas are not affected by other contaminants such as arsenic or iodine and residents were of similar physical and mental health status. Other important confounders (maternal demographics, smoking, reproductive health) were not considered. Covariate data were obtained from a study questionnaire.
  - *Potentially important study-specific confounders:* All key confounders were considered in this study.
    - *Direction/magnitude of effect*: Not applicable.
  - *Basis for rating:* Probably low risk of bias based on indirect evidence that the methods used to collect the information were valid and reliable and direct evidence that key confounders including potential co-exposures were addressed.
- ***Attrition:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Results are complete for the 180 children selected for the study.
  - *Basis for rating:* Definitely low risk of bias based on direct evidence that there was no attrition.
- ***Exposure:***
  - *Rating:* Definitely low risk of bias (++)
  - *Summary:* Drinking water samples (10 mL) were collected from the tube wells of each child's household. Three fasting venous blood samples were also collected. Urine samples were collected in the early morning before breakfast. Fluoride contents in drinking water (W-F), serum (S-F), and urine (U-F) were measured using an ion analyzer EA940 with a fluoride ion-selective electrode (Shanghai constant magnetic electronic technology Co, Ltd, China) according to the China standard GB 7484-87. All reference solutions for the fluoride determinations were double-deionized water. Parallel samples were set for determination and averages were taken. The quantitation limits of this method for W-F, S-F, and U-F were 0.2, 0.012, and 0.5 mg/L, respectively. Recovery rates for this method were in the range of 94.3%–106.4%. The intra- and inter-assay coefficients of variation for fluoride were 2.7% and 6.7%, respectively. Dilution of the urinary fluoride was not addressed.

        - *Direction/magnitude of effect*: Not applicable.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that the exposure was consistently assessed using well-established methods that directly measured exposure.
- ***Outcome:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* A Combined Raven's Test for Rural China (CRT-RC) was taken to evaluate the IQ of each child (++ for methods). The study report stated that all tests were administered at school by a trained examiner who was masked to participants' drinking water fluoride levels (++ for blinding). Overall rating for methods and blinding=++.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that the outcome was blindly assessed using instruments that were valid and reliable in the study population.
- ***Selective Reporting:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:* All results outlined in the abstract, introduction, and methods sections were reported in sufficient detail.
    - *Basis for rating:* Definitely low risk of bias based on direct evidence that all measured outcomes were reported.
- ***Other potential threats:***
    - *Rating:* Definitely low risk of bias (++)
    - *Summary:*
        - *Statistical analyses:* Statistical methods are very well-documented including testing for normality of the data.
        - *Other potential concerns:* None identified.
    - *Basis for rating:* Definitely low risk if bias based on direct evidence that the statistical analyses were appropriate and there were no other potential threats to risk of bias identified.
- ***Basis for classification as lower risk-of-bias study overall:*** Definitely or probably low risk-of-bias ratings in confounding, exposure, and outcome. Study strengths include individual exposure measurements, outcomes blindly assessed, and assessment of potential key confounders including potential co-exposures.

Appendix 5. Results of Fluoride Meta-analyses

What is the strength of the relationship between exposure to fluoride and children's IQ?

Aim 1. To update existing meta-analyses with additional studies

Approach

The approach used to perform a meta-analysis of the associations between exposure to fluoride and children's IQ levels was outlined in the associated protocol (https://ntp.niehs.nih.gov/go/785076). Details are presented below.

The mean-effect meta-analysis included studies that reported effect estimates as mean outcome measures and included measures of uncertainty such as standard deviation (SD), standard error (SE), 95% CI, and number of subjects (N) for at least one exposed and one reference exposure group. If results from multiple exposure groups were reported within a single study, the highest exposure group was considered the "exposed" group and the lowest exposure group was considered the "reference" group. A sensitivity analysis was performed to evaluate the impact of using any exposed group compared to the reference group (**Figure A-25**). This was accomplished by combining the information from the exposure groups using the approach outlined in the Cochrane Handbook for Systematic Reviews (Higgins *et al.* 2019).

When results were not reported for gender-specific groups or age-specific subgroups (<10, ≥10), they were calculated (if possible) by combining groups, following the approach outlined in the Cochrane Handbook for Systematic Reviews (Higgins *et al.* 2019). Similarly, when only mean effects, Ns, and p-values for differences between groups are reported (Lin *et al.* 1991), SDs were calculated using the SE and t-statistic (assuming equal variances) (Higgins *et al.* 2019).

The meta-analysis pooled the standardized mean difference and corresponding 95% CI using a random-effects model. Heterogeneity was assessed by Cochran's Q test (Cochran 1954) and the I^2 statistic. Forest plots were used to display results and to examine possible heterogeneity between studies. Potential publication bias was assessed by developing funnel plots and performing Egger regression on the estimates of effect size (Begg and Mazumdar 1994, Egger *et al.* 2008, Egger *et al.* 1997). If publication bias was believed to be present, trim-and-fill methods (Duval and Tweedie 2000a, b) were used to estimate the number of missing studies affected by publication bias, assess the effect of those studies on the effect estimate, and predict the impact of the hypothetical "missing" studies. To investigate sources of heterogeneity, subgroup analyses were performed by risk-of-bias evaluation, gender, age group, country, type of intelligence assessment, and type of exposure.

There were 46 studies included in the mean-effect meta-analysis (see **Table A5-1**). **Table A-2** presents information on the studies excluded from the analysis because of missing information on the number of subjects and/or the mean or variance of the outcome. Other studies were excluded because of overlapping study populations. For studies with overlapping populations (i.e., multiple studies that use the same cohort), results were selected with the most information considering the following factors: exposure metric, exposure range, exposure period, number of subjects, and statistical adjustment for potential confounders. Other studies were excluded from the mean-effect meta-analysis because they reported individual-level effects.

Summary Results

Table A5-1. Pooled SMDs and 95% CIs for Children's IQ Score and Exposures to Fluoride				
			Heterogeneity	
Analysis	**Number of Studies**	**SMD (95% CI)**	**p-value**	**I^2**
Overall Effect	46	−0.50 (−0.61, −0.39)	<0.001	89%
Subgroup Analyses				
Risk of Bias				
Lower	9	−0.31 (−0.52, −0.10)	<0.001	87%
Higher	37	−0.56 (−0.68, −0.43)	<0.001	88%
Gender[1]				
Males	12	−0.78 (−0.99, −0.56)	<0.001	75%
Females	11	−0.65 (−0.85, −0.45)	0.001	66%
Age Group				
<10 years[1]	10	−0.55 (−0.79, −0.30)	<0.001	83%
≥ 10 years	11	−0.58 (−0.78, −0.38)	<0.001	76%
Country				
China	31	−0.44 (−0.54, −0.33)	<0.001	86%
India	9	−1.02 (−1.54, −0.51)	<0.001	93%
Iran	4	−0.68 (−0.99, −0.38)	0.077	56%
Assessment Type				
CRT-RC tests	23	−0.36 (−0.47, −0.26)	<0.001	81%
Non-CRT-RC tests	23	−0.67 (−0.87, −0.47)	<0.001	90%
Raven's tests	10	-0.76 (−1.10, −0.43)	<0.001	91%
Other tests	13	−0.62 (−0.88, −0.35)	<0.001	90%
Exposure Type				
Water fluoride	28	−0.45 (−0.57, −0.34)	<0.001	85%
Dental fluorosis	7	−0.99 (−1.57, −0.41)	<0.001	96%
Other exposures[2]	11	−0.47 (−0.67, −0.27)	<0.001	85%
Sensitivity Analysis				
Any exposure *vs.* reference	46	−0.46 (−0.58, −0.35)	<0.001	92%
Previous Meta-analyses				
Duan *et al.* (2018)	26	−0.52 (−0.62, −0.42)	<0.001	69%
Choi *et al.* (2012)	27	−0.45 (−0.56, −0.34)	<0.001	80%

Notes:
CI = confidence interval; CRC-RC = Combined Raven's Test–The Rural edition in China; SMD = standardized weighted mean difference
[1]An *et al.* (1992) includes 10-year-old children in the <10 age group (7−10 years reported).
[2]Includes iodine (Ren *et al.* 1989 [translated in Ren *et al.* 2008], Lin *et al.* 1991, Wang *et al.* 2001); arsenic (Zhang *et al.* 1998, Wang *et al.* 2007); aluminum (Sun *et al.* 1991); and non-drinking water fluoride (i.e., fluoride from coal burning (Guo *et al.* 1991 [translated in Guo et al. 2008a], Li *et al.* 1994 [translated in Li et al. 2008b], Li *et al.* 1995, Wang *et al.* 1996 [translated in Wang et al. 2008b], Wang *et al.* 2005, Li *et al.* 2009, Bai *et al.* 2014)).

Overall Effect (Main Analysis)

For the group-level exposure meta-analysis, a comparison on the mean outcome measure (IQ score) was conducted across two exposure groups ("exposed" and "reference"). The random-effects pooled SMD estimated from the 46 studies included in the meta-analysis was −0.50 (95% CI: −0.61, −0.39). There was evidence of heterogeneity (I^2 = 89%, p < 0.001; **Table A5-1** and **Figure A5-1**) and publication bias (funnel plot and Egger's test p < 0.001, Begg's test p = 0.08; **Figure A5-2** and **Figure A5-3**). Eliminating publication bias through trim-and-fill analysis continued to support the finding that exposure to fluoride is associated with lower IQ in children, with an adjusted pooled effect estimate of −0.42 (95% CI: −0.54, −0.30) (**Figure A5-4** and **Figure A5-5**).

Aravind *et al.* (2016), Kundu *et al.* (2015), and Razdan *et al.* (2017) studies were excluded from the main analysis due to uncertainties about the way IQ assessments for children were performed in those studies. A sensitivity analysis was conducted that included these studies (**Figure A-35**).

Subgroup Analyses

Risk of Bias

Subgroup analysis by risk-of-bias evaluation showed that exposure to fluoride is associated with lower IQ scores in children for both higher and lower risk-of-bias studies (**Figure A5-6**), with a more severe effect for the higher risk-of-bias studies. The funnel plots and Egger's and Begg's tests of publication bias showed evidence of publication bias only among higher risk-of-bias studies (**Figure A-1**, **Figure A-2**). Eliminating publication bias through trim-and-fill analysis continued to support that exposure to fluoride is associated with lower IQ scores in children, with an adjusted pooled effect estimate of −0.35 (95% CI: −0.50, −0.21) (**Figure A-4**). There was no evidence of publication bias among lower risk-of-bias studies.

Gender

Subgroup analysis by gender showed that exposure to fluoride is associated with lower IQ scores in both males and females (**Figure A5-11)**. There was a slight suggestion of publication bias in the Egger's test for males, but not females (**Figure A-5** and **Figure A-6**). Eliminating publication bias through trim-and-fill analysis continued to support that exposure to fluoride is associated with lower IQ scores in males, with an adjusted pooled SMD estimate of −0.68 (95% CI: −0.90, −0.46) (**Figure A-8**).

Age Group

Subgroup analysis by age group showed that exposure to fluoride is associated with lower IQ scores in children regardless of age group (<10 years or ≥10 years) (**Figure A5-12**). The funnel plots and Egger's and Begg's tests of publication bias showed evidence of publication bias in children younger than 10 years old (**Figure A-9** and **Figure A-10**). Eliminating publication bias through trim-and-fill analysis for studies in children younger than 10 years old continued to support that exposure to fluoride is associated with lower IQ in children, with an adjusted pooled effect estimate of −0.55 (95% CI: −0.79, −0.30) (**Figure A-10**). There was no suggestion of publication bias in the subgroup analyses for children 10 years old and older.

Country

Subgroup analysis by country showed that exposure to fluoride is associated with lower IQ scores in children in China, India, and Iran (**Figure A5-13**), with the largest effect in India. A funnel plot with the SEs of the SMD plotted against the SMD from each study showed slight evidence of publication bias in India (**Figure A-11**). In addition, Egger's and Begg's tests of publication bias revealed evidence of publication bias for studies in India (p < 0.001, **Figure A-12**). Eliminating publication bias through trim-and-fill analysis for studies in India continued to support that exposure to fluoride is associated with

lower IQ in children, with an adjusted pooled effect estimate of −1.49 (95% CI: −2.15, −0.83) (**Figure A-13**). There was no suggestion of publication bias in the subgroup analyses for China or Iran.

Assessment Type

Subgroup analysis by assessment type showed that exposure to fluoride is associated with lower IQ scores in children tested using non-CRT-RC tests than using CRT-RC tests (**Figure A5-14**). The funnel plots and Egger's and Begg's tests of publication bias showed evidence of publication bias only among non-CRT-RC and Raven's tests (**Figure A-15** and **Figure A-16**). Eliminating publication bias through trim-and-fill analysis for studies with Raven's tests continued to support that exposure to fluoride is associated with lower IQ scores in children, with an adjusted pooled SMD estimate of −1.22 (95% CI: −1.68, −0.75) (**Figure A-19** and **Figure A-20**). There was no suggestion of publication bias in the subgroup analysis for studies using CRT-RC or other types of tests.

Exposure Type

Subgroup analysis by exposure type showed that exposure to fluoride is associated with lower IQ scores in children in studies that reported mean effects by fluoride exposure type (**Figure A5-15**). The funnel plots and Egger's test of publication bias showed evidence of publication bias for water fluoride and dental fluorosis (**Figure A-21** and **Figure A-22**). Eliminating publication bias through trim-and-fill analysis continued to support that exposure to fluoride in water is associated with lower IQ scores in children, with an adjusted pooled SMD estimate of −0.42 (95% CI: −0.53, −0.30) (**Figure A-23** and **Figure A-24**). There was no suggestion of publication bias in the subgroup analysis for studies with other exposures or non-drinking water fluoride exposures.

Overall Analysis

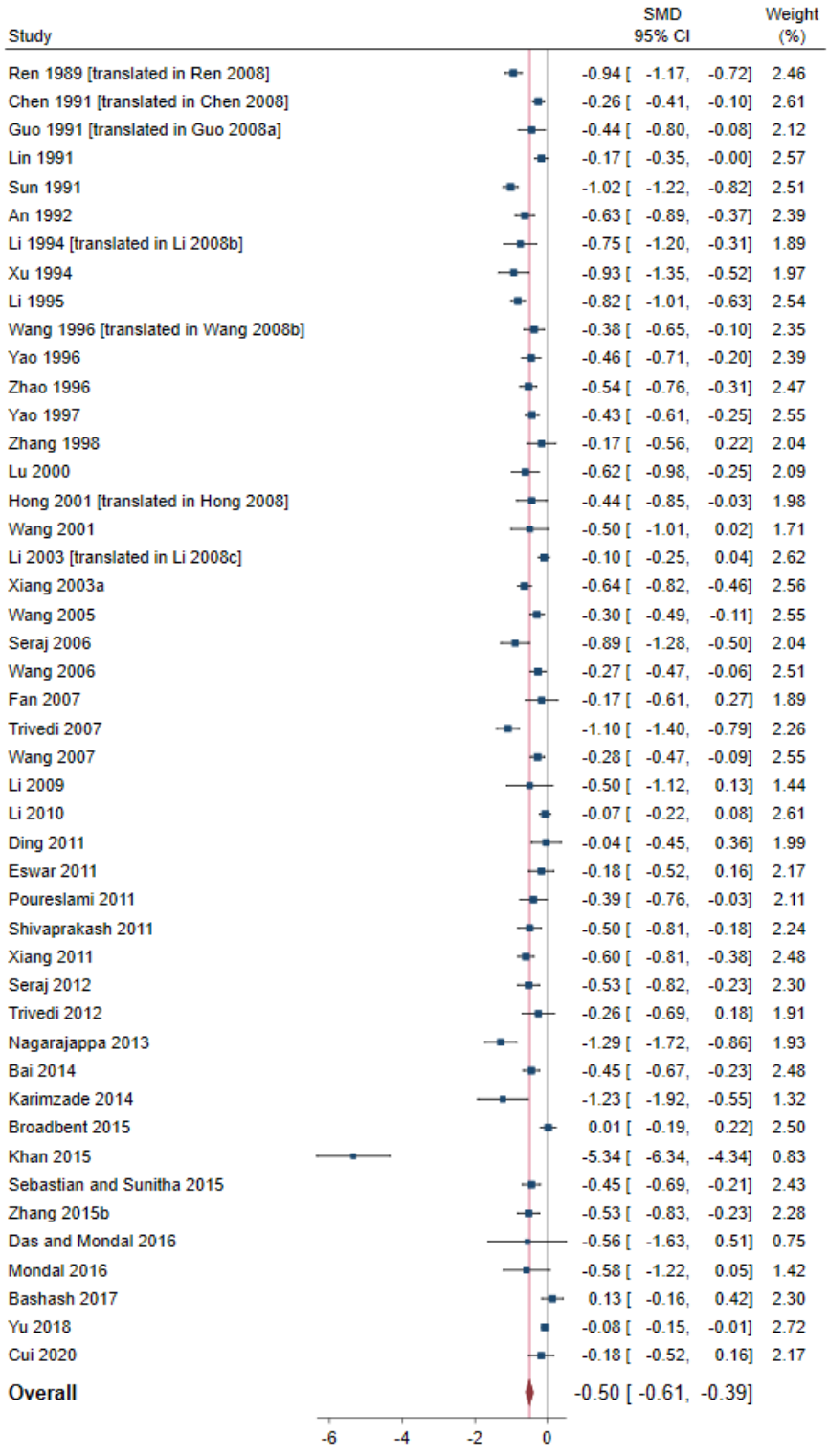


Figure A5-1. Association Between Fluoride Exposure and IQ Scores in Children: Overall Analysis

SMDs for individual studies are shown with solid boxes representing the weight, and the random-effects pooled SMD is shown as a solid diamond. Horizontal lines represent 95% CIs for the study-specific SMDs.

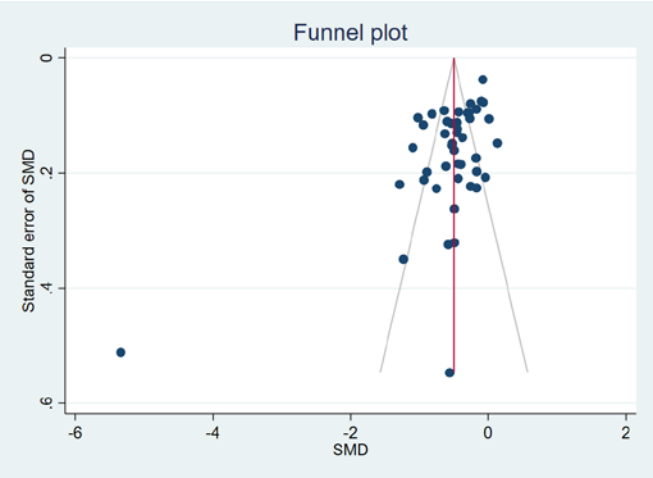


Figure A5-2. Funnel Plot of Included Studies

This funnel plot shows individual studies included in the analysis according to random-effect standardized weighted mean difference (SMD) estimates (x-axis) and the standard error (SE) of each study-specific SMD (y-axis). The solid vertical line indicates the pooled SMD estimate for all studies combined and the dashed lines indicated pseudo 95% confidence limits around the pooled SMD estimate.

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -3.21
      SE of beta1 =     0.649
                z =     -4.95
       Prob > |z| =    0.0000


Begg's test for small-study effects

Kendall's score =   -185.00
    SE of score =   105.617
              z =     -1.76
     Prob > |z| =    0.0815
```

Figure A5-3. Test for Publication Bias

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                           Number of studies =     50
  Model: Random-effects                      observed =     46
 Method: DerSimonian-Laird                   imputed =      4

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.502	-0.611	-0.393
Observed + Imputed	-0.419	-0.537	-0.301

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration                           Number of studies =     56
  Model: Random-effects                      observed =     46
 Method: DerSimonian-Laird                   imputed =      10

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.502	-0.611	-0.393
Observed + Imputed	-0.643	-0.776	-0.511

Figure A5-4. Trim-and-fill Analysis

Left panel shows the random-effects pooled SMD after filling in to the right using a run estimator (the linear estimator to the right showed no change in pooled SMD); right panel shows random-effects pooled SMD after filling in to the left using a linear estimator (the run estimator to the left showed no change in pooled SMD).

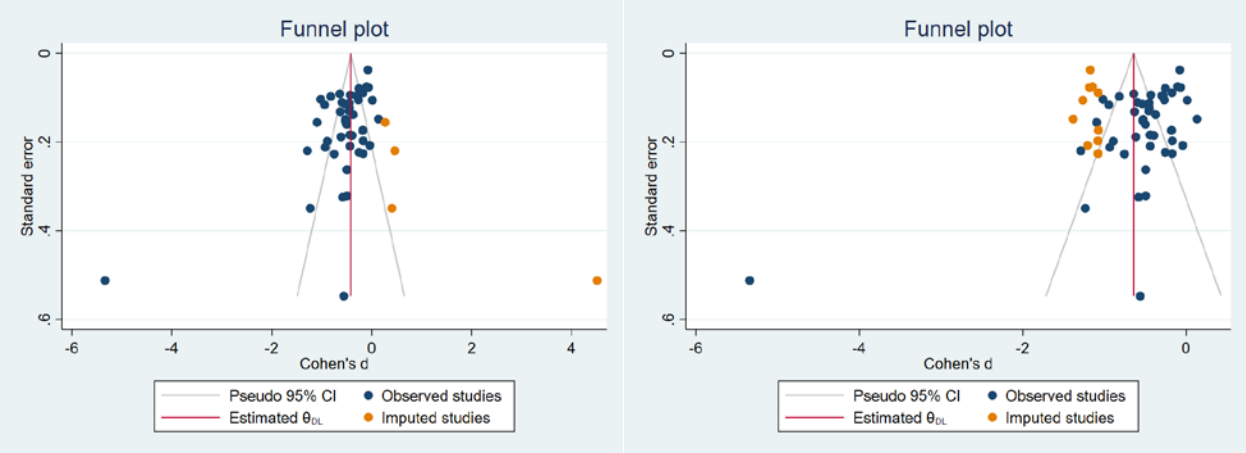


Figure A5-5. Filled-in Funnel Plots to Eliminate Publication Bias

Left panel shows the funnel plot filled in to the right using a run estimator (the linear estimator to the right showed no change in pooled SMD); right panel shows the funnel plot filled in to the left using a linear estimator (the run estimator to the left showed no change in pooled SMD).

Risk-of-bias Subgroup Analysis

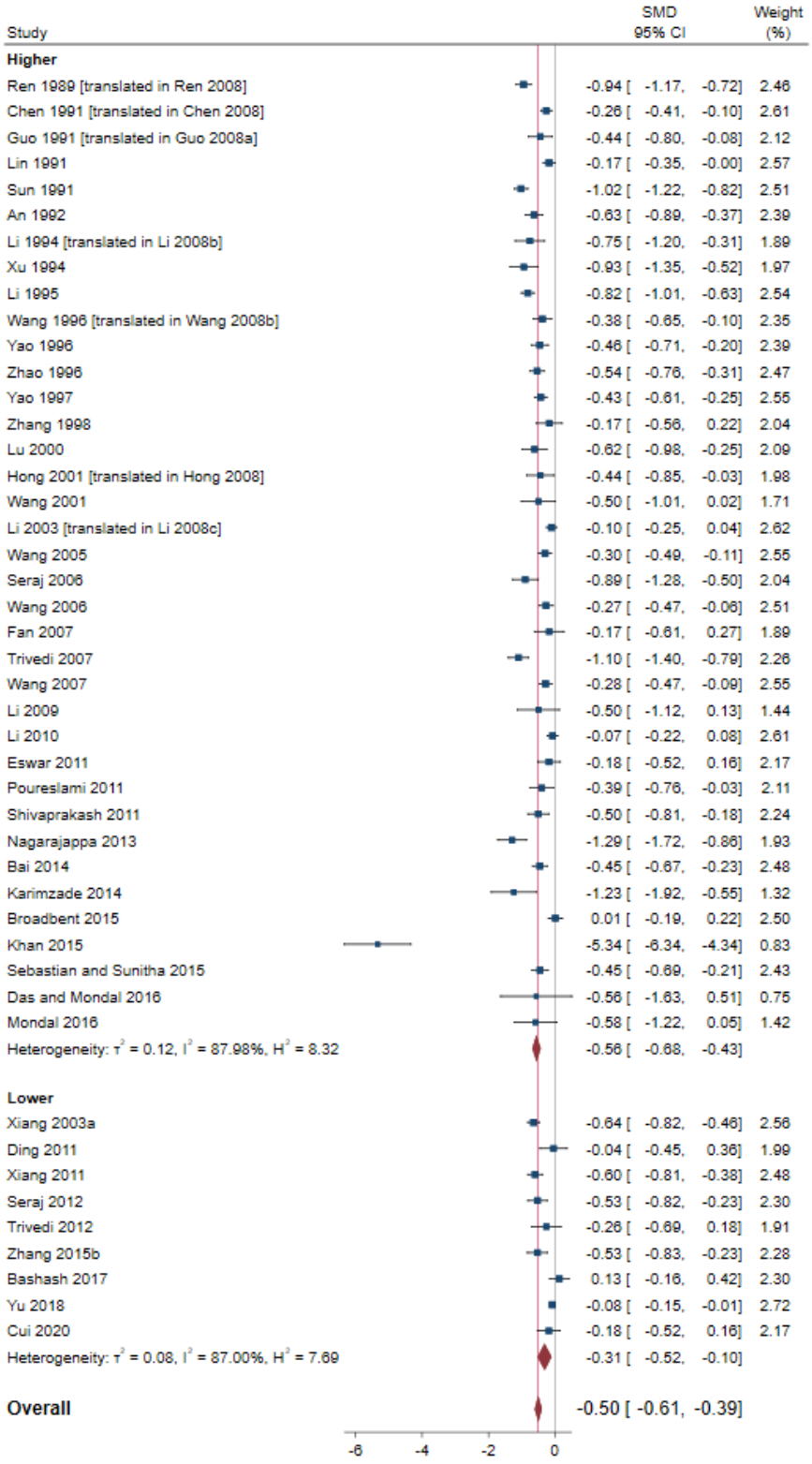


Figure A5-6. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Risk of Bias

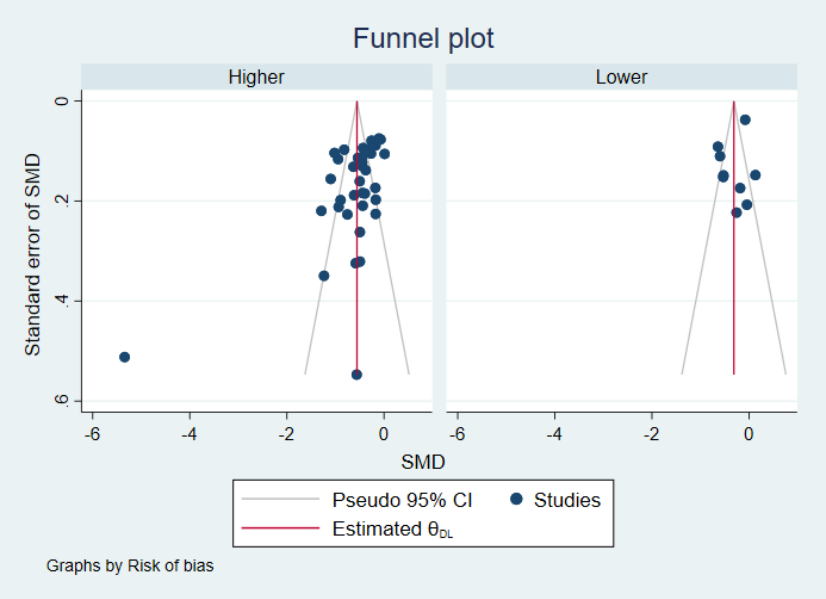


Figure A5-7. Funnel Plot by Risk-of-bias Evaluation

```
Regression-based Egger test for small-study effects
Random-effects model                                  Higher RoB
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -3.63
      SE of beta1 =     0.709
                z =     -5.12
       Prob > |z| =    0.0000

Regression-based Egger test for small-study effects
Random-effects model                                  Lower RoB
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =      0.78
      SE of beta1 =     2.155
                z =      0.36
       Prob > |z| =    0.7169
```

Figure A5-8. Test for Publication Bias by Risk of Bias

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                          Number of studies =     46
  Model: Random-effects                     observed =     37
 Method: DerSimonian-Laird                  imputed =      9

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.556	-0.684	-0.428
Observed + Imputed	-0.354	-0.498	-0.210

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration                          Number of studies =     44
  Model: Random-effects                     observed =     37
 Method: DerSimonian-Laird                  imputed =      7

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.556	-0.684	-0.428
Observed + Imputed	-0.684	-0.831	-0.537

Figure A5-9. Trim-and-fill Analysis for Higher Risk-of-bias Studies

Left panel shows the random-effects pooled SMD after filling in to the right using a run estimator (the linear estimator to the right showed no change in the pooled SMD); right panel shows random-effects pooled SMD after filling in to the left using a linear estimator (the run estimator to the left showed no change in the pooled SMD).

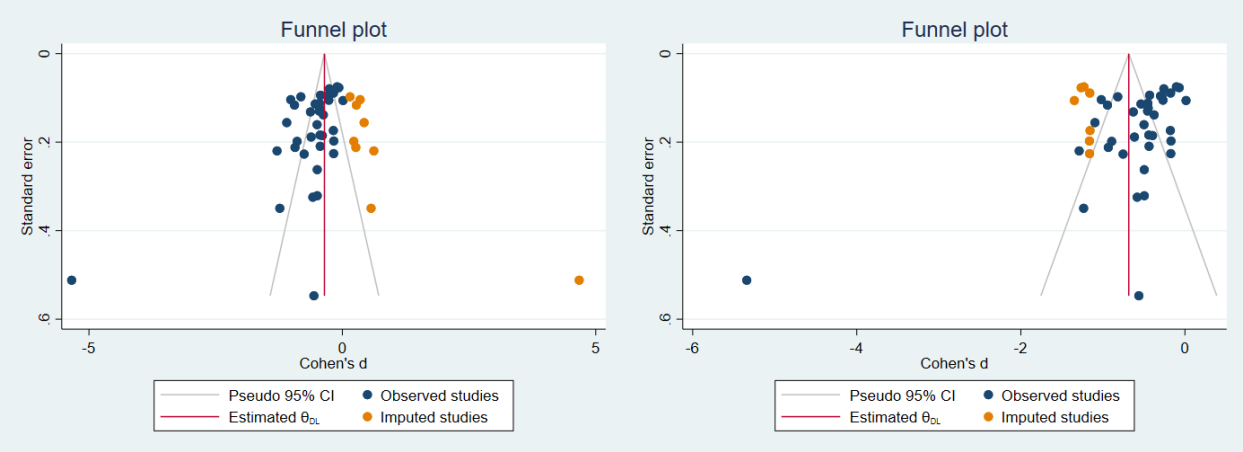


Figure A5-10. Filled-in Funnel Plots to Eliminate Publication Bias for Higher Risk-of-bias Studies

Left panel shows funnel plot filled in to the right using a run estimator (the linear estimator to the right showed no change in the pooled SMD); right panel shows the funnel plot filled in to the left using a linear estimator (the run estimator to the left showed no change in the pooled SMD).

Gender Subgroup Analysis

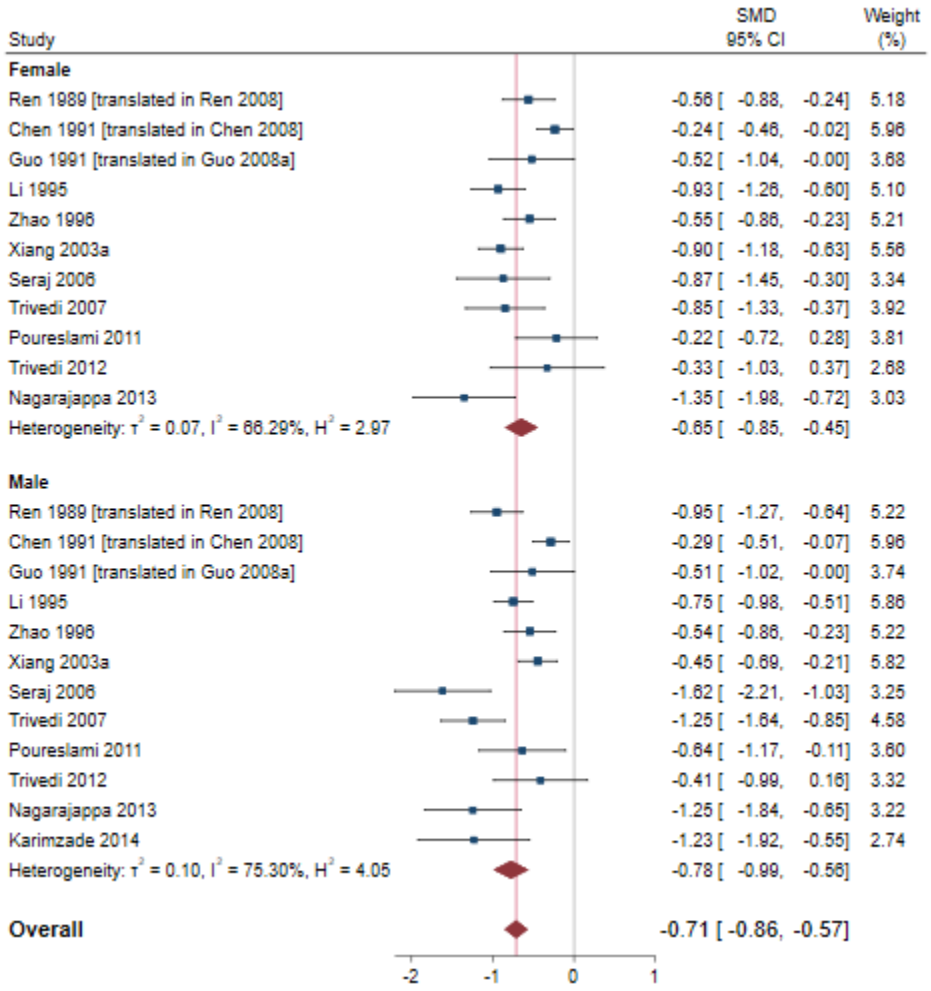


Figure A5-11. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Gender

Age Group Subgroup Analysis

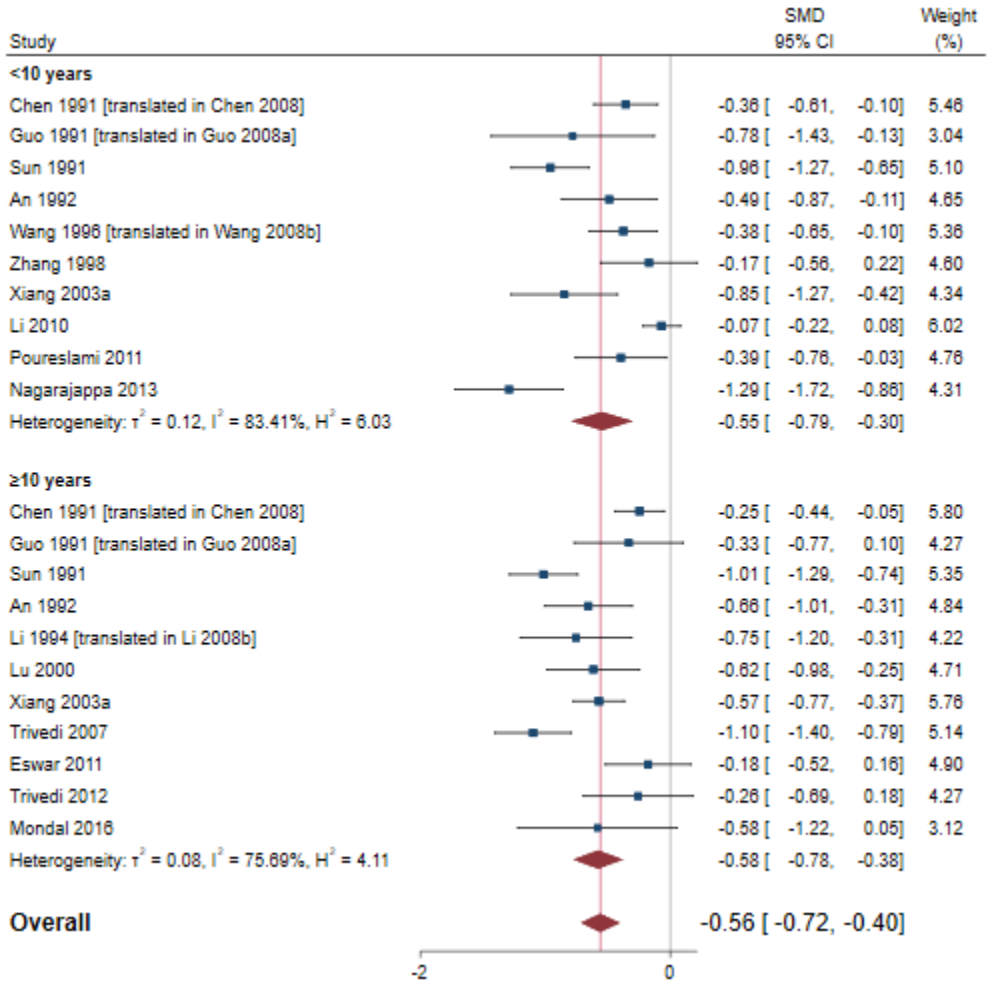


Figure A5-12. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Age Group

Country Subgroup Analysis

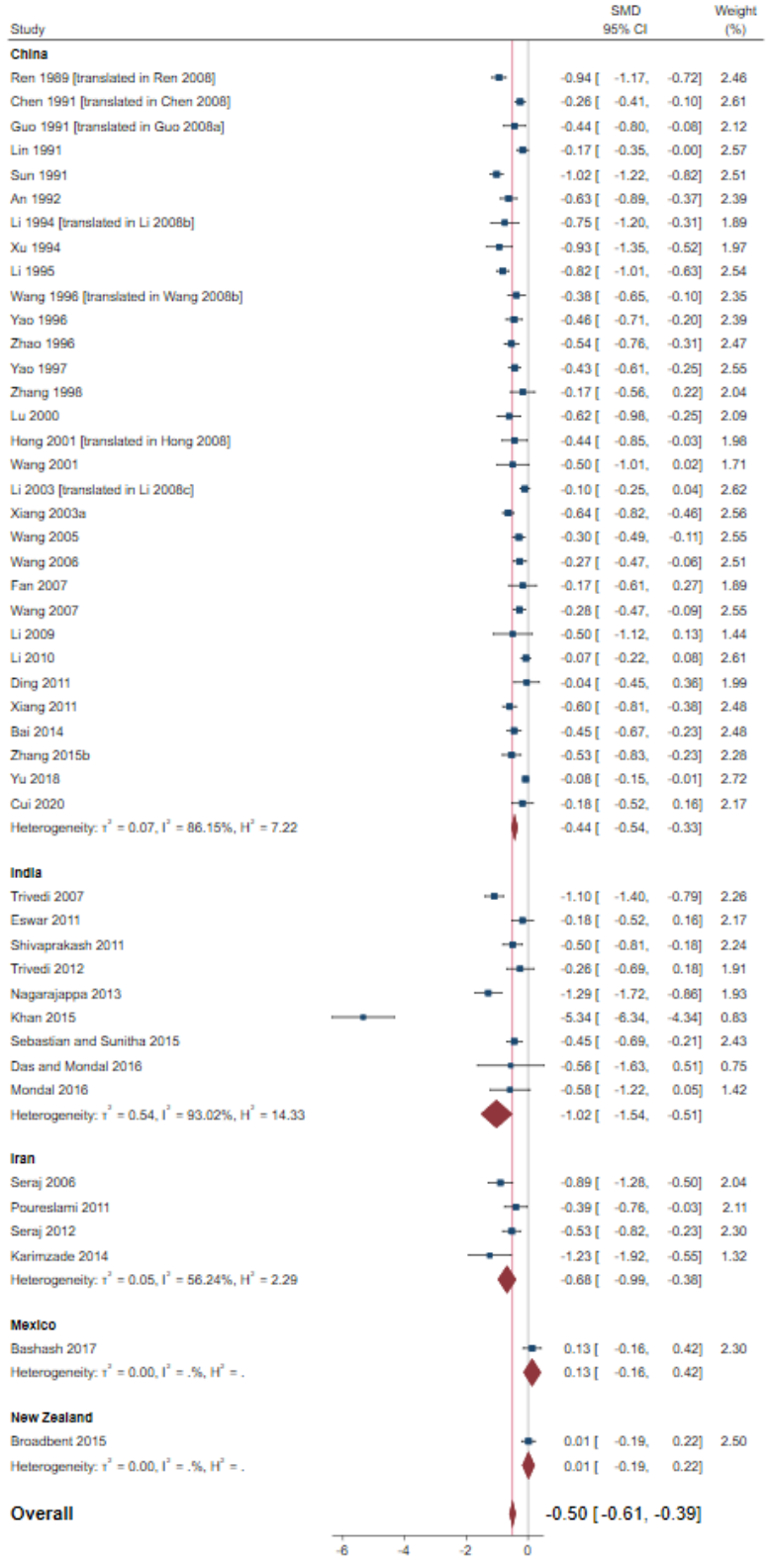


Figure A5-13. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Country

Assessment Type Subgroup Analysis

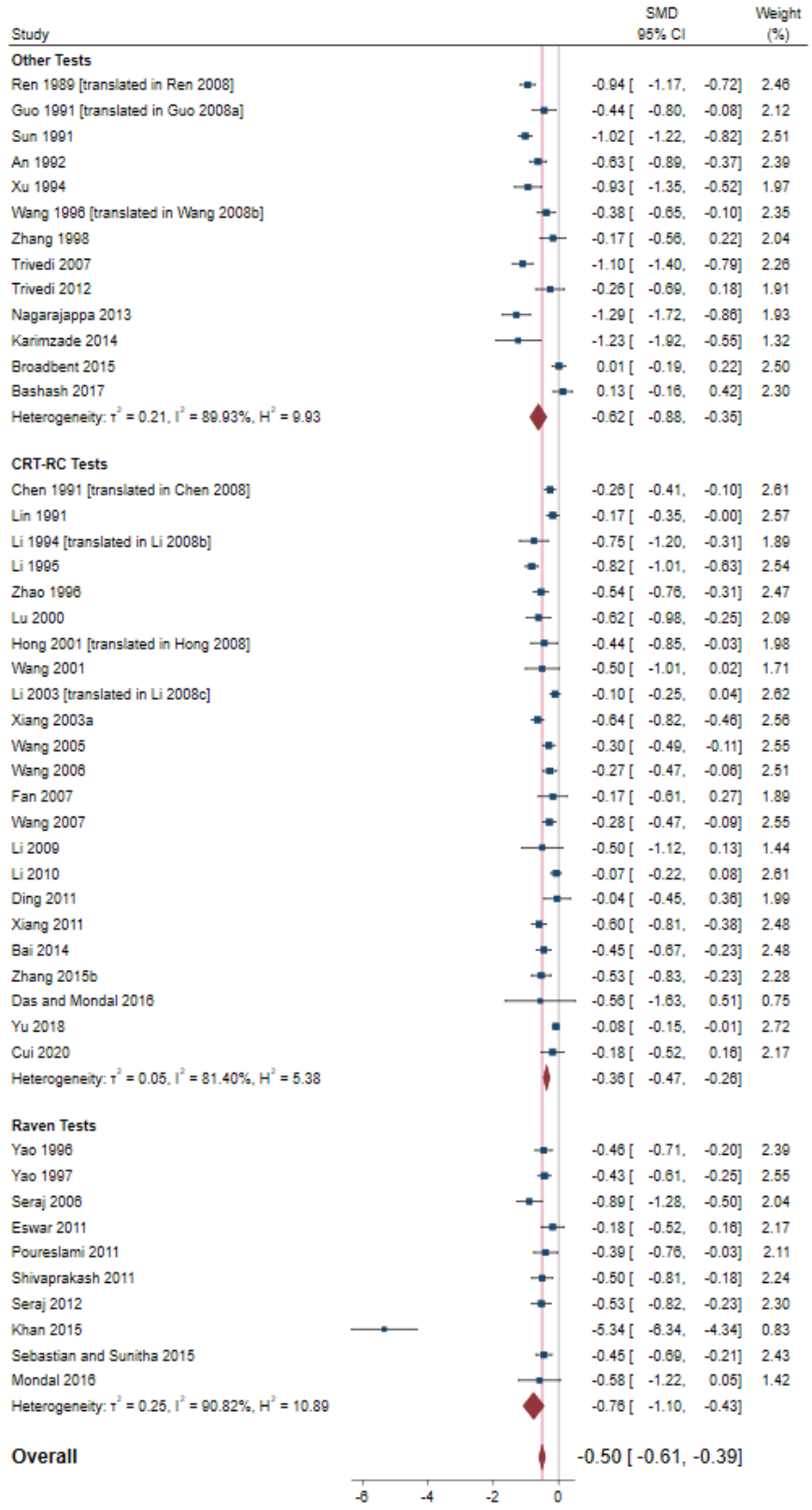


Figure A5-14. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Assessment Type

Exposure Type Subgroup Analysis

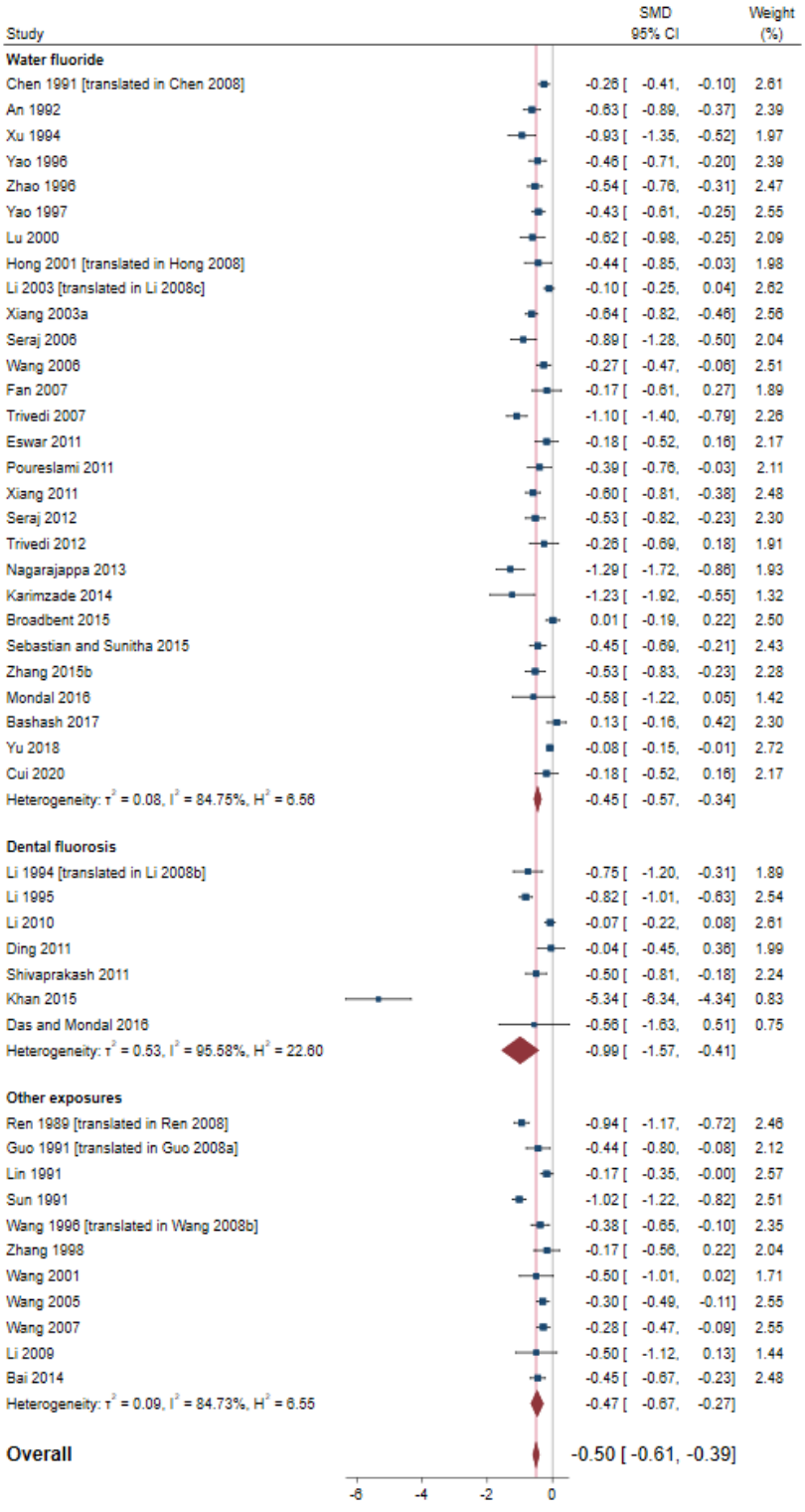


Figure A5-15. Association Between Fluoride Exposure and IQ Scores in Children: Effect by Exposure Type

Exposure types include water, dental fluorosis, and other exposures (iodine, arsenic, aluminum, and fluoride from coal burning).

Aim 2. To conduct new meta-analyses using individual-level exposure data

Approach

The individual-effect meta-analysis included studies that reported effect estimates as betas and included 95% CIs or SEs (see **Table B-1**). Adjusted effect estimates were used in the meta-analysis. If results from multiple models were reported within a single study, the most adjusted results were selected. The study outcomes were evaluated with respect to a 1-mg/L unit increase in exposure. To ensure consistent units across studies, units of exposure were transformed to mg/L as needed. For Bashash *et al.* (2017), Yu *et al.* (2018), and Till *et al.* (2020), units of exposure were transformed to levels ranging from 0.5 to 1 mg/L. For Cui *et al.* (2018), units of exposure were transformed from 1 log mg/L to 1 mg/L. Cui *et al.* (2018) reported an association between IQ and log-transformed exposure. A sensitivity analysis was performed to evaluate the impact of using Cui *et al.* (2018), since the relationship between IQ and exposure evaluated in this study was not linear (as in the other studies included). Yu *et al.* (2018) reported estimates from piecewise linear regression models, with three estimated ranges for urinary fluoride exposure (low 0.01−1.60 mg/L, medium 1.60−2.50 mg/L, and high 2.50−5.54 mg/L) and two estimated ranges for water fluoride exposure (low 0.20−3.40 mg/L and high 3.40–3.90 mg/L). Because these piecewise effect estimates are likely correlated, study-specific pooled effect estimates were used for urinary and water fluoride exposures for the overall effect meta-analysis. A sensitivity analysis was performed to evaluate the impact of using the pooled estimate rather than the piecewise estimates from Yu *et al.* (2018).

Yu *et al.* (2018) and Wang *et al.* (2020b) used the same study cohort of children recruited in 2015 from the rural areas of Tianjin City, China. Only results from Yu *et al.* (2018) were included in the meta-analysis since Wang *et al.* (2020b) used a subset (n = 571) of the original study population from Yu *et al.* (2018) (n = 2,668). A sensitivity analysis was conducted to evaluate the impact of using the effect estimate from Wang *et al.* (2020b) rather than the pooled effect estimate from Yu *et al.* (2018).

Green *et al.* (2019) and Till *et al.* (2020) used the same cohort of 398 mother-child dyads in the Maternal-Infant Research on Environmental Chemicals (MIREC) cohort that reported drinking tap water in 10 Canadian cities. Both studies reported effect estimates for maternal urinary fluoride (MUF) and water fluoride concentrations. In the Green *et al.* (2019) study, 512 mother-child pairs had MUF data (and all covariates) compared to 398 pairs in Till *et al.* (2020). Water fluoride levels were available for 420 pairs in Green *et al.* (2019) compared to 398 pairs in Till *et al.* (2020). Both references reported effect estimates adjusted for maternal education, maternal race, child's sex, HOME total score, and secondhand smoking status in the child's residence. In addition, Till *et al.* (2020) adjusted for child's age at IQ testing. The age range for study subjects was 3–4 years old. For the main analysis, results from Green *et al.* (2019) were included. A sensitivity analysis was conducted using the water fluoride result for formula-fed children and the MUF result from Till *et al.* (2020) since these are most adjusted compared to Green *et al.* (2019). For fluoride intakes, estimates from both studies were used including total fluoride intake from Green *et al.* (2019) and infant formula intake from Till *et al.* (2020).

In the overall effect analysis, for studies reporting multiple measures of fluoride exposure, results associated with measured or estimated individual-level exposures, biomarker levels (such as urinary fluoride), or fluoride intake were prioritized over results associated with water fluoride concentrations. Subgroup analyses were performed that considered all exposure types. All studies used in these analyses with individual-level effects were lower risk of bias.

The overall effect based on studies with individual-level measures of exposure showed that a 1-unit increase in fluoride exposure was associated with a significantly lower IQ score (beta = −1.40; 95% CI: −2.33, −0.47) (**Table A5-2**). A 1-mg/day increase in fluoride intake resulted in significantly lower IQ score (beta = −3.31; 95% CI: −6.12, −0.50). A 1-mg/L increase in water fluoride also resulted in significantly lower IQ score (beta = −4.77; 95% CI: −9.10, −0.45). The results for fluoride intake and water fluoride, however, are based on two studies and should be interpreted with caution.

There was evidence of moderate heterogeneity (I^2 = 46%, p = 0.101; **Table A5-2**, **Figure A5-16,** and **Figure A5-17**) in studies with individual-level urinary exposure levels. Eliminating publication bias through trim-and-fill analysis continued to support that 1-mg/L increases in individual-level urinary or water fluoride were associated with lower IQ scores, with an adjusted pooled effect estimate of −0.82 (95% CI: −1.81, 0.17) (**Figure A5-19**).

A sensitivity analysis to evaluate the impact of using the piecewise estimates from Yu *et al.* (2018) revealed no significant change in the pooled effect estimate (−1.37; 95% CI: −2.38, −0.37) (**Figure B-1**). A sensitivity analysis to evaluate using estimates from Wang *et al.* (2020b) rather than Yu *et al.* (2018) study also revealed no significant change in the pooled effect estimate (−1.24; 95% CI: −1.94, −0.54) (**Figure B-6**). A sensitivity analysis using the water fluoride result for formula-fed children and the MUF result from Till *et al.* (2020) (rather than Green *et al.* (2019)) suggested no significant change in the overall pooled effect estimate (−1.50; 95% CI: −2.44, −0.57) (**Figure B-11**).

Summary Results

Table A5-2. Pooled Effect Estimates and 95% CIs for Children's IQ Scores and Individual-level Exposures to Fluoride

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Effect				
Full-scale IQ	6	−1.40 (−2.33, −0.47)	0.101	46%
Verbal IQ	6	−1.36 (−2.28, −0.45)	0.110	44%
Performance IQ	6	−1.33 (−2.25, −0.42)	0.117	43%
Subgroup Analyses				
Gender				
Males	2	−2.23 (−5.45, 0.99)	0.092	65%
Females	3	−1.64 (−4.80, 1.51)	0.045	68%
Country				
Canada	1	−1.95 (−5.18, 1.29)	NA	
China	4	−0.84 (−1.39, −0.30)	0.342	10%
Mexico	1	−5.00 (−8.53, −1.47)	NA	
Assessment Type				
CRT-RC tests	4	−0.84 (−1.39, −0.30)	0.342	10%
Non-CRT-RC tests	2	−3.39 (−6.37, −0.41)	0.212	36%
Exposure Type				
Urinary fluoride	6	−1.40 (−2.33, −0.47)	0.101	46%
Intake	2	−3.31 (−6.12, −0.50)	0.746	0%
Water fluoride	2	−4.77 (−9.09, −0.45)	0.707	0%

Notes:
CI = confidence interval; NA = not applicable

Overall Analysis

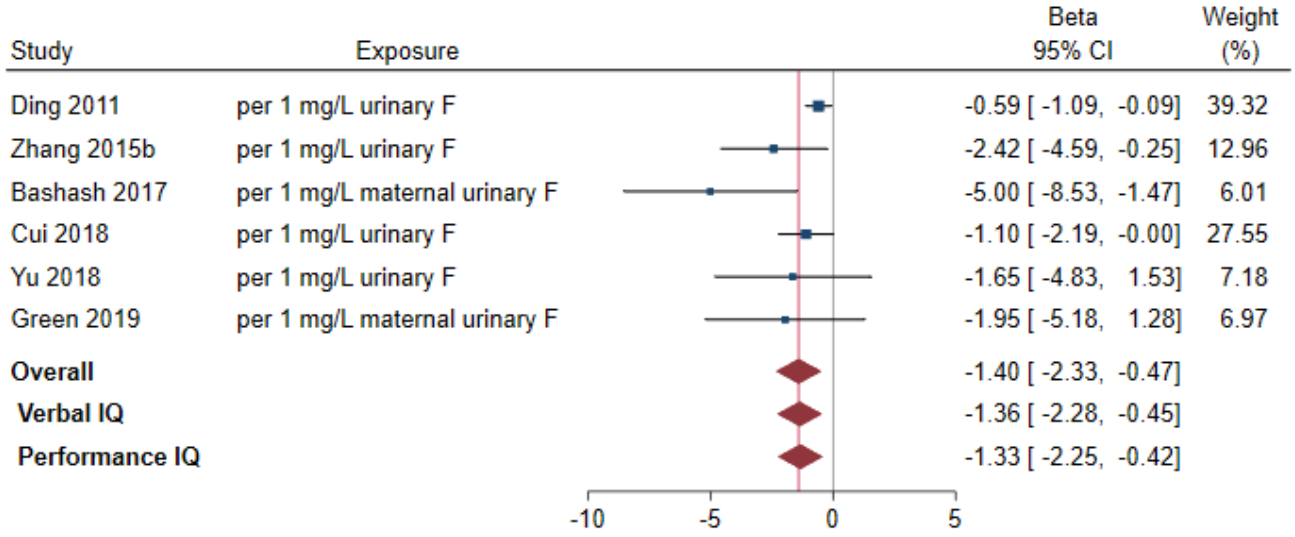


Figure A5-16. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Overall Analysis

Estimates (betas) for individual studies are shown with solid boxes representing the weight, and the pooled estimate is shown as a solid diamond. Horizontal lines represent 95% Cis for the study-specific betas.

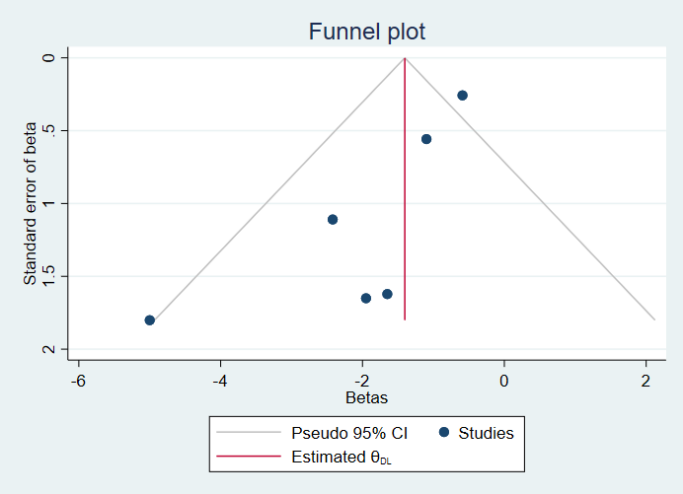


Figure A5-17. Funnel Plot of Included Studies with Individual-level Exposures

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -1.68
      SE of beta1 =     0.628
                z =     -2.67
      Prob > |z| =    0.0075
```

