PROTOCOL FOR A SYSTEMATIC REVIEW OF TRAFFIC-RELATED AIR POLLUTION AND SELECTED HEALTH OUTCOMES

Project Co-Leaders: Brandiese E.J. Beverly, PhD and Kembra Howdeshell, PhD, Office of Health Assessment and Translation (OHAT), DNTP

Summary: NTP is conducting a systematic review to evaluate the evidence for an association between traffic-related air pollution and hypertensive disorders of pregnancy. The protocol is detailed in this document.

Date Original Protocol Published: June 2016

Date Revised Protocol Published: December 12, 2019; reflecting revisions made through August 7, 2019

Documentation of Revisions to the Protocol: The principal revisions are detailed in the Protocol History and Revisions Table on Page 32 including the reasons for each revision. In addition, updated language or new text starts with the word “Revision:”. Strikethrough text indicates original text that has been modified, and new text is marked in bold text.

PROTOCOL FOR A SYSTEMATIC REVIEW OF TRAFFIC-RELATED AIR POLLUTION AND SELECTED HEALTH OUTCOMES

~~**Project Leader:** Kembra Howdeshell, PhD, Office of Health Assessment and Translation (OHAT), DNTP~~
REVISION: Project Co-Leaders: Brandiese E.J. Beverly, PhD and Kembra Howdeshell, PhD, Office of Health Assessment and Translation (OHAT), DNTP

Summary: ~~OHAT is conducting two systematic reviews to evaluate the evidence for an association between traffic-related air pollution and (1) pregnancy-associated hypertension, and (2) neurological development. The protocol is detailed in this document.~~
REVISION: OHAT is conducting a systematic review to evaluate the evidence for an association between traffic-related air pollution and hypertensive disorders of pregnancy. The protocol is detailed in this document.

BACKGROUND AND SIGNIFICANCE

Background

Traffic-related air pollution contributes significantly to ambient air pollution, especially in urban settings (Krzyzanowski *et al.* 2005, HEI 2010). Children are identified as a subpopulation that is particularly sensitive to air pollution. The majority of research on traffic-related air pollution associated with children's health has focused on the development and exacerbation of asthma and other respiratory problems (HEI 2010). However, there is increasing evidence that air pollution may impact many other facets of development including neurological function (Wang *et al.* 2009), cardiovascular development and function (Bilenko *et al.* 2015), metabolism (e.g., obesity incidence) (Calderon-Garciduenas *et al.* 2015), and pregnancy outcomes, such as gestational hypertension (Pedersen *et al.* 2014), congenital birth defects (Farhi *et al.* 2014), impaired fetal growth (Pereira *et al.* 2011, Pereira *et al.* 2012), infant mortality, and preterm birth (Pereira *et al.* 2010). Of special interest, the Centers for Disease Control and Prevention (CDC) recently published a review of traffic-related air pollution associated with childhood cancer (Boothe *et al.* 2014) and is conducting a literature review on traffic-related air pollution associated with adverse pregnancy outcomes focused on fetal growth and preterm birth (Tegan Boehmer, manuscript in preparation). ~~The current OHAT evaluation will evaluate the evidence for an association between traffic-related air pollution and two emerging health outcomes directly or indirectly related to children's health: (1) pregnancy-associated hypertensive disorders and (2) neurological development.~~

REVISION: **Two topics were considered for evaluation of the evidence for an association between traffic-related air pollution and emerging health outcomes directly or indirectly related to children's health: (1) hypertensive disorders of pregnancy and (2) neurological development. This protocol also outlines a separate review of traffic-related air pollution and neurological disorders, but this topic is currently under review by other organizations. Therefore, OHAT will develop a systematic review on hypertensive disorders of pregnancy; the evaluation on neurological development will not be pursued.**

Exposure Measures

Traffic-related air pollution is characterized by two main types of exposure measurements, direct traffic measurements and measurements of traffic-related air pollutants. Direct traffic measurements are

