NTP 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) appreciates the comments provided by the National Academies of Sciences, Engineering, and Medicine (NASEM) Committee in their 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 has 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 final NTP 2022 Monograph includes consideration of comments from that external peer review in addition to the NASEM committee's comments.

In addition, the final monograph removes the meta-analysis that was added at the NASEM Committee's request following their review of the September 2019 version. The meta-analysis is being prepared as a separate journal publication, taking into consideration the NASEM Committee comments on the NTP September 2020 Draft Monograph.

Therefore, this document contains the NTP response to the NASEM Committee Letter Review comments that are directly relevant to the final NTP 2022 Monograph and describes the changes made in response to the committee's comments. The NASEM committee's comments on the meta-analysis, including the section titled "Evaluation of the Meta-Analysis" are not included in this document because the meta-analysis is not part of the NTP 2022 Monograph. Those comments and the response to the comments on the meta-analysis will be released when the manuscript is published.

Other than the meta-analysis, the complete text from the NASEM Letter Review has been included in the pages that follow and are formatted in black text for clarity. The NTP responses begin with the word "Response," are formatted in orange text, and are interspersed within the original NASEM Committee text.

NTP, April 2022

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.

Copyright 2021 by the National Academy of Sciences. All rights reserved. Printed in the United States

of America

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: NTP appreciates the desire of the committee for more specificity in the protocol with respect to laying out all aspects of the systematic review; however, NTP respectfully submits 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. The NTP has added the following text to the methods section of the NTP 2022 Monograph to further clarify the role of the OHAT handbook. "The protocol served as the complete methods followed for the conduct of the systematic review. The OHAT handbook is a source of general systematic review methods that were selected and tailored in developing the 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: Figure 2 (titled "Study Selection Diagram") in the NTP 2022 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 NTP 2022 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: We call attention to the committee of text on pg. 22 of the NTP September 2020 Draft Monograph that discusses the SWIFT-Active screening process and implications for stopping at 98% with respect to possible studies missed. The NTP assumes that the committee means to refer to the term "predicted recall," as the term "percent recall" is not found in the NTP September 2020 Draft Monograph. In the "Evaluation of SWIFT-Active Screener Results" section of the NTP 2022 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: The reasons cited in the NTP September 2020 Draft Monograph for not reanalyzing the animal data have been removed from the NTP 2022 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. The NTP also respectfully disagrees that the NTP September 2020 Draft Monograph implies that "a large number of epidemiology studies generally negates the value of animal studies in hazard identification" as is indicated in the committee's comments. The NTP 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: We agree with the committee's recommendation, and this information has been removed from the NTP 2022 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: The NTP recognizes 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 NTP 2022 Monograph to further explain the rationale for our approach. The NTP searched for and was 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, NTP 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: The NTP accepted this suggestion and has 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 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

[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.

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: 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 NTP 2022 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 NTP 2022 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: We recognize the committee's continued concerns over the consistent application of the risk-of-bias criteria. While a top priority to NTP 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. The NTP applies the risk-of-bias criteria to individual studies and specifically looks 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 NTP 2022 Monograph (previously Appendix 4 in the September 2020 Draft 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 NTP September 2020 Draft Monograph was still in draft form, we felt 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 NTP September 2019 Draft Monograph to the NTP September 2020 Draft 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 NTP September 2020 Draft 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 NTP September 2019 Draft Monograph to the NTP September 2020 Draft 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 NTP 2022 Monograph.
- Das and Modal 2016 – From the NTP September 2019 Draft Monograph to the NTP September 2020 Draft 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: NTP has re-evaluated risk of bias due to potential confounding. After further review, NTP 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 4 of the NTP September 2020 Draft Monograph and Appendix E of the NTP 2022 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: 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: 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 NTP September 2020 Draft Monograph (and in the NTP 2022 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: The sub-bullet "Direction/magnitude of effect" text in Appendix 4 of the NTP September 2020 Draft 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 NTP 2022 Monograph (previously Appendix 4 in the September 2020 Draft 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: 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: Please note that the risk-of-bias rating explanations provided in HAWC and Appendix 4 of the NTP September 2020 Draft 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. 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 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 in Appendix 4 of the NTP September 2020 Draft Monograph and Appendix E of the NTP 2022 Monograph (e.g., "Blinding or other methods to reduce bias are not reported, but correspondence with the study authors indicated that the teachers were blind to the status of fluoride").

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: The NTP appreciates 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. NTP respectfully stands by its 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. NTP also considers 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, the NTP 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 that 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: 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 2018, the NTP states in Appendix 4 of the NTP September 2020 Draft Monograph that, "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." 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. NTP acknowledges in Appendix E of the NTP 2022 Monograph (previously Appendix 4 in the September 2020 Draft Monograph) that there is still potential for knowledge of the area by the outcome assessor, but overall NTP 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: The NTP recognizes 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" 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. The NTP considers self-reporting of a learning disability to be an insensitive method (as stated in Appendix 4 of the NTP September 2020 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), the NTP considers 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: We appreciate the committee's continued concerns over the adequacy of the statistical approaches used in some of the publications reviewed in the NTP 2020 Draft Monograph. We have expanded our comments concerning the statistical methods used in the low risk-of-bias studies in Appendix E of the NTP 2022 Monograph (previously Appendix 4 in the September 2020 Draft 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: The potential impact of clustering is addressed in multiple ways in the NTP 2022 Monograph, that expand previous discussion and analysis. We have revised text in Appendix E of the NTP 2022 Monograph (previously Appendix 4 in the September 2020 Draft 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], and 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 NTP 2022 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: We have revised text in Appendix E of the NTP 2022 Monograph (previously Appendix 4 in the September 2020 Draft 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***.

Responses to comments on the meta-analysis will be released when the manuscript is published.

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: 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 NTP 2022 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: A detailed response to the committee's critique of the meta-analysis will be provided when the manuscript is published. 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 NTP September 2019 Draft 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. 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 NTP 2022 Monograph and will pursue publishing it separately.

Evaluation of the Meta-Analysis

Note: The NASEM committee comments on the "Evaluation of the Meta-Analysis" are not included in this document. Those comments and responses to comments on the meta-analysis will be released when the manuscript is published.

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: The committee correctly states that the data driving the hazard conclusions in the NTP September 2019 and 2020 Draft 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 NTP 2022 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 NTP September 2019 and September 2020 versions of the draft 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: NTP agrees with and appreciates the committee's statements concerning the importance of this assessment and believes that the final 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.