Figure A5-18. Test for Publication Bias for Studies with Individual-level Exposures

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the right

Iteration                          Number of studies =      9
  Model: Random-effects                     observed =      6
 Method: DerSimonian-Laird                  imputed =       3

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Beta	[95% Conf.	Interval]
Observed	-1.403	-2.333	-0.473
Observed + Imputed	-0.818	-1.809	0.172

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                          Number of studies =     10
  Model: Random-effects                     observed =      6
 Method: DerSimonian-Laird                  imputed =       4

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Beta	[95% Conf.	Interval]
Observed	-1.403	-2.333	-0.473
Observed + Imputed	-0.737	-1.671	0.197

Figure A5-19. Trim-and-fill Analysis for Studies with Individual-level Exposures

Left panel shows the random-effects pooled effect estimate after filling in to the right using a liner estimator; right panel shows random-effects pooled effect estimate after filling in to the right using a run estimator. Filling in to the left using a linear or a run estimator showed no change in the pooled slope.

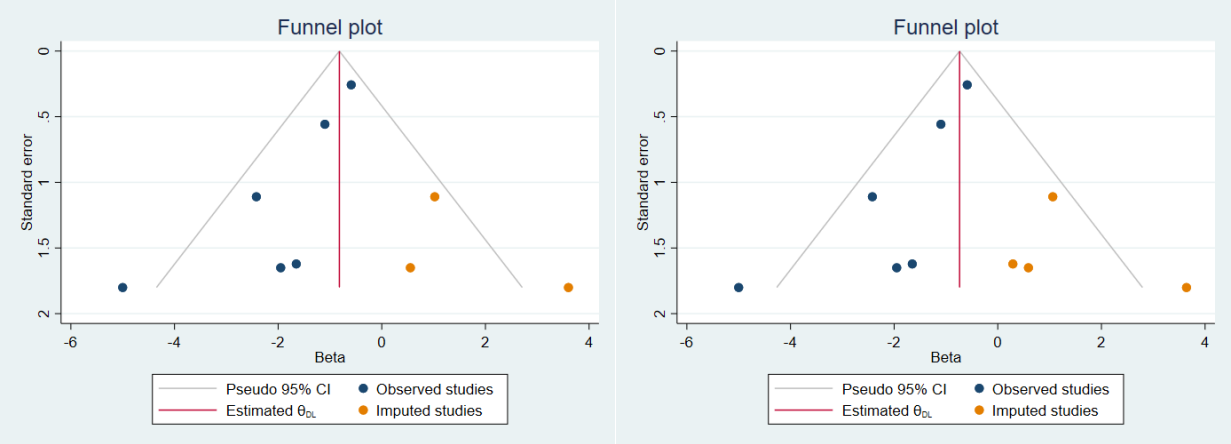


Figure A5-20. Filled-in Funnel Plots to Eliminate Publication Bias for Studies with Individual-level Exposures

Left panel shows funnel plot filled in to the right using a linear estimator; right panel shows the funnel plot filled in to the right using a run estimator. Filling in to the left using a linear or a run estimator showed no change in the pooled effect estimate.

Exposure Type Subgroup Analysis

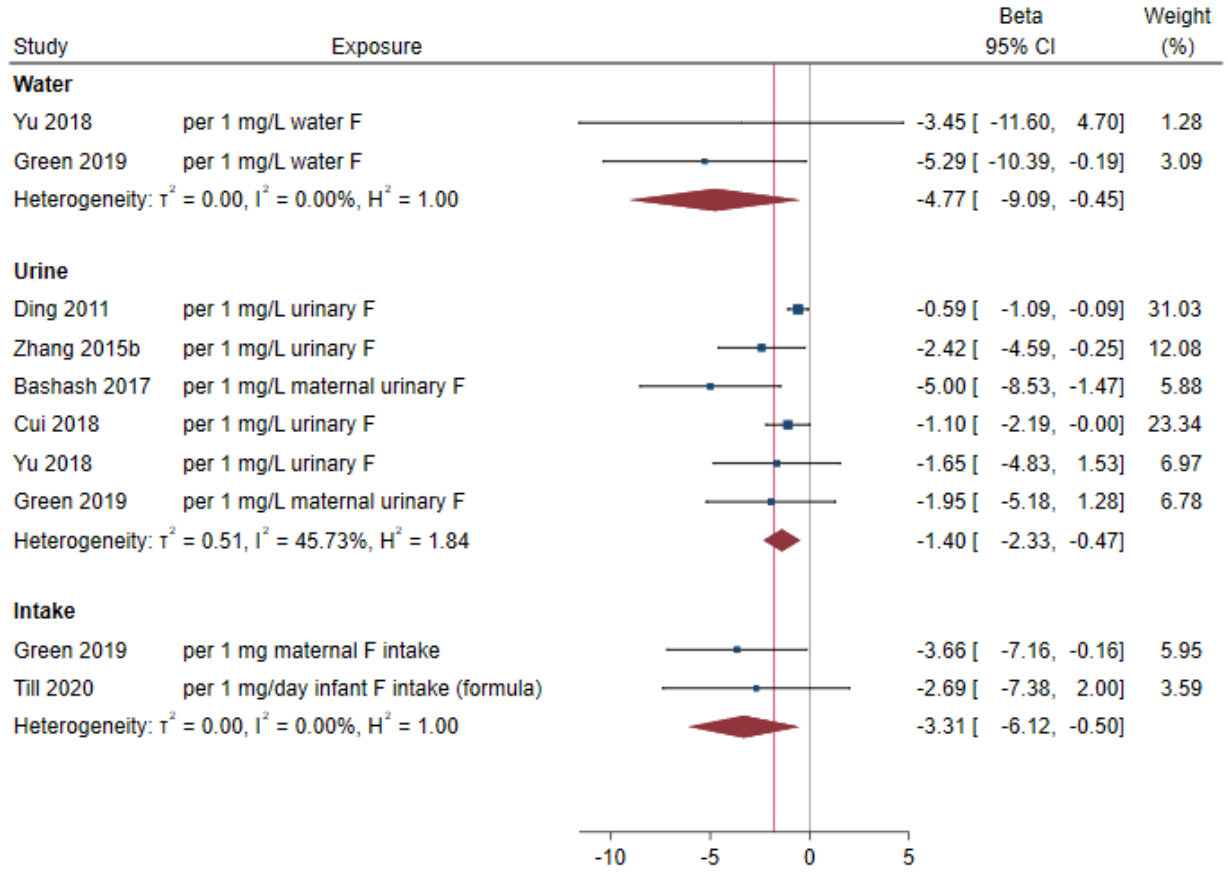


Figure A5-21. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Exposure Type

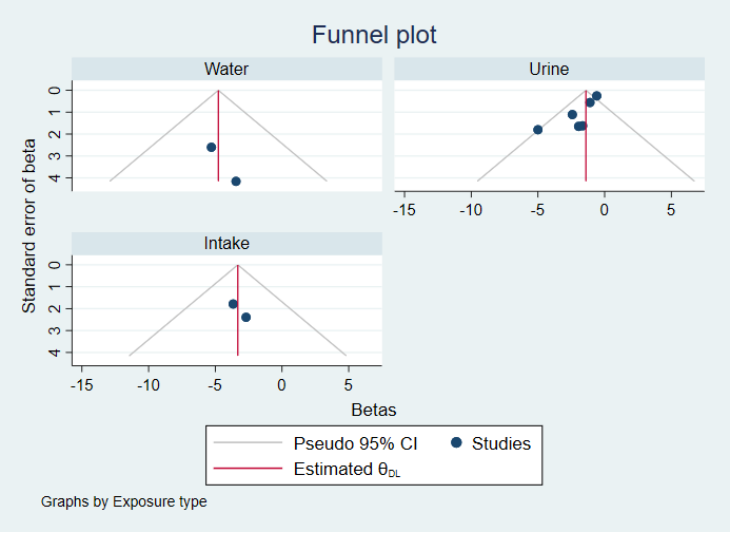


Figure A5-22. Funnel Plot of Included Studies

Note: The analysis for publication bias for studies with fluoride in urine is identical to the overall analyses. The analyses for publication bias for studies with fluoride in water or fluoride intake rely on two studies each and are not shown.

Country Subgroup Analysis

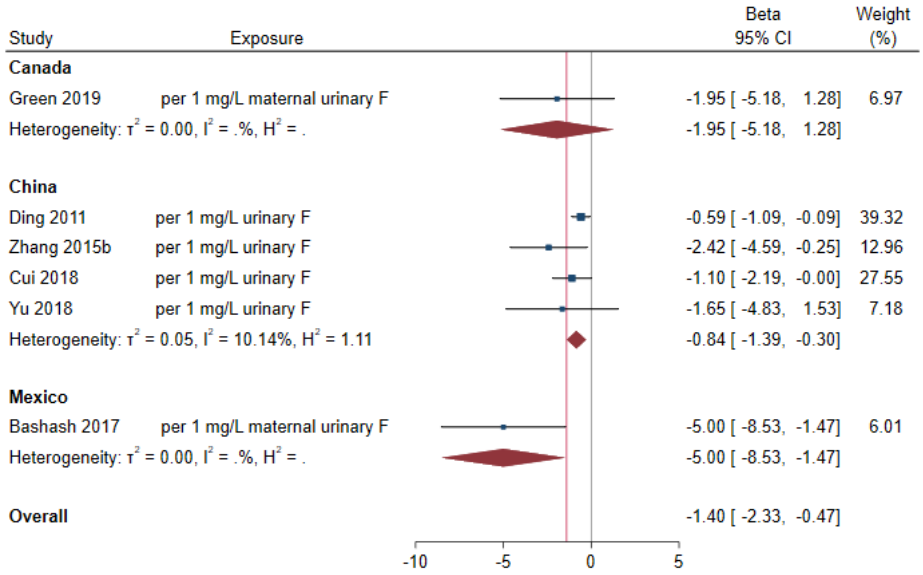


Figure A5-23. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Country

Note: The analyses for publication bias for studies from China, Canada, and Mexico rely on a very small number of studies each and are not shown.

Assessment Type Subgroup Analysis

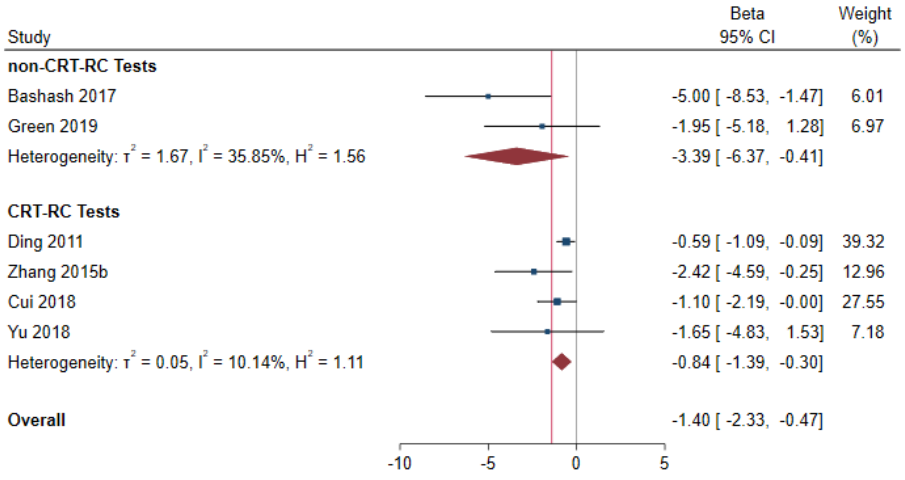


Figure A5-24. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Assessment Type

Note: The analyses for publication bias for CRT-RC studies and non-CRT-RC studies include only four and two studies, respectively, and are not shown.

Gender Subgroup Analysis

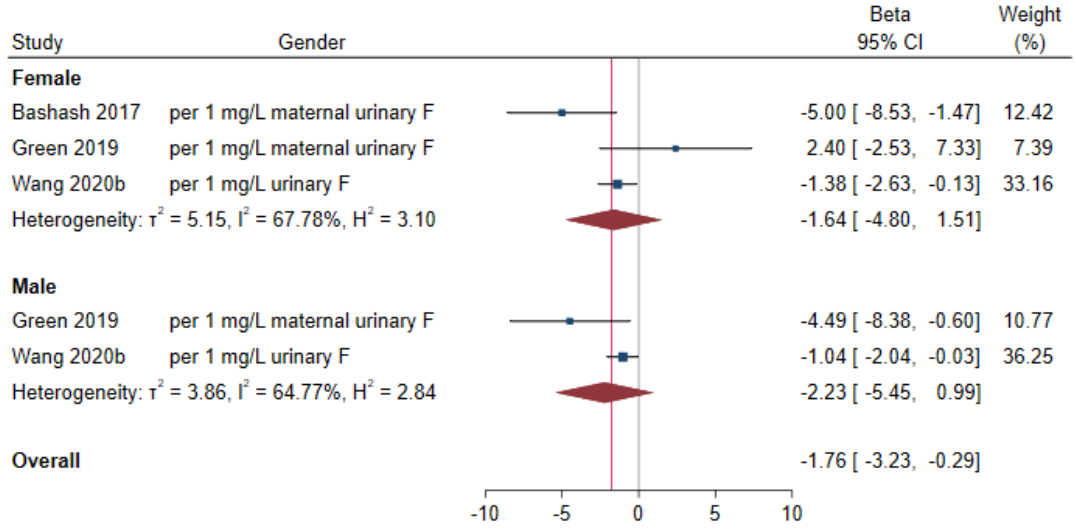


Figure A5-25. Association Between Individual-level Fluoride Exposure and IQ Scores in Children: Effect by Gender

Note: The analysis for publication bias by gender relies on two or three studies each and are not shown.

Dose-response meta-analyses using mean-effect estimates

Approach

For the dose-response meta-analysis using the mean-effect estimates, a one-step approach developed in the protocol (https://ntp.niehs.nih.gov/go/785076) and Crippa *et al.* (2018) was used. The approach uses linear mixed models to analyze all available data, including studies with one nonreference group. For each study, the median or mean fluoride intake for each exposure group was assigned to its corresponding effect estimate. If median or mean intakes by exposure group were not provided, the midpoint of the upper and lower boundaries in every category was assigned as the average intake. If the upper boundary for the highest exposure group was not reported, the boundary was assumed to have the same amplitude as the nearest category. For each study, the SMDs and corresponding SEs are used to compare the differences in mean IQ between the exposed and reference groups. The corresponding SMD for the reference group is set to zero for this analysis. The SMDs and corresponding variances are used to estimate a pooled dose-response curve using a restricted maximum likelihood estimation method. To examine a potential nonlinear relationship between exposure to fluoride and children's IQ levels, quadratic terms and restricted cubic splines were created, and a potential departure from a linear trend was assessed by testing the coefficient of the quadratic term and a second spline equal to zero. Models were compared and best model fit was determined based on the Akaike information criterion (AIC) (Müller *et al.* 2013).

A dose-response meta-analysis using the effect estimates reported in studies with individual-level exposure was considered. However, because of the small number of studies (n = 10), the various types of exposure metrics, and the different types of reported effect estimates that could not be combined, a dose-response meta-analysis of these studies could not be conducted.

Summary Results

Characteristics of the studies that compared mean IQ scores among groups of children with different levels of fluoride exposure are shown in **Table A-1**.

The meta-analysis combining data from 31 studies with fluoride levels in drinking water showed a significantly lower children's IQ score with increasing exposure (**Table A5-3** A5-3). Based on the AIC, the best model fit was achieved when restricted cubic splines with three knots were added to the linear model. However, given the small difference in AICs between the different models, and for ease of interpretability, the linear model results were chosen for the purposes of discussion although results from all models are presented in **Table A5-3.** Based on the linear model, the decrease in mean SMD between exposed and reference groups was −0.14 (95% CI: −0.19, −0.08) (**Table A5-3**). When the analysis was restricted to studies with the "high" group exposed to < 1.5 mg/L fluoride in drinking water (n = 9; 2 lower risk-of-bias studies and 7 higher risk-of-bias studies), the mean SMD became positive and nonsignificant (0.32; 95% CI: −0.57, 1.20). However, when including groups exposed to < 2 mg/L fluoride in drinking water, the mean SMD in children's IQ scores was both negative and statistically significant (SMD = −0.27; 95% CI: −0.36, −0.17) (n = 9; 2 lower risk-of-bias studies and 7 higher risk-of-bias studies).

The meta-analysis combining data from 9 studies with fluoride levels in urine showed a significantly lower children's IQ with increasing exposure (SMD = −0.18; 95% CI: −0.31, −0.05) (**Table A5-3**). Based on AIC, the best model fit was the linear model (**Table A5-3**). There was no improvement in the fit of the model when a quadratic term or restricted cubic splines were added to the linear model (**Table A5-3**). When the analyses were restricted to studies with the "high" group with < 1.5 mg/L fluoride in urine (n = 4; 2 lower risk-of-bias studies and 2 higher risk-of-bias studies), the direction of the effect did not

change, but it was no longer statistically significant (SMD = −0.13; 95% CI: −0.29, 0.03). When the dose-response meta-analysis was extended to include exposed groups with < 2 mg/L urinary fluoride (n = 6; 3 lower risk-of-bias studies and 3 higher risk-of-bias studies), the mean SMD remained negative and not statistically significant (−0.25; 95% CI: −0.71, 0.22).

Dose-Response Meta-analysis Using Mean Effects—Model Selection

Table A5-3. Model Comparison for Dose-response Meta-analysis for Children's IQ Scores (SMDs) and Exposures to Fluoride: Parameter Estimates and Model Fit[1]

Analysis	No. of Studies/ No. of Observations	Linear Model[2]	Quadratic Model[3]	Restricted Cubic Splines Model[4]
Water Fluoride				
All data	31/49	−0.14 (−0.19, −0.08) 113.6	−0.23 (−0.32, −0.14) 0.02 (0.01, 0.03) 110	−0.24 (−0.40, −0.08) 0.38 (−0.06, 0.81) 101.7
<1.5 mg/L	9/12	0.32 (−0.57, 1.20) 26.1	1.97 (−0.98, 4.92) −1.26 (−3.23, 0.71) 28.8	0.81 (−0.37, 1.99) −19.37 (−42.19, 3.44) 22.4
<2 mg/L	9/17	−0.27 (−0.36, −0.17) 44.6	0.64 (−1.04, 2.32) −0.40 (−1.09, 0.29) 31.6	0.23 (−0.71, 1.17) −4.63 (−12.47, 3.21) 27.6
<4 mg/L	24/38	−0.17 (−0.25, −0.09) 81.7	−0.28 (−0.64, 0.08) 0.03 (−0.10, 0.16) 79.5	−0.26 (−0.48, −0.03) 0.36 (−1.06. 1.79) 73.9
Urinary Fluoride				
All data	9/22	−0.18 (−0.31, −0.05) 61.1	−0.17 (−0.46, 0.12) −0.004 (−0.05, 0.04) 71.6	−0.12 (−0.31, 0.07) −0.11 (−0.35, 0.13)) 68.7
<1.5 mg/L	4/7	−0.13 (−0.29, 0.03) −0.3	−0.63 (−1.28, 0.01) 0.31 (−0.06, 0.68) 4.7	−0.30 (−0.55, −0.04) 5.09 (−1.00, 11.18) −0.9
<2 mg/L	6/11	−0.09 (−0.22, 0.03) −2.3	−0.25 (−0.71, 0.22) 0.07 (−0.13, 0.27) 5.8	−0.13 (−0.34, 0.09) 0.25 (−0.87, 1.37) 2.9

Notes:
AIC = Akaike information criterion; SMD = standardized mean difference
[1]Parameter estimates are changes in SMDs (beta [95% CI]); model fit is represented by the AIC.
[2]The estimates represent change in SMD for the linear model and AIC, respectively.
[3]The estimates represent change in SMD for the linear term, change in SMD for quadratic term, and AIC, respectively.
[4]The estimates represent change in SMD for the first spline term, change in SMD for the second spline term, and AIC, respectively.

Attachment A. Subgroup and Sensitivity Analyses (Aim 1)

Table A-1. Characteristics of Studies Included in the Mean-effect Meta-analysis

Reference[1]	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Range			
Ren *et al.* (1989) [translated in Ren *et al.* 2008]	China	8–14	No fluoride measurement High fluoride and low iodine village/low iodine village	Not specified	Wechsler Intelligence Scale for Children	Higher	Gender; iodine
Chen *et al.* (1991) [translated in Chen *et al.* 2008][w]	China	7–14	Drinking water Endemic fluorosis village/nonendemic village	0.89 mg/L (nonendemic); 4.55 mg/L (endemic)	Chinese Standardized Raven Test	Higher	Age; gender
Guo *et al.* (1991) [translated in Guo *et al.* 2008a]	China	7–13	Serum Coal burning-related fluoride endemic area/control area using wood	0.1044 ± 0.0652 mg/L (control); 0.1483 ± 0.0473 mg/L (endemic)	Chinese Binet Intelligence Test	Higher	Age; gender; SES
Lin *et al.* (1991) [w*]	China	7–14	Drinking water High fluoride and low iodine village/low iodine village/control area with iodine supplementation	0.34 mg/L (low iodine village); 0.88 mg/L (high fluoride, low iodine village)	Combined Raven's Test for Rural China	Higher	SES
Sun *et al.* (1991)	China	6.5–12	No fluoride measurement Endemic (aluminum-fluoride endemic toxicosis)/nonendemic	Fluorosis: 98.36% (endemic)	Japan's Shigeo Kobayashi's 50-point scoring method	Higher	Age
An *et al.* (1992)[w]	China	7–16	Drinking water High fluoride/nonhigh fluoride area	0.6−1.0 mg/L (nonhigh); 2.1−7.6 mg/L (high)	Wechsler Intelligence Scale for Children-Revised	Higher	Age; race; SES
Li *et al.* (1994) [translated in Li *et al.* 2008b]	China	12–13	Grain (cooked by burning high-fluoride coal) High fluoride group III (dental fluorosis present)/high fluoride group II (dental fluorosis present)/high fluoride group I (no dental fluorosis)/control group (no dental fluorosis)	0.5 mg/kg (reference); 4.7 mg/kg (group I); 5.2 mg/kg (group II); 31.6 mg/kg (group III)	Proofing test	Higher	Age; gender; SES

Table A-1. Characteristics of Studies Included in the Mean-effect Meta-analysis

Reference[1]	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Range			
Xu *et al.* (1994)[w*]	China	8–14	Drinking water Low- and high-fluoride and iodine regions/control region	0.8 mg/L (reference); 0.8 mg/L (low iodine); 0.5 mg/L (low fluoride, low iodine); 0.38 mg/L (low fluoride); 0.5 mg/L (low fluoride, high iodine); 2.0 mg/L (high fluoride, low iodine); 1.8 mg/L (high fluoride); 3.9 mg/L (high fluoride, high iodine);	Binet-Simon Scale	Higher	–
Li *et al.* (1995)[u]	China	8–13	Urine Dental fluorosis index (DFI) Fluorosis area due to soot from coal burning/nonfluorosis area	1.02 mg/L; DFI: <0.4 (nonfluorosis area); 1.81 mg/L; DFI: 0.8 (slight fluorosis area); 2.01 mg/L; DFI: 2.5 (medium fluorosis area); 2.69 mg/L; DFI: 3.2 (severe fluorosis area)	China Rui Wen Scaler for Rural Areas	Higher	Gender
Wang *et al.* (1996) [translated in Wang *et al.* 2008b][w]	China	4–7	Drinking water (well) High fluoride region/low fluoride region	0.58−1.0 mg/L (low); >1.0−8.6 mg/L (high)	Wechsler Preschool and Primary Scale of Intelligence	Higher	Age; gender
Yao *et al.* (1996)[w]	China	8–12	Drinking water Endemic fluorosis area/nonendemic area	1 mg/L (nonendemic); 2 mg/L (slightly endemic); 11 mg/L (severely endemic)	Raven Test – Associative Atlas	Higher	Iodine; SES
Zhao *et al.* (1996)[w]	China	7–14	Drinking water High fluoride village (Sima)/low fluoride village (Xinghua)	0.91 mg/L (low) 4.12 mg/L (high)	China Rui Wen Scaler for Rural Areas	Higher	Age; SES
Yao (1997)[w*]	China	7–12	Drinking water Fluorosis area without water improvements/fluorosis area with water improvements/nonfluorosis area	0.4 mg/L (nonfluorosis area); 0.33 mg/L (fluorosis area with water improvement); 2 mg/L (fluorosis area without water improvement)	Raven's Standard Progressive Matrices (China's Rural Version)	Higher	Iodine; SES

Table A-1. Characteristics of Studies Included in the Mean-effect Meta-analysis

Reference[1]	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Range			
Zhang *et al.* (1998)[w*]	China	4–10	Drinking water High fluoride and high arsenic group/high fluoride group/reference	0.58 mg/L (reference); 0.8 mg/L (high fluoride, high arsenic); 0.8 mg/L (high fluoride)	Shigeo Kobayashi 50-pt. test	Higher	Age; arsenic
Lu *et al.* (2000)[w,u]	China	10–12	Urine, drinking water High fluoride area/low fluoride area	0.37 ± 0.04 mg/L drinking water (low); 1.43 ± 0.64 mg/L urine (low); 3.15 ± 0.61 mg/L drinking water (high); 4.99 ± 2.57 mg/L urine (high)	Chinese Combined Raven Test-C2	Higher	SES
Hong *et al.* (2001) [translated in Hong *et al.* 2008][w*]	China	8–14	Drinking water High fluoride and iodine regions/reference	0.75 mg/L (reference); 2.90 mg/L (high fluoride); 2.85 mg/L (high fluoride, high iodine); 2.94 mg/L (high fluoride, low iodine); 0.48 mg/L (low fluoride, low iodine)	Chinese Standardized Raven Test	Higher	Iodine; SES; demographics
Wang *et al.* (2001)[w]	China	8–12	Drinking water Investigative point (high fluoride)/control point (low fluoride)	0.5 mg/L (low); 2.97 mg/L (high)	Combined Raven's Test for Rural China	Higher	–
Li *et al.* (2003) [translated in Li *et al.* 2008c]	China	6–13	No fluoride measurement Endemic fluorosis areas/reference	Not specified	Chinese Standardized Raven Test	Higher	–

Table A-1. Characteristics of Studies Included in the Mean-effect Meta-analysis

Reference[1]	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Range			
Xiang *et al.* (2003a)[w*]	China	8–13	Drinking water Endemic fluorosis areas/nonendemic fluorosis area	0.36 ± 0.15 mg/L (reference); 0.75 ± 0.14 mg/L (endemic fluorosis area group A); 1.53 ± 0.27 mg/L (endemic fluorosis area group B); 2.46 ± 0.3 mg/L (endemic fluorosis area group C); 3.28 ± 0.25 mg/L (endemic fluorosis area group D); 4.16 ± 0.22 mg/L (endemic fluorosis area group E); 2.47 ± 0.79 mg/L (high fluoride)	Combined Raven's Test for Rural China	Lower	Age; gender; iodine; lead; SES
Wang *et al.* (2005)	China	8–12	Drinking water High fluoride group/reference	0.48 mg/L (reference); 8.31 mg/L (high)	Chinese Combined Raven Test-C2	Higher	SES
Seraj *et al.* (2006)[w]	Iran	7–11	Drinking water High fluoride area/low fluoride area	0.4 ppm (low); 2.5 ppm (high area)	Raven Test	Higher	Gender
Wang *et al.* (2006)[w]	China	8–12	Drinking water Area severely affected by fluorosis/reference	0.73 ± 0.28 mg/L (reference); 5.54 ± 3.88 mg/L (high)	Combined Raven's Test for Rural China	Higher	–
Fan *et al.* (2007)[w]	China	7–14	Drinking water High fluoride area/low fluoride area	1.03 mg/L (low); 3.15 mg/L (high)	Chinese Combined Raven Test-C2	Higher	–
Trivedi *et al.* (2007)[w]	India	12–13	Drinking water High fluoride area/low fluoride area	2.01 ± 0.009 mg/L (low); 5.55 ± 0.41 mg/L (high)	questionnaire prepared by Professor JH Shah	Higher	Age; gender
Wang *et al.* (2007)[w]	China	8–12	Drinking water High fluoride area/reference	0.5 ± 0.2 mg/L (reference); 8.3 ± 1.9 mg/L (high)	Combined Raven's Test for Rural China	Higher	Age; gender; arsenic; SES

Table A-1. Characteristics of Studies Included in the Mean-effect Meta-analysis

Reference[1]	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Range			
Li *et al.* (2009)[u*]	China	8–12	Urine Endemic fluorosis region caused by coal burning (severe/medium/mild/reference) Degree of dental fluorosis (severe/medium/mild/very mild/suspected/normal)	0.962 ± 0.517 mg/L (reference) 1.235 ± 0.426 mg/L (mild endemic region) 1.670 ± 0.663 mg/L (medium endemic region) 2.336 ± 1.128 mg/L (severe endemic region) 0.867 ± 0.233 mg/L (normal fluorosis) 1.094 ± 0.355 mg/L (suspected fluorosis) 1.173 ± 0.480 mg/L (very mild fluorosis) 1.637 ± 0.682 mg/L (mild fluorosis) 2.005 ± 0.796 mg/L (medium fluorosis) 2.662 ± 1.093 mg/L (severe fluorosis)	Combined Raven's Test for Rural China	Higher	Age; gender
Li *et al.* (2010)	China	7–10	No fluoride measurement Dental fluorosis children/nondental fluorosis children	Not specified	Combined Raven's Test for Rural China	Higher	Gender