relatively simple indicators of exposure to traffic exposure, which are relatively easy to obtain, using, for example, geographic information system (GIS) methods. Direct traffic measurements may include distance to main road, length of main streets with a buffer zone around homes or schools, traffic volume, etc. However, direct traffic measurements may lack precision because they often do not account for meteorology, dispersion and terrain (HEI 2010). Traffic-related air pollutants are defined as those primarily emitted by mobile sources due to fossil fuel combustion; mobile sources include passenger cars, diesel trucks and buses, and "non-road" equipment (*e.g.*, recreational vehicles, lawn and garden equipment, *etc.*) (http://www.epa.gov/oms/toxics.htm). Several major air pollutants regulated by the United States Environmental Protection Agency (US EPA) are used as pollutant surrogates for traffic-related air pollution, such as CO, NOx (usually in the form of nitrogen dioxide (NO2)) and particulate matter. Particulate matter (PM) that is most specific to vehicular exhaust includes: black carbon (BC) and elemental carbon (EC), which are present in diesel exhaust; ultrafine PM (particles of ≤0.1 µm diameter), and coarse PM (particles of >2.5 to <10 µm), which includes road dust and brake and tire wear. Traffic also contributes to PM2.5 (particles of ≤2.5 µm) pollution, although it is not a source of primary PM emissions (HEI 2010). Mobile source air toxics are also compounds (individual chemicals and mixtures) released during fossil fuel combustion, which are known or suspected to induce cancer or other serious health and environmental effects. In February 2007, the US EPA issued a rule to reduce 1162 hazardous air pollutants from mobile sources, and identified 8 mobile source air toxics as key: benzene, 1,3-butadiene, formaldehyde, acetaldehyde, acrolein, polycyclic organic matter, naphthalene, and diesel exhaust (http://www3.epa.gov/otaq/toxics-regs.htm#02262007). Mobile source toxics that are frequency evaluated in human observational studies of the health effects associated with traffic-related air pollution include: benzene, diesel exhaust and polycyclic aromatic hydrocarbons (PAHs). Traffic exposure metrics used in this systematic review are outlined in **Table 1**.

Table 1. Traffic Exposure Metrics Used in This Systematic Review

Metric	Examples
Direct Measurements of Traffic[a]	
Distance or length	Distance to nearest main road or highway; distance to street canyons (streets lined with tall buildings on either side); length of main streets within buffer zone around homes or schools
Traffic density	Density on nearest road; average density estimated from road networks with buffer zones around homes or schools; street canyons with buffers
Traffic-Related Air Pollutants[b]	
Measurements[c]	Measurements of traffic-related air pollutants are included if the monitors or other sources of measurement used to estimate the pollutant-exposure surrogate could reasonably be traffic-related (e.g., roadway-specific monitoring or subjects lived within short distances of fixed monitors)
Modelling	Dispersion modelling of traffic-related air pollutants; other models used to estimate traffic-related air pollution (e.g., land-use regression model)

[a]Studies with self-reporting of direct measures of traffic will be included; however, they will be identified as Definitely High Risk of Bias (Appendix 3). [b]Carbon monoxide (CO), black carbon (BC), elemental carbon (EC), nitrogen oxides (NOx) and particulate matter (PM; specifically, PM2.5 (fine PM), ultrafine PM and coarse PM); and mobile source toxics: benzene, diesel exhaust, and polycyclic aromatic hydrocarbons (PAH). [c]Studies utilizing personal air monitoring and/or urine biomarkers will be used as long as some description of traffic is included in the methods.

~~Pregnancy-induced hypertension~~

REVISION: Hypertensive Disorders of Pregnancy

Recent research suggests that ambient air pollution, including traffic-related pollutants, is significantly associated with pregnancy-induced hypertension, which may influence pregnancy outcomes (Wu *et al.* 2011, Dadvand *et al.* 2014). An association between ambient air pollution and hypertension has been reported in the general population in the United States (Coogan *et al.* 2012), China (Guo *et al.* 2010) and Denmark (Sorensen *et al.* 2012). There have been several studies assessing the association between ambient air pollution and pregnancy-induced hypertension in the past 6 years, including two recent systematic reviews and meta-analyses (Hu *et al.* 2014, Pedersen *et al.* 2014). Only one of the systematic reviews included the full-range of exposure measures included in the OHAT evaluation (Pedersen *et al.* 2014). The OHAT review will be unique from the two recent systematic reviews because it will include consideration of experimental animal data and it will conduct a risk-of-bias evaluation of each study.

~~Pregnancy-associated hypertension~~ complicates approximately 10% of pregnancies worldwide (Roberts *et al.* 2005, Roberts *et al.* 2011). ~~Pregnancy-associated hypertension~~ has its onset at ≥ 20 weeks of gestation and ranges from hypertension alone (gestational hypertension) through proteinuria (i.e., protein in the urine) with the possibility of multi-organ dysfunction (preeclampsia) to seizures (eclampsia). Preeclampsia may also develop in women with pre-existing chronic hypertension. Preeclampsia is associated with multiple maternal complications, which can lead to death, including: edema, intravascular coagulation, renal and liver dysfunction, stroke, and placenta abruption (rupture) or infarction (stress) (Duley 2009). Maternal hypertension has a direct relationship to fetal and infant health as preeclampsia has been highly associated with intrauterine growth restriction, perinatal and neonatal mortality, premature birth, and associated prematurity-related neonatal diseases.