Table A-1. Characteristics of Studies Included in the Mean-effect Meta-analysis

Reference[1]	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Range			
Ding *et al.* (2011)[u*]	China	7–14	Urine Dental fluorosis (questionable/moderate/mild/very mild/normal) Mean urinary fluoride levels (10 groups)	0.80 ± 0.55 mg/L (normal dental fluorosis); 1.11 ± 0.74 mg/L (very mild dental fluorosis); 1.31 ± 0.78 mg/L (mild dental fluorosis); 1.46 ± 0.79 mg/L (moderate dental fluorosis); 1.13 ± 0.73 mg/L (questionable dental fluorosis); 0.26 mg/L (group 1); 0.45 mg/L (group 2); 0.56 mg/L (group 3); 0.66 mg/L (group 4); 0.75 mg/L (group 5); 0.89 mg/L (group 6); 1.08 mg/L (group 7); 1.33 mg/L (group 8); 1.74 mg/L (group 9); 2.96 mg/L (group 10)	Combined Raven's Test for Rural China	Lower	Age; arsenic; iodine; lead; SES; demographics
Eswar *et al.* (2011)[w]	India	12–14	Drinking water High fluoride villages/low fluoride villages	0.29 mg/L (low); 2.45 mg/L (high)	Standard Progressive Matrices	Higher	Age; gender
Poureslami *et al.* (2011)[w]	Iran	7–9	Drinking water Endemic dental fluorosis city/reference city	0.41 mg/L (reference); 2.38 mg/L (endemic)	Persian version of Raven's Matrices Test	Higher	Gender
Shivaprakash *et al.* (2011)[w]	India	7–11	Drinking water Fluorosis severity groups (mild/moderate/severe)/all fluorosis/no fluorosis	<0.5 ppm (no fluorosis); 2.5–3.5 ppm (mild fluorosis); 2.5–3.5 ppm (moderate fluorosis); 2.5–3.5 ppm (severe fluorosis); 2.5–3.5 ppm (all fluorosis)	Raven's Colored Progressive Matrices	Higher	Health factors; SES
Xiang *et al.* (2011)	China	8–13	Serum Quartiles (Q4/Q3/Q2 and Q1)	<0.05 mg/L (Q1 and Q2 reference); 0.05–0.08 mg/L (Q3); >0.08 mg/L (Q4)	Combined Raven's Test for Rural China	Lower	Age; gender; iodine; lead; SES

Table A-1. Characteristics of Studies Included in the Mean-effect Meta-analysis

Reference[1]	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Range			
Seraj *et al.* (2012)[w]	Iran	6–11	Drinking water High/medium/low fluoride levels	0.8 ± 0.3 mg/L (normal); 3.1 ± 0.9 mg/L (medium); 5.2 ± 1.1 mg/L (high)	Raven's Colored Progressive Matrices	Lower	Age; gender; SES
Trivedi *et al.* (2012)[w]	India	12–13	Ground water High fluoride area/low fluoride area	0.84 ± 0.38 mg/L (low); 2.3 ± 0.87 mg/L (high)	questionnaire prepared by Professor JH Shah	Lower	Gender; SES
Nagarajappa *et al.* (2013)[w]	India	8–10	Drinking water High fluoride area/low fluoride area	0.5 mg/L (low); 2.4–3.5 mg/L (high)	Seguin Form Board Test	Higher	Gender; SES; demographics
Bai *et al.* (2014)[u*]	China	8–12	Urine Coal-burning-borne fluorosis areas (seriously-affected/lightly-affected/reference)	0.54 mg/L (reference); 0.81 mg/L (lightly-affected area); 1.96 mg/L (seriously-affected area)	Chinese Combined Raven Test-C2	Higher	SES
Karimzade *et al.* (2014)[w]	Iran	9–12	Drinking water High fluoride area/low fluoride area	0.25 mg/L (low); 3.94 mg/L (high)	Iranian version of the Raymond B Cattell test	Higher	Gender
Broadbent *et al.* (2015)[w*]	New Zealand	7–13	Drinking water Area with community water fluoridation (high)/area without community water fluoridation (low) Fluoride tablet use (ever/never) Fluoride toothpaste use (always/sometimes/never)	0.0–0.3 mg/L (low); 0.7–1.0 mg/L (high) 0 mg (never used fluoride tablets); 0.5 mg (ever used fluoride tablets) Range not specified for fluoride toothpaste use (always/sometimes/never)	Wechsler Intelligence Scale for Children-Revised	Higher	Gender; SES; low birth weight; breastfeeding
Khan *et al.* (2015)	India	6–11	Drinking water High fluoride areas (Unnao)/low fluoride areas (Tiwariganj) Fluorosis grades (normal/very mild/mild/moderate/severe)	0.19 ppm (Tiwariganj); 2.41 ppm (Unnao) Range not specified by fluorosis grades	Raven's Colored Progressive Matrices	Higher	Health factors; SES
Sebastian and Sunitha (2015)[w*]	India	10–12	Drinking water Low fluoride villages/normal fluoride villages/high fluoride villages	0.40 mg/L (low); 1.2 mg/L (normal); 2.0 mg/L (high)	Raven's Colored Progressive Matrices	Higher	Age; gender; SES

Table A-1. Characteristics of Studies Included in the Mean-effect Meta-analysis

Reference[1]	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Range			
Zhang *et al.* (2015b)[w*]	China	10–12	Urine, drinking water, serum High fluoride areas/reference	1.10 ± 0.67 mg/L urine; 0.63 (0.58–0.68)mg/L water; 0.06 ± 0.03 serum (reference); 2.40 ± 1.01 mg/L urine; 1.40 (1.23–1.57) mg/L water; 0.18 ± 0.11 serum (high)	Combined Raven's Test for Rural China	Lower	Age; gender; arsenic; iodine; drinking water fluoride; SES; thyroid hormone levels; COMT genotype
Das and Mondal (2016)[u]	India	6–18	Urine, drinking water Dental fluorosis (severe/moderate/mild/very mild/normal)	2.91 ± 1.76 mg/L urine; 0.069 ± 0.021 mg/kg-d drinking water (normal dental fluorosis); 2.50 ± 2.39 mg/L urine; 0.064 ± 0.004 mg/kg-d drinking water (questionable dental fluorosis); 2.58 ± 1.31 mg/L urine; 0.060 ± 0.036 mg/kg-d drinking water (very mild dental fluorosis); 2.95 ± 1.44 mg/L urine; 0.060 ± 0.030 mg/kg-d drinking water (mild dental fluorosis); 4.82 ± 3.57 mg/L urine; 0.099 ± 0.063 mg/kg-d drinking water (moderate dental fluorosis); 3.81 ± 2.51 mg/L urine; 0.093 ± 0.040 mg/kg-d drinking water (severe dental fluorosis)	Combined Raven's Test for Rural China	Higher	–
Mondal *et al.* (2016)[w]	India	10–14	Drinking water High fluoride areas/low fluoride areas	Not reported (low); 0.33–18.08 mg/L (high)	Raven Standard Theoretical Intelligence Test	Higher	SES

Table A-1. Characteristics of Studies Included in the Mean-effect Meta-analysis

Reference[1]	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Range			
Bashash *et al.* (2017) [u]	Mexico	6–12	Urine	<0.80 mg/L (reference); ≥0.80 mg/L (high)	Wechsler Abbreviated Scale of Intelligence	Lower	Age; gender; weight at birth; parity; gestational age; maternal characteristics (smoking history, marital status, age at delivery, IQ, education, cohort)
Yu *et al.* (2018)[u*,w]	China	7–13	Drinking water High fluoride/normal fluoride Urine (per 0.5-mg/L increment in each range) High/medium/low fluoride ranges	≤1 mg/L (normal); >1 mg/L (high) 0.01–1.60 mg/L (low range urinary); 1.60–2.50 mg/L (medium range urinary); 2.50–5.54 mg/L (high range urinary)	Combined Raven's Test for Rural China	Lower	Age; gender; health factors; SES
Cui *et al.* (2020)[u]	China	7–12	Urine High/medium/low fluoride levels	<1.6 mg/L (low); 1.6–2.5 mg/L (medium); >=2.5 mg/L (high)	Combined Raven's Test	Lower	Gender; arsenic; iodine

Notes:

COMT = catechol-O-methyltransferase; RoB = risk of bias; SES = socioeconomic status

[1]A "w" superscript indicates studies included in the mean effect dose-response meta-analysis using fluoride in water; a "u" superscript indicates studies included in the mean effect dose-response meta-analysis using fluoride in urine; "*" indicates studies included in the mean effect dose-response meta-analysis at levels < 1.5 mg/L.

Table A-2. Studies Excluded from Mean-effect Meta-analysis	
Reference, Country	**Reason for Exclusion**
Qin *et al.* (1990)[translated in Qin *et al.* 2008], China	Missing mean or SD of outcome measure
Yang *et al.* (1994) [translated in Yang *et al.* 2008], China	Overlapping population with Wang *et al.* (2001); Table 2 in Yang *et al.* (1994) seemed incomplete
Wang *et al.* (2005b) [translated in Wang *et al.* 2008a], China	Missing mean or SD of outcome measure
Rocha-Amador *et al.* (2007), Mexico	Missing mean or SD of outcome measure
Liu *et al.* (2000) [translated in Liu *et al.* 2008], China	Overlapping population with Lu *et al.* (2000)
Sudhir *et al.* (2009), India	Missing mean or SD of outcome measure
He and Zhang (2010), China	Missing mean or SD of outcome measure
Kang *et al.* (2011), China	Missing mean or SD of outcome measure
Saxena *et al.* (2012), India	Missing mean or SD of outcome measure
Wang *et al.* (2012), China	Overlapping population with Xiang *et al.* (2003a); used in individual-level meta-analysis
Singh *et al.* (2013), India	Missing mean or SD of outcome measure
Wei *et al.* (2014), China	Missing mean or SD of outcome measure
Choi *et al.* (2015), China	Cognitive functions other than IQ
Kundu *et al.* (2015), India	Unusual IQ scores; used only for sensitivity analysis
Aravind *et al.* (2016), India	Unusual IQ scores; used only for sensitivity analysis
Razdan *et al.* (2017), India	Unusual IQ scores; used only for sensitivity analysis
Cui *et al.* (2018), China	Missing mean or SD of outcome measure; used in individual-level meta-analysis
Green *et al.* (2019), Canada	Missing mean or SD of outcome measure; used in individual-level meta-analysis
Soto-Barreras *et al.* (2019), Mexico	Missing mean or SD of outcome measure
Zhao *et al.* (2019), China	Overlapping population with Yu *et al.* (2018), but smaller sample size
Zhou *et al.* (2019), China	Overlapping population with Yu *et al.* (2018), but smaller sample size
Till *et al.* (2020), Canada	Missing mean or SD of outcome measure; used in individual-level meta-analysis
Wang *et al.* (2020b), China	Missing mean or SD of outcome measure; used in individual-level meta-analysis
Zhao *et al.* (2020), China	Overlapping population with Yu *et al.* (2018), but smaller sample size

Notes:
SD = standard deviation

Table A-3. Characteristics of Studies Included in the Individual-level Meta-analysis

Reference	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Range			
Ding *et al.* (2011)	China	7–14	Urine	0.10–3.55 mg/L	Combined Raven's Test for Rural China	Lower	Age; arsenic; iodine; lead; SES; demographics
Zhang *et al.* (2015b)	China	10–12	Urine High fluoride area/reference	1.10 ± 0.67 mg/L (reference); 2.40 ± 1.01 mg/L (high fluoride area)	Combined Raven's Test for Rural China	Lower	Age; gender; arsenic; iodine; drinking water fluoride; SES; thyroid hormone levels; COMT genotype
Bashash *et al.* (2017)	Mexico	6–12	Urine	0.18–2.8 mg/L	Wechsler Abbreviated Scale of Intelligence	Lower	Age; gender; weight at birth; parity; gestational age; maternal characteristics (smoking history, marital status, age at delivery, IQ, education, cohort)
Cui *et al.* (2018)	China	7–12	Urine	0.8–2.0 mg/L	Combined Raven's Test for Rural China	Lower	Age; maternal education; smoking in family member; stress; anger; dopamine receptor-2 polymorphism
Yu *et al.* (2018)	China	7–13	Urine, drinking water	0.01–5.54 mg/L (urine) 0.20–3.90 mg/L (water)	Combined Raven's Test for Rural China	Lower	Age; gender; maternal education; paternal education; low birth weight
Green *et al.* (2019)	Canada	3–4	Maternal urine, maternal fluoride intake, drinking water	0.51 ± 0.36 mg/L (urine) 0.54 ± 0.44 mg/day (maternal daily fluoride intake) 0.31 ± 0.23 mg/L (water)	Wechsler Primary and Preschool Scale of Intelligence-III	Lower	Gender; city; maternal education; race/ethnicity; HOME score; prenatal secondhand smoke exposure

Table A-3. Characteristics of Studies Included in the Individual-level Meta-analysis

Reference	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Range			
Till *et al.* (2020)	Canada	3–4	Residence, maternal urine, infant fluoride intake from formula, drinking water Fluoridated/nonfluoridated areas	0.64–0.70 (fluoridated), 0.38–0.42 mg/L (nonfluoridated) (urine) 0.12–0.34 (fluoridated), 0.02–0.08 mg/day (nonfluoridated) (infant formula fluoride intake) 0.58 (fluoridated), 0.13 (nonfluoridated) (water)	Wechsler Primary and Preschool Scale of Intelligence-III	Lower	Age; gender; maternal education; maternal race; HOME total score; secondhand smoke status in the child's house
Wang *et al.* (2020b)	China	7–13	Urine, drinking water	0.01–5.54 mg/L (urine) 0.20–3.90 mg/L (water)	Combined Raven's Test for Rural China	Lower	Age; gender; body mass index; maternal education; paternal education; household income; low birth weight

Notes:
COMT = catechol-O-methyltransferase; RoB = risk of bias; SES = socioeconomic status; HOME = Home Observation for Measurement of the Environment

Effects by Risk-of-bias Evaluation

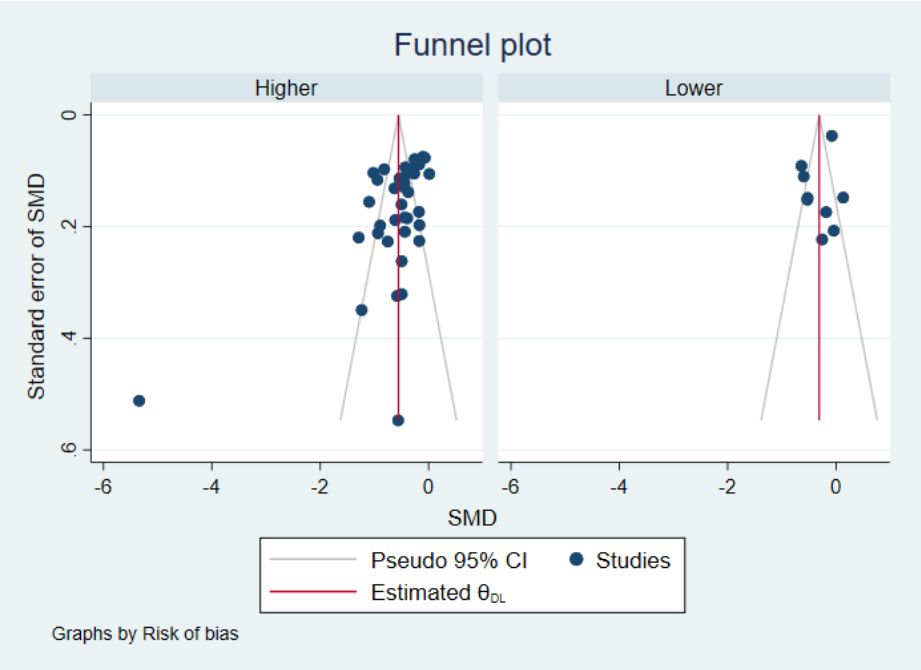


Figure A-1. Funnel Plot by Risk-of-bias Evaluation

```
Regression-based Egger test for small-study effects
Random-effects model                                  Higher RoB
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -3.63
      SE of beta1 =     0.709
                z =     -5.12
       Prob > |z| =    0.0000

Regression-based Egger test for small-study effects
Random-effects model                                  Lower RoB
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =      0.78
      SE of beta1 =     2.155
                z =      0.36
       Prob > |z| =    0.7169
```

Figure A-2. Test for Publication Bias by Risk of Bias

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                          Number of studies =     46
  Model: Random-effects                   observed =     37
 Method: DerSimonian-Laird                imputed =      9

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.556	-0.684	-0.428
Observed + Imputed	-0.354	-0.498	-0.210

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration                          Number of studies =     44
  Model: Random-effects                   observed =     37
 Method: DerSimonian-Laird                imputed =      7

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.556	-0.684	-0.428
Observed + Imputed	-0.684	-0.831	-0.537

Figure A-3. Trim-and-fill Analysis for High Risk-of-bias Studies

Filling in to the right using a linear estimator or to the left using a run estimator showed no change in the pooled SMD.

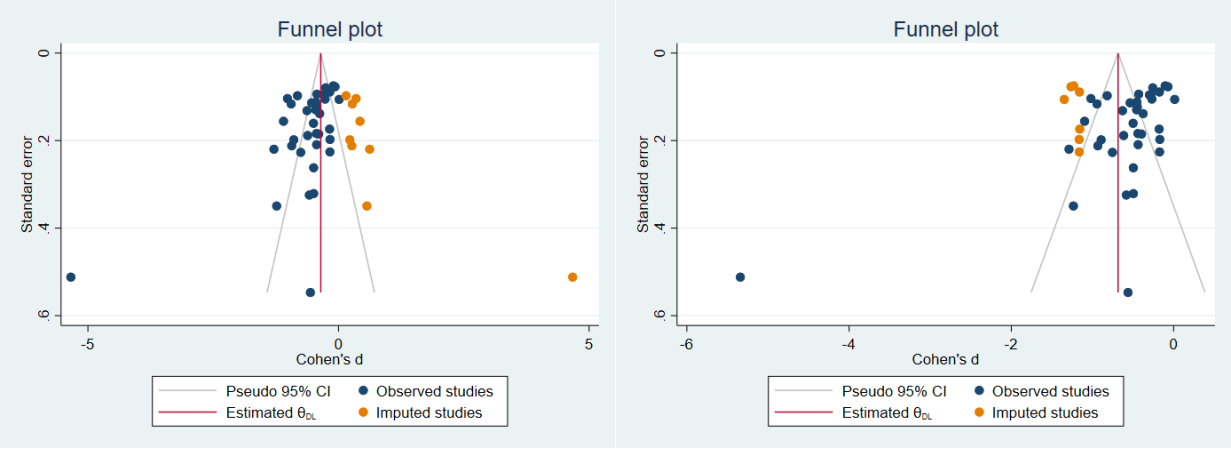


Figure A-4. Filled-in Funnel Plots for High Risk-of-bias Studies

Left panel shows the random-effects pooled SMD after filling in to the right using a run estimator (the linear estimator to the right showed no change in the pooled SMD); right panel shows random-effects pooled SMD after filling in to the left using a linear estimator (the run estimator to the left showed no change in the pooled SMD).

Effects by Gender

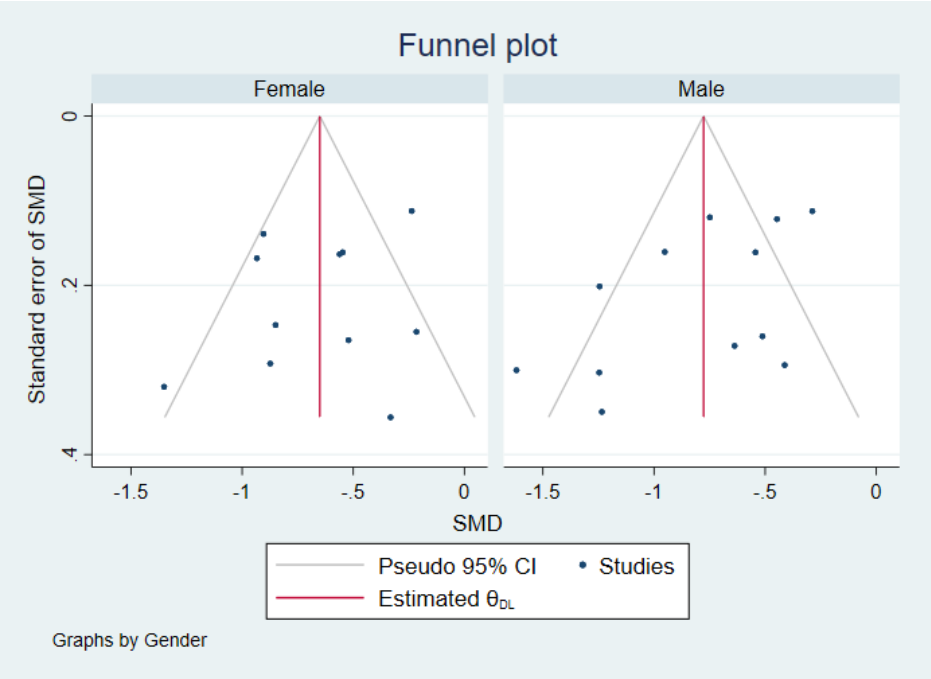


Figure A-5. Funnel Plots by Gender

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird
                                              Females
H0: beta1 = 0; no small-study effects
            beta1 =     -0.90
      SE of beta1 =     1.423
                z =     -0.63
      Prob > |z| =    0.5282

Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird
                                              Males
H0: beta1 = 0; no small-study effects
            beta1 =     -2.60
      SE of beta1 =     1.225
                z =     -2.12
      Prob > |z| =    0.0338
```

Figure A-6. Test for Publication Bias by Gender

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the right

Iteration                          Number of studies =     14
  Model: Random-effects                   observed =     12
 Method: DerSimonian-Laird                imputed =      2

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.777	-0.994	-0.560
Observed + Imputed	-0.681	-0.900	-0.461

Figure A-7. Trim-and-fill Analysis for Studies in Boys Using Linear and Run Estimators

Filling in to the right using a run estimator or to the left using a linear or a run estimator showed no change in the pooled SMD.

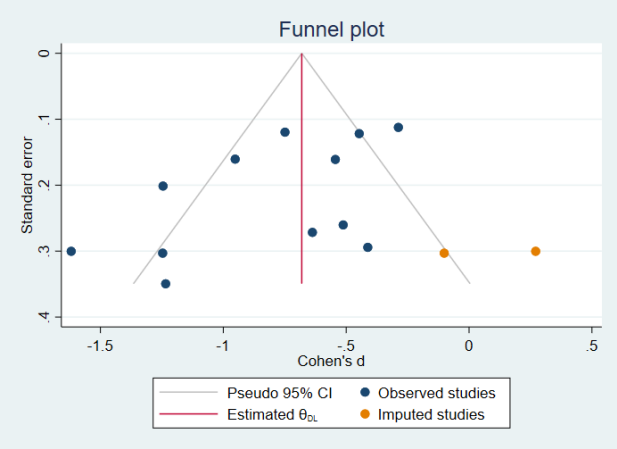


Figure A-8. Filled-in Funnel Plots for Studies in Boys

Panel shows funnel plot filled in to the right using a linear estimator. Filling in to the right using a run estimator or to the left using a linear or a run estimator showed no change in the pooled SMD.

Effects by Age Group

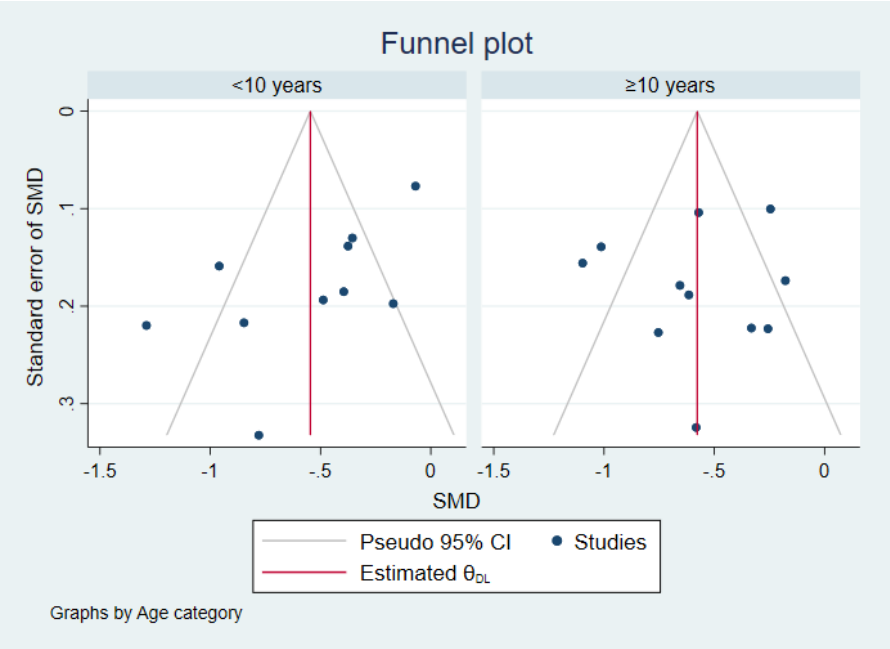


Figure A-9. Funnel Plot by Age Group

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird                          < 10 years

H0: beta1 = 0; no small-study effects
            beta1 =     -3.53
      SE of beta1 =     1.631
                z =     -2.16
       Prob > |z| =    0.0307


Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird                          ≥ 10 years

H0: beta1 = 0; no small-study effects
            beta1 =      0.32
      SE of beta1 =     1.883
                z =      0.17
       Prob > |z| =    0.8638
```

Figure A-10. Test for Publication Bias by Age Group

Note: Although suggestive of publication bias in the less-than-10 age group, filling in to the right or left using a linear or run estimator showed no change in the pooled SMD.

Effects by Country

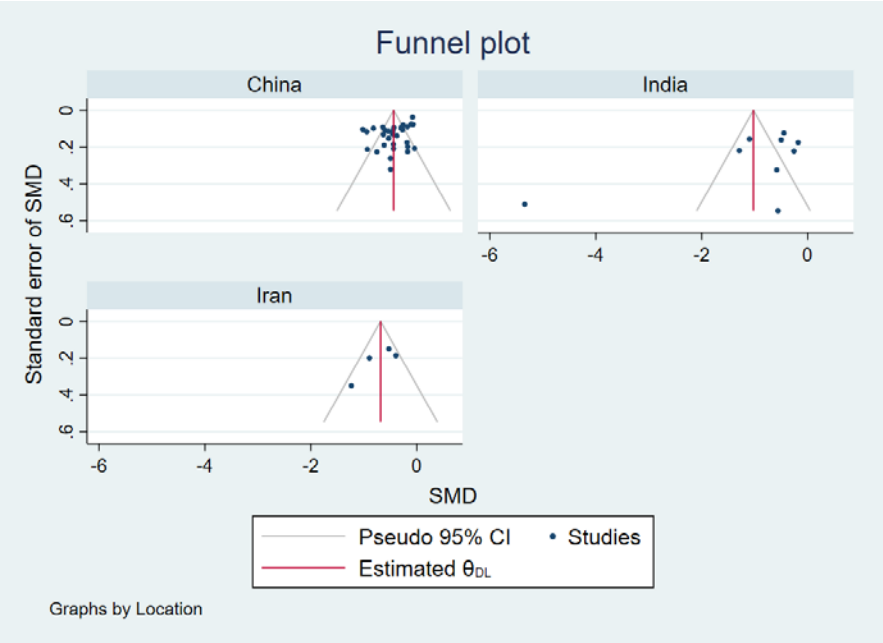


Figure A-11. Funnel Plot by Country

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird
                                        China

H0: beta1 = 0; no small-study effects
            beta1 =     -0.83
      SE of beta1 =     0.875
                z =     -0.95
       Prob > |z| =    0.3399
                                        India

H0: beta1 = 0; no small-study effects
            beta1 =     -5.64
      SE of beta1 =     1.770
                z =     -3.18
       Prob > |z| =    0.0015
                                        Iran

H0: beta1 = 0; no small-study effects
            beta1 =     -3.78
      SE of beta1 =     2.062
                z =     -1.83
       Prob > |z| =    0.0669
```

Figure A-12. Test for Publication Bias by Country

Nonparametric trim-and-fill analysis of publication bias **India**
Linear estimator, imputing on the left

Iteration Number of studies = 12
Model: Random-effects observed = 9
Method: DerSimonian-Laird imputed = 3

Pooling
Model: Random-effects
Method: DerSimonian-Laird

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-1.023	-1.540	-0.506
Observed + Imputed	-1.491	-2.154	-0.828

Figure A-13. Trim-and-fill Analysis for Studies in India

Filling in to the right using a linear or a run estimator or to the left using a run estimator showed no change in the pooled SMD.

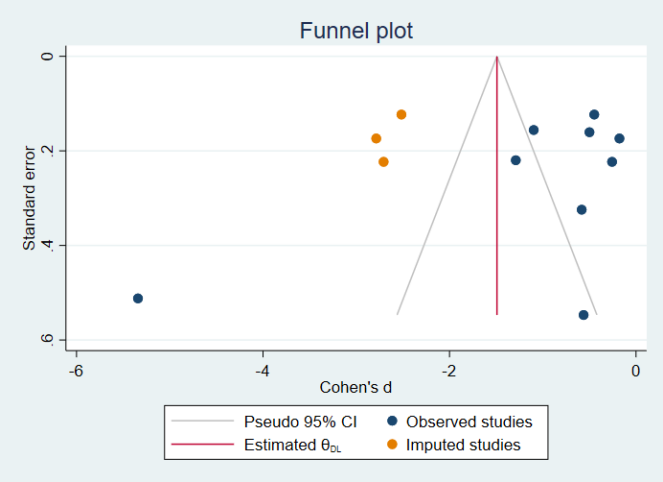


Figure A-14. Filled-in Funnel Plot for Studies in India

Panel shows funnel plot filled in to the left using a linear estimator. Filling in to the right using a linear or a run estimator or to the left using a run estimator showed no change in the pooled SMD.

Effect by Assessment Type

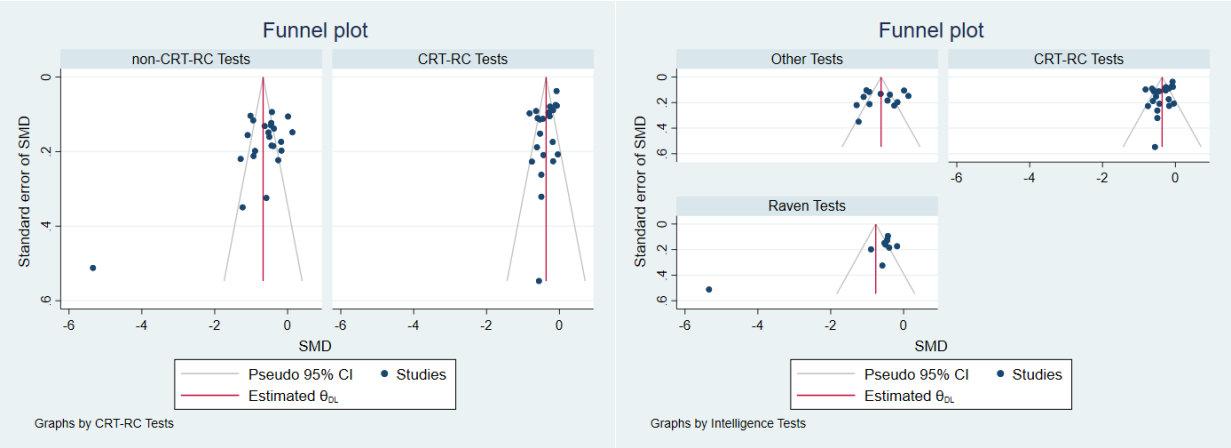


Figure A-15. Funnel Plot by CRT-RC-type Test

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird
```

CRT-CR Tests

```
H0: beta1 = 0; no small-study effects
            beta1 =     -0.93
      SE of beta1 =     0.707
                z =     -1.31
       Prob > |z| =    0.1904
```

non CRT-RC Tests

```
H0: beta1 = 0; no small-study effects
            beta1 =     -6.91
      SE of beta1 =     1.277
                z =     -5.41
       Prob > |z| =    0.0000
```

Raven's Tests

```
H0: beta1 = 0; no small-study effects
            beta1 =     -9.17
      SE of beta1 =     1.488
                z =     -6.16
       Prob > |z| =    0.0000
```

Other Tests

```
H0: beta1 = 0; no small-study effects
            beta1 =     -2.28
      SE of beta1 =     2.322
                z =     -0.98
       Prob > |z| =    0.3259
```

Figure A-16. Test for Publication Bias by Assessment Type

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration                           Number of studies =     30
  Model: Random-effects                      observed =     23
 Method: DerSimonian-Laird                   imputed =       7

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.672	-0.874	-0.469
Observed + Imputed	-0.920	-1.152	-0.687

Figure A-17. Trim-and-fill Analysis in Non-CRT-RC Tests

Filling in to the right using a linear or a run estimator or to the left using a run estimator showed no change in the pooled SMD.

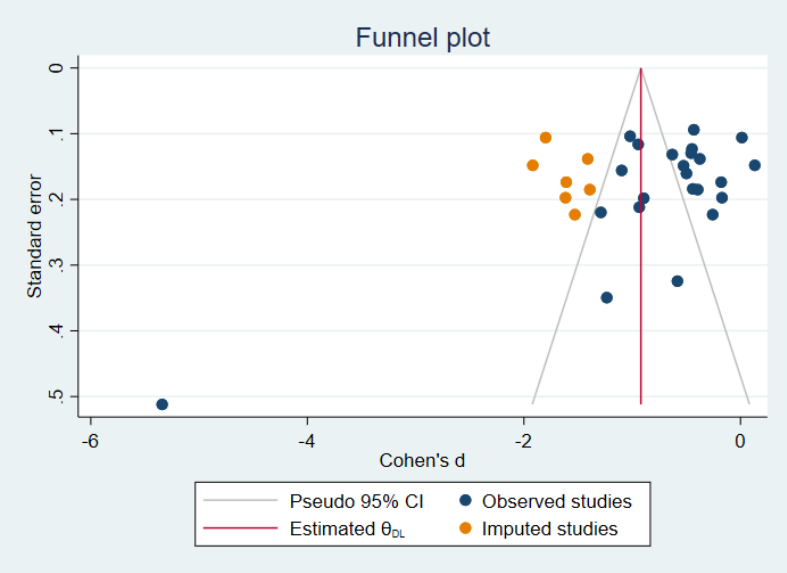


Figure A-18. Filled-in Funnel Plot to Eliminate Publication Bias in Non-CRT-RC Tests

Panel shows funnel plot filled in to the left using a linear estimator. Filling in to the right using a linear or a run estimator or to the left using a run estimator showed no change in the pooled SMD.

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration                           Number of studies =     14
  Model: Random-effects                      observed =     10
 Method: DerSimonian-Laird                   imputed =      4

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.763	-1.098	-0.428
Observed + Imputed	-1.215	-1.676	-0.754

Figure A-19. Trim-and-fill Analysis for Raven-type Tests

Filling in to the right using a linear or a run estimator or to the left using a run estimator showed no change in the pooled SMD.

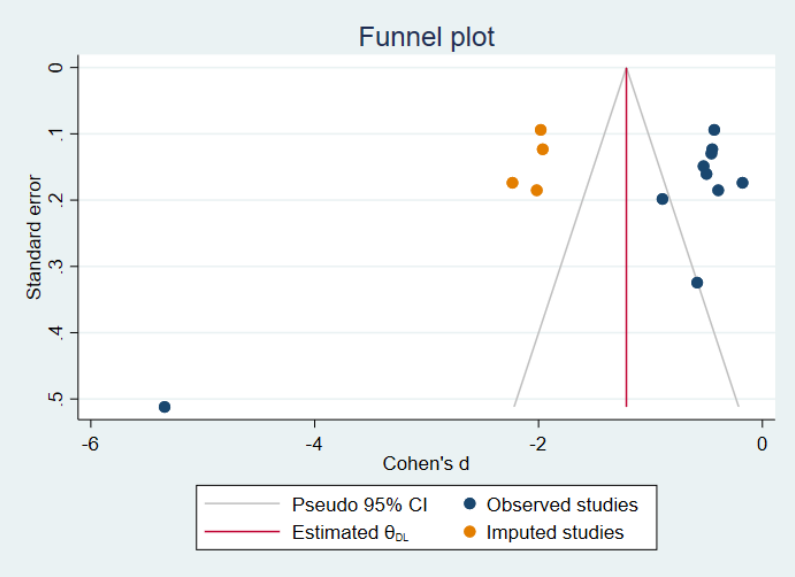


Figure A-20. Filled-in Funnel Plot to Eliminate Publication Bias for Raven-type Tests

Panel shows funnel plot filled in to the left using a linear estimator. Filling in to the right using a linear or a run estimator or to the left using a run estimator showed no change in the pooled SMD.

Effect by Exposure Type

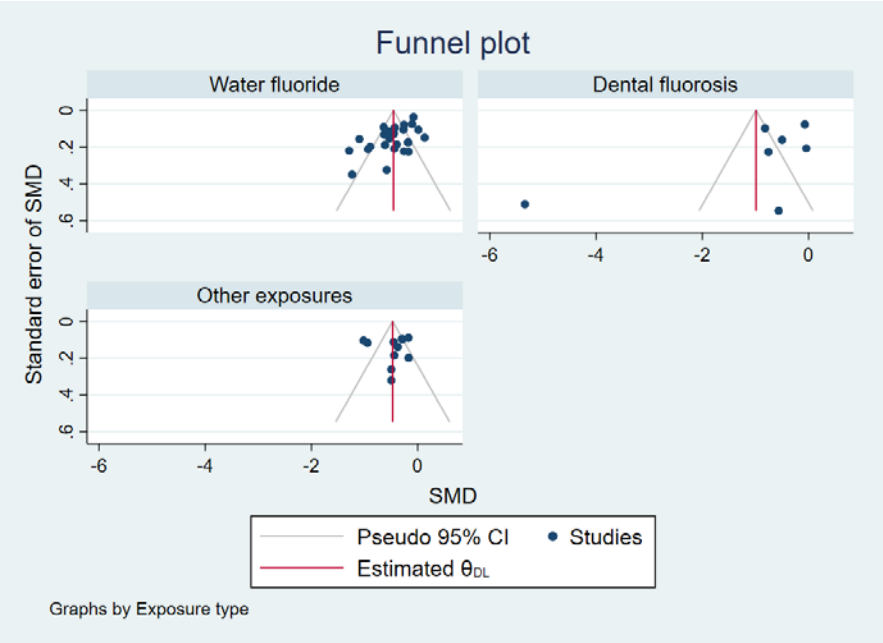


Figure A-21. Funnel Plot by Exposure Type

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird
                                        Water fluoride

H0: beta1 = 0; no small-study effects
            beta1 =     -2.44
      SE of beta1 =     0.837
                z =     -2.92
       Prob > |z| =    0.0035
                                        Dental fluorosis
H0: beta1 = 0; no small-study effects
            beta1 =     -5.81
      SE of beta1 =     1.649
                z =     -3.52
       Prob > |z| =    0.0004

                                        Other exposures
H0: beta1 = 0; no small-study effects
            beta1 =      0.21
      SE of beta1 =     1.609
                z =      0.13
       Prob > |z| =    0.8954
```

Figure A-22. Test for Publication Bias by Exposure Type

Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Water fluoride

Iteration	Number of studies = 30
Model: Random-effects	observed = 28
Method: DerSimonian-Laird	imputed = 2

Pooling
Model: Random-effects
Method: DerSimonian-Laird

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.455	-0.573	-0.336
Observed + Imputed	-0.415	-0.533	-0.296

Dental fluorosis

Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration	Number of studies = 9
Model: Random-effects	observed = 7
Method: DerSimonian-Laird	imputed = 2

Pooling
Model: Random-effects
Method: DerSimonian-Laird

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.989	-1.566	-0.412
Observed + Imputed	-1.510	-2.467	-0.553

Figure A-23. Trim-and-fill Analysis for Water Fluoride and Dental Fluorosis Exposures

For water fluoride, filling in to the right using a linear estimator or to the left using a linear or a run estimator showed no change in the pooled SMD. For dental fluorosis, filling in to the right using a linear or a run estimator or to the left using a run estimator showed no change in the pooled SMD.

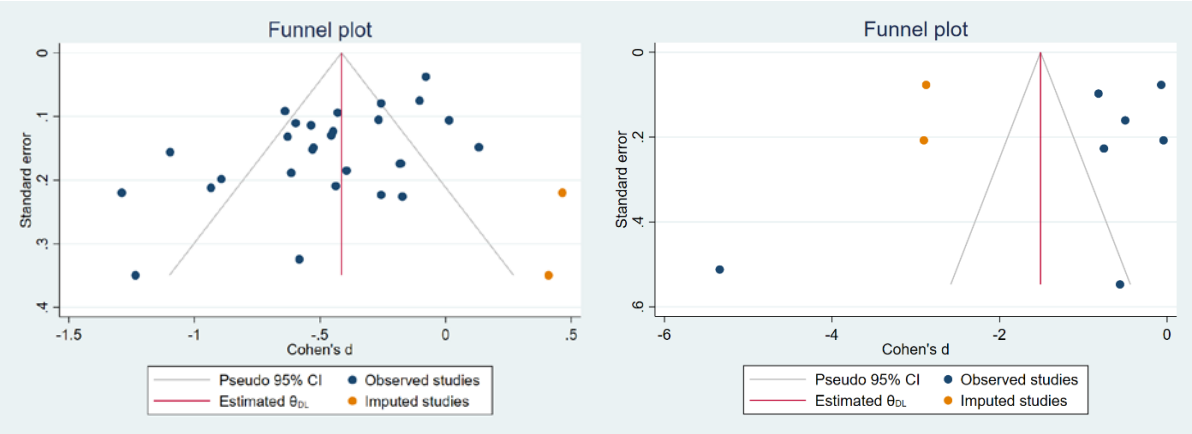


Figure A-24. Filled-in Funnel Plots to Eliminate Publication Bias for Water Fluoride (Left Panel) and Dental Fluorosis (Right Panel) Studies

For water fluoride, panel shows funnel plot filled in to the right using a run estimator. Filling in to the right using a linear estimator or to the left using a linear or a run estimator showed no change in the pooled SMD. For dental fluorosis, panel shows funnel plot filled in to the left using a linear estimator. Filling in to the right using a linear or a run estimator or to the left using a run estimator showed no change in the pooled SMD.

Sensitivity Analysis: Any Exposure Group Compared to Reference Group

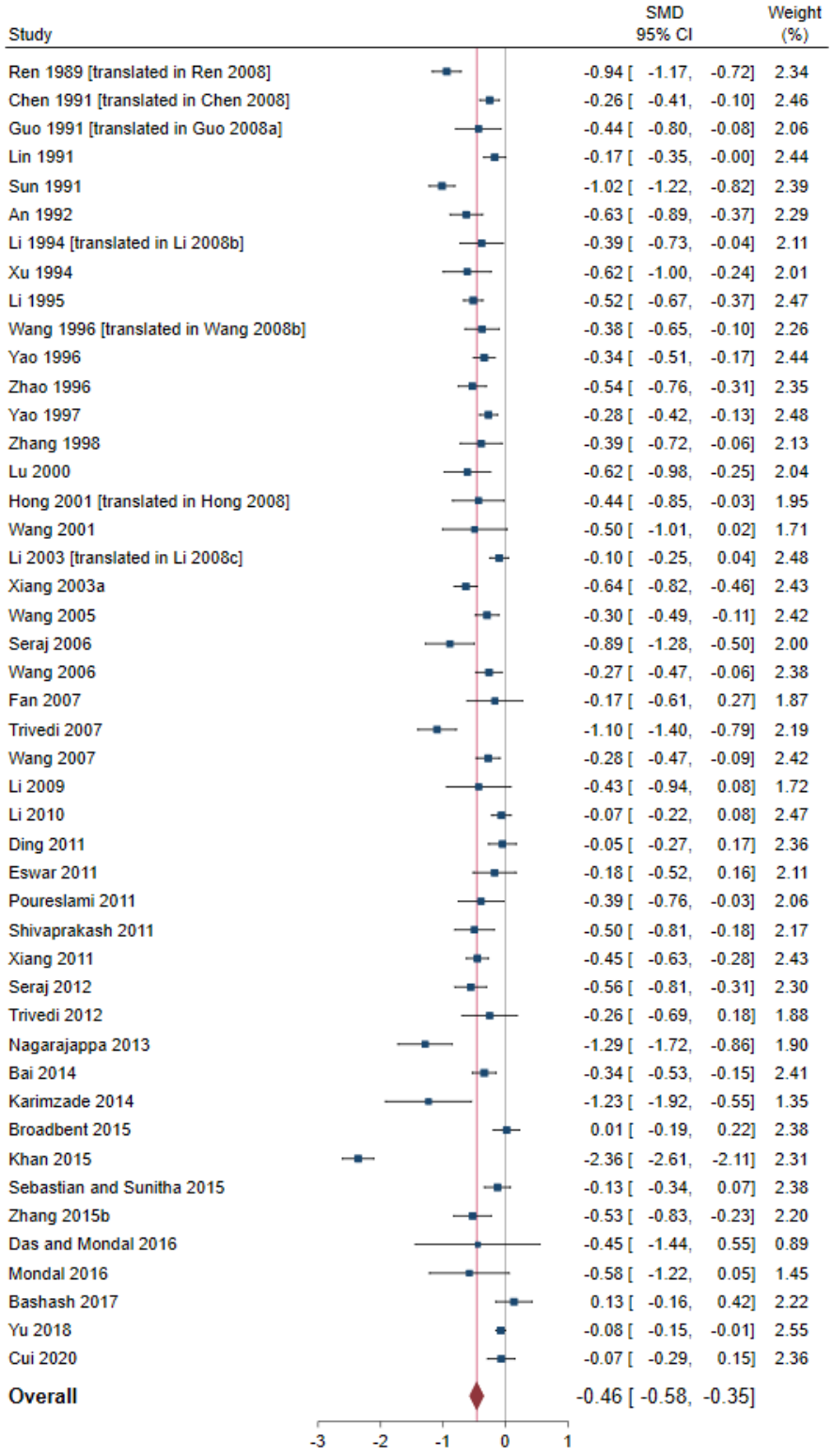


Figure A-25. Association Between Fluoride Exposure and IQ Scores in Children Using Any Exposure Versus Reference

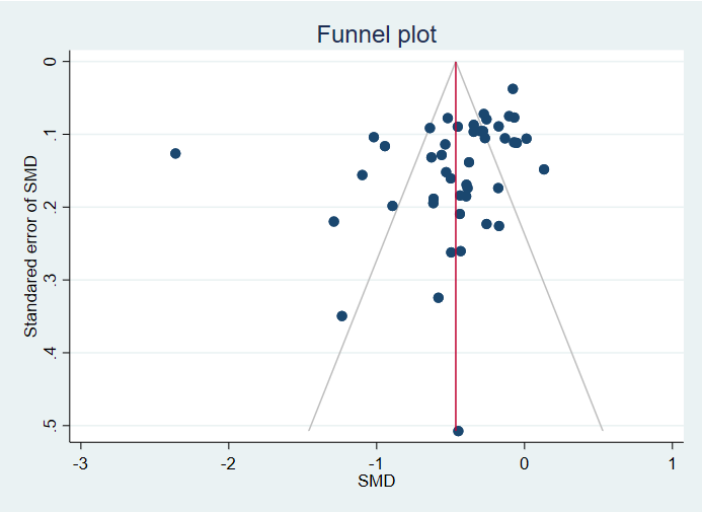


Figure A-26. Funnel Plot in Sensitivity Analysis Using Any Exposure Versus Reference

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -1.42
      SE of beta1 =     0.805
                z =     -1.76
       Prob > |z| =    0.0785

Begg's test for small-study effects

Kendall's score =   -237.00
    SE of score =   105.617
                z =     -2.25
     Prob > |z| =    0.0255
```

Figure A-27. Test for Publication Bias in Sensitivity Analysis Using Any Exposure Versus Reference

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                           Number of studies =     52
  Model: Random-effects                      observed =     46
 Method: DerSimonian-Laird                   imputed =      6

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.465	-0.580	-0.349
Observed + Imputed	-0.340	-0.475	-0.206

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration                           Number of studies =     58
  Model: Random-effects                      observed =     46
 Method: DerSimonian-Laird                   imputed =      12

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.465	-0.580	-0.349
Observed + Imputed	-0.608	-0.738	-0.479

Figure A-28. Trim-and-fill Analysis in Sensitivity Analysis Using Any Exposure Versus Reference

Filling in to the right using a linear estimator or to the left using a run estimator showed no change in the pooled SMD.