REVISION: Hypertensive disorders of pregnancy;

REVISION: Hypertensive disorders of pregnancy

Neurological Development[1]

Research on the association between air pollution and neurological development in children has expanded dramatically in the past 5 years. There is evidence that traffic-related air pollutants can enter the circulation and reach the brain (ultrafine PM; reviewed in (Genc *et al.* 2012). Traffic-related air pollutants such as PAHs, NO2 and PM2.5 have been associated with decreased intelligence quotient, language development and motor function in children (Kicinski *et al.* 2015, Sunyer *et al.* 2015). A number of studies have also evaluated the association between air pollutants commonly emitted from traffic and neurological disorders, such as attention deficit hyperactivity disorder and autism (Kalkbrenner *et al.* 2015, Raz *et al.* 2015). The body of literature on behavior effects of air pollutants in animals has also grown, especially for exposures such as diesel exhaust (Thirtamara Rajamani *et al.* 2013, Yokota *et al.* 2013), CO (Cheng *et al.* 2012), and PAHs (Li *et al.* 2012). It has been hypothesized that many of the neurobehavioral outcomes may be related the induction of oxidative stress, inflammation and modification of neurotransmitters and other biochemical signals by air pollutants during development. The current evaluation will focus on developmental exposure to traffic-related air pollution exposure associated with and neurobehavioral outcomes (*e.g.*, intelligence quotient, language,

[1] **REVISION**: NTP will not pursue a systematic review on TRAP and neurological development.

motor skills) as well as neurological disorders, such as attention deficit/hyperactivity disorder and autism.

Significance

~~The two proposed OHAT evaluations will build upon and extend the literature synthesis efforts of HEI, EPA, CDC, and others to understand the effects of traffic-related air pollution on 1) pregnancy-associated hypertensive disorders and 2) neurological development of children~~. In particular, the OHAT systematic reviews will: (1) focus on health outcomes that have not been extensively reviewed; (2) consider a wide range of exposure measures for traffic-related air pollution; (3) incorporate evidence from experimental animal studies; and (4) use the OHAT method on systematic review for reaching conclusions on confidence in the evidence identified from the systematic review (http://ntp.niehs.nih.gov/go/38673). Finally, data management will be conducted in a manner that permits public sharing of the data extracted from included studies.

REVISION: The proposed OHAT evaluation will build upon and extend the literature synthesis efforts of HEI, EPA, CDC, and others to understand the effects of traffic-related air pollution on hypertensive disorders of pregnancy.

OVERALL OBJECTIVE AND SPECIFIC AIMS

~~The overall objective of this evaluation is to develop hazard identification conclusions about whether traffic-related air pollution is associated with pregnancy-associated hypertensive disorders or neurological development in children by integrating levels of evidence from human and experimental animal studies~~.

REVISION: The overall objective of this evaluation is to develop hazard identification conclusions about whether traffic-related air pollution is associated with hypertensive disorders of pregnancy by integrating levels of evidence from human and experimental animal studies.

Specific aims:

- ~~Identify literature reporting the effects of traffic-related air pollution exposure characterized by direct traffic measurements (e.g., distance to main road, length of main streets with a buffer zone around homes or schools, traffic volume, etc.) and air pollutants associated with traffic emissions (e.g., CO, NOx, BC, EC, PM2.5, ultrafine PM, and coarse PM as well as benzene, diesel exhaust, and PAH (Table 1) on two health outcomes: 1) pregnancy-associated hypertension and associated disorders and 2) neurological development in humans and animals (experimental and wildlife)~~
- **REVISION: Identify literature reporting the effects of traffic-related air pollution exposure characterized by direct traffic measurements (e.g., distance to main road, length of main streets with a buffer zone around homes or schools, traffic volume, etc.) and air pollutants associated with traffic emissions (e.g., CO, NOx, BC, EC, PM2.5, ultrafine PM, and coarse PM as well as benzene, diesel exhaust, and PAH (Table 1) on hypertensive disorders of pregnancy in humans and animals (experimental and wildlife)**
- Extract data on potential health effects from relevant studies (data extraction files of the included studies will be shared upon release of final report)
- Assess the internal validity ("risk of bias") of individual human and animal studies using predefined criteria
- Synthesize the evidence using a narrative approach or meta-analysis (if appropriate) considering limitations on data integration (i.e., heterogeneity, sample size, etc.)
- Rate confidence in the body of evidence for human and animal studies, separately, according to one of four statements: (1) High, (2) Moderate, (3) Low, or (4) Very Low or No Evidence Available
- Translate confidence ratings into level of evidence of health effects for human and animal studies, separately, according to one of four statements: (1) High, (2) Moderate, (3) Low, or (4) Inadequate
- Use the level of evidence ratings for human and animal data to reach one of five possible hazard identification categories: (1) Known, (2) Presumed, (3) Suspected, (4) Not Classifiable, or (5) Not Identified To be a Hazard to Humans