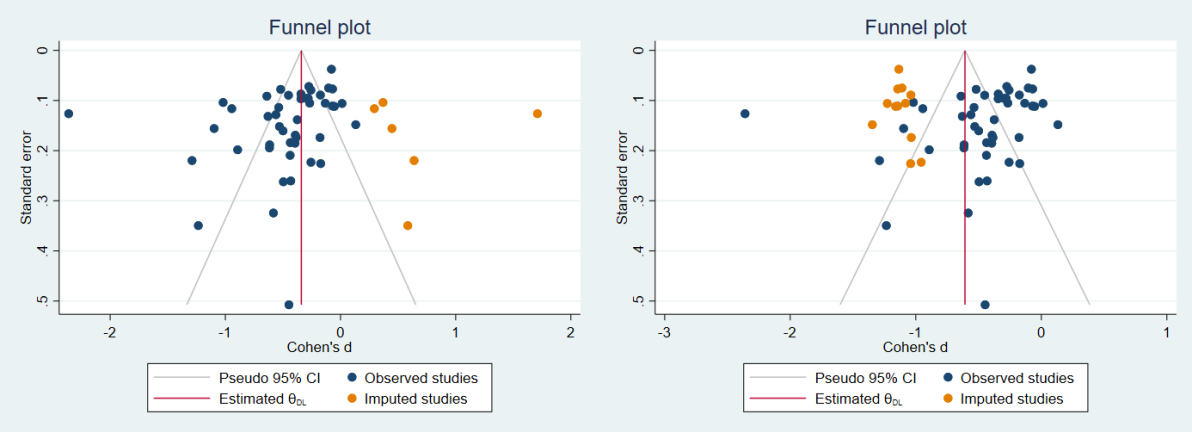


Figure A-29. Filled-in Funnel Plots to Eliminate Publication Bias in Sensitivity Analysis Using Any Exposure Versus Reference

Left panel shows the random-effects pooled SMD after filling in to the right using a run estimator (the linear estimator to the right showed no change in the pooled SMD); right panel shows random-effects pooled SMD after filling in to the left using a linear estimator (the run estimator to the left showed no change in the pooled SMD).

Sensitivity Analysis: Any Exposure Group Compared to Reference Group Including Aravind *et al.* (2016), Kundu *et al.* (2015), and Razdan *et al.* (2017)

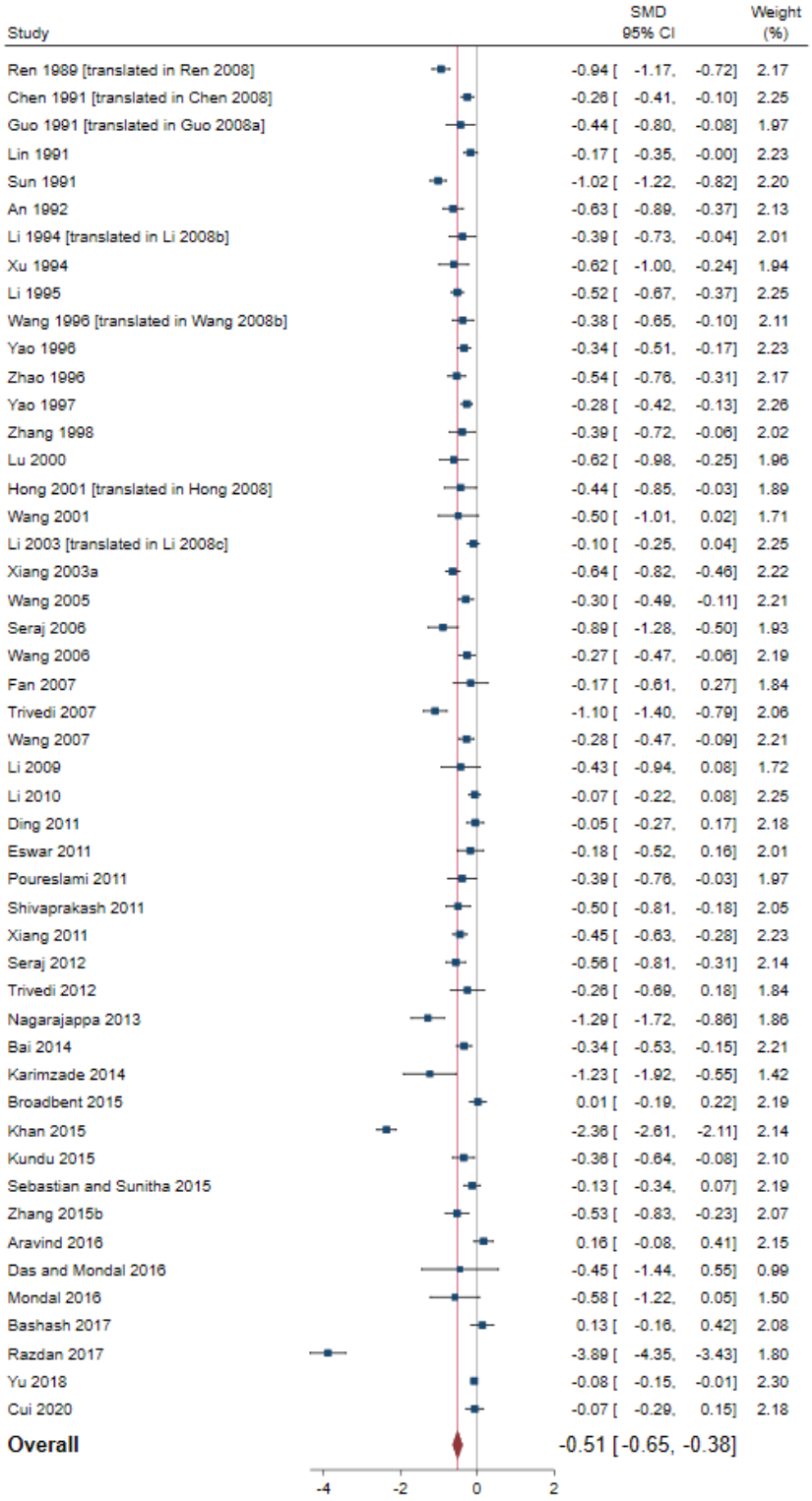


Figure A-30. Association Between Fluoride Exposure and IQ Scores in Children in Analysis Using Any Exposure Group Versus Reference [Including Aravind *et al.* (2016), Kundu *et al.* (2015), and Razdan *et al.* (2017)]

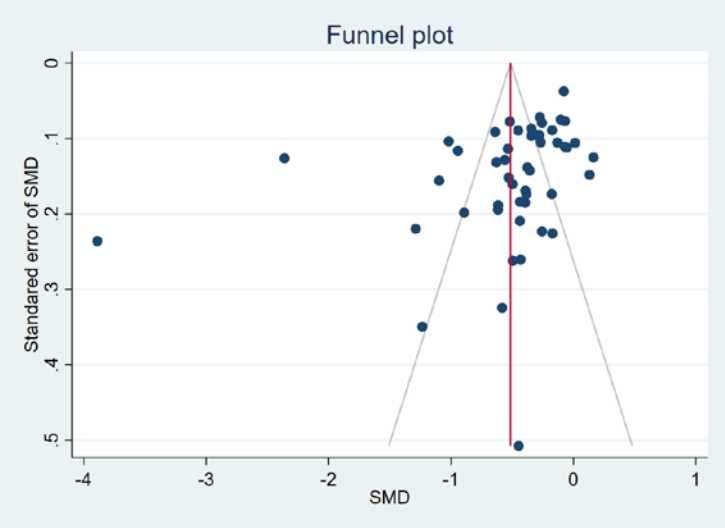


Figure A-31. Funnel Plot in Sensitivity Analysis Using Any Exposure Versus Reference [Including Aravind *et al.* (2016), Kundu *et al.* (2015), and Razdan *et al.* (2017)]

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -2.46
      SE of beta1 =     0.890
                z =     -2.77
      Prob > |z| =    0.0056
```

Figure A-32. Test for Publication Bias in Sensitivity Analysis Using Any Exposure Versus Reference [Including Aravind *et al.* (2016), Kundu *et al.* (2015), and Razdan *et al.* (2017)]

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                           Number of studies =     56
  Model: Random-effects                      observed =     49
 Method: DerSimonian-Laird                   imputed =      7

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------------------------
             Studies |   Cohen's d    [95% Conf. Interval]
---------------------+--------------------------------------------------------
            Observed |     -0.514      -0.645      -0.383
  Observed + Imputed |     -0.329      -0.485      -0.173
------------------------------------------------------------------------------
```

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration                           Number of studies =     65
  Model: Random-effects                      observed =     49
 Method: DerSimonian-Laird                   imputed =      16

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------------------------
             Studies |   Cohen's d    [95% Conf. Interval]
---------------------+--------------------------------------------------------
            Observed |     -0.514      -0.645      -0.383
  Observed + Imputed |     -0.730      -0.880      -0.579
------------------------------------------------------------------------------
```

Figure A-33. Trim-and-fill Analysis in Sensitivity Analysis Using Any Exposure Versus Reference [Including Aravind *et al.* (2016), Kundu *et al.* (2015), and Razdan *et al.* (2017)]

Filling in to the right using a linear estimator or to the left using a run estimator showed no change in the pooled SMD.

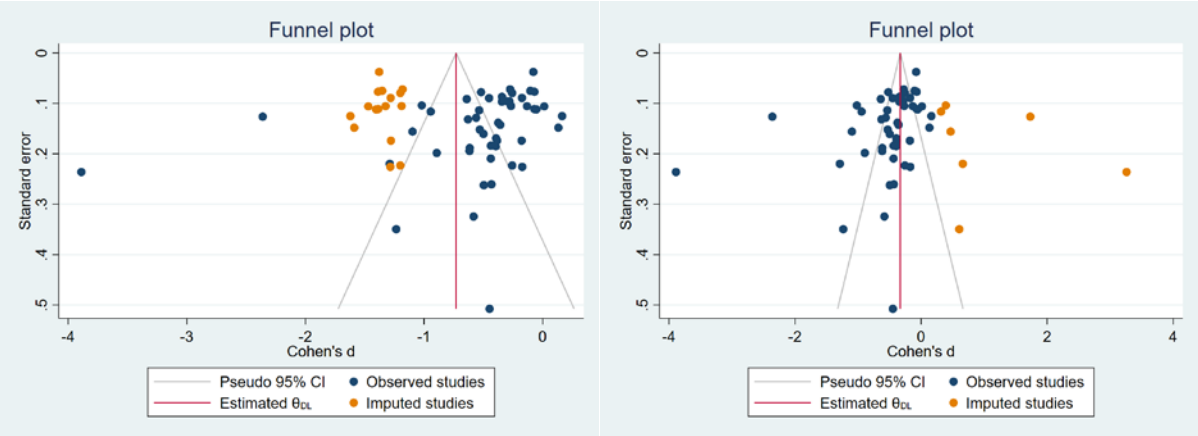


Figure A-34. Filled-in Funnel Plots to Eliminate Publication Bias in Sensitivity Analysis Using Any Exposure Versus Reference

Left panel shows the random-effects pooled SMD after filling in to the right using a run estimator; right panel shows random-effects pooled SMD after filling in to the left using a linear estimator. Filling in to the right using a linear estimator or to the left using a run estimator showed no change in the pooled SMD.

Sensitivity Analysis: Highest Exposure Group Compared to Reference Group Including Aravind *et al.* (2016), Kundu *et al.* (2015), and Razdan *et al.* (2017)

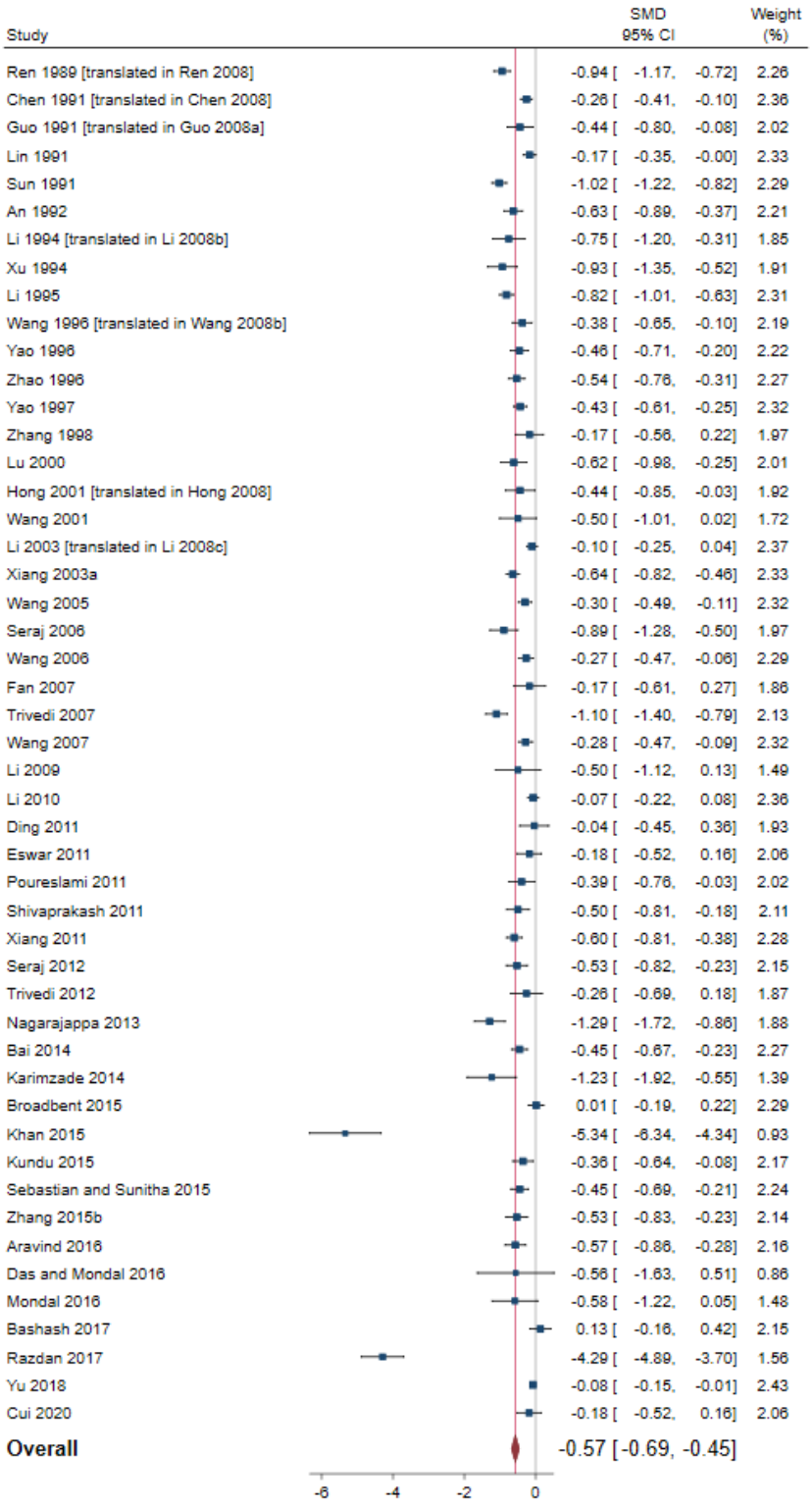


Figure A-35. Association Between Fluoride Exposure and IQ Scores in Children in Analysis Using Highest Exposure Versus Reference [Including Aravind *et al.* (2016), Kundu *et al.* (2015), and Razdan *et al.* (2017)]

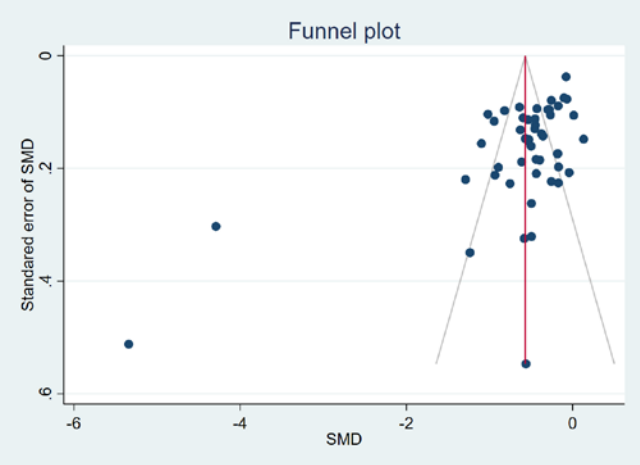


Figure A-36. Funnel Plot in Sensitivity Analysis Using Highest Exposure Versus Reference [Including Aravind *et al.* (2016), Kundu *et al.* (2015), and Razdan *et al.* (2017)]

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -4.47
      SE of beta1 =     0.696
                z =     -6.43
      Prob > |z| =    0.0000
```

Figure A-37. Test for Publication Bias in Sensitivity Analysis Using Highest Exposure Versus Reference [Including Aravind *et al.* (2016), Kundu *et al.* (2015), and Razdan *et al.* (2017)]

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                           Number of studies =     54
  Model: Random-effects                      observed =     49
 Method: DerSimonian-Laird                   imputed =      5

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------------------------
             Studies |   Cohen's d    [95% Conf. Interval]
---------------------+--------------------------------------------------------
            Observed |     -0.570      -0.693      -0.447
  Observed + Imputed |     -0.422      -0.563      -0.281
------------------------------------------------------------------------------
```

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration                           Number of studies =     66
  Model: Random-effects                      observed =     49
 Method: DerSimonian-Laird                   imputed =      17

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------------------------
             Studies |   Cohen's d    [95% Conf. Interval]
---------------------+--------------------------------------------------------
            Observed |     -0.570      -0.693      -0.447
  Observed + Imputed |     -0.838      -0.999      -0.676
------------------------------------------------------------------------------
```

Figure A-38. Trim-and-fill Analysis in Sensitivity Analysis Using Highest Exposure Versus Reference [Including Aravind *et al.* (2016), Kundu *et al.* (2015), and Razdan *et al.* (2017)]

Filling in to the right using a linear estimator or to the left using a run estimator showed no change in the pooled SMD.

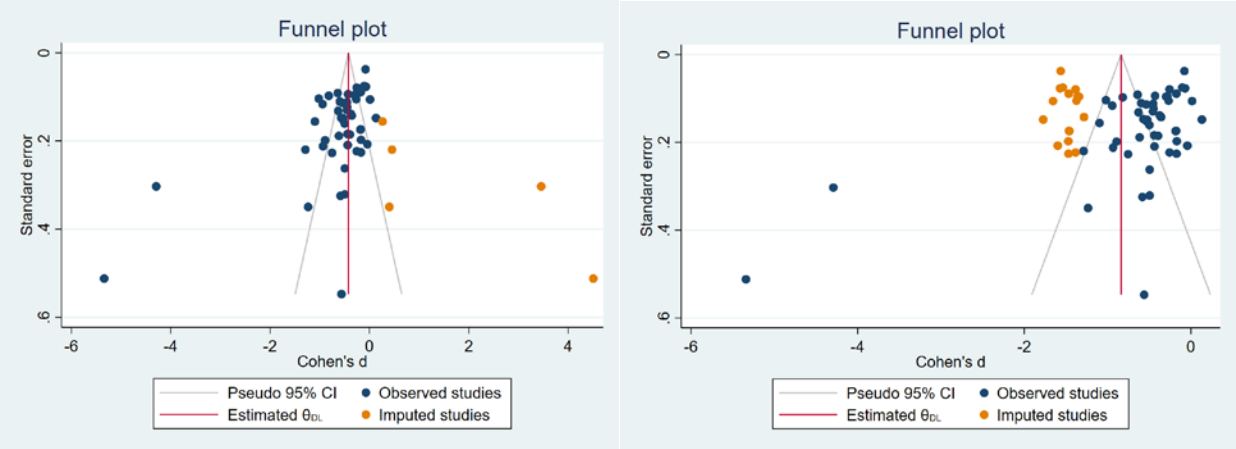


Figure A-39. Filled-in Funnel Plots to Eliminate Publication Bias in Sensitivity Analysis Using Highest Exposure Versus Reference [Including Aravind *et al.* (2016), Kundu *et al.* (2015), and Razdan *et al.* (2017)]

Left panel shows the random-effects pooled SMD after filling in to the right using a run estimator; right panel shows random-effects pooled SMD after filling in to the left using a linear estimator. Filling in to the right using a linear estimator or to the left using a run estimator showed no change in the pooled SMD.

Sensitivity Analysis: Excluding Lin *et al.* (1991)[1]

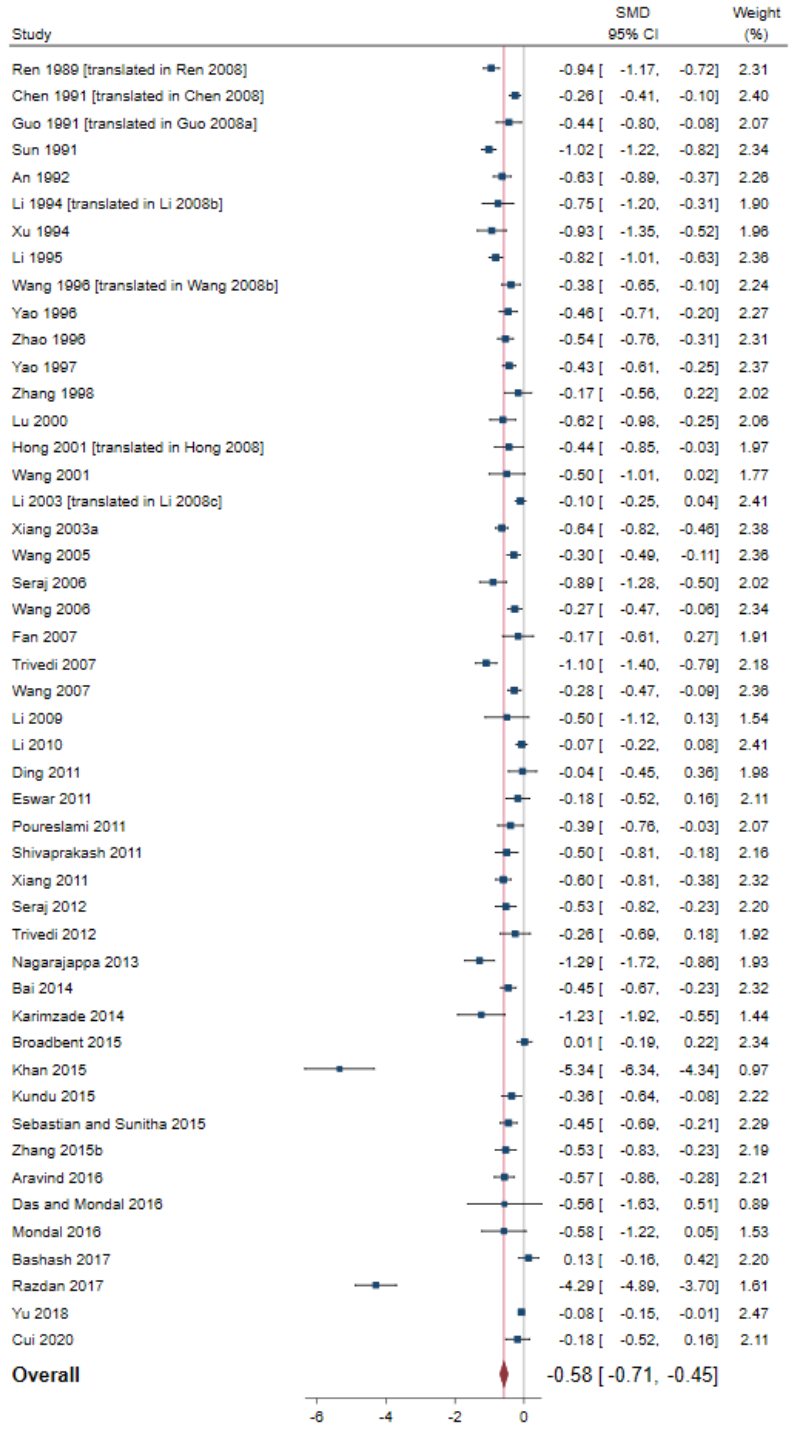


Figure A-40. Association Between Fluoride Exposure and IQ Scores in Children in Analysis Using Highest Exposure Versus Reference [Excluding Lin *et al.* (1991)]

[1]Lin *et al.* (1991): ICF calculated standard errors based on p-values.

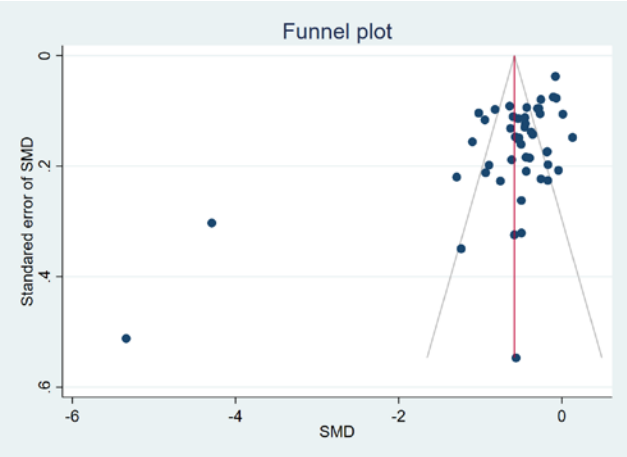


Figure A-41. Funnel Plot in Sensitivity Analysis Using Highest Exposure Versus Reference [Excluding Lin *et al.* (1991)]

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -4.47
      SE of beta1 =     0.711
                z =     -6.29
      Prob > |z| =    0.0000
```

Figure A-42. Test for Publication Bias in Sensitivity Analysis Using Highest Exposure Versus Reference [Excluding Lin *et al.* (1991)]

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                           Number of studies =     53
  Model: Random-effects                      observed =     48
 Method: DerSimonian-Laird                   imputed =      5

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------------------------
             Studies |   Cohen's d    [95% Conf. Interval]
---------------------+--------------------------------------------------------
            Observed |     -0.581      -0.707      -0.454
  Observed + Imputed |     -0.429      -0.573      -0.284
------------------------------------------------------------------------------
```

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the left

Iteration                           Number of studies =     65
  Model: Random-effects                      observed =     48
 Method: DerSimonian-Laird                   imputed =      17

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird

------------------------------------------------------------------------------
             Studies |   Cohen's d    [95% Conf. Interval]
---------------------+--------------------------------------------------------
            Observed |     -0.581      -0.707      -0.454
  Observed + Imputed |     -0.855      -1.021      -0.689
------------------------------------------------------------------------------
```

Figure A-43. Trim-and-fill Analysis in Sensitivity Analysis Using Highest Exposure Versus Reference [Excluding Lin *et al.* (1991)]

Filling in to the right using a linear estimator or to the left using a run estimator showed no change in the pooled SMD.

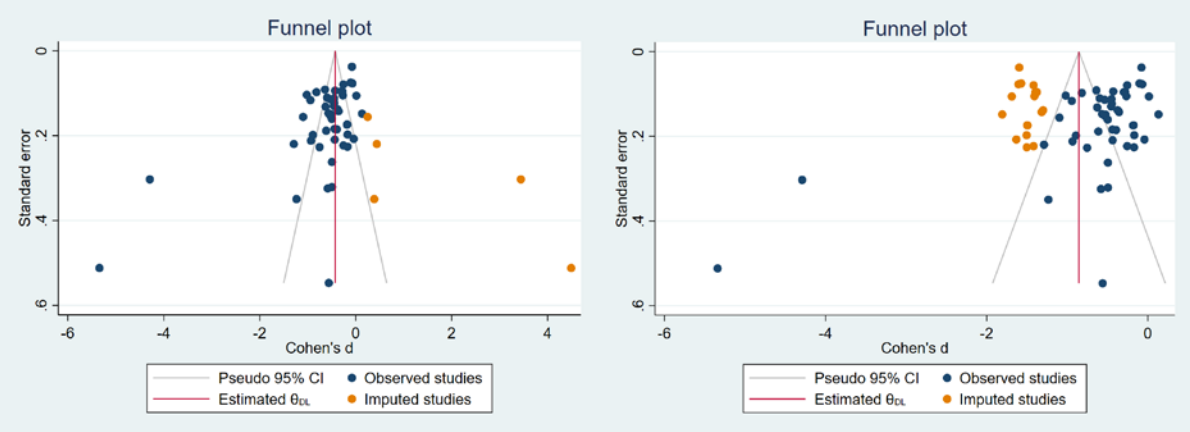


Figure A-44. Filled-in Funnel Plots to Eliminate Publication Bias in Sensitivity Analysis Using Highest Exposure Versus Reference [Excluding Lin *et al.* (1991)]

Left panel shows the random-effects pooled SMD after filling in to the right using a run estimator; right panel shows random-effects pooled SMD after filling in to the left using a linear estimator. Filling in to the right using a linear estimator or to the left using a run estimator showed no change in the pooled SMD.

Sensitivity Analysis: Excluding Aravind *et al.* (2016), Kundu *et al.* (2015), Razdan *et al.* (2017), and Khan *et al.* (2015)

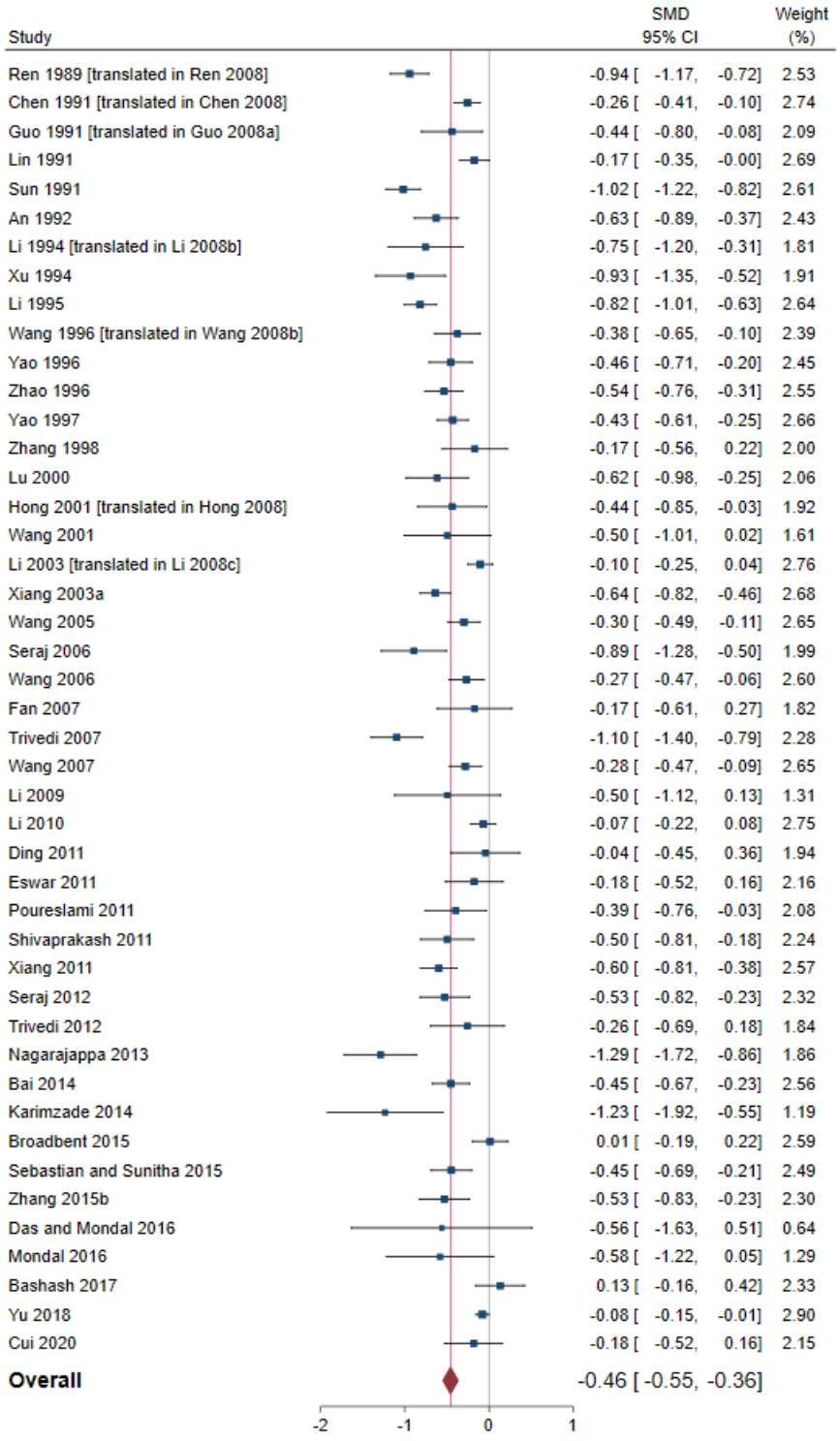


Figure A-45. Association Between Fluoride Exposure and IQ Scores in Children in Analysis Using Highest Exposure Versus Reference [Excluding Aravind *et al.* (2016), Kundu *et al.* (2015), Razdan *et al.* (2017), and Khan *et al.* (2015)]

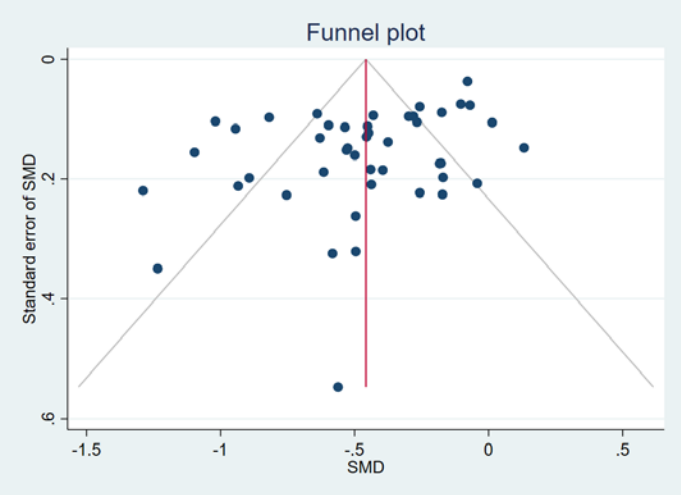


Figure A-46. Funnel Plot in Sensitivity Analysis Using Highest Exposure Versus Reference [Excluding Aravind *et al.* (2016), Kundu *et al.* (2015), Razdan *et al.* (2017), and Khan *et al.* (2015)]

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -1.30
      SE of beta1 =     0.659
                z =     -1.98
       Prob > |z| =    0.0481
```

Figure A-47. Test for Publication Bias in Sensitivity Analysis Using Any Exposure Versus Reference [Excluding Aravind *et al.* (2016), Kundu *et al.* (2015), Razdan *et al.* (2017), and Khan *et al.* (2015)]

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                           Number of studies =     48
  Model: Random-effects                      observed =     45
 Method: DerSimonian-Laird                   imputed =       3

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Cohen's d	[95% Conf.	Interval]
Observed	-0.458	-0.554	-0.361
Observed + Imputed	-0.417	-0.514	-0.319

Figure A-48. Trim-and-fill Analysis in Sensitivity Analysis Using Highest Exposure Versus Reference [Excluding Aravind *et al.* (2016), Kundu *et al.* (2015), Razdan *et al.* (2017), and Khan *et al.* (2015)]

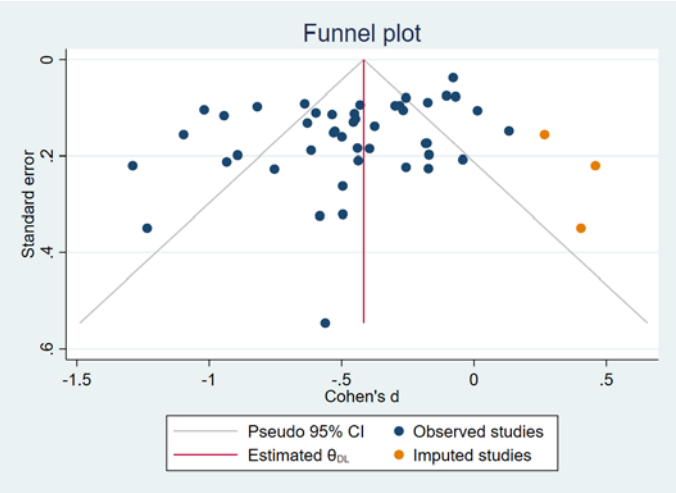


Figure A-49. Filled-in Funnel Plot to Eliminate Publication Bias in Sensitivity Analysis Using Highest Exposure Versus Reference [Excluding Aravind *et al.* (2016), Kundu *et al.* (2015), Razdan *et al.* (2017), and Khan *et al.* (2015)]

Panel shows the random-effects pooled SMD after filling in to the right using a run estimator. Filling in to the right using a linear estimator or to the left using a linear or a run estimator showed no change in the pooled SMD.

Attachment B. Subgroup and Sensitivity Analyses (Aim 2)

Table B-1. Characteristics of Studies Included in the Individual-level Meta-analysis

Reference	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric)	Range			
Ding *et al.* (2011)	China	7–14	Urine	0.10−3.55 mg/L	Combined Raven's Test for Rural China	Lower	Age; arsenic; iodine; lead; SES; demographics
Zhang *et al.* (2015b)	China	10–12	Urine	1.10 ± 0.67 mg/L (reference); 2.40 ± 1.01 mg/L (high fluoride area)	Combined Raven's Test for Rural China	Lower	Age; gender; arsenic; iodine; drinking water fluoride; SES; thyroid hormone levels; COMT genotype
Bashash *et al.* (2017)	Mexico	6–12	Urine	0.18−2.8 mg/L	Wechsler Abbreviated Scale of Intelligence	Lower	Age; gender; weight at birth; parity; gestational age; maternal characteristics (smoking history, marital status, age at delivery, IQ, education, and cohort)
Cui *et al.* (2018)	China	7–12	Urine	0.8−2.0 mg/L	Combined Raven's Test for Rural China	Lower	Age; maternal education; smoking in family member; stress; anger; dopamine receptor-2 polymorphism
Yu *et al.* (2018)	China	7–13	Urine, drinking Water	0.01−5.54 mg/L (urine) 0.20−3.90 mg/L (water)	Combined Raven's Test for Rural China	Lower	Age; gender; maternal education; paternal education; low birth weight

Table B-1. Characteristics of Studies Included in the Individual-level Meta-analysis

Reference	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric)	Range			
Green *et al.* (2019)	Canada	3–4	Maternal urine, maternal fluoride intake, drinking water	0.51 ± 0.36 mg/L (urine) 0.54 ± 0.44 mg/day (maternal daily fluoride intake) 0.31 ± 0.23 mg/L (water)	Wechsler Primary and Preschool Scale of Intelligence-III	Lower	Gender; city; maternal education; race/ethnicity; HOME score; prenatal secondhand smoke exposure
Till *et al.* (2020)	Canada	3–4	Residence (fluoridated/nonfluoridated cities), maternal urine, infant fluoride intake from formula, drinking water	0.64−0.70 mg/L (fluoridated), 0.38−0.42 mg/L (nonfluoridated) (urine) 0.12−0.34 mg/day (fluoridated), 0.02−0.08 mg/day (nonfluoridated) (infant formula fluoride intake) 0.58 mg/L (fluoridated), 0.13 mg/L (nonfluoridated) (water)	Wechsler Primary and Preschool Scale of Intelligence-III	Lower	Age; gender; maternal education; maternal race; HOME total score; second-hand smoke status in the child's house
Wang *et al.* (2020b)	China	7–13	Urine, drinking water	0.01−5.54 mg/L (urine) 0.20−3.90 mg/L (water)	Combined Raven's Test for Rural China	Lower	Age; gender; body mass index; maternal education; paternal education; household income; low birth weight

Sensitivity Analysis for Individual-level Studies: No Pooling for Yu *et al.* (2018)

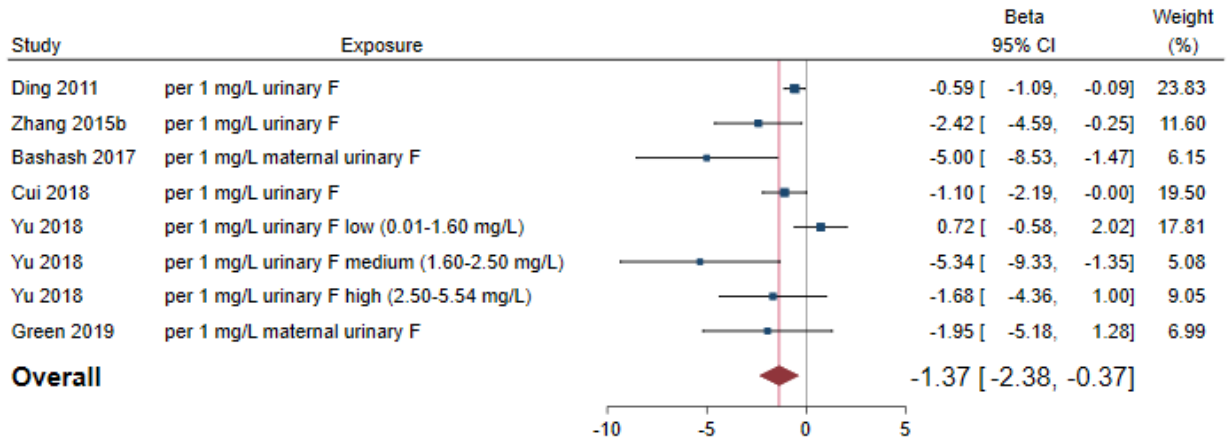


Figure B-1. Association Between Individual-level Fluoride Exposure and IQ Scores in Children [No Pooling for Yu *et al.* (2018)]

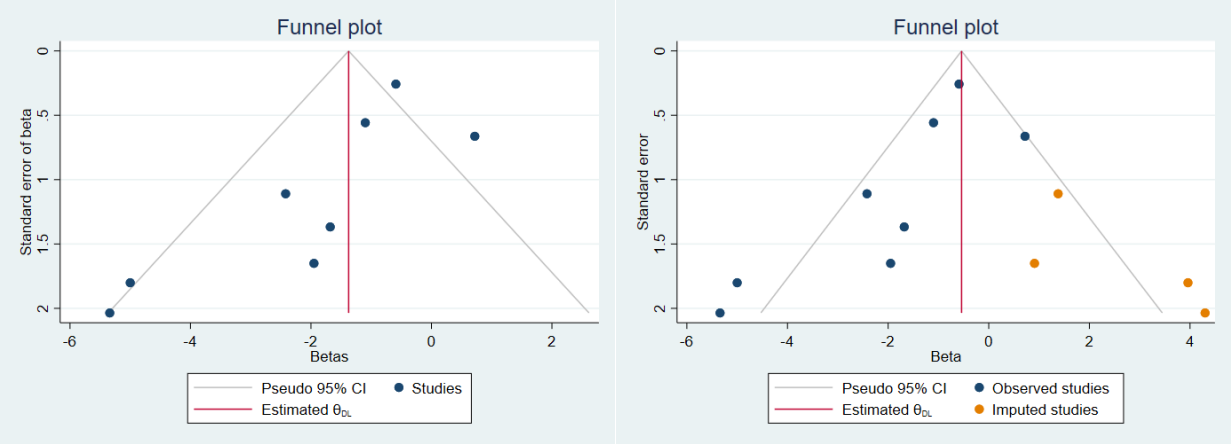


Figure B-2. Funnel Plots of Studies with Individual-level Exposures and Filled-in Funnel Plot to Eliminate Publication Bias [No Pooling for Yu *et al.* (2018)]

Right panel shows funnel plot filled in to the right using a linear estimator. Filling in to the left using a linear or a run estimator showed no change in the pooled effect estimate.

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -2.21
      SE of beta1 =     0.801
                z =     -2.75
      Prob > |z| =    0.0059
```

Figure B-3. Test for Publication Bias for Studies with Individual-level Exposures

Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the right

Iteration — Number of studies = 12
Model: Random-effects — observed = 8
Method: DerSimonian-Laird — imputed = 4

Pooling
Model: Random-effects
Method: DerSimonian-Laird

Studies	Beta	[95% Conf.	Interval]
Observed	-1.374	-2.379	-0.369
Observed + Imputed	-0.540	-1.567	0.486

Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration — Number of studies = 9
Model: Random-effects — observed = 8
Method: DerSimonian-Laird — imputed = 1

Pooling
Model: Random-effects
Method: DerSimonian-Laird

Studies	Beta	[95% Conf.	Interval]
Observed	-1.374	-2.379	-0.369
Observed + Imputed	-1.188	-2.220	-0.155

Figure B-4. Trim-and-fill Analysis for Studies with Individual-level Exposures

Filling in to the left using a linear or a run estimator showed no change in the pooled effect estimate.

Sensitivity Analysis for Individual-level Studies: Effect by Exposure Type, No Pooling for Yu *et al.* (2018)

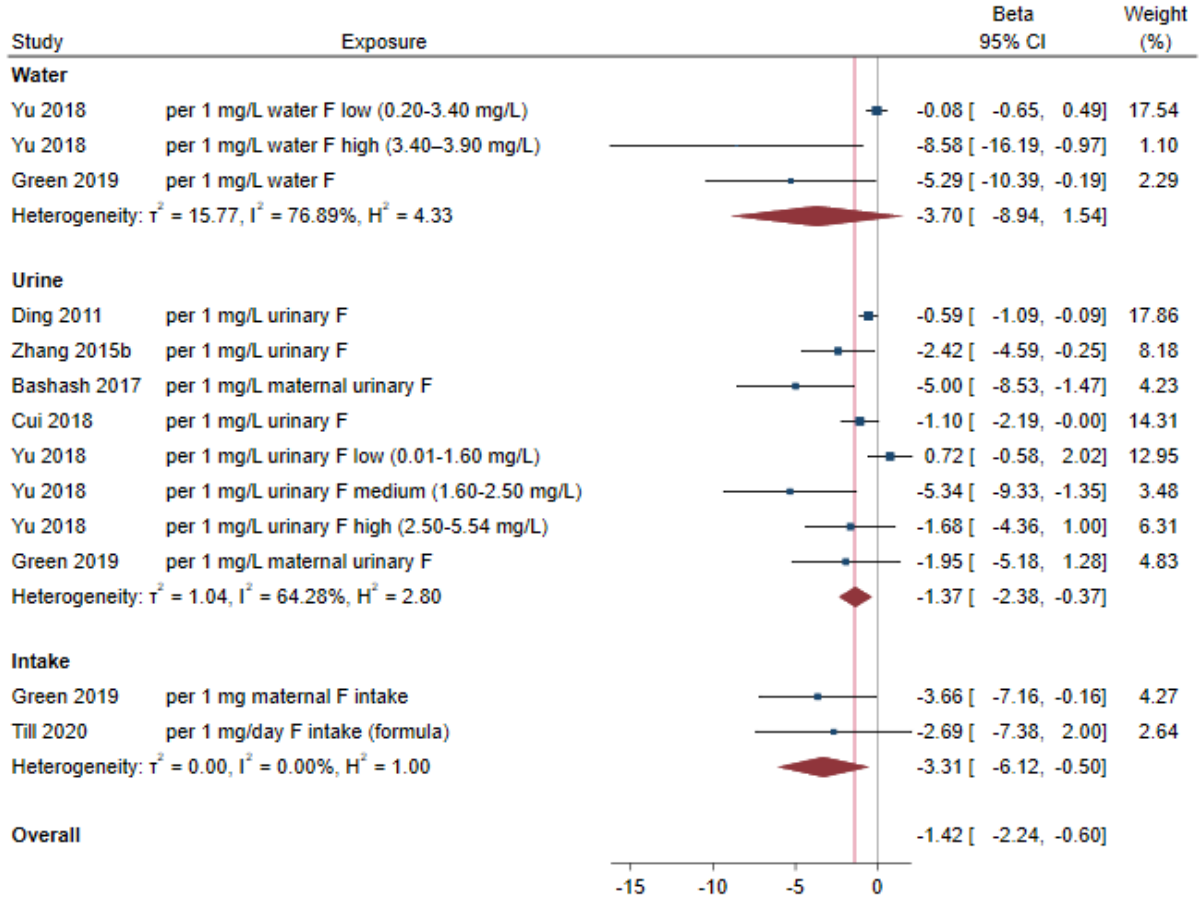


Figure B-5. Association Between Fluoride Exposure and IQ Scores in Children for Individual-level Exposures [No Pooling for Yu *et al.* (2018)]

Note: The analysis for publication bias for studies with fluoride in urine is identical to the overall analyses. The analyses for publication bias for studies with fluoride in water or fluoride intake rely on two studies each and are not shown.

Sensitivity Analysis for Individual-level Studies: Using Wang *et al.* (2020b) Versus Yu *et al.* (2018)

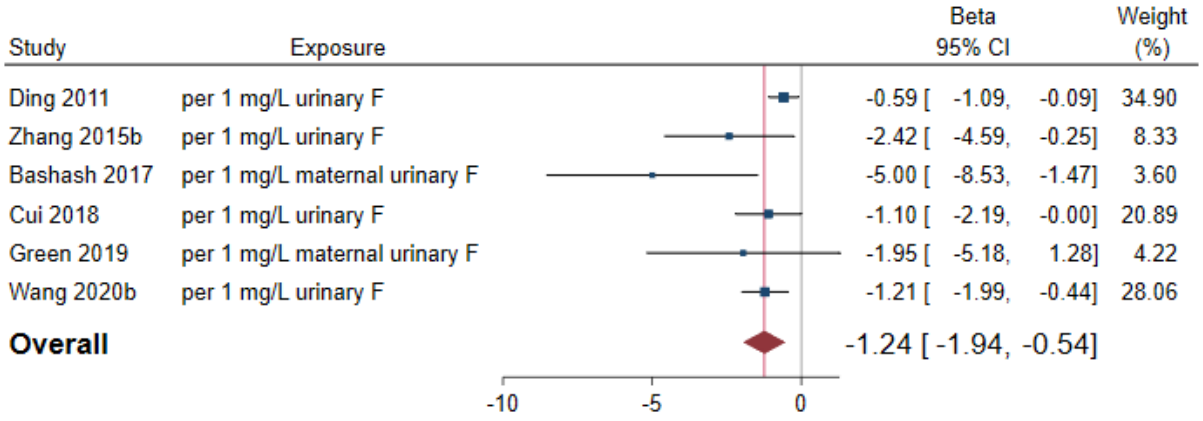


Figure B-6. Association Between Fluoride Exposure and IQ Scores in Children for Individual-level Exposures [Using Wang *et al.* (2020b) Estimates in Place of Yu *et al.* (2018) Estimates]

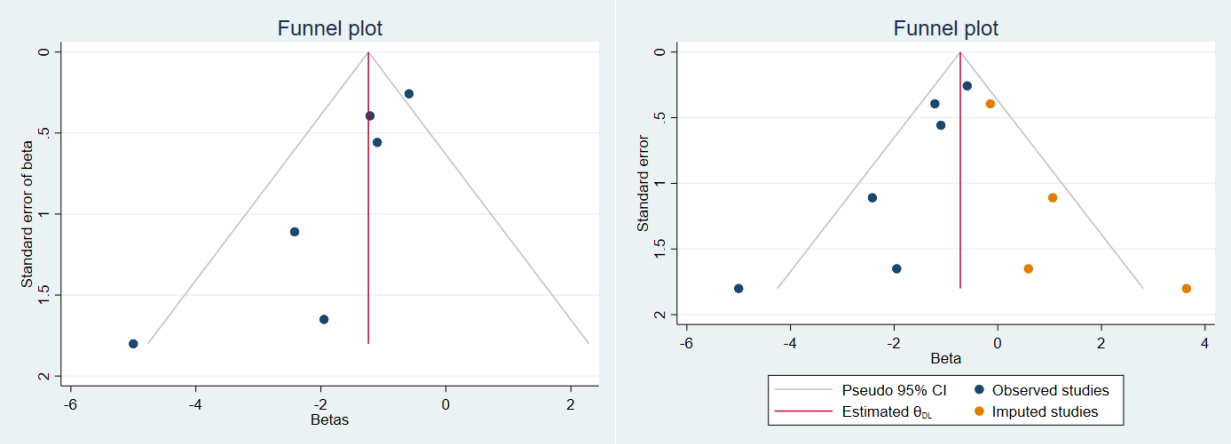


Figure B-7. Funnel Plot of Studies with Individual-level Exposures and Filled-in Funnel Plot to Eliminate Publication Bias [Using Wang *et al.* (2020b) Estimates in Place of Yu *et al.* (2018) Estimates]

Right panel shows funnel plot filled in to the right using a run estimator. Filling in to the left using a linear or a run estimator showed no change in the pooled effect estimate.

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -1.97
      SE of beta1 =     0.712
                z =     -2.77
      Prob > |z| =    0.0056
```

Figure B-8. Test for Publication Bias for Studies with Individual-level Exposures [Using Wang *et al.* (2020b) Estimates in Place of Yu *et al.* (2018) Estimates]

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the right

Iteration                          Number of studies =      9
  Model: Random-effects                   observed =      6
 Method: DerSimonian-Laird                imputed =      3

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Beta	[95% Conf.	Interval]
Observed	-1.240	-1.939	-0.540
Observed + Imputed	-0.900	-1.676	-0.123

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                          Number of studies =     10
  Model: Random-effects                   observed =      6
 Method: DerSimonian-Laird                imputed =      4

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Beta	[95% Conf.	Interval]
Observed	-1.240	-1.939	-0.540
Observed + Imputed	-0.722	-1.411	-0.034

Figure B-9. Trim-and-fill Analysis for Studies with Individual-level Exposures [Using Wang *et al.* (2020b) Estimates in Place of Yu *et al.* (2018) Estimates]

Filling in to the left using a linear or a run estimator showed no change in the pooled effect estimate.

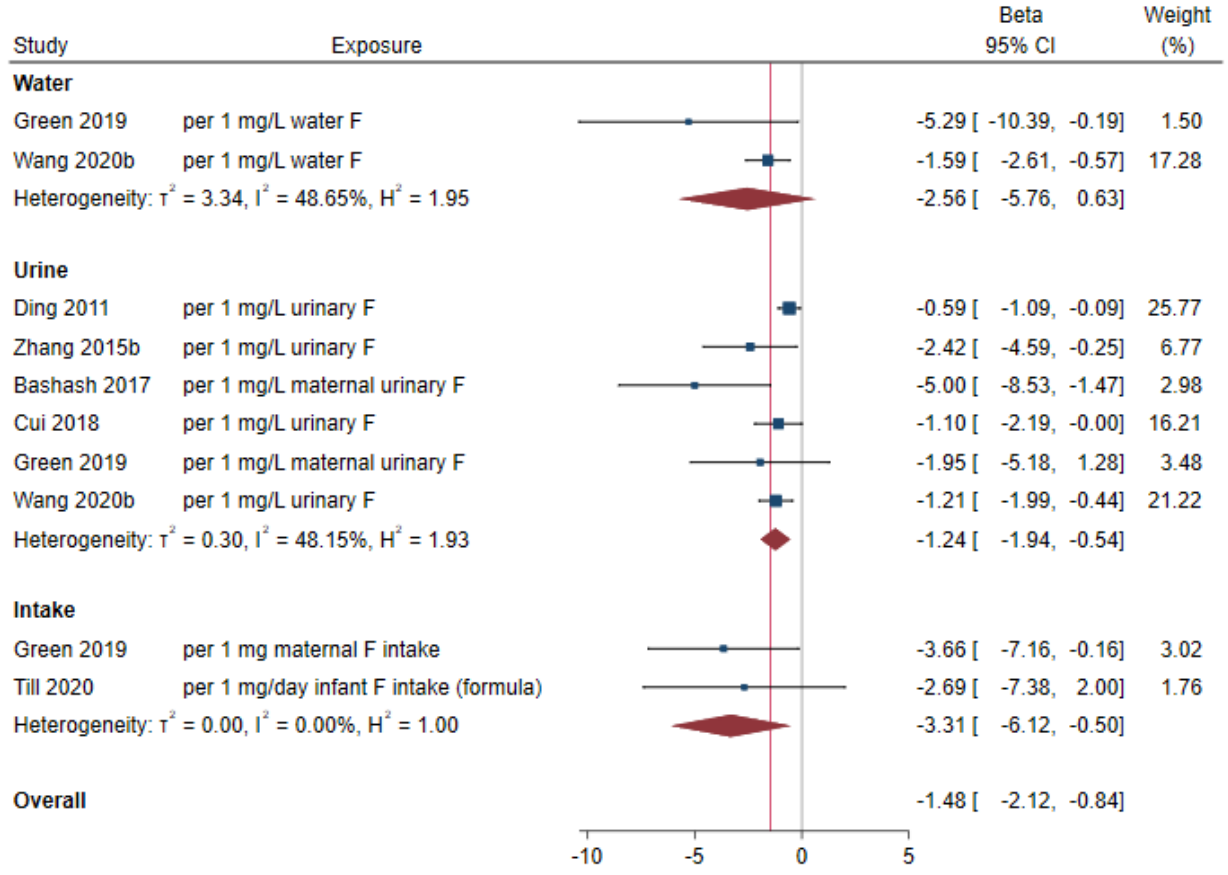


Figure B-10. Association Between Fluoride Exposure and IQ Scores in Children by Exposure for Individual-level Exposures [Using Wang *et al.* (2020b) Estimates in Place of Yu *et al.* (2018) Estimates]

Note: The analysis for publication bias for studies with fluoride in urine is identical to the overall analysis. The analyses for publication bias for studies with fluoride in water or fluoride intake rely on two studies each and are not shown.

Sensitivity Analysis for Individual-level Studies: Using Till *et al.* (2020) Versus Green *et al.* (2019)

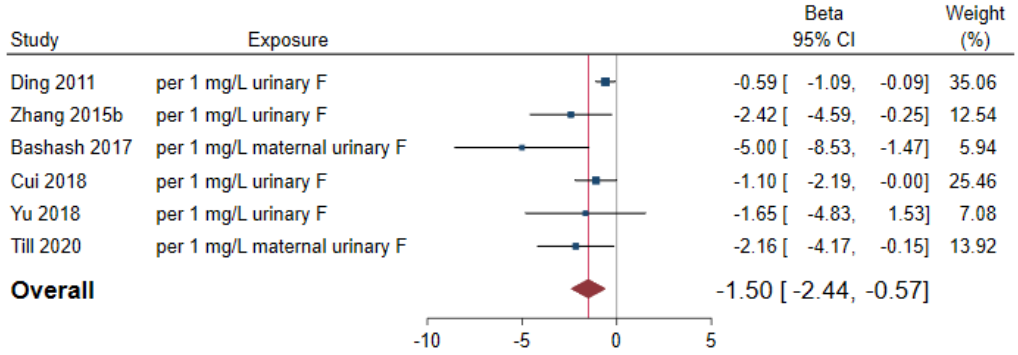


Figure B-11. Association Between Fluoride Exposure and IQ Scores in Children for Individual-level Exposures [Using Till *et al.* (2020) Estimates in Place of Green *et al.* (2019) Estimates]

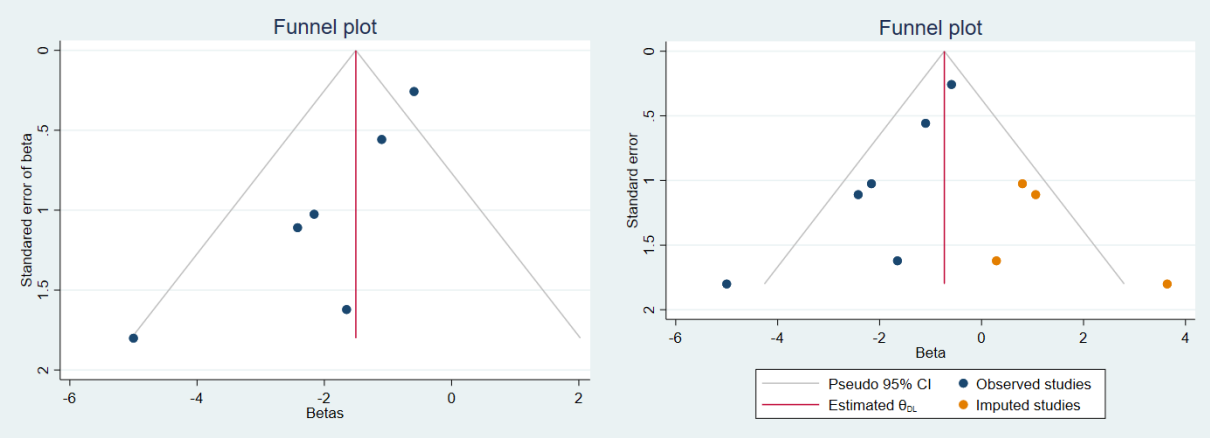


Figure B-12. Funnel Plot of Included Studies with Individual-level Exposures and Filled-in Funnel Plot to Eliminate Publication Bias [Using Till *et al.* (2020) Estimates in Place of Green *et al.* (2019) Estimates]

Right panel shows funnel plot filled in to the right using a run estimator. Filling in to the left using a linear or a run estimator showed no change in the pooled effect estimate.

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -1.94
      SE of beta1 =     0.655
                z =     -2.96
      Prob > |z| =    0.0031
```

Figure B-13. Test for Publication Bias for Studies with Individual-level Exposures [Using Till *et al.* (2020) Estimates in Place of Green *et al.* (2019) Estimates]

Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the right

Iteration — Number of studies = 9
Model: Random-effects — observed = 6
Method: DerSimonian-Laird — imputed = 3

Pooling
Model: Random-effects
Method: DerSimonian-Laird

Studies	Beta	[95% Conf.	Interval]
Observed	-1.504	-2.439	-0.570
Observed + Imputed	-0.808	-1.774	0.159

Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration — Number of studies = 10
Model: Random-effects — observed = 6
Method: DerSimonian-Laird — imputed = 4

Pooling
Model: Random-effects
Method: DerSimonian-Laird

Studies	Beta	[95% Conf.	Interval]
Observed	-1.504	-2.439	-0.570
Observed + Imputed	-0.730	-1.649	0.188

Figure B-14. Trim-and-fill Analysis for Studies with Individual-level Exposures [Using Till *et al.* (2020) Estimates in Place of Green *et al.* (2019) Estimates]

Filling in to the left using a linear or a run estimator showed no change in the pooled effect estimate.

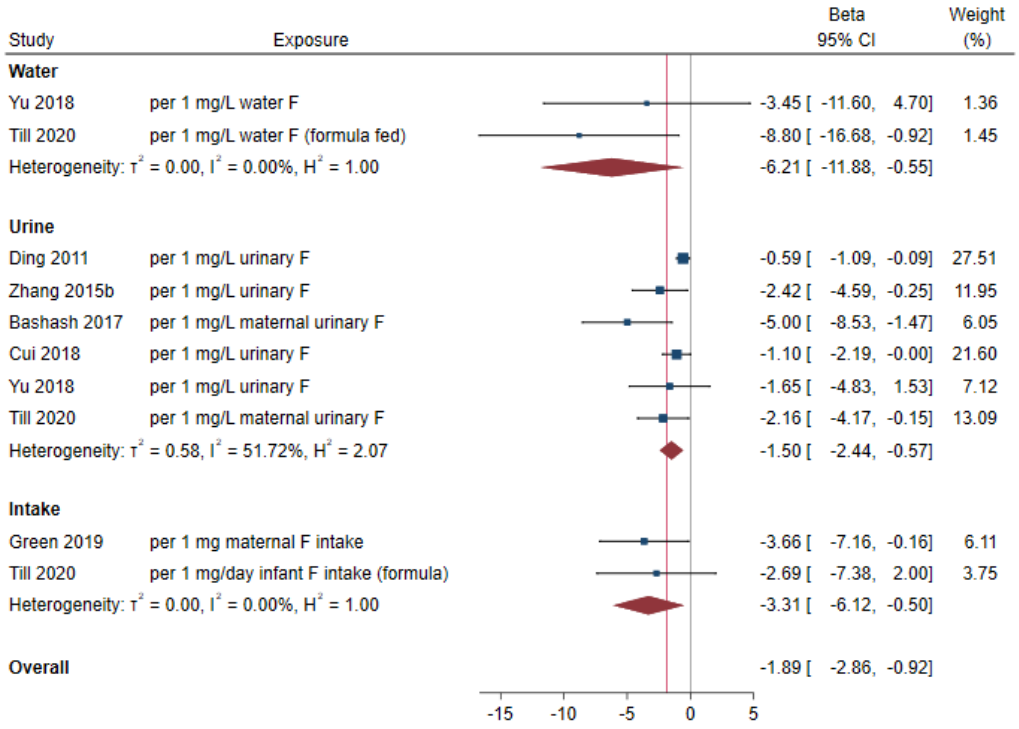


Figure B-15. Association Between Fluoride Exposure and IQ Scores in Children by Exposure Type for Individual-level Exposures [Using Till *et al.* (2020) Estimates in Place of Green *et al.* (2019) Estimates]

Note: The analysis for publication bias for studies with fluoride in urine is identical to the overall analysis. The analyses for publication bias for studies with fluoride in water or fluoride intake rely on two studies each and are not shown.

Sensitivity Analysis for Individual-level Studies: Verbal IQ

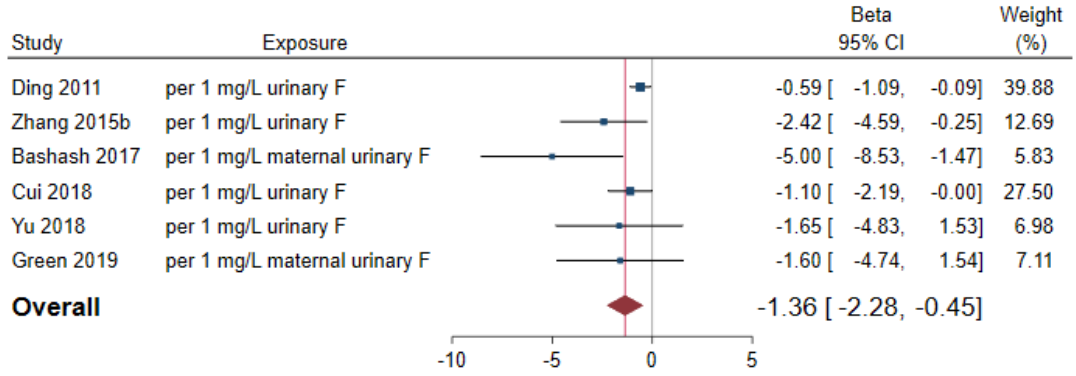


Figure B-16. Association between Fluoride Exposure and IQ Scores in Children Using Verbal IQ Score for Green *et al.* (2019)

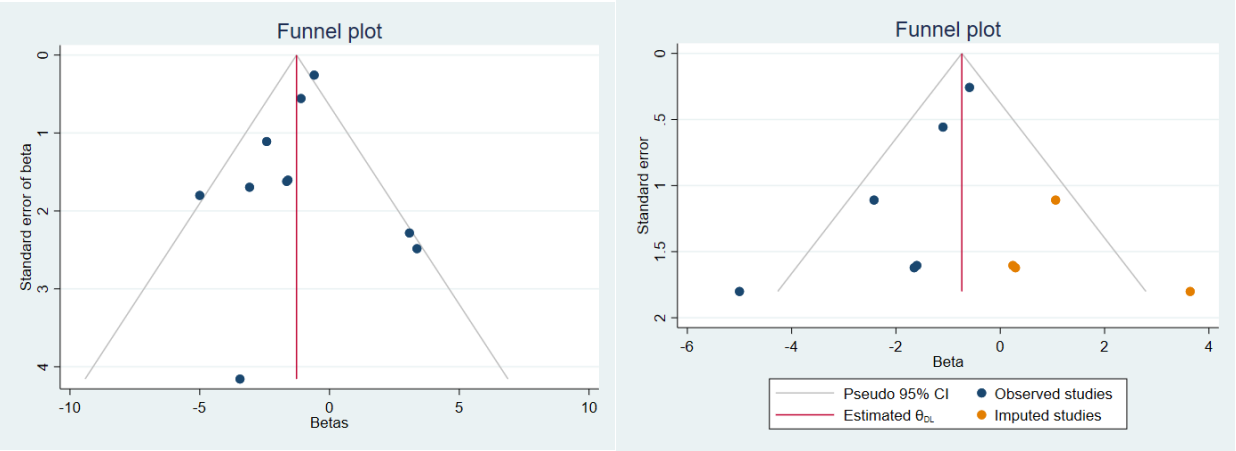


Figure B-17. Funnel Plot of Included Studies with Individual-level Exposures and Filled-in Funnel Plot to Eliminate Publication Bias [Using Verbal IQ Score for Green *et al.* (2019)]

Right panel shows funnel plot filled in to the right using a run estimator. Filling in to the left using a linear or a run estimator showed no change in the pooled effect estimate.

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -1.62
      SE of beta1 =     0.629
                z =     -2.58
      Prob > |z| =    0.0098
```

Figure B-18. Test for Publication Bias for Studies with Individual-level Exposures [Using Verbal IQ Score for Green *et al.* (2019)]

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the right

Iteration                          Number of studies =      9
  Model: Random-effects                     observed =      6
 Method: DerSimonian-Laird                  imputed =       3

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Beta	[95% Conf.	Interval]
Observed	-1.365	-2.278	-0.452
Observed + Imputed	-0.814	-1.788	0.159

```
Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration                          Number of studies =     10
  Model: Random-effects                     observed =      6
 Method: DerSimonian-Laird                  imputed =       4

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Beta	[95% Conf.	Interval]
Observed	-1.365	-2.278	-0.452
Observed + Imputed	-0.737	-1.654	0.180

Figure B-19. Trim-and-fill Analysis for Studies with Individual-level Exposures [Using Verbal IQ Score for Green *et al.* (2019)]

Filling in to the left using a linear or a run estimator showed no change in the pooled effect estimate.

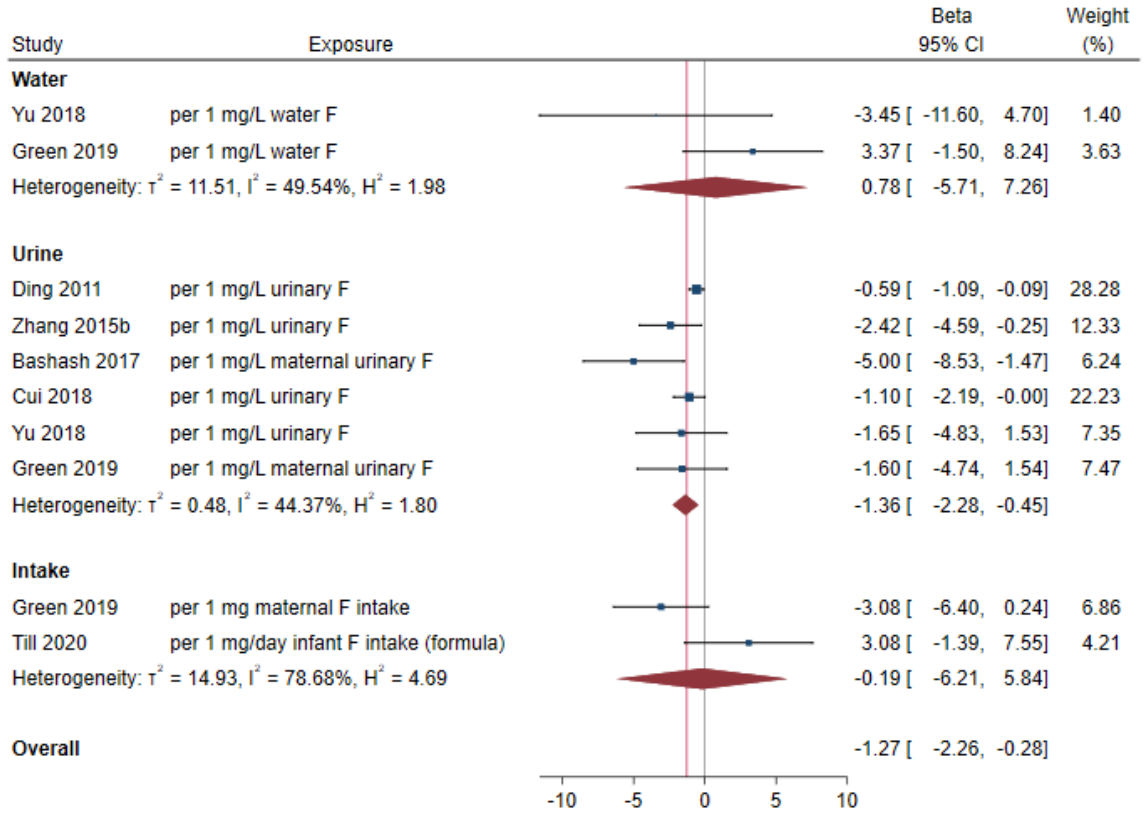


Figure B-20. Association Between Fluoride Exposure and IQ Scores in Children by Exposure Type Using Verbal IQ Score for Green *et al.* (2019)

Note: The analysis for publication bias for studies with fluoride in urine is identical to the overall analysis. The analyses for publication bias for studies with fluoride in water or fluoride intake rely on two studies each and are not shown.

Sensitivity Analysis for Individual-level Studies: Performance IQ

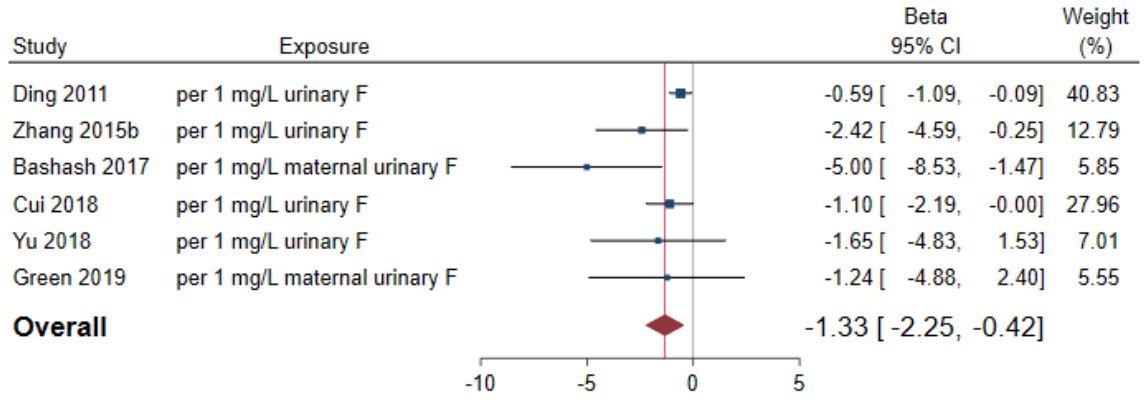


Figure B-21. Association Between Fluoride Exposure and IQ Scores in Children Using Performance IQ Score for Green *et al.* (2019)

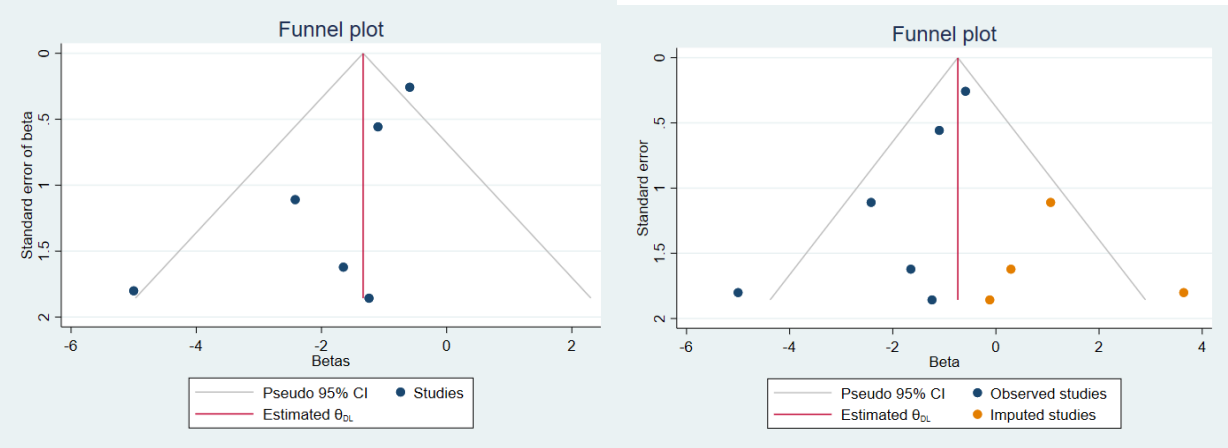


Figure B-22. Funnel Plot of Included Studies with Individual-level Exposures and Filled-in Funnel Plot to Eliminate Publication Bias [Using Performance IQ Score for Green *et al.* (2019)]

Right panel shows funnel plot filled in to the right using a run estimator. Filling in to the left using a linear or a run estimator showed no change in the pooled effect estimate.

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -1.52
      SE of beta1 =     0.623
                z =     -2.44
      Prob > |z| =    0.0147
```

Figure B-23. Test for Publication Bias for Studies with Individual-level Exposures [Using Performance IQ Score for Green *et al.* (2019)]

Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the right

Iteration — Number of studies = 9
Model: Random-effects — observed = 6
Method: DerSimonian-Laird — imputed = 3

Pooling
Model: Random-effects
Method: DerSimonian-Laird

Studies	Beta	[95% Conf.	Interval]
Observed	-1.335	-2.248	-0.421
Observed + Imputed	-0.777	-1.759	0.204

Nonparametric trim-and-fill analysis of publication bias
Run estimator, imputing on the right

Iteration — Number of studies = 10
Model: Random-effects — observed = 6
Method: DerSimonian-Laird — imputed = 4

Pooling
Model: Random-effects
Method: DerSimonian-Laird

Studies	Beta	[95% Conf.	Interval]
Observed	-1.335	-2.248	-0.421
Observed + Imputed	-0.739	-1.662	0.184

Figure B-24. Trim-and-fill Analysis for Studies with Individual-level Exposures [Using Performance IQ Score for Green *et al.* (2019)]

Filling in to the left using a linear or a run estimator showed no change in the pooled effect estimate.

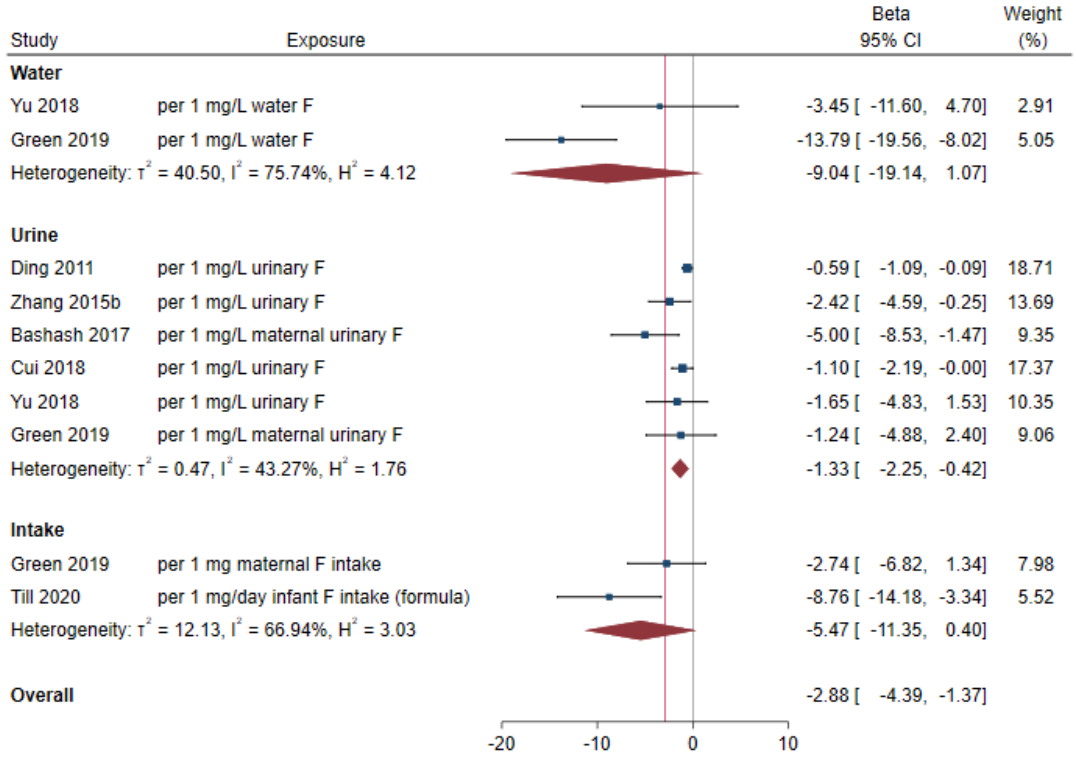


Figure B-25. Association Between Fluoride Exposure and IQ Scores in Children by Exposure Type Using Performance IQ Score for Green *et al.* (2019)

Note: The analysis for publication bias for studies with fluoride in urine is identical to the overall analysis. The analyses for publication bias for studies with fluoride in water or fluoride intake rely on two studies each and are not shown.

Sensitivity Analysis for Individual-level Studies: Excluding Cui *et al.* (2018)

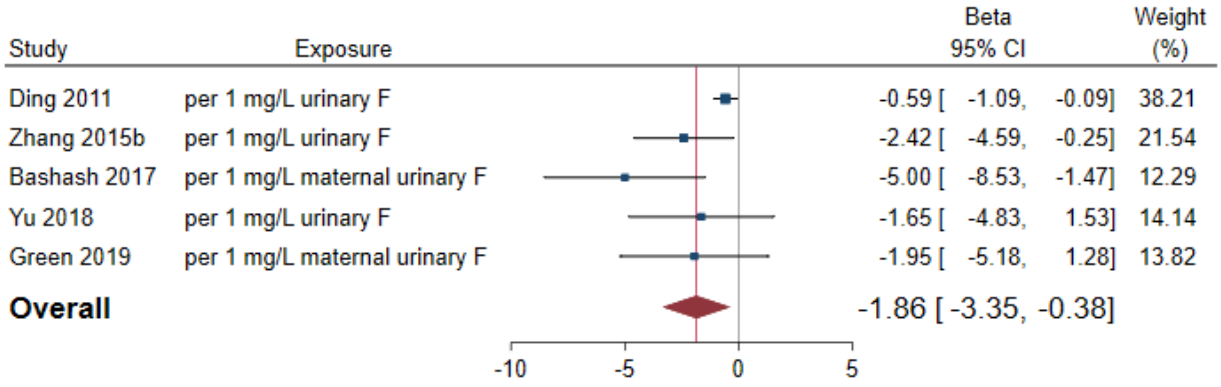


Figure B-26. Association Between Fluoride Exposure and IQ Scores in Children [Excluding Cui *et al.* (2018)]

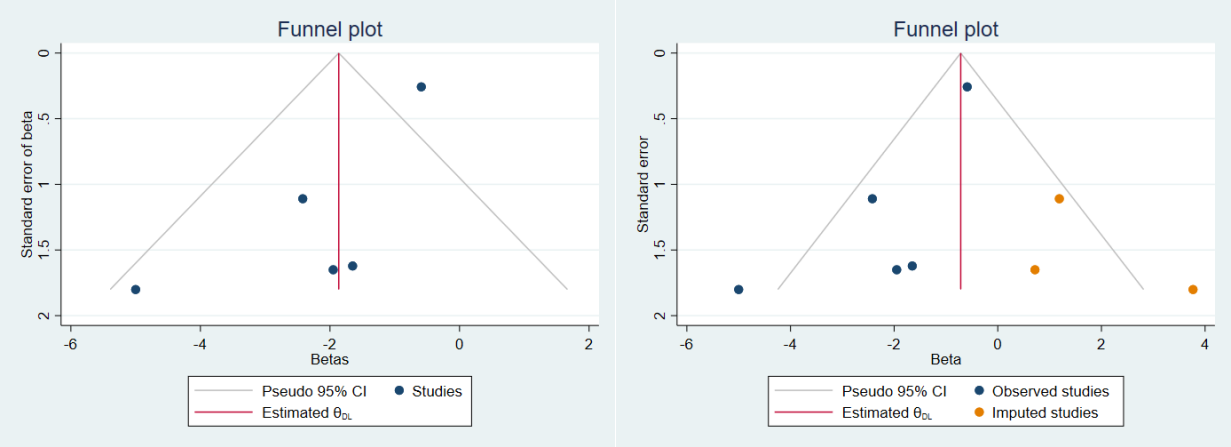


Figure B-27. Funnel Plot and Filled-in Funnel Plot to Eliminate Publication Bias of Included Studies with Individual-level Exposures [Excluding Cui *et al.* (2018)]

Right panel shows filling in to the right using a linear estimator. Filling in to the right using a run estimator or to the left using a linear or run estimator showed no change in the pooled effect estimate.

```
Regression-based Egger test for small-study effects
Random-effects model
Method: DerSimonian-Laird

H0: beta1 = 0; no small-study effects
            beta1 =     -1.68
      SE of beta1 =     0.638
                z =     -2.63
      Prob > |z| =    0.0085
```

Figure B-28. Test for Publication Bias for Studies with Individual-level Exposures [Excluding Cui *et al.* (2018)]

```
Nonparametric trim-and-fill analysis of publication bias
Linear estimator, imputing on the right

Iteration                           Number of studies =      8
  Model: Random-effects                      observed =      5
 Method: DerSimonian-Laird                   imputed =      3

Pooling
  Model: Random-effects
 Method: DerSimonian-Laird
```

Studies	Beta	[95% Conf.	Interval]
Observed	-1.864	-3.351	-0.377
Observed + Imputed	-0.715	-2.056	0.625

Figure B-29. Trim-and-fill Analysis for Studies with Individual-level Exposures [Excluding Cui *et al.* (2018)]

Filling in to the right using a run estimator or to the left using a linear or run estimator showed no change in the pooled effect estimate.

NIEHS/DNTP Response to the

REVIEW OF THE REVISED NTP MONOGRAPH ON THE SYSTEMATIC REVIEW OF FLUORIDE EXPOSURE AND NEURODEVELOPMENTAL AND COGNITIVE HEALTH EFFECTS: A LETTER REPORT

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

The National Toxicology Program (NTP) and the Division of the National Toxicology Program (DNTP) at the National Institute of Environmental Health Sciences (NIEHS) appreciates the comments provided by the National Academies of Sciences, Engineering, and Medicine (NASEM) Committee in its review of the September 2020 revised draft of the NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects.

The NASEM Committee reviews of the draft NTP monographs on fluoride (September 2019 and September 2020) determined that, "Overall the revised monograph seems to include a wealth of evidence and a number of evaluations that support its main conclusion, but the monograph falls short of providing a clear and convincing argument that supports its assessments…" Thus, we have removed the hazard assessment step and added "State of the Science" to the title to indicate the change. The monograph was retitled the "NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review" and underwent additional peer review by five external experts. The prepublication 2022 NTP Monograph includes consideration of comments from that external peer review and from key stakeholders across HHS in addition to the NASEM Committee's comments.

In addition, the prepublication 2022 NTP Monograph removes the meta-analysis that was added at the NASEM Committee's request following their review of the 2019 draft NTP Monograph. The meta-analysis is being prepared as a separate journal publication, taking into consideration the NASEM Committee's comments on the 2020 draft NTP Monograph.

The NIEHS/DNTP separated the NASEM Committee Letter Review comments that were applicable to the prepublication 2022 NTP Monograph from the comments that are focused on the meta-analysis because these are now two separate, distinct evaluations. For each set of comments, NIEHS/DNTP prepared responses and described changes made in response to the comments. The NASEM Committee's comments and responses that were directly relevant to the meta-analysis are in the document titled "Sup01_Meta-analysis" (see Word file Sup01_Meta-analysis – NASEM_comments_on_meta-analysis_only_and_NIEHS_DNTP_response.docx) and are not included in this document. In particular, the section titled "*Evaluation of the Meta-Analysis*" is not included here.

This document contains a subset of the overall NASEM Committee's comments and NIEHS/DNTP responses that are related to the prepublication 2022 NTP Monograph. For clarity, the complete text from the NASEM Letter Review, with the exception of the meta-analysis relevant comments, have been included in the pages that follow and is formatted in black text. The responses begin with the word "**Response**," are formatted in blue font, and are interspersed within the original NASEM Committee text.

REVIEW OF THE REVISED NTP MONOGRAPH ON THE SYSTEMATIC REVIEW OF FLUORIDE EXPOSURE AND NEURODEVELOPMENTAL AND COGNITIVE HEALTH EFFECTS: A LETTER REPORT

Committee to Review the Revised NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects

Board on Environmental Studies and Toxicology Division on Earth and Life Studies

A Consensus Study Report of

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

This project was supported by Contract HHSN263201800029I between the National Academy of Sciences and the National Institute of Environmental Health Sciences. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project.

International Standard Book Number-13: 978-0-309-16269-2 International Standard Book Number-10: 0-309-16269-6 Digital Object Identifier: https://doi.org/10.17226/26030

This publication is available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2021. *Review of the Revised NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Letter Report.* Washington, DC: The National Academies Press. https://doi.org/10.17226/26030.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Division on Earth and Life Studies
Board on Environmental Studies and Toxicology

January 26, 2021

Mary S. Wolfe, PhD
Deputy Division Director for Policy
Director, Office of Liaison, Policy, and Review National Toxicology Program
111 T.W. Alexander Drive Keystone Building, MD A2-03 Research Triangle
Park, NC 27709

Dear Dr. Wolfe,

At your request, the National Academies of Sciences, Engineering, and Medicine (the National Academies) convened the Committee to Review the Revised NTP Monograph on Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects. The committee was asked to determine whether substantive concerns raised in the National Academies 2020 report *Review of the Draft NTP Monograph: Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* have been sufficiently addressed by revisions of the monograph and whether the evidence presented by NTP in the revised monograph supports its conclusions. Overall, the committee appreciates the efforts to revise the monograph to address concerns previously raised. Although the monograph is much improved in many important ways, the committee still has concerns as expressed in the comments in this letter report.

Given the strong views of water-fluoridation advocates who are concerned with preventing dental caries and their systemic sequelae and the equally strong views of antifluoridation advocates who contend that fluoride exposure poses a threat to health, preparing a report that can withstand the scrutiny of both sides is extremely challenging. The report must present its methods clearly, document the results transparently, and provide the rationale for conclusions in such a way that even those who disagree with them will appreciate that the process by which they were derived is clear and was implemented without error. The question is not whether this committee or the multiple audiences come to the same conclusions but rather whether the methods and analysis documented in the monograph support NTP's conclusions.

According to the committee's task statement, the committee's primary focus was "to determine whether the evidence as presented by NTP in its revised monograph supports its conclusions." As documented in this letter report, the committee had difficulty in following various aspects of the reported methods, identified a few worrisome remaining inconsistencies, was not able to find some key data used in the meta-analysis, and had concern about the wording of some conclusions. Even though the evidence provided appears to show consistent indications of an association between exposure to high fluoride concentrations and cognitive deficits in children, the monograph falls short of providing a clear and convincing argument that supports its assessment. It also needs to emphasize that much of the evidence presented comes from studies that involve

relatively high fluoride concentrations and that the monograph cannot be used to draw conclusions regarding low fluoride exposure concentrations (less than 1.5 mg/L), including those typically associated with drinking water fluoridation.

Sincerely,

David A. Savitz, *Chair* Committee to Review the Revised NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects

REVIEW OF THE REVISED NTP MONOGRAPH ON THE SYSTEMATIC REVIEW OF FLUORIDE EXPOSURE AND NEURODEVELOPMENTAL AND COGNITIVE HEALTH EFFECTS: A LETTER REPORT

In 2019, the National Toxicology Program (NTP) released the draft monograph *Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* (NTP 2019a).[1] The draft monograph summarized the findings of the systematic review and concluded that "fluoride is *presumed to be a cognitive neurodevelopmental hazard to humans*. This conclusion is based on a consistent pattern of findings in human studies across several different populations showing that higher fluoride exposure is associated with decreased IQ or other cognitive impairments in children" (NTP 2019a, p. 59). Given the controversies surrounding the risks and benefits associated with fluoride exposure and to ensure the integrity of its evaluation, NTP asked the National Academies of Sciences, Engineering, and Medicine (the National Academies) to review the draft monograph.

The National Academies committee that was convened to address the request identified deficiencies in the analysis of various aspects of some of the studies and in the analysis, summary, and presentation of the data in the draft monograph (NASEM 2020). The committee provided many suggestions for improvement and concluded that NTP had not adequately supported its conclusions. It noted that the committee's finding did not mean that NTP's conclusions were incorrect; rather, further analysis or reanalysis would be needed to support the conclusions. Taking the committee's suggestions into consideration, NTP revised the draft monograph.

STATEMENT OF TASK AND COMMITTEE APPROACH

NTP asked the National Academies to review the revised monograph (NTP 2020a) to ensure that it was responsive to the committee's recommendations and, more important, adequately supported its conclusions. Attachment A provides the verbatim statement of task. The committee that reviewed the draft monograph was reconvened to review the revised monograph; Attachment B provides biographic information on the committee.

To complete its task, the committee held several virtual meetings, one of which included a public session at which NTP provided an overview of the changes that had been made in the draft monograph. The committee reviewed the revised monograph, including the newly added appendixes with details of lower risk-of-bias studies and the meta-analysis; NTP responses to the committee's recommendations; the revised protocol; and public comments submitted to the committee. It is important to note that the committee did not conduct its own independent evaluation of the evidence, nor did it conduct a data audit; both were outside its scope. The committee reviewed the revised monograph and determined whether the evidence as presented in it supported NTP's main conclusion that "fluoride is *presumed to be a cognitive neurodevelopmental hazard to humans*" (NTP 2020a, p. 80). Each section below provides the committee's assessment of NTP responses to substantive issues previously raised (NASEM 2020) regarding methods, animal evidence, human evidence, and communication. Attachment C summarizes the substantive issues previously raised and NTP's responses. The committee

[1] Referred to hereafter as the draft monograph. The revised version released in 2020 is referred to as the revised monograph.

provides many recommendations for improving the revised monograph and has highlighted in boldface, italics some particularly critical ones, but all are important to address.

METHODS

In its previous review, the committee raised several issues associated with the general methods of NTP's systematic review process. The issues were concerning because they decreased the transparency of the process and the probability of reproducing the findings and did not align with some general best practices for systematic review. The committee finds that NTP has addressed many of the issues regarding methods in its revisions of the draft monograph but notes that some further improvements would be useful. A brief overview of suggested improvements is provided below; other methodologic issues raised in the previous review that are not discussed here have been adequately addressed in the revised monograph. The committee considers the remaining issues related to the systematic review methods to be minor with the exception of the comment below concerning NTP's process for upgrading and downgrading the body of evidence (NTP 2020b, Table 5).

First, the role of the Office of Health Assessment and Translation (OHAT) handbook (NTP 2015, 2019b,c) has been explicitly added to the revised monograph. Two statements in the revised monograph—on pp. ii and 6 (footnote)—describe the OHAT handbook as a source of general systematic review methods that are selected and tailored to the project in the prespecified protocol. Although the statement clarifies the general role of the handbook, the committee finds that it does not address the committee's previous recommendation to set the expectation for how closely the process described in the handbook will be followed in the protocol and in the eventual systematic review. For example, the handbook section "Key Questions and Analytical Framework" that guides development of the population, exposure, comparator, and outcomes (PECO) statement is not included in the fluoride protocol or the revised monograph. As the committee recommended in its previous review, NTP should treat each systematic review protocol as a stand-alone document that contains all the information necessary for understanding of the planning and conduct of the review, and these expectations should be explicitly stated in the protocol. The committee did not find that revisions of the protocol adequately addressed this recommendation.

Response: Agree (change made)

- We appreciate the desire of the Committee for more specificity in the protocol with respect to laying out all aspects of the systematic review; however, we respectfully submit that the detail provided in the protocol followed for both the systematic review and meta-analysis are well within, and in many aspects exceed, standard practice in the field. We added the following text to the *Methods* section of the prepublication 2022 NTP Monograph to further clarify the role of the OHAT handbook.

  "The protocol served as the complete set of methods followed for the conduct of this systematic review. The OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673) is a source of general systematic review methods that were selected and tailored in developing this protocol. Options in the OHAT handbook that were not specifically referred to in the protocol were not part of the methods for the systematic review."

Second, several recommendations in the committee's previous review that might have increased the overall transparency of the monograph do not appear to have been addressed, such as reporting the excluded studies at the title and abstract step (also recommended in the OHAT handbook) and adding to the protocol clear definitions for each factor that contributes to increasing or decreasing confidence in the body of evidence and key considerations that warrant upgrading or downgrading the body of evidence (NTP 2020b, Table 5, p. 18). ***The committee found that such omissions decrease the reproducibility and transparency of the systematic review process and should be viewed as a deficiency that should be addressed.***

Response: Agree (change made)

- Figure 2 (titled "Study Selection Diagram") in the prepublication 2022 NTP Monograph has been transformed into an interactive Tableau® figure (https://hawcproject.org/summary/visual/assessment/405/Figure-2/). The list of studies excluded at the title and abstract stage can now be accessed through this interactive Tableau® figure. The section of the prepublication 2022 NTP Monograph titled "*Confidence Rating: Assessment of Body of Evidence*" has been expanded to provide short descriptions and key considerations for each factor considered for downgrading or upgrading confidence in the human body of evidence.

Third, NTP has added text to the revised monograph regarding the use of the SWIFT-Active Screener tool to priority-rank studies for screening and to set stopping rules. However, the committee recommends that a more detailed explanation of some terminology be added to eliminate any confusion that might arise given the novelty of the use of such tools. For example, the term *percent recall* might lack consistent interpretation, and it would be helpful to define it to clarify the implications of stopping at a set recall, such as 98% estimated recall, and the implication of the potential number of missed studies at the set stopping point.

Response: Agree (change made)

- We call attention to the Committee of text on pg. 22 of the 2020 draft NTP Monograph that discusses the SWIFT-Active screening process and implications for stopping at 98% with respect to possible studies missed. We assume that the Committee means to refer to the term "predicted recall," as the term "percent recall" is not found in the 2020 draft NTP Monograph. In the *Evaluation of SWIFT-Active Screener Results* section of the prepublication 2022 NTP Monograph, the use of the term "predicted recall" has been supplemented with a layman description of the concept.

ANIMAL EVIDENCE

The committee appreciates that NTP agrees that there were problems with the risk-of-bias analyses of the animal studies, and it agrees with NTP's decision that devoting further effort to refining the analyses is not worthwhile but has concerns regarding the reasons provided by NTP for not reanalyzing any of the animal data. NTP provided the following reasons in the revised monograph: "(1)…a more critical risk-of-bias assessment would result in fewer relevant animal studies judged to be of high quality; (2)…the highest quality experimental animal study reviewed for this monograph (McPherson et al. 2018) did not find effects of fluoride on learning, memory or motor activity in the critical ≤20 ppm in drinking water concentration range; and (3)…[there are] a large number of human epidemiology studies directly addressing neurobehavioral and cognitive effects of fluoride in children" (NTP 2020a, p. 58). Although the committee agrees with the first reason to the extent that a reanalysis would probably not find any low risk-of-bias studies, it is inappropriate for NTP to highlight one specific study (McPherson et al. 2018) as a rationale for not reassessing all the animal literature. Regarding the third reason, the committee disagrees that a large number of epidemiologic studies generally negates the value of animal studies in hazard determination. Instead, NTP should clarify that a large number of relevant epidemiologic studies can be used as a primary source of evidence to support a conclusion in its hazard identification scheme for integrating human and animal data to reach a final rating of the overall evidence.

Response: Disagree (no change)

- The reasons cited in the 2020 draft NTP Monograph for not reanalyzing the animal data have been removed from the prepublication 2022 NTP Monograph. We appreciate the Committee's objection to citing one high quality study as part of the reason to not carry further reviews of the experimental animal literature. Note that the significance of McPherson et al. (2018) in this context is that it was performed at the NIEHS specifically to address deficiencies in prior studies identified in the NTP 2016 systematic review of the animal literature. We also respectfully disagree that the 2020 draft NTP Monograph implies that "a large number of epidemiologic studies generally negates the value of animal studies in hazard determination" as is indicated in the Committee's comments. The prepublication 2022 NTP Monograph does not state or imply that epidemiological studies negate the value of animal studies.

In the revised monograph, NTP has added a disclaimer about the animal evidence but left the original discussion unchanged. ***The committee strongly recommends that NTP not publish the monograph with the original text that states that evidence of effects on activity or motor function invalidate observations of learning or memory deficits.*** If taken out of context, that text could be interpreted incorrectly or raise questions about the scientific validity of the monograph more generally. For example, Yang et al. (2018) was grouped with studies that were classified as high risk of bias because in addition to finding learning deficits by using the Morris water maze, it found open-field effects. However, the Morris water maze data are highly unlikely to be affected by the minor open-field differences found in that study not only because swimming is different from ambulation and rearing but because there were no differences among groups in learning the task over 5 days of testing. Differences emerged only on retesting 10 and 20 days later and then were not significant on days 30, 40, and 50. It is implausible that rats with any kind of activity effect would learn the Morris water maze equally well, show deficits on only some retest days, and then fail to show further deficits because of an open-field effect. That example shows that the monograph

overgeneralizes concerns about activity without examining the learning data in sufficient depth to determine their validity. Instead, the monograph dismisses all data on the basis of a sweeping indictment that no learning differences can be used if activity differences are found. That view is not scientifically justifiable.

The committee strongly recommends that NTP revise the monograph text that states that a change in motor activity necessarily complicates interpretation of learning and memory tests and that the absence of an evaluation of motor activity is automatically problematic.[2] First, the mere observation of a change in motor activity does not automatically undermine a learning and memory effect, nor does the absence of statements about the general health of the animals undercut validity, as the monograph asserts. Second, the absence of a motor-activity test does not necessarily invalidate a learning and memory effect if the test has an internal control for activity. The central issue is whether the learning and memory method alone or in combination with other indexes dissociates learning from performance in a way that allows a correct interpretation of animal learning and memory.

Response: Agree (change made)

- We agree with the Committee's recommendation, and this information has been removed from the prepublication 2022 NTP Monograph.

[2] Text that needs to be edited includes p. 58, last paragraph, lines 4–7, and p. 59, last paragraph, lines 4–13.

HUMAN EVIDENCE

The committee provided many suggestions in its previous report (NASEM 2020) to address deficiencies that it identified in the analysis of the human evidence provided in the draft monograph (NTP 2019a). The headings in this section represent the overarching concerns that the committee raised in its previous report, and the text provides the committee's assessment of NTP's responses to the concerns and the revisions made in the draft monograph.

Potential for Biased Selection of Studies

NTP has done excellent work in responding to concerns about a potentially biased selection of studies. The expansion of the literature search to include several Chinese databases strengthens NTP's review and strengthens the overall process that it has used to support its conclusions. In a few respects, NTP could improve the process even further, and these are discussed below.

First, the databases that NTP chose for searching the Chinese literature were selected on the basis of their covering "studies previously identified from other sources" (NTP 2020b, p. 6). Although that approach might be appropriate, it would have been helpful for NTP to provide a few brief details about the quality or scope of the two new Chinese databases. For example, NTP chose such databases as PubMed and BIOSIS for a reason—for example, fairly extensive coverage of journals or some quality-control standards. Do the same reasons or qualities also apply to the CNKI and Wanfang databases? ***NTP should also address the concern that selecting databases on the basis of studies already identified might perpetuate, rather than ameliorate, biases resulting from the initial search.***

Response: Disagree (no change)

- We recognize the desire of the Committee for further information on the databases selected. Details were added to the *Supplemental Chinese Database Literature Search* section of the prepublication 2022 NTP Monograph to further explain the rationale for our approach. We searched for and were unable to find definitive guidance on the most comprehensive, highest quality, or otherwise most appropriate non-English-language databases for health studies of fluoride. Therefore, we chose databases (CNKI and Wanfang) that identified non-English-language studies that we were aware of—"seed" studies previously identified from other resources. It is standard practice to use seed studies to test search strings and explore the value of databases. An informationist requires some means of judging whether a database has the appropriate content. Note that the CNKI and Wanfang databases are large and recognized by information scientists in the United States. We recognize that the coverage, scope, and completeness of the various search engines providing access to the Chinese literature is somewhat opaque. Therefore, we explored more than 15 databases to identify databases that indexed the seed studies. The CNKI and Wanfang databases contained the highest proportion of seed studies (>50%).
- We found this the most effective approach to ensure that databases selected were able to identify at least some references that were appropriate to the topic. Preliminary searches were performed on all of the databases considered, understanding that optimization of search strings would then be necessary for each database. Further optimization of the search string was only applied to databases where at least one previously identified seed study was found. We find it unlikely that not finding seed studies would make it more likely that these databases contained potentially missing studies. Therefore, we respectfully disagree with the Committee's concern that this approach may have further perpetuated a potential bias in our

initial search.

- Furthermore, we took steps to ensure that a consistent peer-review standard was applied to the included human studies identified in the CNKI and Wanfang databases and to all of the relevant human studies published in non-English languages. An epidemiologist fluent in Chinese and an informationist conducted searches for publicly available information on peer-review practices of all non-English language journals (n = 30) in which human studies were published that had been included as relevant for this review. If publicly available information was not available on peer-review practices, we contacted the journals in Chinese and requested additional information. Through this process, we confirmed that 28 out of the 30 non-English journals in which relevant human studies were published have peer-review practices (described on the website, listed in a major bibliographic database with known peer-review standards, and/or confirmed directly). Publicly available details of the peer-review procedures of two journals (Chinese Primary Health Care and Lit Inf Prev Med, renamed Preventive Medicine Tribune) were limited and we did not receive responses to our inquiries. There were only three relevant studies that were published in these journals (Yao et al. 1996; Yao et al. 1997; and Hong et al. 2001[a]) and we had previously rated all of them as high risk-of-bias studies. A note was added to the rationale for the "other potential threats to internal validity" risk-of-bias question for each of these studies in the Health Assessment and Workspace Collaborative (HAWC) to reflect that they were published in a journal with an unclear peer-review process.

Second, the monograph states that "newly-retrieved human references were reviewed to identify studies that might impact conclusions with priority given to identifying and translating null studies" (NTP 2020a, p. 10). It is somewhat understandable that NTP would want to focus on null studies because these studies would most likely affect NTP's conclusions. However, that statement provides questionable justification, given NTP's primary mission—an unbiased review of the literature, which means including all relevant studies whether positive or negative. ***NTP needs to consider all eligible studies identified in the new literature search.***

Response: Agree (change made)

- We accepted this suggestion and have taken additional steps to translate and extract data from all non-English language studies identified from the Chinese database searches that were not included previously. As a result, eight additional studies have been incorporated into the systematic review (six on IQ in children, one on other neurodevelopmental or cognitive effects in children, and one on cognitive effects in adults). All eight are high risk-of-bias studies, and the addition of these eight studies has not resulted in any changes to the confidence ratings or any substantive updates to discussions in the monograph. We have updated the text in the *Literature Search* section to reflect that the search of Chinese databases was conducted to identify studies that may have been missed in previous searches because non-English-language studies are not always indexed in the main databases used for

[a]Yao L, Zhou J, Wang S, Cui K, Lin F. 1996. Analysis of TSH levels and intelligence of children residing in high fluorosis areas. Lit Inf Prev Med 2(1): 26-27.

Yao Y. 1997. Comparative assessment of the physical and mental development of children in endemic fluorosis area with water improvement and without water improvement. Lit Inf Prev Med 3(1): 42-43.

Hong FG, Cao YX, Yang D, Wang H. 2001. [Research on the effects of fluoride on child intellectual development under different environmental conditions]. Chin Prim Health Care 15(3): 56-57.

this systematic review.

Lack of Independence of Studies

NTP recognizes that the monograph evaluates and describes multiple publications from the same study. It also indicates some uncertainty about a few publications that cannot be attributed to a parent study, given insufficient published details. The revised monograph states that it addressed the independence issue, but the exact process used for selection of a single publication remains unclear, and in the meta-analysis, two reports on the same population are inappropriately included as described below. It would be useful for the monograph to identify clearly which publications were derived from which study to minimize concerns about potential selection bias;

doing so would also help to define the publications selected for the meta-analysis. NTP might consider editing the monograph to differentiate studies from publications or papers. That revision can be achieved by restricting the term *study* to the original body of research conducted with a defined population during a specified time and using the terms *publications* and *papers* to refer to the published work drawn from a study.

Response: Disagree (edited for clarity)

- We assure the Committee that all attempts were made to determine when a single study population was the source material for more than one report. In the prepublication 2022 NTP Monograph, we have added details to clearly define the approach used in the document, and we have gone through the monograph to ensure that appropriate distinctions are made. “Study population” refers to a defined population on which an original body of research was conducted. The published work drawn from that original body of research is often referred to as a “study.” In addition, IQ studies and studies on other neurodevelopmental effects in children that report on the same study populations have now been identified in Tables 6 and 7 of the monograph. Also note that the prepublication 2022 NTP Monograph clarifies that the terms “study” and “publication” are used interchangeably to refer to a published work drawn from an original body of research conducted on a defined population.

Inconsistent Application of Risk-of-Bias Criteria

In response to the committee's concern regarding the risk-of-bias assessment, NTP has added Appendix 4, which provides its rationale for classifying studies relative to their estimated risk of bias. The new appendix is helpful and adds transparency, but inconsistencies remain in the application of risk-of-bias criteria to individual studies, particularly in NTP's evaluation of how various studies handled major confounders, co-exposures, and outcomes. An example concerns the handling of co-exposure to arsenic and lead. According to the protocol, a cross- sectional study is rated as having a probably low risk of bias on confounding if there is direct evidence that appropriate adjustments for arsenic and lead were made; the monograph requires the studies to address arsenic and lead, if applicable. Barberio et al. (2017) did not adjust for arsenic and lead, nor did the authors discuss co-exposures; however, it was rated as having a probably low risk of bias. The committee also identified several studies whose classification changed in revisions in the draft monograph without any justification provided (Sudhir et al.
2009; Trivedi et al. 2012; Das and Modal 2016).

Response: Agree (change made)

- We recognize the Committee's continued concerns over the consistent application of the risk-of-bias criteria. While a top priority to us as well, it is important to emphasize to the Committee that the risk-of-bias criteria laid out in the protocol are not an algorithm or a scoring system. Each study describes a unique set of circumstances. We apply the risk-of-bias criteria to individual studies and specifically look across studies to ensure that the criteria are consistently applied, with the understanding that scientific judgement is needed, and risk-of-bias judgements are made on a case-by-case basis.
- Barberio et al. (2017) used data from the Canadian Health Measures Survey which consists of a nationally representative sample of Canadians. Because most Canadians (~89%) receive water from municipal water supplies, which monitor for levels of lead and arsenic, we assumed that co-exposure to lead and arsenic in drinking water was not applicable to this study (which follows the guidance in the protocol). However, we agree that this reasoning should have been more explicitly explained, and we have added further details to the confounding risk-of-bias domain discussion for this study in *Appendix E* of the prepublication 2022 NTP Monograph (previously *Appendix 4* in the 2020 draft NTP Monograph).
- Below, we provide justifications for why the three studies identified by the Committee changed in risk-of-bias classification. Many of the changes occurred after implementing the Committee's recommendations from the first peer review with regard to risk of bias. However, because the 2020 draft NTP Monograph was still in draft form, we consider that if the reasoning for the risk-of-bias ratings was clearly explained in the appendix, reasons for changing ratings of individual studies between drafts was not appropriate.
    - Sudhir et al. 2009 – From the 2019 draft NTP Monograph to the 2020 draft NTP Monograph, the confounding rating changed from *probably high risk of bias* to *probably low risk of bias*. Because of this rating change, the overall risk-of-bias status of the study changed from high to low risk of bias. The change in the confounding rating is based on the use of groundwater quality maps to identify areas where arsenic could be a concern. The following explanation of this approach was added to the 2020 draft NTP Monograph: "*In order to identify areas of China, India, and Mexico where arsenic is a concern, groundwater quality maps were evaluated*

(https://www.gapmaps.org/Home/Public#) (Podgorski and Berg 2020). If no arsenic measurements were available for the area, the arsenic groundwater quality predictions from the global arsenic 2020 map were used (Podgorski and Berg 2020). If an area had less than 50% probability of having arsenic levels greater than 10 µg/L (the WHO guideline concentration), the area was considered not to have an issue with arsenic that needed to be addressed by the study authors."

- Trivedi et al. 2012 – From the 2019 draft NTP Monograph to the 2020 draft NTP Monograph, the confounding and exposure assessment ratings changed from *probably high risk of bias* to *probably low risk of bias*. Because of these rating changes, the overall risk-of-bias status of the study changed from high to low risk of bias. The change in the confounding rating is based on the use of groundwater quality maps to identify areas where arsenic could be a concern. The change in the exposure assessment rating is based on additional information obtained via author inquiry regarding the availability of groundwater fluoride levels and urine fluoride levels for all children for which IQ was assessed. Additional details are provided in *Appendix E* of the prepublication 2022 NTP Monograph.
- Das and Modal 2016 – From the 2019 draft NTP Monograph to the 2020 draft NTP Monograph, the outcome assessment rating changed from *probably low risk of bias* to *probably high risk of bias*. Because of this rating change, the overall risk-of-bias status of the study changed from low to high risk of bias. The change is based on the determination that the study authors administered the Combined Raven's Test for Rural China (CRT-RC) on an Indian population; however, this test is validated in a Chinese population not an Indian population and there is no information provided to indicate it was validated in the study population.

Evaluation of Confounding Insufficient, Difficult to Understand, or Applied Inconsistently

The revised monograph articulates a formal approach for assessing confounding by defining what it considers to be key confounders (that is, children's age, sex, and socioeconomic status) and other potential confounders. The addition of Appendix 4 makes it easier to follow how individual studies were assessed for risk of bias and confounding, but the committee still considers NTP's evaluation of confounding insufficient and sometimes inconsistently applied.
For example, Cui et al. (2020), which was rated as having a probably high risk of bias for confounding and was included with the lower risk-of-bias studies, presented a univariate comparison of IQ by high vs low fluoride exposure without any adjustment for confounders. According to the protocol, the study should have been rated as having a definitely high risk of bias for confounding and included with the higher risk-of-bias studies.

Response: Disagree (no change)

- We have re-evaluated risk of bias due to potential confounding. After further review, we would like to clarify to the Committee that Cui et al. (2020) did not meet the protocol's definition for *definitely high risk of bias* due to confounding, which requires direct evidence that important covariates, known confounders, and co-exposures differed between the groups and were not taken into account. Therefore, it is appropriate for the Cui et al. (2020) study to receive a *probably high risk of bias* rating for the confounding domain. As stated in *Appendix E* of the prepublication 2022 NTP Monograph (previously *Appendix 4* of the 2020 draft NTP Monograph), the *probably high risk of bias* rating is based on "indirect evidence" that age was not addressed as a confounder, and it may be related to both IQ and exposure. If there was direct evidence that age differed by exposure or IQ level, the study would have received a *definitely high risk of bias* rating, and the study would have been considered high risk of bias overall.

An example of inconsistent application of criteria to classify confounding is the adjustment for smoking and lead exposure. Specifically, Broadbent et al. (2015) is rated as having a probably high risk of bias on confounding, but other studies, such as Trivedi et al. (2012), were not similarly ranked.

Response: Disagree (no change)

- We respectfully disagree that this is a compelling example of inconsistent application of criteria to classify confounding. The primary reason the confounding domain in Broadbent et al. (2015) was rated *probably high risk of bias* was that it did not address age (a key confounder for all studies), and there was indirect evidence that age was not addressed as a confounder and that it may be related to both IQ and exposure (IQ was measured in children with an age range of 7-13 years with no information on the ages in the different groups or similarities between the groups), which justifies a rating of *probably high risk of bias* for confounding. Although Trivedi et al. (2012) also did not directly address age, they provided indirect evidence that children living in low and high fluoride villages were of similar ages based on the grades included in the study population (6^{th} and 7^{th} grade), which justifies a rating of *probably low risk of bias* for confounding.

Another example of inconsistent application of confounding assessment concerns Valdez- Jimenez et al. (2017); here, the issue was the unbalanced and unexplained demographic characteristics of the study population.

Response: Disagree (no change)

- We are unable to respond directly, as we find the exact concern unclear. Please note that Valdez-Jimenez et al. (2017) was rated *probably high risk of bias* for confounding in the 2020 draft NTP Monograph (and in the prepublication 2022 NTP Monograph) primarily based on indirect evidence that there was a potential for co-exposure with arsenic that was not addressed.

In Appendix 4, NTP attempted to clarify the direction and magnitude of bias due to confounding, although supporting text is often unclear. For several studies, NTP added a paragraph on the potential direction of bias due to lack of adjustment for arsenic exposure but then provided an argument to justify its absence as a confounder (see, for example, Sudhir et al. 2009). As noted, the committee did not conduct a full audit but examined some illustrative papers and still found reasons for concern.

Response: Disagree (edited for clarity)

- **Response:** The sub-bullet "Direction/magnitude of effect" text in *Appendix E* of the prepublication 2022 NTP Monograph (previously *Appendix 4* in the 2020 draft NTP Monograph) explains the conceptual impact of potential confounding concerns. In Sudhir et al. (2009), for example, the "Direction/magnitude of effect" text explains that the presence of arsenic would potentially bias away from the null if arsenic were present along with fluoride, and the text before and after this sub-bullet clearly states that arsenic is not considered an issue in this study. In *Appendix E* of the prepublication 2022 NTP Monograph, the "Direction/magnitude of effect" sub-bullet text has been revised to clearly state that the impacts on direction/magnitude of effect are conceptional concerns that depend on whether the specific issue applied. If a potential confounder is not considered an issue in a study, this determination is clearly stated in the "Direction/magnitude of effect" sub-bullet.

Possibility of Exposure Misclassification

The revised monograph addresses methodologic issues concerning potential exposure misclassification in light of the various types of exposure measures—for example, child and mother spot urines, serum, drinking water, urine, and residence—considered in the studies. Specifically, Appendix 4 addresses the potential direction and magnitude of bias due to exposure misclassification, if applicable. Thus, the committee's prior concerns regarding exposure misclassification appear to have been adequately addressed.

Response: Agree (no change requested)

- We appreciate the Committee's positive feedback.

Need for Further Consideration of Blinding

In its previous review, the committee recommended that NTP consider more carefully the effect of not intentionally blinding outcome assessors when evaluating the human studies. In its response, NTP indicated that when authors did not directly provide evidence of examiner blinding, it contacted the authors for information. It is unclear how the risk-of-bias information has been updated regarding blinding on the basis of any new information that was received. Specifically, Health Assessment and Workspace Collaborative records identify only whether and when authors were contacted but not what information was obtained or how it might have changed risk-of-bias ratings.

Response: Disagree (edited for clarity)

- Please note that the risk-of-bias rating explanations provided in HAWC and *Appendix E* in the prepublication 2022 NTP Monograph (previously *Appendix 4* of the 2020 draft NTP Monograph) previously noted whether an author responded and whether the response provided affected the risk-of-bias rating. To provide information more clearly on author inquiries and how information provided by the authors was used in the risk-of-bias analysis, we have also made updates to the HAWC study profiles for each human study and to *Appendix E*. Please note the following:
    - When author inquiries were conducted, they are noted in the study profiles (e.g., "Author was contacted in September 2017 to obtain information for RoB assessment").
    - If the author did not respond, it is noted in the study profile (e.g., "No response was received to email request for clarification").
    - If the author responded and provided additional information that informed a rating decision in the risk-of-bias analysis, it is now noted in the study profiles and *Appendix E* which risk-of-bias questions were impacted (e.g., "Additional information provided by the authors informed the rating decision for the following risk-of-bias domains: Detection [outcome assessment]"). Additional details on the information provided by authors can be found in the risk-of-bias explanation rating in HAWC and *Appendix E* (e.g., "*Correspondence with the study authors indicated that the outcome assessors were blind to the children's fluoride status*").

NTP also stated that it "verified that the lower risk-of-bias studies did not provide direct evidence of imprecision or lack of blinding" (NTP 2020c). However, that approach assumes that authors will always reveal in their manuscripts a lack of blinding and other weaknesses in their study design. A more conservative approach would be to assume that there was no blinding of outcome assessors unless it was specified in the manuscript and that a designation of probably high risk of bias for this criterion (at a minimum) would be more appropriate when the blinding status was not explicitly stated. That approach would follow the one described in the protocol in which NTP states that "studies should be considered 'probably high RoB' unless specific direct or indirect evidence of blinding is provided" (NTP 2020b, p. 13).

Response: Disagree (no change)

- We appreciate the Committee's recommendation regarding assessor blinding; however, it fails to account for the standard practice of considering both direct and indirect evidence and

judging the two types of evidence accordingly. We respectfully stand by our decision to consider risk of bias from assessor blinding as "definitely high" if there is direct evidence of lack of assessor blinding and "definitely low" if there is direct evidence of assessor blinding. Direct evidence is the strongest evidence and justifies the "definite" ratings. We also consider indirect evidence of whether assessors were blind to the exposure status of individuals when assessing outcomes. For example, in studies with a cross-sectional design in which exposure and outcome were measured simultaneously, it was considered more likely that the outcome assessor did not know the exposure status of individuals when assessing outcome. Therefore, simultaneous measurement of exposure and outcome was considered indirect evidence of assessor blinding and was rated *probably low risk of bias* if the outcome was otherwise assessed appropriately. Further, we would like to clarify that, if authors do not report information regarding blinding of outcome assessors, a rating of *not reported* is applied, which is equal to a *probably high risk of bias* rating in concern—effectively, in the absence of information, the default rating is *probably high risk of bias*. Study authors are then contacted for missing information and the rating is only changed if authors provide additional details indicating whether assessors were blind to exposure status.

Appendix 4 in the revised monograph outlines details of each lower risk-of-bias study and includes outcome-assessor blinding, if known, and any information gathered from direct contact with manuscript authors. In several cases in which assessor blinding was not known, risk of bias for confidence in the outcome assessment was considered low because of the cross-sectional design in which exposure and outcome were measured simultaneously or when all children resided in the same geographic area. The committee considers that an acceptable approach.
However, in studies in which children were tested in schools or other facilities in areas where low and high fluoride concentrations of different localities were being compared (see, for example, Cui et al. 2018), there is an increased risk of bias because examiners might make assumptions about children in the different areas. A designation of probably high risk of bias (at a minimum) would be more appropriate in those cases given the approach described in the protocol noted above.

Response: Disagree (no change)

- As mentioned in our previous response, simultaneous measurement of exposure and outcome was considered indirect evidence of assessor blinding and was rated *probably low risk of bias* if the outcome was otherwise assessed appropriately.
- To address the Committee's specific concern about Cui et al. (2018), we state in *Appendix E* of the prepublication 2022 NTP Monograph (previously *Appendix 4* of the 2020 draft NTP Monograph) that, "*Blinding or other methods to reduce bias were not reported. Although it was unlikely that the outcome assessor would have knowledge of the child's urine fluoride levels, there was potential that they would know whether the child was from an endemic or non-endemic area if the IQ tests were conducted at the child's school, and there was no information provided on how the IQ tests were administered.*" Also, in response to an author inquiry, the study author noted that the cross-sectional nature of the study with outcome and exposure assessed at the same time made the outcome assessors effectively blind to the exposure. We acknowledge in *Appendix E* of the prepublication 2022 NTP Monograph that there is still potential for knowledge of the area by the outcome assessor, but overall, we determined that there was sufficient indirect evidence of assessor blinding to support a rating of *probably low risk of bias* for blinding.

Flawed Measures of Neurodevelopmental and Cognitive Outcome

The committee raised a concern in its previous review about studies that were classified as having lower risk of bias when measurement of a neurodevelopmental or cognitive outcome was flawed. NTP's response indicated that it did not change the draft monograph but verified that the lower risk-of-bias studies did not provide direct evidence of imprecision in their outcome measurement. However, the committee remains concerned about the application of the protocol definitions to rate studies. For example, Barberio et al. (2017) assessed outcomes that rely on parent or child self-report of diagnosis of learning disability or attention deficit hyperactivity disorder. According to the protocol, that study would be rated as either probably or definitely high risk-of-bias because the method was not listed in Table 6 (NTP 2020b, p. 21), but NTP failed to address whether there is direct evidence that a self-reported diagnosis has been validated as a reliable outcome measure. That evidence would allow one to distinguish which category (probably or definitely high risk of bias) would be most appropriate. ***Because the outcome measure is critical for the interpretation of the findings, the committee recommends that NTP apply its criteria in a more consistent manner and specifically address whether there is direct evidence of the sensitivity and precision of self-reported neurodevelopmental outcomes.***

Response: Disagree (no change)

- We recognize the Committee's continued concern on risk of bias for outcome assessment tools. However, the Committee may be misunderstanding the definition of direct evidence and the different types of evidence needed for each situation. Direct evidence is required for either a *definitely low risk of bias* or *definitely high risk of bias* rating. Direct evidence that the neurodevelopmental or cognitive function outcome was assessed using well-established, validated assessment methods and direct evidence that assessors were blind to exposure status are required for a *definitely low risk of bias* rating on outcome. Similarly, direct evidence that the outcome assessment method was imprecise or insensitive or direct evidence of a lack of assessor blinding is required for a *definitely high risk of bias* rating on outcome. We consider self-reporting of a learning disability to be an insensitive method (as stated in *Appendix E* of the prepublication 2022 NTP Monograph, previously *Appendix 4* of the 2020 draft NTP Monograph), but in the absence of direct evidence that the outcome assessment method is an insensitive or imprecise method (i.e., a known, previous demonstration that the instrument was not reliable in the study subjects or similar population), we consider this concern to result in *probably high risk of bias* for outcome assessment and not *definitely high risk of bias*.

Lack of Rigorous Statistical Review

The committee recognizes that NTP made substantial efforts to improve the statistical reviews of the lower risk-of-bias studies. Each study was reviewed by a senior statistician, and summaries of the analytic methods were added to the study descriptions in Appendix 4 in the section "Other potential threats." However, the summaries provided for a few publications were only a single sentence—"Statistical analyses used were appropriate for the study" (Sudhir et al. 2009; Barberio et al. 2017; Bashash et al. 2017, 2018)—and two other summaries mentioned only log-transformations (Choi et al. 2015) or that tests of normality were performed (Zhang et al. 2015). For those

publications, NTP should have provided more evidence to support its conclusion that the analyses were appropriate. It is also concerning that NTP assumed that the analyses in Soto-Barreras et al. (2019) were appropriate despite few details provided in the manuscript regarding their methods.

Response: Disagree (edited for clarity)

- We appreciate the Committee's continued concerns over the adequacy of the statistical approaches used in some of the publications reviewed in the 2020 draft NTP Monograph. We have expanded our comments concerning the statistical methods used in the low risk-of-bias studies in *Appendix E* of the prepublication 2022 NTP Monograph (previously *Appendix 4* in the 2020 draft NTP Monograph).

The committee also finds that NTP did not adequately address the issue of clustering. Most of the attention to clustering pertained to the examples provided in the committee's previous review. Although it was important for NTP to review those examples, they were meant to highlight the issue and were not meant to serve as a comprehensive list of problematic studies. In fact, when reviewing Appendix 4 in the revised monograph, the committee found several other studies whose analyses failed to account for clustering. Of most concern are the studies that used fluoride concentration measured at the community level as the exposure—see, for example, Seraj et al. (2012), Till et al. (2020), Trivedi et al. (2012), and Wang et al. (2012). When everyone in a community is subject to the same exposure, the standard error of the difference in means between high-exposure and low-exposure groups increases multiplicatively by the square root of a variance inflation factor (VIF) equal to [1 + (n - 1)r], where n is the number of persons in each community and r is the correlation in outcomes (such as IQ score) between members of the same community (Murray 1998; Donner and Klar 2000; Feng et al. 2001). The same phenomenon occurs in randomized control trials that assign treatment to groups of persons. Thus, unless within-community clustering is accounted for in the analysis—for example, through a random- effects model—standard-error estimates will be too small and confidence intervals (CIs) too narrow. Those errors could have a substantial effect on the meta-analysis, which requires valid estimates of within-study variability. The same issue applies to analyses that use community- level exposure to estimate slopes in a regression model. For individual-level exposures, such as urinary fluoride concentration, the VIF is probably smaller than one would see for community- level exposures because some communities might contain people in multiple exposure groups.

Response: Agree (change made)

- **Response:** The potential impact of clustering is addressed in multiple ways in the prepublication 2022 NTP Monograph, that expand previous discussion and analysis. We have revised text in *Appendix E* of the prepublication 2022 NTP Monograph (previously *Appendix 4* in the 2020 draft NTP Monograph) to clearly specify which low risk-of-bias studies addressed clustering when that was a feature of the study design or statistical analysis. We have also reached out to the study authors to request additional information as suggested by the comment and addressed the impact of any information provided. As suggested by the Committee, lack of accounting for clustering has little impact in studies with individual-level exposure levels (e.g., urinary fluoride levels) that also account for many important confounders that often capture the cluster (city) effect.
- The potential impact of clustering is illustrated by Bashash et al. (2017) who accounted for clustering at the cohort level by using cohort as a fixed effect in the models. In addition, the

models accounted for many important confounders, which are also likely to reflect the cohort effects. The similarity between the unadjusted and the adjusted effect estimates β (95% CI)=(−2.37 [(−4.45, −0.29)] and −2.50 ([−4.12, −0.59)], respectively) reflects the minimal impact of accounting for the cohort effect.

- In addition, for the studies referenced in the comment (Seraj et al. 2012; Till et al. 2020; Trivedi et al. 2012; Wang et al. 2012), the number of clusters is relatively small. In such cases, there is "typically not enough information to accurately estimate group-level variation. As a result, multi-level models in this setting typically gain little beyond classical varying-coefficient models" (Gelman and Hill 2006).
- The above response applies to the prepublication 2022 NTP Monograph, additional response specific to the meta-analysis will be released when the manuscript is published.

However, it is still important to account for clustering in the analysis because one would expect most people in a community to be in the same exposure group. ***NTP should note specifically whether each study applied an analytic approach that addressed clustering when that was a feature of the design***.

In the case of Green et al. (2019), NTP learned from the investigators that accounting for city-level clustering via a random-effects model "showed similar results to the main model." More details should be provided regarding the similarity of results because although overall conclusions might not have changed, the results of the meta-analysis could be affected by incorrect exposure-effect or standard-error estimates.

Response: Agree (change made)

- We have revised text in *Appendix E* of the prepublication 2022 NTP Monograph (previously *Appendix 4* in the 2020 draft NTP Monograph) to clearly specify which low risk-of-bias studies addressed clustering when that was a feature of the study design or statistical analysis.
- In the case of Green et al. (2019), we contacted the study authors and received the results from models using city as a random intercept. The overall adjusted effect estimates with city as a fixed effect and with city as a random effect were not significantly different from each other: β (95% CI) =−1.95 (−5.19, 1.28) and −2.20 (−5.39, 0.98), respectively.

The statistical review conducted by NTP also failed to identify a study that did not properly account for the sampling design. Yu et al. (2018) used a hierarchical stratified sampling design but did not indicate that sampling weights were used in the analysis. Thus, both point estimates (means and regression coefficients) and standard errors were likely biased (Lohr, 2019). ***NTP should examine the studies included in the meta-analysis in greater depth to determine whether each study properly accounted for its design because not doing so could invalidate the meta-analysis results***.

Response: Agree (change made)

- To address the part of the comment we agree with "***NTP should examine the studies included in the meta-analysis in greater depth to determine whether each study properly accounted for its design"***, we revised text in *Appendix E* of the prepublication 2022 NTP

Monograph (previously *Appendix 4* in the 2020 draft NTP Monograph) to note specifically whether studies accounted for the sampling strategy. In cases where the publication used stratified or clustered sampling designs but did not mention whether the sampling strategy was accounted for in the analysis (e.g., Cui et al. 2018, 2020; Yu et al. 2018), we contacted study authors to specifically ask for this information. If they responded, we updated the information in *Appendix E* (noting that the information came from correspondence with the authors). If authors did not respond, we noted that we contacted them but did not receive a response.

- A detailed response to the Committee's critique of the meta-analysis is provided in "Sup01_Meta-analysis".

Need to Juxtapose Results of Broadly Comparable Studies

In its previous review, the committee expressed concern about selective consideration and presentation of results from the various studies. That approach can convey inaccurate impressions regarding consistency unless the findings are derived from studies that are comparable or aligned with respect to study population, exposure measurement, and outcome ascertainment. Some text in the revised monograph continues to be impressionistic and haphazard in citing various findings from studies and does not provide a clear rationale for why some findings are reported and others are not. The committee notes that reporting findings that are most or least supportive of a finding does not necessarily indicate bias and that this issue might be more editorial than substantive in that the text is not the basis for drawing conclusions. However, it does constitute a concern with transparent communication.

Response: Agree (change made)

- **Response:** We appreciate the Committee's comments on this point and have carefully re-evaluated the information presented in the monograph. We have detected an imbalance in the presentation toward highlighting flaws and limitations in the studies and have attempted to address this in the prepublication 2022 NTP Monograph. In a few instances, we have added details to the main data table summarizing the results of the IQ studies in children (Table 6) to account for all outcomes reported.

The critical information regarding comparison of study results comes from the new meta-analysis, which seeks to extract and integrate comparable findings from selected studies as discussed further below. The overall approach appears to be sound in comparing mean IQ scores for the most and least highly exposed to fluoride even though the absolute fluoride concentrations are not comparable among studies. ***Because the meta- analysis is so critical to the conclusions that are drawn, NTP should provide the data that were used from each study to enable the reader to understand and evaluate what was done***.

The values that were used to determine the standardized mean differences (SMDs) could not be found in the revised monograph, nor was there a figure that showed the pattern of results from studies restricted to the lower exposure ranges. . A more detailed assessment of the meta-analysis is provided in the next section.

Response: Disagree (no change)

- A detailed response to the Committee's critique of the meta-analysis is provided in "Sup01_Meta-analysis". However, we take issue with the Committee's assertion that the meta-analysis is critical to the conclusions drawn. Indeed, we reached the same hazard conclusions in the 2019 draft NTP Monograph, which lacked a meta-analysis, as we did in the 2020 draft NTP Monograph, in which we included a meta-analysis at the Committee's recommendation. Because of the extensive comments on the meta-analysis, and consistent with the original decision to not perform one because of the uncertainty over the precision of the findings of many of the high risk-of-bias studies, we have removed the meta-analysis from the prepublication 2022 NTP Monograph and will pursue publishing it separately.

Evaluation of the Meta-Analysis

Note: The NASEM Committee's comments in the "Evaluation of the Meta-Analysis" section are not reproduced here as they are not directly relevant to the prepublication 2022 NTP Monograph. See "Sup01_Meta-analysis" for the meta-analysis-relevant comments and responses.

COMMUNICATION

Overall, NTP has done a good job of identifying and extracting the underlying epidemiologic information that it needs to evaluate the possible neurodevelopmental effects of fluoride. With a few exceptions, the major problem with the report is not related to missing or misinterpreted information, but rather with how the underlying research and its evaluations are presented by NTP. As detailed in many of the preceding comments, NTP's protocols and its evaluations of the research are sometimes difficult to follow. As NTP is aware, the issue of fluoride toxicity and safety is highly contentious. To be widely accepted, any analysis concerning the issue needs to be performed and presented with exceptional care and with exceptional clarity. Overall, the revised monograph seems to include a wealth of evidence and a number of evaluations that support its main conclusion, but the monograph falls short of providing a clear and convincing argument that supports its assessment, given the lack of details in several places and the lack of clarity on several substantive issues.

Much of the evidence presented in the report comes from studies that involve relatively high fluoride concentrations. Little or no conclusive information can be garnered from the revised monograph about the effects of fluoride at low exposure concentrations (less than 1.5 mg/L). ***NTP therefore should make it clear that the monograph cannot be used to draw any conclusions regarding low fluoride exposure concentrations, including those typically associated with drinking-water fluoridation***. Drawing conclusions about the effects of low fluoride exposures (less than 1.5 mg/L) would require a full dose–response assessment, which would include at a minimum more detailed analyses of dose–response patterns, models, and model fit; full evaluations of the evidence for supporting or refuting threshold effects; assessment of the differences in exposure metrics and intake rates; more detailed analyses of statistical power and uncertainty; evaluation of differences in susceptibility; and detailed quantitative analyses of effects of bias and confounding of small effect sizes. Those analyses fall outside the scope of the NTP monograph, which focuses on hazard identification and not dose– response assessment. ***Given the substantial concern regarding health implications of various fluoride exposures, comments or inferences that are not based on rigorous analyses should be avoided.***

Response: Disagree (no change)

- The Committee correctly states that the data driving the hazard conclusions in the 2019 and 2020 draft NTP Monographs primarily reflect high exposures (i.e., >1.5 mg/L in drinking water, along with other fluoride sources including food, beverages, and oral hygiene products). The extent to which community artificial water fluoridation contributes to high fluoride exposures is not addressed in the prepublication 2022 NTP Monograph, although some studies evaluated individuals with high fluoride exposures that were associated at least in part with community water fluoridation (e.g., Green et al. 2019). Both the 2019 and 2020 versions of the draft NTP Monograph concluded that the findings concerning children's IQ, where exposures were equivalent to or below 1.5 mg/L, were inconsistent and therefore unclear.

NTP CONCLUSION

As noted above, the committee focused on determining whether the evidence as presented in the revised monograph supported NTP's main conclusion that "fluoride is *presumed to be a cognitive neurodevelopmental hazard to humans*" (NTP 2020a, p. 80). The revised monograph is

much improved from the initial draft that the committee reviewed. The addition of the meta-analysis substantially increases the support for NTP's main conclusion. However, the committee is still concerned about the presentation of the data, the methods, and the analyses in the revised monograph and finds that the monograph falls short of providing a clear and convincing argument that supports its assessment. The committee urges NTP to improve the clarity of the document. The monograph has great importance in the discussion about effects of fluoride on neurodevelopmental and cognitive health effects and will likely influence exposure guidelines or regulations. Thus, it is extremely important for it to be able to withstand scientific scrutiny by those who have vastly different opinions on the risks and benefits associated with fluoride exposure. ***The committee strongly recommends that NTP improve the revised monograph by seriously considering the suggestions that are provided in this letter report to improve its clarity and transparency.***

Response: Agree (change made)

- We agree with and appreciate the Committee's statements concerning the importance of this assessment and consider that the prepublication 2022 NTP Monograph has been improved in clarity and transparency through responses to the Committee's criticisms of earlier drafts.

Attachments
A Statement of Task
B Committee Membership
C Key Issues and NTP Response
D Bibliography
E Acknowledgment of Reviewers

ATTACHMENT A STATEMENT OF TASK

An ad hoc committee of the National Academies of Sciences, Engineering, and Medicine will review the revised National Toxicology Program (NTP) *Monograph on Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects*. The committee will consider whether NTP's revisions have addressed the substantive concerns raised in the National Academies 2020 report *Review of the Draft NTP Monograph: Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects*. The primary focus of the committee will be to determine whether the evidence as presented by NTP in its revised monograph supports its conclusions.

ATTACHMENT B COMMITTEE MEMBERSHIP

DAVID A. SAVITZ (NAM) (*Chair*), Brown University School of Public Heath, Providence, RI
KEVIN M. CROFTON, R3Fellows, Durham, NC
AKHGAR GHASSABIAN, New York University School of Medicine, New York, NY **JUDITH B. KLOTZ**, Drexel University Dornsife School of Public Health, Philadelphia, PA **JULEEN LAM**, California State University, East Bay, Hayward, CA
PAMELA J. LEIN, University of California, Davis, CA
GERMAINE M. BUCK LOUIS, George Mason University, Fairfax, VA **MICHAEL L. PENNELL**, The Ohio State University, Columbus, OH **CRAIG STEINMAUS**, University of California, Berkeley, Oakland, CA
CHARLES V. VORHEES, University of Cincinnati College of Medicine, Cincinnati, OH
KIMBERLY YOLTON, Cincinnati Children's Hospital Medical Center and University of Cincinnati College of Medicine, Cincinnati, OH

STAFF

ELLEN K. MANTUS, Project Director
RADIAH ROSE-CRAWFORD, Manager, Editorial Projects
BENJAMIN ULRICH, Program Assistant

BIOGRAPHIES

David A. Savitz (NAM) (*Chair*) is professor of epidemiology and associate dean for research of the Brown University School of Public Health, with joint appointments in obstetrics and gynecology and pediatrics at the Alpert Medical School. He was vice president of research at the university from 2013 to 2017. His epidemiologic research has addressed a wide array of important public-health issues, including environmental hazards in the workplace and community, reproductive-health outcomes, and environmental influences on cancer. He has worked extensively on health effects of nonionizing radiation, pesticides, drinking-water treatment byproducts, and perfluorinated compounds. Before joining Brown University, Dr.
Savitz held appointments as the Charles W. Bluhdorn Professor of Community and Preventive Medicine at Mount Sinai School of Medicine and professor at the University of North Carolina School of Public Health. He was president of the Society for Epidemiologic Research and the Society for Pediatric and Perinatal Epidemiologic Research and was a North American regional councilor for the International Epidemiological Association. Dr. Savitz was elected to the National Academy of Medicine in 2007. He received an MS in preventive medicine from Ohio State University and a PhD in epidemiology from the University of Pittsburgh Graduate School of Public Health.

Germaine M. Buck Louis is dean of the College of Health and Human Services of George Mason University. Her research has addressed a mixture of environmental exposures, including endocrine disruptors, stress, diet, and physical activity in relation to a spectrum of reproductive outcomes in men and women. She was an early pioneer in the application of the exposome research paradigm for understanding environmental influences on human fecundity and fertility impairments. Before joining the university, Dr. Louis was the director of the Division of Intramural Population Health Research in the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health, where she led population- health scientists in

designing research aimed at enhancing the health and well-being of fetuses, pregnant women, children, and young adults. She has served the National Academies, Pan American Health Organization, US Environmental Protection Agency, and World Health Organization in various roles. She is a former president of the Society of Pediatric and Perinatal Epidemiologic Research and of the Society for Epidemiologic Research and has served on the boards of the American College of Epidemiology and the International Society for Environmental Epidemiology. Dr. Louis received a PhD in epidemiology from the State University of New York at Buffalo.

Kevin M. Crofton is principal and consultant at R3Fellows, LLC. Previously, he worked for more than 35 years as a developmental neurotoxicologist in the US Environmental Protection Agency (EPA) Office of Research and Development. Dr. Crofton has also served as an adjunct associate professor at Duke University, the University of North Carolina, and North Carolina State University. His research interests include developmental neurotoxicity with an emphasis on the consequences of endocrine disruption for neurodevelopment. He recently received the EPA Distinguished Career Service Award. Dr. Crofton received an MS in toxicology from Miami University and a PhD in toxicology from the University of North Carolina at Chapel Hill.

Akhgar Ghassabian is an investigator and assistant professor in the Departments of Pediatrics, Population Health, and Environmental Medicine of the New York University (NYU) School of Medicine. Her research focuses on identifying environmental exposures that contribute to the etiology of developmental disabilities in childhood. Before joining NYU, Dr. Ghassabian was the intramural research training award fellow at the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health. During her doctoral and postdoctoral training, Dr. Ghassabian was involved in birth-cohort studies in Europe and in the United States. She was a collaborator on European epidemiologic consortia examining the effect of nutrition and air pollution on children's neurodevelopment. Dr. Ghassabian was the recipient of the Rubicon Award from the Netherlands Organization for Scientific Research in 2014 and the Robin/Guze Young Investigator Award from the American Psychopathological Association in 2019. She obtained an MD from Tehran University of Medical Sciences and a PhD in epidemiology from Erasmus University Rotterdam, the Netherlands.

Judith B. Klotz is an affiliate faculty member in the Department of Environmental and Occupational Health of the Drexel University Dornsife School of Public Health and an adjunct associate professor in the Department of Epidemiology of the Rutgers School of Public Health. She is a member of the Health Effects Committee of the New Jersey Drinking Water Quality Institute and of the Public Health Standing Committee of the Science Advisory Board, both advisory groups of the New Jersey Department of Environmental Protection. She served as environmental scientist and program manager in environmental health and in cancer surveillance in the New Jersey Department of Health from 1984 to 2003 and focused especially on toxic substances in drinking water and the environmental epidemiology of cancer and reproductive outcomes. Dr. Klotz has served on several National Academies committees, including the Committee on Fluoride in Drinking Water and the Committee on the Review of the Styrene Assessment in the National Toxicology Program 12th Report on Carcinogens. She received an MS in genetics from the University of Michigan and a DrPH in environmental health sciences from Columbia University.

Juleen Lam is an assistant professor in the Department of Health Sciences of the California State University, East Bay. She is also an affiliate researcher in the Department of Obstetrics, Gynecology and Reproductive Sciences of the University of California, San Francisco, School of

Medicine. Her research interests are in environmental epidemiology, evaluation of population exposures to environmental contaminants, assessment and communication of environmental risks, and reproductive and developmental health. She specializes in analysis of environmental- health data and development and application of risk-assessment methods. Dr. Lam has been involved in the development of systematic review methods for environmental-health data and has had a pivotal role in implementing, publishing, and disseminating these approaches in academic and government settings. She is a member of the US Environmental Protection Agency Board of Scientific Counselors Chemical Safety for Sustainability Subcommittee. She served on the National Academies Committee to Review DOD's Approach to Deriving an Occupational Exposure Limit for TCE. She received an MS in environmental engineering management from George Washington University and an MHS in biostatistics and PhD in environmental-health policy from the Johns Hopkins University Bloomberg School of Public Health.

Pamela J. Lein is a professor of neurotoxicology in the Department of Molecular Biosciences of the University of California, Davis, School of Veterinary Medicine. Her research interests are in how environmental stressors interact with genetic susceptibilities to influence the risk and severity of neurodevelopmental disorders and neurodegeneration. Because altered patterns of connectivity are associated with neurologic deficits, her research focuses on investigating how environmental contaminants, chemical convulsants, and inflammation perturb neuronal connectivity as determined by using biochemical, morphogenic, and electrophysiologic end points. Her group is also developing biomarkers of organophosphate neurotoxicity and testing novel therapeutic approaches for protecting against the neurodegenerative effects associated with neurotoxic proconvulsants. Dr. Lein was a member of the National Academies Committee to Review Report on Long-Term Health Effects on Army Test Subjects. She received an MS in environmental health from East Tennessee State University and a PhD in pharmacology and toxicology from the State University of New York at Buffalo.

Michael L. Pennell is associate professor in the Division of Biostatistics in the College of Public Health of Ohio State University. His research interests are in nonparametric Bayes, first hitting time models for survival analysis; design and analysis of group randomized trials; joint modeling outcomes of different scales; statistical methods in toxicologic risk assessment; and statistical applications in biomedical research, including cancer control, pathology, and veterinary medicine. Dr. Pennell has served as an ad hoc member of the US Environmental Protection Agency (EPA) Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel, the EPA Science Advisory Board on trichloroethylene and Libby amphibole asbestos, and the Chemical Safety Advisory Subcommittee for 1-bromopropane. He served on the National Academies Committee to Evaluate the IRIS Protocol for Inorganic Arsenic. He received an MS and a PhD in biostatistics from the University of North Carolina at Chapel Hill.

Craig Steinmaus is an associate adjunct professor of epidemiology at the University of California, Berkeley (UCB). He is also a public-health medical officer III in the California Environmental Protection Agency (CalEPA) and is the UCB director of the Arsenic Health Effects Research Group. He is a board-certified physician with over 12 years of patient-care experience. His epidemiologic research has involved studies of drinking-water contaminants with a focus on early-life exposure and other factors conferring susceptibility. He also teaches graduate courses on epidemiology, causal inference, and systematic review at UCB and at the University of California, San Francisco. Dr. Steinmaus has served on several study sections of the National Institutes of Health and Centers for Disease Control and Prevention and is a full member of the Cancer, Heart, and Sleep Epidemiology, A study section. His work in the CalEPA water toxicology section has involved systematic reviews and risk assessments of drinking-water agents, including nitrate, arsenic, copper, perchlorate, fluoride, chromium, and trihalomethanes. He received an MD from the University of California, Davis, School of Medicine and an MPH in environmental-health sciences from UCB.

Charles V. Vorhees is a professor in the University of Cincinnati College of Medicine. He is co-director of the Animal Behavior Core and program director of the Teratology Training Program. He is on the graduate faculty of the graduate programs in neuroscience and molecular and developmental biology. His research focuses on brain development and behavior. He was a founding member of the Neurobehavioral Teratology Society in 1977 and was elected president in 1984–1985 and 2012–2013. Dr. Vorhees has served on multiple scientific advisory committees for the US Food and Drug Administration, US Environmental Protection Agency, and National Institutes of Health. He was on the National Academies Subcommittee on Reproductive and Developmental Toxicants. Dr. Vorhees obtained an MA and a PhD in neurobiology from Vanderbilt University.

Kimberly Yolton is a professor in Cincinnati Children's Hospital Medical Center (CCHMC) and the University of Cincinnati College of Medicine and director of research in the Department of General and Community Pediatrics. She is a developmental psychologist and epidemiologist with over 25 years of experience in studying the effects of prenatal and early-life exposures on neurobehavior from infancy through childhood and directs the longitudinal Health Outcomes and Measures of the Environment (HOME) Study. She was formerly the director of a follow-up clinic serving high-risk infants and young children and has extensive experience with infants and children who were prenatally exposed to substances of abuse, who were born prematurely or at low birth weight, or who come from disadvantaged home environments. She was involved in the initial development of the NICU Network Neurobehavioral Scale (NNNS), a specialized neurobehavioral assessment tool used with healthy and high-risk newborns, and conducts frequent training on the proper administration, scoring, and interpretation of the instrument for research and clinical purposes. She has been affiliated with the National Institutes of Health– funded Neonatal Research Network for over 25 years at two sites as an examiner, Gold Standard reviewer for intelligence testing, follow-up principal investigator, and steering-committee member. She often collaborates with investigators regarding neurobehavioral assessment and staff training strategies to acquire the most appropriate outcome measures with the highest standards of reliability and validity. She earned a PhD in child development and developmental psychology from Ohio State University and completed a 3-year National Research Service Award in Pediatric Environmental Health at CCHMC.

ATTACHMENT C

This attachment summarizes the substantive issues raised in the committee's previous report (NASEM 2020) concerning the general systematic review methods and the evaluation of the human evidence. Because NTP decided to base its conclusions on the human evidence, it did not re-evaluate the animal evidence to address the committee's previous concerns. Instead, it added a disclaimer to the revised monograph and left the original text unchanged. For that reason, the committee's concerns regarding the animal evidence are not listed here.

Committee Issue on Methods and Communication	NTP Response
NTP added foreword to monograph and text to protocol to clarify relationship.	
NTP added text to protocol and monograph to clarify literature search strategy and to clarify assessment of animal data.	
Absence of exclusion–inclusion criteria from protocol	No information provided.
Lack of justification for some decisions SWIFT-Active screener to justify approach.	NTP added information to the monograph on
Inconsistencies between protocol and monograph concerning critical confounders to evaluate.	NTP clarified text in protocol and monograph
Communication concerning how monograph can be used (or not) to inform water fluoridation concentrations	No information provided.

Committee Issue on Evaluation of Human Evidence	NTP Response
Potential for Biased Selection of Studies	NTP conducted supplemental searches of Chinese databases and identified additional studies.
NTP revised the monograph to indicate the multiple publications on the same population. However, when conducting the meta-analysis, NTP included more than one publication for a single study population in at least one case.	
Inconsistent Application of Risk-of Bias Criteria	NTP added Appendix 4.
Evaluation of Confounding Insufficient, Difficult to Understand, or Applied Inconsistently	NTP revised text to identify clearly key confounders that applied to all study populations. NTP added Appendix 4.
Possibility of Exposure Misclassification	NTP added Appendix 4.
Need for Further Consideration of Failure to Blind Examiners	NTP added Appendix 4.
Flawed Measures of Neurodevelopmental and Cognitive Outcomes	NTP verified lower risk-of-bias studies that did not provide direct evidence of imprecision or lack of blinding.
NTP examined studies identified by committee and included discussion in Appendix 4.	

Need to Juxtapose Results across Broadly Comparable Studies	NTP conducted subgroup analyses as part of meta-analysis to address heterogeneity in the data and further analyze consistency of data.
Need to Consider Conducting Meta-Analysis meta-analysis using individual-level exposure data.	NTP updated meta-analyses and conducted new
Lack of Support for Conclusion that Effects Occur at Higher Fluoride Doses	NTP conducted dose–response analysis as part of meta-analysis.

ATTACHMENT D

Altman, D.G., and J.M. Bland. 2003. Interaction revised: The difference between two estimates. BMJ 326(7382):219.

Barberio, A.M., C. Quinonez, F.S. Hosein, and L. McLaren. 2017. Fluoride exposure and reported learning disability among Canadian children: Implications for community water fluoridation. Can. J. Public Health 108(December):229-239.

Bashash, M., D. Thomas, H. Hu, E.A. Martinez-Mier, B.N. Sanchez, N. Basu, K.E. Peterson, A.S. Ettinger, R. Wright, Z. Zhang, Y. Liu, L. Schnaas, A. Mercado-Garcia, M.M. Tellez-Rojo, and M. Hernandez-Avila. 2017. Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6-12 years of age in Mexico. Environ. Health Perspect. 125(9):1-12.

Bashash, M., M. Marchand, H. Hu, C. Till, E.A. Martinez-Mier, B.N. Sanchez, N. Basu, K.E. Peterson, R. Green, L. Schnaas, A. Mercado-Garcia, M. Hernandez-Avila, and M.M. Tellez-Rojo. 2018. Prenatal fluoride exposure and attention deficit hyperactivity disorder (ADHD) symptoms in children 6-12 years of age in Mexico City. Environ. Int. 121(Pt1):658-666.

Broadbent, JM, W.M. Thomson, T.E. Moffitt, and R. Poulton. 2015. Community water fluoridation and intelligence response. Am. J. Public Health 105(1):3-4.

Choi, A.L., Sun G., Zhang Y., Grandjean P. 2012. Developmental fluoride neurotoxicity: A systematic review and meta-analysis. Environ. Health. Perspect. 120(10):1362-1368.

Choi, A.L., Y. Zhang, G. Sun, D.C. Bellinger, K. Wang, X.J. Yang, J.S. Li, Q. Zheng, Y. Fu, and P. Grandjean. 2015. Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study. Neurotoxicol. Teratol. 47(Jan-Feb):96-101.

Cui, Y., B. Zhang, J. Ma, Y. Wang, L. Zhao, C. Hou, J. Yu, Y. Zhao, Z. Zhang, J. Nie, T. Gao, G. Zhou, H. Liu. 2018. Dopamine receptor D2 gene polymorphism, urine fluoride, and intelligence impairment of children in China: A school-based cross-sectional study. Ecotoxicol. Environ. Saf. 165(December):270-277.

Cui, Y. J. Yu, B. Zhang, B. Guo, T. Gao, and H. Liu. 2020. The relationships between thyroid-stimulating hormone and/or dopamine levels in peripheral blood and IQ in children with different urinary iodine concentrations. Neurosci. Lett. 729(June):134981.

Das, K., and N.K. Mondal. 2016. Dental fluorosis and urinary fluoride concentration as a reflection of fluoride exposure and its impact on IQ level and BMI of children of Laxmisagar, Simlapal Block of Bankura District, W.B., India. Environ. Monit. Assess. 188:218.

Ding, Y., H. Sun, H. Han, W. Wang, X. Ji, X. Liu, and D. Sun. 2011. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. J. Hazard. Mater. 186:1942-1946.

Donner, A., and N. Klar. 2000. Design and Analysis of Cluster Randomization Trials in Health Research. London: Hodder Arnold Publication.

Feng, Z., P. Diehr, A. Peterson, and D. McLerran. 2001. Selected statistical issues is group randomized trials. Annu. Rev. Public Health. 22(May):167-181.

Green R., B. Lanphear, R. Hornung, D. Flora, E.A. Martinez-Mier, R. Neufeld, P. Ayotte, G. Muckle, and C. Till. 2019. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. JAMA Pediatr. 173(10):940-948.

Lohr, S. 2019. Sampling Design and Analysis, 2nd Edition. Boca Raton, FL: CRC Press.McPherson, C.A., G. Zhang, R. Gilliam, S.S. Brar, R. Wilson, A. Brix, C. Picut, and G.J. Harry. 2018. An evaluation of neurotoxicity following fluoride exposure from gestational through adult ages in Long-Evans hooded rats. Neurotoxicol. Res. 34(4):781-798.

McPherson, C.A., G. Zhang, R. Gilliam, S.S. Brar, R. Wilson, A. Brix, C. Picut, and G.J. Harry. 2018. An evaluation of neurotoxicity following fluoride exposure from gestational through adult ages in Long-Evans hooded rats. Neurotoxicol. Res. 1-18.

Murray, D. 1998. Design and Analysis of Group Randomized Trials. New York: Oxford University Press.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2020. Review of the Draft NTP Monograph: Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects. Washington, DC: The National Academies Press.

NTP (National Toxicology Program). 2015. New OHAT Handbook for Conducting Systematic Reviews. Office of Health Assessment and Translation, Division of the NTP, National Institute of Environmental Health Sciences, National Institutes of Health, US Department of Health and Human Services. Available: https://ntp.niehs.nih.gov/update/2015/1/ohat-handbook/index.html.

NTP. 2019a. Draft NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects. Office of Health Assessment and Translation, Division of the NTP, National Institute of Environmental Health Sciences, National Institutes of Health, US Department of Health and Human Services.

NTP. 2019b. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Research Triangle Park, NC: National Institute of Environmental Health Sciences.

NTP. 2019c. Updates and Clarification to the OHAT Approach for Systematic Review and Evidence Integration Research Triangle Park, NC: National Institutes of Health. Available: https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookclarificationmarch2019_508.pdf.

NTP. 2020a. Draft NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects. Office of Health Assessment and Translation, Division of the NTP, National Institute of Environmental Health Sciences, National Institutes of Health, US Department of Health and Human Services.

NTP. 2020b. Protocol for Systematic Review of Effects of Fluoride Exposure on Neurodevelopment. Date of Second Revised Protocol Published: September 16, 2020.

NTP. 2020c. Review of the Draft NTP Monograph: Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Consensus Study Report of the National Academies of Science, Engineering, and Medicine (NASEM 2020)—Response to Comments.

Seraj, B., M. Shahrabi, M. Shadfar, R. Ahmadi, M. Fallahzadeh, H.F. Eslamlu, and M.J. Kharazifard. 2012. Effect of high water fluoride concentration on the intellectual development of children in Makoo, Iran. J. Dent. 9(3):221-229.

Soto-Barreras U., K.Y. Escalante-Villalobos, B. Holguin-Loya, B. Perez-Aguirre, A. Nevarez-Rascon, R.E. Martinez-Martinez, and J. P. Loyola-Rodriguez. 2019. Effect of fluoride in drinking water on dental caries and IQ in children. Fluoride 52(3 Pt 3):474-482.

Sudhir, K. M., G.N. Chandu, G.M. Prashant, and V.V. Reddy. 2009. Effect of fluoride exposure on intelligence quotient (IQ) among 3-15 year old school children of known endemic area of fluorosis, Nalgonda District, Andhra Pradesh. J. Indian Assoc. Public Health Dent. 2009(13):88-94.

Till, C., R. Green, D. Flora, R. Hornung, E.A. Martinez-Mier, M. Blazer, L. Farmus, P. Ayotte, G. Muckle, and B. Lanphear. 2020. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. Environ. Int. 134(January):105315.

Trivedi, M., N. Sangai, R. Patel, M. Payak, and S. Vyas. 2012. Assessment of groundwater quality with special reference to fluoride and its impact on IQ of schoolchildren in six sans of the Mundra Region, Kachchh, Gujurat, India. Fluoride 45(4):377-383.

Valdez Jimenez, L., O.D. Lopez Guzman, M. Cervantes Flores, R. Costilla-Salazar, J. Calderon Hernandez, Y. Alcaraz Contreras, and D.O. Rocha-Amador. 2017. In utero exposure to fluoride and cognitive development delay in infants. Neurotox. 59(March):65-70.

Wang G., M. Gao, M. Ahang, M. Yang, and Q. Xiang. 2012. Correlation between total fluoride intake and children's IQ. J. Southeast Univ. Med. Ed. 31(6):743-746.

Wang, M., L. Liu, H. Li, Y. Li, H. Liu, C. Hou, Q. Zeng, P. Li, Q. Zhao, L. Dong, G. Zhou, X. Yu, L. Liu, Q. Guan, S. Zhang, and A. Wang. 2020. Thyroid function, intelligence, and low-moderate fluoride exposure among Chinese school-age children. Environ. Int. 134(January):105229.

Xiang, Q. Y. Liang, B. Chen, and L. Chen. 2011. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. Fluoride 44:191-194.

Xiang, Q., Y. Liang, L. Chen, C. Wang, B. Chen. X. Chen, and M. Zhou. 2003. Effect of fluoride in drinking water on children's intelligence. Fluoride 36:84-94.

Yang, L., P. Jin, X. Wang, Q. Zhou, X. Lin, and S. Xi. 2018. Fluoride activates microglia, secretes inflammatory factors and influences synaptic neuron plasticity in the hippocampus of rats. Neurotoxicology 69(December):108-120.

Yu, X., J. Chen, Y. Li, H. Liu, C. Hou, Q. Zeng, Y. Cui, L. Zhao, P. Li, Z. Zhou, S. Pang, S. Tang, K. Tian, Q. Zhao, L. Dong, C. Xu, X. Zhang, S. Zhang, L. Liu, and A. Wang. 2018. Threshold effects of moderately excessive fluoride exposure on children's health: A potential association between dental fluorosis and loss of excellent intelligence. Environ. Int. 118(September):116-124.

Zhang, S., X. Zhang, H. Liu, W. Au, Z. Guan, Q. Zeng, C. Jiang, H. Gao, C. Zhang, R. Lei, T. Zia, Z. Wang, L. Yang,Y. Chen, X. Wu, Y. Cui, L. Yu, and A. Wang. 2015. Modifying effect of COMT gene polymorphism and a predictive role for proteomics analysis in children's intelligence in endemic fluorosis area in Tianjin, China. Toxicol. Sci. 144(2):238-245.

ATTACHMENT E ACKNOWLEDGMENT OF REVIEWERS

This consensus letter report was reviewed in draft form by persons chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the National Academies of Sciences, Engineering, and Medicine in making each published report as sound as possible and to ensure that it meets institutional standards of quality, objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.
We thank the following for their review of this report:

Ana Navas-Acien, Columbia University **David Bellinger**, Harvard Medical School **Weihsueh Chiu**, Texas A&M University
David Dorman, North Carolina State University **Jayanth Kumar**, California Department of Public Health **Karen Robinson**, Johns Hopkins University

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations of this report, nor did they see the final draft before its release. The review of the report was overseen by **Jonathan Samet** (NAM), Colorado School of Public Health, who was responsible for making certain that an independent examination of the report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content rests entirely with the authoring committee and the National Academies.

NIEHS/DNTP Response to the

REVIEW OF THE REVISED NTP MONOGRAPH ON THE SYSTEMATIC REVIEW OF FLUORIDE EXPOSURE AND NEURODEVELOPMENTAL AND COGNITIVE HEALTH EFFECTS: A LETTER REPORT

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

The National Toxicology Program (NTP) and the Division of the National Toxicology Program (DNTP) at the National Institute of Environmental Health Sciences (NIEHS) appreciates the comments provided by the National Academies of Sciences, Engineering, and Medicine (NASEM) Committee in its review of the September 2020 revised draft of the NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects.

The NASEM Committee reviews of the draft NTP monographs on fluoride (September 2019 and September 2020) determined that, "Overall the revised monograph seems to include a wealth of evidence and a number of evaluations that support its main conclusion, but the monograph falls short of providing a clear and convincing argument that supports its assessments…" Thus, the NTP removed the hazard assessment step and added "State of the Science" to the title to indicate the change. The monograph was retitled as "NTP Monograph on the State of the Science Concerning Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects: A Systematic Review" and underwent additional peer review by five external experts. The prepublication 2022 NTP Monograph includes consideration of comments from that external peer review and from key stakeholders across HHS in addition to the NASEM Committee's comments.

In addition, the prepublication 2022 NTP Monograph removed the meta-analysis that was added at the NASEM Committee's request following its review of the September 2019 draft NTP Monograph. The meta-analysis is being prepared as a separate journal publication, taking into consideration the NASEM Committee's comments on the September 2020 draft NTP Monograph.

The NIEHS/DNTP separated the NASEM Committee Letter Review comments that were applicable to the prepublication 2022 NTP Monograph from the comments that are focused on the meta-analysis because these are now two separate, distinct evaluations. For each set of comments, NIEHS/DNTP prepared responses and described changes made in response to the comments. The NASEM Committee's comments and responses that were directly relevant to the prepublication 2022 NTP Monograph are in the document titled "Sup01_Monograph" (see Word file Sup01_Monograph_NASEM_comments_on_monograph_only_Feb_2021_and_NIEHS_ DNTP_ response.docx) and are not included here.

This document contains a subset of the overall NASEM Committee's comments and NIEHS/DNTP responses that are related to the meta-analysis. Therefore, all of the comments from the "Evaluation of the Meta-Analysis" section are included, along with the portion of meta-analysis-relevant comments from the "Lack of Rigorous Statistical Review" and "Need to Juxtapose Results of Broadly Comparable Studies" sections. For clarity, the complete text from comments relevant to the meta-analysis from the NASEM Letter Review have been included in the pages that follow and is formatted in black text. Formatting has been applied to aid in reading and page numbers have been added starting with "Sup01_Meta-analysis." The responses begin with the word "Response," are formatted in blue font, and are interspersed within the original NASEM Committee text.

Lack of Rigorous Statistical Review

Note: The NASEM Committee's comment in the first paragraph of this section is on the prepublication 2022 NTP Monograph and is not reproduced here as it is not directly relevant to the meta-analysis. See "Sup01_Monograph" for the monograph-relevant comments and responses.

The committee also finds that NTP did not adequately address the issue of clustering. Most of the attention to clustering pertained to the examples provided in the committee's previous review. Although it was important for NTP to review those examples, they were meant to highlight the issue and were not meant to serve as a comprehensive list of problematic studies. In fact, when reviewing Appendix 4 in the revised monograph, the committee found several other studies whose analyses failed to account for clustering. Of most concern are the studies that used fluoride concentration measured at the community level as the exposure—see, for example, Seraj et al. (2012), Till et al. (2020), Trivedi et al. (2012), and Wang et al. (2012). When everyone in a community is subject to the same exposure, the standard error of the difference in means between high-exposure and low-exposure groups increases multiplicatively by the square root of a variance inflation factor (VIF) equal to [1 + (n - 1)r], where n is the number of persons in each community and r is the correlation in outcomes (such as IQ score) between members of the same community (Murray 1998; Donner and Klar 2000; Feng et al. 2001). The same phenomenon occurs in randomized control trials that assign treatment to groups of persons. Thus, unless within-community clustering is accounted for in the analysis—for example, through a random- effects model—standard-error estimates will be too small and confidence intervals (CIs) too narrow. Those errors could have a substantial effect on the meta-analysis, which requires valid estimates of within-study variability. The same issue applies to analyses that use community-level exposure to estimate slopes in a regression model. For individual-level exposures, such as urinary fluoride concentration, the VIF is probably smaller than one would see for community-level exposures because some communities might contain people in multiple exposure groups.

However, it is still important to account for clustering in the analysis because one would expect most people in a community to be in the same exposure group. ***NTP should note specifically whether each study applied an analytic approach that addressed clustering when that was a feature of the design***.

In the case of Green et al. (2019), NTP learned from the investigators that accounting for city-level clustering via a random-effects model "showed similar results to the main model." More details should be provided regarding the similarity of results because although overall conclusions might not have changed, the results of the meta-analysis could be affected by incorrect exposure-effect or standard-error estimates.

Note: The NASEM Committee's comments in the three paragraphs above are relevant to both the prepublication 2022 NTP Monograph and meta-analysis and are therefore included here. See "Sup01_Monograph" for the monograph-relevant responses.

Response: Agree (change made)

- We have revised text in Appendix E of the prepublication 2022 NTP Monograph (previously Appendix 4 in the September 2020 draft NTP Monograph) to note specifically whether each low risk-of-bias study applied an analytic approach that addressed clustering when that was a feature of the study design. We have also reached out to the study authors to request additional information and addressed the impact of any information provided.
- For the studies referenced in the comment (Seraj et al. 2012; Till et al. 2020; Trivedi et al. 2012; Wang et al. 2012), the number of clusters in each is relatively small. In such cases, there is "typically not enough information to accurately estimate group-level variation. As a result, multi-level models in this setting typically gain little beyond classical varying-coefficient models" (Gelman and Hill, 2006). For these studies and all low risk-of-bias studies, in the *Other potential threats* domain in Appendix E of the prepublication 2022 NTP Monograph, we now discuss whether clustering was specifically addressed and/or whether there were other factors that may have increased or reduced our concern for clustering-related bias.
- As suggested by the Committee, lack of accounting for clustering has little impact in studies with individual-level exposure measures (e.g., urinary fluoride levels) that also account for many important covariates that often capture the cluster (city or cohort) effect. The minimal impact of clustering is illustrated by Bashash et al. (2017) who accounted for clustering at the cohort level by using cohort as a fixed effect in the linear regression models. In addition, these models adjusted for many important covariates, which are also likely to reflect the cohort effects. The minimal impact of clustering is reflected in the similarity between the unadjusted and adjusted effect estimates (β [95% CI] = −2.37 [−4.45, −0.29] and −2.50 [−4.12, −0.59], respectively).
- In the case of Green et al. (2019), we contacted the study authors and received the results from models using city as a random intercept. The authors shared that the overall adjusted effect estimates with city as a fixed effect and with city as a random effect were not significantly different from each other (β [95% CI] = −1.95 [−5.19, 1.28] and −2.20 [−5.39, 0.98], respectively).

  To be responsive to the Committee's comment, we added a sensitivity analysis to the meta-analysis manuscript using the effect estimate from the random-effect models from both Bashash et al. (2017) and Green et al. (2019) to assess the impact of accounting for clustering. The results of that new sensitivity analysis compared to the main overall effect estimate have been added to eTable 6 (see excerpt of eTable 6 below).

  Excerpt of eTable 6. Regression Slopes Meta-analysis

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimate				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Sensitivity Analysis *using estimates from random effect models for Green et al. (2019)[113] and Bashash et al. (2017)[112]*				
Full-scale IQ	9	−1.80 (−2.80, −0.80)	<0.001	76%

- In summary, for 12 of the 13 low risk-of-bias studies included in the meta-analysis, the potential impact of clustering was considered minimal due to use of individual-level data and/or adjustment for important covariates (including cohort or city) or similarities between study areas with respect to certain characteristics (e.g., SES). In one of these 12 low risk-of-bias studies (Xiang et al. 2003a), clustering at the village level was not considered in the analytic approach; however, only two villages were compared. A dose-response relationship was observed with the unexposed village as the reference, and a dose-response relationship was observed within the "exposed" village. The dose-response relationship within the "exposed" village suggests that the effect is not driven by between-village differences only, thus reducing the concern for the effect being biased due to lack of accounting for the differences between the two villages and likely minimizing the impact on the effect estimates. In summary, the potential impact of clustering was considered minimal and unlikely to appreciably bias the observed effect estimates.
- In the remaining low risk-of-bias study (Trivedi et al. 2012), comparative analyses did not account for clustering of children within the high- and low-fluoride villages (six villages in total). The study did not use individual exposure levels and did not account for clustering; therefore, the lack of these considerations is likely to bias the standard error of the difference in mean IQ levels between the high- and low-fluoride villages. This bias is likely to make the differences appear stronger than they actually are; however, the clustering bias is not considered sufficient to fully account for the reported differences. Therefore, the lack of accounting for clustering was not considered a major concern, as it would not likely change the nearly 5-point difference in IQ scores reported in the study between the high- and low-fluoride villages. The study is considered low risk of bias overall because it has low potential for bias for the three key risk-of-bias questions (confounding, exposure characterization, and outcome assessment), and clustering was not considered a major concern.

  To be responsive to the Committee's comment, we added a sensitivity analysis to the meta-analysis manuscript that excluded Trivedi et al (2012) from the *mean-effects meta-analysis* (both the overall effect analysis and the low risk-of-bias subgroup analysis) to assess the impact of clustering. Excluding Trivedi et al. (2012) did not change the results appreciably. The results of this sensitivity analysis compared to the main overall effect estimate are shown below.

  Excerpt of eTable 3. Sensitivity Analyses for Mean-effects Meta-analysis: Pooled SMDs and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimates				
Overall effect	55	−0.46 (−0.55, −0.37)	<0.001	87%
Low risk of bias	10	−0.22 (−0.39, −0.05)	<0.001	83%
Sensitivity Analyses excluding Trivedi et al. (2012)[40]				
Overall effect	54	−0.46 (−0.56, −0.37)	<0.001	87%
Low risk of bias	9	−0.22 (−0.40, −0.04)	<0.001	85%

The statistical review conducted by NTP also failed to identify a study that did not properly account for the sampling design. Yu et al. (2018) used a hierarchical stratified sampling design but did not indicate that sampling weights were used in the analysis. Thus, both point estimates (means and regression coefficients) and standard errors were likely biased (Lohr, 2019). ***NTP should examine the studies included in the meta-analysis in greater depth to determine whether each study properly accounted for its design because not doing so could invalidate the meta-analysis results***.

Response: Agree (change made)

- We agree with the first part of the recommendation that says we should examine whether the studies properly accounted for sampling design, and we address this part of the comment in the final bullet of this response.
- However, we disagree that using effect estimates from studies that did not properly account for the sampling design invalidates the results of the meta-analysis. The purpose of the meta-analysis is to combine results from multiple studies with a variety of features to examine the data collectively and more precisely quantify the overall (pooled) association. The meta-analysis also allows for exploration of sources of heterogeneity through stratified analyses. Our risk-of-bias assessment carefully considered failures to account for sampling strategy or clustering in determining study-specific potential for bias. Our analyses stratify results by risk-of-bias status to evaluate the potential impact on the overall effect estimates from studies that have high potential for bias versus studies that have low potential for bias.
- We also performed new sensitivity analyses excluding the results from the studies that did not account for complex sampling strategies (Yu et al. 2018; Zhang et al. 2015b) and the pooled effect estimate for full-scale IQ did not change appreciably (see excerpt of eTable 6 below).
- We also performed new sensitivity analyses using unadjusted effect estimates when available. The results showed that including unadjusted effect estimates in the meta-analysis did not change the results appreciably (see excerpt of eTable 6 below).
- While the true tradeoff between bias and uncertainty is unknown, including some unadjusted or insufficiently adjusted effect estimates can still be informative and lead to a more precise pooled estimate due to the number of studies (Higgins et al. 2019).

Excerpt of eTable 6. Regression Slopes Meta-analysis

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimate				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Sensitivity Analyses				
Excluding Yu et al. (2018)[3] and Zhang et al. (2015b)[110]				
Full-scale IQ	7	−1.76 (−2.90, −0.62)	<0.001	82%
Using unadjusted estimates from Bashash et al. (2017),[112] Cui et al. (2018),[76] Green et al. (2019)[113], Yu et al. (2018)[3]				
Full-scale IQ	9	−1.82 (−2.81, −0.83)	<0.001	76%

- To address the part of the comment we agree with "***NTP should examine the studies included in the meta-analysis in greater depth to determine whether each study properly accounted for its design"***, we revised text in Appendix E of the prepublication 2022 NTP Monograph (previously Appendix 4 in the September 2020 draft NTP Monograph) to note specifically whether studies accounted for the sampling strategy. In cases where the publication used stratified or clustered sampling designs but did not mention whether the sampling strategy was accounted for in the analysis (e.g., Cui et al. 2018, 2020; Yu et al. 2018), we contacted study authors to specifically ask for this information. If they responded, we updated the information in Appendix E (noting that the information came from correspondence with the authors). If authors did not respond, we noted that we contacted them but did not receive a response.

Need to Juxtapose Results of Broadly Comparable Studies

Note: The NASEM Committee's comment in the first paragraph of this section is on the prepublication 2022 NTP Monograph and is not reproduced here as it is not directly relevant to the meta-analysis. See "Sup01_Monograph" for the monograph-relevant comments and responses.

The critical information regarding comparison of study results comes from the new meta-analysis, which seeks to extract and integrate comparable findings from selected studies as discussed further below. The overall approach appears to be sound in comparing mean IQ scores for the most and least highly exposed to fluoride even though the absolute fluoride concentrations are not comparable among studies. ***Because the meta- analysis is so critical to the conclusions that are drawn, NTP should provide the data that were used from each study to enable the reader to understand and evaluate what was done***.

Note: This comment was about conclusions made on the 2020 draft NTP Monograph, but because it mentions the meta-analysis, we have responded here. While we appreciate the Committee's support of the methods used in the meta-analysis, we strongly disagree with the Committee's assertion that the meta-analysis is critical to the conclusions drawn in the monograph. NTP reached the same hazard conclusions in the 2019 draft NTP Monograph, which lacked a meta-analysis, as we did in the 2020 revision in which we included a meta-analysis at the Committee's recommendation.

With removal of the hazard assessment from the 2020 draft NTP Monograph, our focus shifted to providing a qualitative confidence assessment of the relevant literature of fluoride exposure and neurodevelopmental and cognitive health effects in children and adults which is presented in the prepublication 2022 NTP Monograph. In contrast, the updated meta-analysis manuscript provides a quantitative assessment of the studies examining fluoride exposure and IQ in children. After considering the scope and nature of the NASEM Committee's comments, we determined that without the hazard assessment section to integrate these two assessments, the confidence assessment of the complete evidence base on neurodevelopmental and cognitive health effects in children and adults is a broad and distinct issue from the specific focus of the meta-analysis on IQ in children. It is our view that the topic is of such high public health importance that the

integration of the confidence assessment of the complete evidence base on neurodevelopmental and cognitive health effects would be better done as a collective effort by the public health community in a larger conversation about the appropriate method and timing of population exposures to fluoride to benefit oral health.

The values that were used to determine the standardized mean differences (SMDs) could not be found in the revised monograph, nor was there a figure that showed the pattern of results from studies restricted to the lower exposure ranges. A more detailed assessment of the meta-analysis is provided in the next section.

Response: Agree (change made)

- At the Committee's suggestion, we have added eTable 2 (excerpt provided below), which presents the specific results used from each study including the values used to determine the standardized mean differences (SMDs). Specifically, eTable 2 presents means, standard deviations, sample sizes, regression slopes with 95% confidence intervals, and exposure levels. The source of the results (e.g., table, figure) from each study publication is also listed.
- The results of the studies restricted to lower exposure ranges are presented in eTable 4, eTable 5, eFigure 17, and eFigure 18, which are all provided in supplemental materials.

Excerpt of eTable 2. Study Characteristics and Study-specific Effect Estimates Included in the Meta-analyses and Sensitivity Analyses

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Bashash et al. (2017)[112,me, u, rs] *Prospective Cohort*	Mexico	6–12	Maternal urine Reference/high fluoride levels (based on children urinary fluoride)	<0.80 mg/L (reference) ≥0.80 mg/L (high)	77, 95.37 (10.31) 112, 96.80 (11.16)	77, 95.37 (10.31) 112, 96.80 (11.16)	−2.50 (−4.12, −0.59) per 0.5 mg/L maternal urinary F	Abstract, Table 3
Razdan et al. (2017)[73,sa] *Cross-sectional*	India	12–14	Drinking water Low/high fluoride levels	0.6 ppm (low) 4.99 ppm (high)	69, 38.61 (6.34) 75, 13.95 (5.14)			Table 2
Valdez Jiménez et al. (2017)[74sa] *Prospective Cohort*	Mexico	Infancy	Maternal urine, drinking water	Urine: 1.9 ± 1.0 mg/L (1st trimester) 2.0 ± 1.1 mg/L (2nd trimester) 2.7 ± 1.1 mg/L (3rd trimester) Water: 2.6 ± 1.1 mg/L (1st trimester) 3.1 ± 1.1 mg/L (2nd trimester) 3.7 ± 1.0 mg/L (3rd trimester)			Bayley MDI: −19.05 (8.9) per 1 log10 mg/L maternal urinary F (1st trimester) −19.34 (7.46) per 1 log10 mg/L maternal urinary F (2nd trimester)	Table 2, Table 4
Cui et al. (2018)[76,rs] *Cross-sectional*	China	7–12	Urine	Boys: 1.3 (0.9–1.7)[d] mg/L Girls: 1.2 (0.9–1.6)[d] mg/L			−2.47 (−4.93, −0.01) per 1 log urinary F	Table 2
Yu et al. (2018)[3,me, w, u*, rs] *Cross-sectional*	China	7–13	Maternal urine Low/medium/high fluoride ranges Drinking water Normal/high fluoride	Urine: 0.01–1.60 mg/L (low) 1.60–2.50 mg/L (medium) 2.50–5.54 mg/L (high) Water: ≤1 mg/L (normal) >1 mg/L (high) Overall: 0.01–5.54 mg/L (urine) 0.20–3.90 mg/L (water)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	0.36 (−0.29, 1.01) per 0.5 mg/L maternal urinary F	Table 1, Table 3

Evaluation of the Meta-Analysis

The committee found the meta-analysis to be a valuable addition to the monograph and acknowledges the tremendous amount of work that was required. The meta-analysis applied

standard, broadly accepted methods, and the data shown in Figure A5-1 and the related evaluations are especially informative (NTP 2020a, p. 235). As noted in the revised monograph, 44 of the 46 studies represented in that figure had effect estimates to the left of zero—results that indicate an association between higher fluoride exposures and lower IQ. Those results highlight the marked consistency in the current epidemiologic literature on fluoride and childhood IQ. The subgroup analyses also add considerable strength to the monograph. Despite those improvements, there are areas in which further clarification or revision is needed. Because the revised monograph provides the first opportunity to review and comment on the meta-analysis, the committee offers more detailed suggestions here than in the other sections of this letter report.

One area that needs attention is data transparency. Although the results of each study in the meta-analysis are presented in figures, it is difficult to understand where each of the data points comes from and what each data point represents. Many of the publications used in the meta- analysis provide a number of results or present results in several ways. For example, Bashash et al. (2017) provide results for both child and maternal urinary fluoride concentrations. It is difficult to determine which results were selected for the overall meta-analysis or for each subgroup analysis. In addition to the figures in the revised monograph, ***NTP should add a table that provides more information on each study result, including the actual result used from each study (SMDs, regression coefficients, and CIs), any data that NTP might have used to calculate the results (for example, means, standard deviations, and sample sizes), and other key information (for example, exposure concentrations of the high- and low-fluoride groups, the method used to assess exposure and outcome, which populations overlap, and information obtained from study authors).*** Table A-1 includes some of that information but does not include the actual results that NTP selected for the meta-analysis. Overall, adding a table that includes the critical information on each study result would allow readers to identify which result from each study was used and support a better understanding of why NTP selected the results that it did for inclusion in the meta-analysis.

Response: Agree (change made)

- We have updated Table A-1 (now called Table 1. Characteristics of Studies Included in the Meta-analysis) (excerpt provided below) and added eTable 2 (excerpt provided below).
- As requested by the Committee, eTable 2 presents the specific results from each study that were used in each meta-analysis (i.e., the *mean-effects meta-analysis*, the *dose-response mean-effects meta-analysis*, and the *regression slopes meta-analysis*). This includes the actual values from each study that were used to calculate the SMDs, regression slopes, and confidence intervals. Specifically, eTable 2 presents the study design, study location, age range of children, assessment (metric and exposure groups), fluoride exposure levels for each group, means, standard deviations, sample sizes, and regression slopes with 95% confidence intervals. The source of the results (e.g., table, figure) from each study publication is also listed.
- Table 1. Characteristics of Studies Included in the Meta-analysis (excerpt provided below) includes the study design, study location, age range of children, exposure assessment (metric and exposure groups), fluoride exposure levels for each group, intelligence assessment, the overall risk of bias rating, and covariates that were adjusted for in the statistical models.

- In addition, the complete study details are available and downloadable in Excel format from HAWC, a publicly accessible online database. All of the information suggested in the comment is available in HAWC including results (e.g., SMDs, CIs), data for calculations (means, SDs, sample sizes), other key information (e.g., exposure concentrations), and the full study risk-of-bias assessment (see https://hawcproject.org/assessment/405/ for information related to the meta-analysis along with the complete details of the prepublication 2022 NTP Monograph).

  Excerpt of eTable 2. Study Characteristics and Study-specific Effect Estimates Included in the Meta-analyses and Sensitivity Analyses

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Bashash et al. (2017)[112,me, u, rs] *Prospective Cohort*	Mexico	6–12	Maternal urine Reference/high fluoride levels (based on children urinary fluoride)	<0.80 mg/L (reference) ≥0.80 mg/L (high)	77, 95.37 (10.31) 112, 96.80 (11.16)	77, 95.37 (10.31) 112, 96.80 (11.16)	−2.50 (−4.12, −0.59) per 0.5 mg/L maternal urinary F	Abstract, Table 3
Razdan et al. (2017)[73,sa] *Cross-sectional*	India	12–14	Drinking water Low/high fluoride levels	0.6 ppm (low) 4.99 ppm (high)	69, 38.61 (6.34) 75, 13.95 (5.14)			Table 2
Valdez Jiménez et al. (2017)[74sa] *Prospective Cohort*	Mexico	Infancy	Maternal urine, drinking water	Urine: 1.9 ± 1.0 mg/L (1st trimester) 2.0 ± 1.1 mg/L (2nd trimester) 2.7 ± 1.1 mg/L (3rd trimester) Water: 2.6 ± 1.1 mg/L (1st trimester) 3.1 ± 1.1 mg/L (2nd trimester) 3.7 ± 1.0 mg/L (3rd trimester)			Bayley MDI: −19.05 (8.9) per 1 log10 mg/L maternal urinary F (1st trimester) −19.34 (7.46) per 1 log10 mg/L maternal urinary F (2nd trimester)	Table 2, Table 4
Cui et al. (2018)[76,rs] *Cross-sectional*	China	7–12	Urine	Boys: 1.3 (0.9–1.7)[d] mg/L Girls: 1.2 (0.9–1.6)[d] mg/L			−2.47 (−4.93, −0.01) per 1 log urinary F	Table 2
Yu et al. (2018)[3,me, w, u*, rs] *Cross-sectional*	China	7–13	Maternal urine Low/medium/high fluoride ranges Drinking water Normal/high fluoride	Urine: 0.01–1.60 mg/L (low) 1.60–2.50 mg/L (medium) 2.50–5.54 mg/L (high) Water: ≤1 mg/L (normal) >1 mg/L (high) Overall: 0.01–5.54 mg/L (urine) 0.20–3.90 mg/L (water)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	0.36 (−0.29, 1.01) per 0.5 mg/L maternal urinary F	Table 1, Table 3

  Excerpt of Table 1. Characteristics of Studies Included in the Meta-analysis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Ren et al. (1989)[66] [translated in Ren et al. 2008][me, o] *Cross-sectional*	China	8–14	No fluoride measurement Low iodine village/high fluoride and low iodine village	Not specified	Wechsler Intelligence Scale for Children	High	Sex; iodine
Chen et al. (1991)[68] [translated in Chen et al. 2008][me, w] *Cross-sectional*	China	7–14	Drinking water Nonendemic/endemic fluorosis village	0.89 mg/L (nonendemic) 4.55 mg/L (endemic)	Chinese Standardized Raven Test	High	Age; sex
Guo et al. (1991)[70] [translated in Guo et al. 2008a][me, o] *Cross-sectional*	China	7–13	Serum Reference area using wood/coal burning-related fluoride endemic area	0.1044 ± 0.0652 mg/L (reference) 0.1483 ± 0.0473 mg/L (endemic)	Chinese Binet Intelligence Test	High	Age; sex; SES

As part of its meta-analysis, NTP presents several subgroup and sensitivity analyses. The committee finds them very informative; several are directly responsive to some of the committee's previous concerns. However, ***NTP should also include subgroup or sensitivity analyses that respond to the committee's concerns about blinding, complex sampling designs, and statistical analyses that account for clustered study designs.*** Those analyses would

include subgroup analyses that separate studies that did and did not blind the outcome assessors, a sensitivity analysis that omits studies with complex sampling designs that did not mention the use of sampling weights, and a sensitivity analysis that omits studies that used community-level exposures but did not account for clustering. Alternatively, NTP could perform a sensitivity analysis in which the standard errors of the studies that did not account for clustering are multiplied by an estimate of the VIF. Other subgroup analyses that should be considered are ones that compare prenatal and postnatal exposures. The additional subgroup or sensitivity analyses noted could help to alleviate some of the committee's current concerns.

Response: Agree (change made)

- We conducted additional sensitivity analyses to address all three concerns raised by the Committee: blinding, complex sampling designs, and clustering (the latter two are also addressed in previous responses). Several additional sensitivity analyses are shown in the table below, with one excluding Cui et al. (2018) to respond to the Committee's concerns about blinding. To address the Committee's concerns about complex sampling designs, we conducted a sensitivity analysis excluding Yu et al. (2018) and Zhang et al. (2015b). To address the Committee's concerns about clustering, we performed three sensitivity analyses—one using the unadjusted effect estimates and one using the estimates from the random effect models from Bashash et al. (2017) and Green et al. (2019) (see excerpt of eTable 6 below), and one excluding Trivedi et al. (2012) from the *mean-effects meta-analysis* (see excerpt of eTable 3 below).
- The additional sensitivity analyses had minimal impact on the pooled effect estimate for full-scale IQ (results shown in excerpts below).

Excerpt of eTable 6. Regression Slopes Meta-analysis

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity p-value	Heterogeneity I^2
Overall Estimate				
Full-scale IQ	9	−1.81 (−2.80, −0.81)	<0.001	77%
Sensitivity Analyses				
Excluding Cui et al. (2018)[76]				
Full-scale IQ	8	−1.89 (−3.03, −0.74)	<0.001	80%
Excluding Yu et al. (2018)[3] and Zhang et al. (2015b)[110]				
Full-scale IQ	7	−1.76 (−2.90, −0.62)	<0.001	82%
Using unadjusted estimates from Bashash et al. (2017),[112] Cui et al. (2018),[76] Green et al. (2019)[113], Yu et al. (2018)[3]				
Full-scale IQ	9	−1.82 (−2.81, −0.83)	<0.001	76%
Using estimates from random effect models for Green et al. (2019)[113] and Bashash et al. (2017)[112]				
Full-scale IQ	9	−1.80 (−2.80, −0.80)	<0.001	76%

Excerpt of eTable 3. Sensitivity Analyses for Mean-effects Meta-analysis: Pooled SMDs and 95% CIs for Children's IQ Score and Exposures to Fluoride

Analysis	Number of Studies	Beta (95% CI)	Heterogeneity	
			p-value	I^2
Overall Estimates				
Overall effect	55	−0.46 (−0.55, −0.37)	<0.001	87%
Low risk of bias	10	−0.22 (−0.39, −0.05)	<0.001	83%
Sensitivity Analyses excluding Trivedi et al. (2012)[40]				
Overall effect	54	−0.46 (−0.56, −0.37)	<0.001	87%
Low risk of bias	9	−0.22 (−0.40, −0.04)	<0.001	85%

Another major concern of the committee in its first review was that NTP might have been including multiple results from a given study population. In its meta-analysis protocols (NTP2020b, p. 83), NTP implies that only one result from each population was used. The section of the meta-analysis of "individual-level exposure data" (NTP 2020a, Appendix 5, p. 246) includes a good discussion of two overlapping sets of publications (Yu et al. 2018/Wang et al. 2020 and Green et al. 2019/Till et al. 2020) and the process used to select one result from each set.

However, NTP appears to have included at least one set of overlapping publications—Xiang (2003) and Xiang (2011) (Figure A5-1)—in the overall meta-analysis of mean effects. ***NTP should review all its analyses to ensure that overlapping publications are not included in any single meta-analysis***. That exercise is especially important given that the issue of "double counting" was a substantive concern of the committee in its first review.

Response: Agree (change made)

- We agree with the Committee that, according to our protocol, we should not have included a set of overlapping populations from Xiang et al. (2003) and Xiang et al. (2011) in the same meta-analysis. Therefore, we have removed the Xiang et al. (2011) assessment of IQ associated with serum fluoride levels from the meta-analyses.
- We further reviewed all the analyses and found no other overlapping populations used in a single meta-analysis.

Another issue involves the overall organization of the meta-analysis protocols and results. Information on the meta-analysis protocols and information on the meta-analysis results are presented in several places. That approach forces the reader to go back and forth between sections and between documents to determine what was done or to obtain a clear picture of the meta-analysis findings. For example, some methods are described in the protocol, some in the revised monograph (NTP 2020a, pp. 48-51), and some in Appendix 5. In addition, NTP presents an exhaustive set of forest plots, funnel plots, Egger and Begg test results, and trim and fill plots and results. NTP can be applauded for developing so many data displays and being so transparent here. However, much of the information is not that helpful, and it is difficult to wade through it, given the sheer volume. Some of the information could be eliminated, summarized, or presented more succinctly or at least provided in a separate

document, website, or appendix.
Overall, some coalescing and reorganization of the meta-analysis protocols and results would make the meta-analysis easier to follow and easier to interpret.

Response: Agree (change made)

- We have considered this comment when presenting the meta-analysis results. All forest plots, funnel plots, Egger and Begg test results, trim-and-fill plots, and sensitivity analyses results are presented in supplemental materials, and the number of meta-analysis figures has been appropriately reduced from 103 to 27.

NTP provides a reasonably thorough and appropriate evaluation of publication bias. ***In addition to what it has presented, it should mention the weaknesses of the tests used to evaluate that bias***. One weakness is that the evaluation of the funnel plot involves mostly a subjective interpretation, which can be especially troublesome when the number of studies is small. Another weakness is the possibility that positive results from the funnel plot and the Egger and Begg tests might be caused by something other than publication bias. In addition, NTP uses the phrase "eliminating publication bias" when it refers to the results of the trim and fill analyses (see, for example, NTP 2020a, p. 49). However, because the tests for publication bias are not 100% specific, it is not known exactly what is being eliminated by the trim and fill process. The committee suggests that a better phrase might be "adjusting for possible publication bias." In summary, acknowledging the weaknesses of the tests that were used to evaluate publication bias would make the report more transparent.

Response: Agree (change made)

- We agree with the Committee's overall comment that, "NTP provides a reasonably thorough and appropriate evaluation of publication bias."
- We also agree with the Committee's recommendation to use the phrase "adjusting for possible publication bias" instead of "eliminating publication bias" and revised the language throughout the manuscript.
- Finally, we agree that the limitations of the tests used to evaluate publication bias should be mentioned and have added the following to the *Discussion* section:

  "There are also several limitations to the existing approaches for evaluating potential for publication bias. The funnel plot asymmetry is a subjective assessment and is recommended only when at least 10 studies are included in the meta-analysis.[64] Furthermore, the Egger regression test and Begg's rank tests[25-27] may suffer from inflated type I power and limited power in certain situations.[65]"

NTP notes that 44 of the 46 studies (96%) in its meta-analysis of childhood IQ have effect estimates to the left of zero. That finding should be emphasized more, and its meaning with respect to evaluating and quantifying heterogeneity should be mentioned. To assess heterogeneity, NTP primarily used the Cochran's Q test. However, heterogeneity can also be assessed by providing a count or percentage of the number of studies to the right or left of the null value. Some would consider that a much simpler, more intuitive, and perhaps more useful

way of assessing heterogeneity, especially in light of the marked differences between the studies in design, study populations, exposure and outcome assessment methods, and statistical analyses. Although that approach should not be used as the sole basis of conclusions, it can be a useful first step in exploring why heterogeneity might exist. For example, Figure A5-1 appears to show that Broadbent et al. (2015) and Bashash et al. (2017) are two major contributors to the heterogeneity seen in the overall meta-analysis, and they should be clearly identified in the monograph. NTP does note that there were two studies with effect estimates to the right of the null (NTP 2020a, p. 49, last full paragraph), but a key reference (Bashash et al. 2017) is missing.

Response: Agree (change made)

- We agree with the Committee's suggestion and have revised the *Results* section to emphasize the large number of studies that show effect estimates to the "left of zero" (i.e., to the left of null), as follows:

  "*The pattern of results across the 55 studies was consistent; 52 (95%) reported an inverse association with SMDs ranging from −5.34 (95% CI: −6.34, −4.34) to −0.04 (95% CI: −0.45, 0.36)*".

- We have revised the *Results* section to include clear references to the studies with effect estimates to the right of the null, as follows:

  "*The three studies with a non-negative association reported SMD estimates of 0.01 (95% CI: −0.19, 0.21),[6] 0.01 (95% CI: −0.19, 0.22),[38] and 0.13 (95% CI: −0.16, 0.42).[5]*"

In addition to identifying the studies, NTP should explore whether there might be an obvious or likely reason for the results of those two studies to tend to differ from the results of the others. For example, the Bashash et al. (2017) result used in the meta-analysis of SMDs appears to be for the cross-sectional evaluation of children's urinary fluoride concentrations. However, the study also presents prospective results that use maternal prenatal urinary fluoride concentrations, and, unlike the cross-sectional results, the prospective results indicate a fairly strong adverse relationship—a relationship that is much more consistent with that in the other studies used in the meta-analysis. ***The rationale for choosing one result over the other should be provided because such decisions can affect the results of the meta-analysis***.

Response: Agree (change made)

- As mentioned in a previous comment, we have added a new table, eTable2 (excerpt provided below), which clearly indicates which results were used from each publication and in which analysis they were used (i.e., *mean-effects meta-analysis* or *regression slopes meta-analysis*). In the example of Bashash et al. (2017), the cutoff based on the children's urinary fluoride concentrations (CUF) (0.80 mg/L) was used to determine the reference and exposed groups needed in the *mean-effects meta-analysis*. The estimate from the association between individual-level maternal urinary fluoride (MUF) concentrations and IQ was used in the *regression slopes meta-analysis*.

- In response to this comment, we have added text to the supplemental materials to identify likely reasons that the results from three studies (Bashash et al. 2017, Broadbent et al. 2015, and Green et al. 2019) differ from the results of the other studies, as follows:

"The three studies with non-negative associations reported SMD estimates of 0.01 (95% CI: −0.19, 0.21),[113] 0.01 (95% CI: −0.19, 0.22),[25] and 0.13 (95% CI: −0.16, 0.42).[112] Two of the three studies with non-negative SMDs compare mean IQs in children living in fluoridated vs. non-fluoridated areas in Canada,[113] or in New Zealand.[25] No other studies included in the main mean-effects meta-analysis made comparisons between fluoridated vs. non-fluoridated areas. In both studies, levels of fluoride in water were low, even in communities with fluoridated drinking water, likely limiting the power to detect an effect.

In Bashash et al.,[112] the SMD compares mean IQ scores in children with urinary fluoride levels below vs. above 0.80 mg/L in Mexico.[112] Unlike other studies in the mean-effects meta-analysis which compared mean IQ scores between fluoridated vs. non-fluoridated areas, or areas with high vs. low fluoride exposures (see eTable 2), the Bashash et al.[112] study was not designed to measure fluoride exposure by geographical area. However, since the mean IQ scores were provided in the manuscript for children with urinary fluoride levels below vs. above 0.80 mg/L, we included them in this analysis. It's worth noting that there was no significant difference when comparing MUF levels between the groups of children with urinary fluoride levels above or below 0.80 mg/L, however when children's IQs were regressed against MUF, a statistically significant inverse association was found."

Excerpt of eTable 2. Study Characteristics and Study-specific Effect Estimates Included in the Meta-analyses and Sensitivity Analyses

Reference[a] *Study Design*	Study Location	Age Range (Years)	Assessment (Metric, Exposure Groups)	Fluoride Exposure Levels	Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Dose-response Mean-effects Meta-analysis N, Mean (SD) [Reference] [Exposed]	Regression Slopes Meta-analysis Slope (SE) or 95% CI per Unit Change Fluoride	Source
Bashash et al. (2017)[112,me, u, rs] *Prospective Cohort*	Mexico	6–12	Maternal urine Reference/high fluoride levels (based on children urinary fluoride)	<0.80 mg/L (reference) ≥0.80 mg/L (high)	77, 95.37 (10.31) 112, 96.80 (11.16)	77, 95.37 (10.31) 112, 96.80 (11.16)	−2.50 (−4.12, −0.59) per 0.5 mg/L maternal urinary F	Abstract, Table 3
Razdan et al. (2017)[73,sa] *Cross-sectional*	India	12–14	Drinking water Low/high fluoride levels	0.6 ppm (low) 4.99 ppm (high)	69, 38.61 (6.34) 75, 13.95 (5.14)			Table 2
Valdez Jiménez et al. (2017)[74sa] *Prospective Cohort*	Mexico	Infancy	Maternal urine, drinking water	Urine: 1.9 ± 1.0 mg/L (1st trimester) 2.0 ± 1.1 mg/L (2nd trimester) 2.7 ± 1.1 mg/L (3rd trimester) Water: 2.6 ± 1.1 mg/L (1st trimester) 3.1 ± 1.1 mg/L (2nd trimester) 3.7 ± 1.0 mg/L (3rd trimester)			Bayley MDI: −19.05 (8.9) per 1 log10 mg/L maternal urinary F (1st trimester) −19.34 (7.46) per 1 log10 mg/L maternal urinary F (2nd trimester)	Table 2, Table 4
Cui et al. (2018)[76,rs] *Cross-sectional*	China	7–12	Urine	Boys: 1.3 (0.9–1.7)[d] mg/L Girls: 1.2 (0.9–1.6)[d] mg/L			−2.47 (−4.93, −0.01) per 1 log urinary F	Table 2
Yu et al. (2018)[3,me, w, u*, rs] *Cross-sectional*	China	7–13	Maternal urine Low/medium/high fluoride ranges Drinking water Normal/high fluoride	Urine: 0.01–1.60 mg/L (low) 1.60–2.50 mg/L (medium) 2.50–5.54 mg/L (high) Water: ≤1 mg/L (normal) >1 mg/L (high) Overall: 0.01–5.54 mg/L (urine) 0.20–3.90 mg/L (water)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	1636, 107.40 (13.00) 1250, 106.40 (12.30)	0.36 (−0.29, 1.01) per 0.5 mg/L maternal urinary F	Table 1, Table 3

Finally, ***NTP should review the process it used to exclude study results from its meta-analysis***. For example, Table A-2 says Green (2019) was excluded because of "missing mean or SD of outcome measure; used in individual level meta-analysis." However, means and SDs are available (Green 2019, Table 1), and at least two other studies (Ding 2011; and Zhang 2015) are used in both the mean-effect and individual-level meta-analyses.

Response: Agree (change made)

- We have revised the mean-effects meta-analyses and the mean-effects dose-response meta-analyses to include SMDs calculated from the means and standard deviations in Green et al. (2019). We also agree with the Committee that Ding et al. (2011) and Zhang et al. (2015) were correctly included in both the mean-effects and regression slopes meta-analyses.
- We have further reviewed our process and assessed whether any other information was excluded when it should have been included, as described in Appendix 6 of the protocol. We found no other instances where we excluded information when it should have been included.
- The studies that were included in the mean-effects and regression slopes meta-analyses are noted with superscript letters in the first column of Table 1 (excerpt provided below). The footnote to the table defines the superscript letters as follows:

  "An "me" superscript indicates that the studies included in the mean-effects meta-analysis; an "o" superscript indicates a study included in "other" exposures mean-effects meta-analysis (see Table 2 footnote); a "w" superscript indicates studies included in the mean-effects dose-response meta-analysis using fluoride in water; a "u" superscript indicates studies included in the mean-effects dose-response meta-analysis using fluoride in urine; "" indicates studies included in the mean-effects dose-response meta-analysis at levels < 1.5 mg/L; an "rs" superscript indicates studies included in the regression slopes meta-analysis.*"

  Excerpt of eTable 1. List of Excluded Studies from Mean-effects Meta-analysis

Reference, Country	Reason for Exclusion
Qin et al. (1990)[45] [translated in Qin et al. 2008], China	Missing mean or SD of outcome measure
Yang et al. (1994)[47] [translated in Yang et al. 2008], China	Overlapping population with Wang et al. (2001)[49]; Table 2 in Yang et al. (1994)[47] seemed incomplete
Wang et al. (2005b)[50] [translated in Wang et al. 2008a], China	Missing mean or SD of outcome measure
Rocha-Amador et al. (2007)[52], Mexico	Missing mean or SD of outcome measure
Liu et al. (2000)[53] [translated in Liu et al. 2008] , China	Overlapping population with Lu et al. (2000)[55]
Sudhir et al. (2009)[56], India	Missing mean or SD of outcome measure
He and Zhang (2010)[57], China	Missing mean or SD of outcome measure
Xiang et al. (2011)[58], China	Overlapping population with Xiang et al. (2003a)[59]
Saxena et al. (2012)[60], India	Missing mean or SD of outcome measure
Wang et al. (2012)[61], China	Overlapping population with Xiang et al. (2003a)[59]

Excerpt of relevant sections of Table 1. Characteristics of Studies Included in the Meta-analysis

Reference[a] *Study Design*	Study Location	Age Range (Years)	Fluoride Exposure		Intelligence Assessment	Overall RoB Rating	Confounders Considered
			Assessment (Metric, Exposure Groups)	Levels			
Ding et al. (2011)[32me, u*, rs] *Cross-sectional*	China	7–14	Dental fluorosis (normal/ questionable/very mild/mild/moderate) Urine Mean urinary fluoride levels (10 groups)	0.80 ± 0.55 mg/L (normal) 1.13 ± 0.73 mg/L (questionable) 1.11 ± 0.74 mg/L (very mild) 1.31 ± 0.78 mg/L (mild) 1.46 ± 0.79 mg/L (moderate) 0.26 mg/L (group 1) 0.45 mg/L (group 2) 0.56 mg/L (group 3) 0.66 mg/L (group 4) 0.75 mg/L (group 5) 0.89 mg/L (group 6) 1.08 mg/L (group 7) 1.33 mg/L (group 8) 1.74 mg/L (group 9) 2.96 mg/L (group 10) 0.10–3.55 mg/L	Combined Raven's Test for Rural China	Low	Age; arsenic; iodine; lead; SES; demographics
Zhang et al.(2015b)[33me, w*, u, rs] *Cross-sectional*	China	10–12	Urine, drinking water, serum Reference/high fluoride areas	Urine: 1.10 ± 0.67 mg/L (reference) 2.40 ± 1.01 mg/L (high) Water: 0.63 (0.58–0.68) mg/L (reference) 1.40 (1.23–1.57) mg/L (high) Serum: 0.06 ± 0.03 (reference) 0.18 ± 0.11 (high)	Combined Raven's Test for Rural China	Low	Age; sex; arsenic; iodine; drinking water fluoride; SES; thyroid hormone levels; COMT genotype
Bashash et al. (2017)[5me, u, rs] *Prospective Cohort*	Mexico	6–12	Maternal urine Reference/high fluoride (based on children urinary fluoride)	<0.80 mg/L (reference) ≥0.80 mg/L (high)	Wechsler Abbreviated Scale of Intelligence	Low	Age; sex; weight at birth; parity; gestational age; maternal characteristics (smoking history, marital status, age at delivery, IQ, education, cohort)
Yu et al. (2018)[11me, w, u*, rs] *Cross-sectional*	China	7–13	Maternal urine Low/medium/high fluoride ranges Drinking water Normal/high fluoride	Urine: 0.01–1.60 mg/L (low) 1.60–2.50 mg/L (medium) 2.50–5.54 mg/L (high) Water: ≤1 mg/L (normal) >1 mg/L (high) Overall: 0.01–5.54 mg/L (urine) 0.20–3.90 mg/L (water)	Combined Raven's Test for Rural China	Low	Age; sex; health factors; SES
Green et al. (2019)[6me, w*, u*, rs] *Prospective Cohort*	Canada	3–4	Maternal urine, drinking water, maternal fluoride intake Nonfluoridated/fluoridated area	Urine: 0.40 ± 0.27 mg/L (nonfluoridated) 0.69 ± 0.42 mg/L (fluoridated) Water: 0.13 ± 0.06 mg/L (nonfluoridated) 0.59 ± 0.08 mg/L (fluoridated) Intake: 0.30 ± 0.26 mg/day (nonfluoridated) 0.93 ± 0.43 mg/day (fluoridated) Overall: 0.51 ± 0.36 mg/L (urine) 0.54 ± 0.44 mg/day (intake) 0.31 ± 0.23 mg/L (water)	Wechsler Primary and Preschool Scale of Intelligence-III	Low	Sex; city; maternal education; race/ethnicity; HOME score; prenatal secondhand smoke exposure

Notes:
COMT = catechol-O-methyltransferase; RoB = risk of bias; SES = socioeconomic status; HOME = Home Observation for Measurement of the Environment
[a]An "me" superscript indicates that the studies included in the *mean-effects meta-analysis*; an "o" superscript indicates a study included in "other" exposures mean-effects analysis (see Table 2 footnote); a "w" superscript indicates studies included in the *mean-effects dose-response meta-analysis* using fluoride in water; a "u" superscript indicates studies included in the *mean-effects dose-response meta-analysis* using fluoride in urine; "*" indicates studies included in the *mean-effects dose-response meta-analysis* at levels < 1.5 mg/L; an "rs" superscript indicates studies included in the *regression slopes meta-analysis*.
[b]Additional exposure regions including iodine levels were not included in the analysis.
[c]Additional exposure regions including arsenic levels were not included in the analysis.
[d]Median (q1−q3).

The committee identified several minor points concerning the meta-analysis, and these are provided below.

- NTP notes that pooled SMDs and pooled relative risks were considered significantly different when their 95% CIs did not overlap (NTP 2020b, p. 85). That approach can provide many false-negative results because significant differences can occur when CIs overlap.

Statistical significance should instead be determined by hypothesis tests, such as those described in Altman and Bland (2003).

Response: Agree (change made)

- Although we consider the stated approach for comparison of confidence intervals to be a widely used and accepted approach in the scientific community (Schenker and Gentleman 2001), we have taken the Committee's recommendation and have not used the overlap between confidence intervals for formal significance testing in describing our results. While we acknowledge that the method of examining overlap can be more conservative (i.e., rejects the null hypothesis less often), the method is still reliable at correctly detecting significant differences when the confidence intervals clearly do not overlap (Schenker and Gentleman 2001).

Almost all the forest and funnel plots are difficult to see because they are too narrow. They should be expanded horizontally. An example of a forest plot that is much easier to read is Figure 2 in Choi et al. (2012).

Response: Agree (change made)

- We agree with the Committee's suggestion and have revised the forest plot for the *mean-effects meta-analysis*, which appears as Figure 2 (shown below). We have also revised four additional figures that were too narrow and difficult to see due to the large number of studies included. These figures present the forest plots of the meta-analyses stratified by risk of bias, country, IQ assessment type, and exposure, and these are included in the supplemental materials.

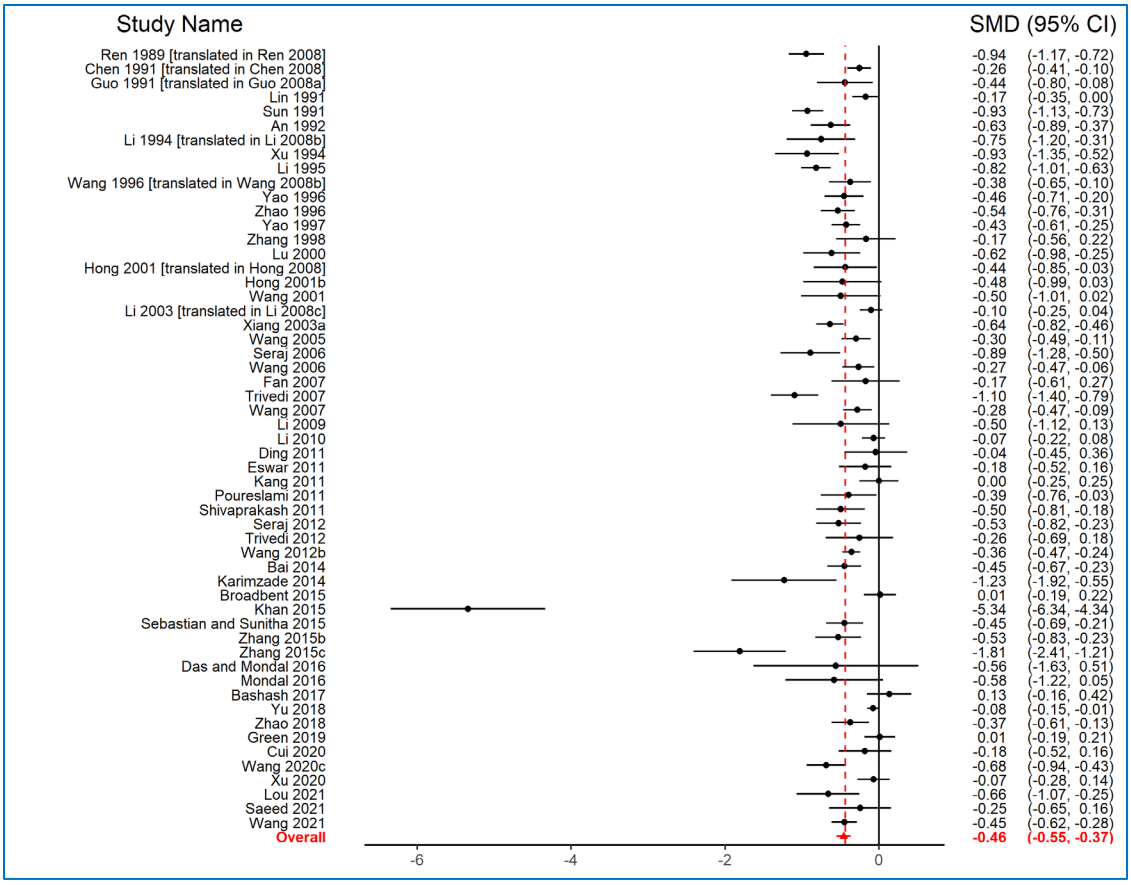


Figure 2. Association Between Fluoride Exposure and IQ Scores in Children

Forest plot for random-effects meta-analysis of the association between fluoride exposure and child's IQ scores. Effect size is expressed as the standardized weighted mean difference for heteroscedastic population variances (SMD). The random-effects pooled SMD is shown as a solid triangle. Horizontal lines represent 95% CIs for the study specific SMDs.

- Labeling the Aim 2 meta-analysis as a "meta-analysis using individual-level exposure data" is somewhat misleading because it is not clear that all the studies used in it involved individual exposure data. Some might have used ecologic exposure data or the types of clustered exposure data discussed above. Aim 2 actually appears to be a meta-analysis of regression slopes, and labeling it as such would be more appropriate.

Response: Agree (change made)

- We followed the Committee's suggestion and relabeled "meta-analysis using individual-level exposure data" to be the "*regression slopes meta-analysis*."
- The change is reflected throughout the *Methods*, *Results*, and *Discussion* sections of the manuscript as well as the *Results* section of the supplemental materials.

- NTP notes that the pooled SMD in its main meta-analysis after applying the trim and fill method is -0.42 (95% CI: -0.54, 0.30) (NTP 2020a, p. 49). NTP should confirm that the CI is correct and that the upper confidence limit is not -0.30.

Response: Agree (change made)

- We corrected the typo in the meta-analysis manuscript to reflect the negative upper limit of the confidence interval. However, since that time, the literature search in the manuscript was updated and this number is no longer relevant.
- The revised trim-and-fill analysis for the *mean-effects meta-analysis* is reflected in the *Results* section as follows:

  "*Adjusting for possible publication bias through trim-and-fill analysis suggested the imputation of seven additional studies to the right side, with an adjusted pooled SMD of –0.36 (95% CI: −0.46, −0.26)*".

- If possible, NTP should summarize its meta-analysis results for SMDs by putting the results in a format that is easier to interpret. For example, if the typical standard deviation for a commonly used IQ test is 15 IQ points, a pooled SMD of -0.50 would be expected to represent about a 7.5-point decrease in IQ. Expressing the major results as estimated IQ points, rather than as just SMDs, would make the results easier for people unfamiliar with SMDs to interpret.

Response: Disagree (no change)

- We agree that this suggestion would make the results easier to interpret. However, we disagree that the suggested approach would appropriately convert the pooled SMD estimate to estimated IQ points. As opposed to the 15 IQ points suggested, the multiplier for the pooled SMD would more appropriately be the standard deviation in the denominator of the pooled SMD from the meta-analysis. This is specific to the meta-analysis and not the same as the standard deviation of the IQs in the population for which the IQ tests were designed. Because the meta-analysis includes studies with different study populations, we decided not to convert the SMDs to IQ points. Therefore, we have not presented the results in the format suggested by the Committee.

- The rationale for excluding the PhD thesis by Thomas from the NTP review of meta-analysis should be provided.[3.]

Response: Disagree (no change)

- We were confused as to why this was recommended by the NASEM Committee for two reasons: (1) our protocol states that unpublished, non-peer-reviewed papers were not included in the systematic review, and (2) the work related to fluoride exposure during pregnancy and children's IQ in the ELEMENT cohort was later published by Bashash (2017) in a peer-reviewed journal and is included in the meta-analysis.
- In addition, the study flow diagram has been updated to an interactive version that provides a list of reasons why studies were excluded. It will be included when the manuscript is submitted for publication, which will allow readers to identify decisions for specific studies.

[3] See https://deepblue.lib.umich.edu/handle/2027.42/110409.