To address our overall objective we developed a PECO statement (Population, Exposure(s), Comparator(s), and Outcome(s)) (Table 2, Table 3), which is used as an aid to develop the evaluation question, develop the search terms, and the inclusion/exclusion criteria for our systematic review (Higgins and Green 2011, AHRQ 2014).

Table 2. Population, Exposure, Comparator, and Outcome (PECO) Statement for ~~Pregnancy-associated Hypertension~~ REVISION: Hypertensive Disorders of Pregnancy	
PECO Element	**Evidence**
Population	Female humans or mammalian animals who were pregnant at exposure and outcome assessment
Exposure	Exposure to traffic-related air pollution, including direct traffic measurements (e.g., distance to main road, length of main streets with a buffer zone around homes or schools, traffic volume) and air pollutants associated with traffic emissions (*e.g.*, CO, NOx, BC, EC, PM2.5, ultrafine PM, and coarse PM) as well as mobile source toxics (*e.g.*, benzene, diesel exhaust, and PAH); air pollutants measured with modelling or exposure measurements that estimate the pollutant-exposure surrogate could reasonably be traffic-related (e.g., roadway-specific monitoring or subjects lived within short distances of fixed monitors) Inhalation route of exposure (human, animal), including nasal instillation (animal)
Comparator	A comparison population exposed to lower levels (or no exposure/exposure below detection levels) of traffic-related air pollution
Outcomes	Gestational hypertension, preeclampsia, eclampsia, HELLP (hemolysis, elevated liver enzymes, low platelet count) syndrome, blood pressure measurements

Table 3. Population, Exposure, Comparator, and Outcome (PECO) Statement for Neurological Outcomes[1]	
PECO Element	**Evidence**
Population	**Human:** Children (<18 years old) with exposure in utero or during childhood **Animal:** Neonatal, juvenile or adult mammals with exposure in utero or during childhood
Exposure	Exposure to traffic-related air pollution, including direct traffic measurements (e.g., distance to main road, length of main streets with a buffer zone around homes or schools, traffic volume) and air pollutants associated with traffic emissions (*e.g.*, CO, NOx, BC, EC, PM2.5, ultrafine PM, and coarse PM) as well as mobile source toxics (*e.g.*, benzene, diesel exhaust, and PAH); air pollutants measured with modelling or exposure measurements that estimate the pollutant-exposure surrogate could reasonably be traffic-related (e.g., roadway-specific monitoring or subjects lived within short distances of fixed monitors) Inhalation route of exposure (human, animal), including nasal instillation (animal)
Comparator	A comparison population exposed to lower levels (or no exposure/exposure below detection levels) of traffic-related air pollution
Outcomes	Any health effect related to neurological development (*e.g.*, learning, memory, motor skills) or neurocognitive or neuro-developmental dysfunction (*e.g.*,

[1] **REVISION**: NTP will not pursue a systematic review on TRAP and neurological development.

Table 3. Population, Exposure, Comparator, and Outcome (PECO) Statement for Neurological Outcomes[1]	
	attention deficit hyperactivity disorder, autism spectrum disorder (ASD))

The overall objective and PECO statements were based on a series of problem formulation steps that included (1) input from an evaluation team with expertise in air pollution, toxicology, epidemiology, systematic review, and information science; (2) deliberation with NTP staff and consultation with scientists at other Federal agencies; and (4) a public review of a concept document by the NTP Board of Scientific Counselors at the 16-18 April 2014 meeting (http://ntp.niehs.nih.gov/go/9741).

Key Questions and Analytical Framework

The overall objective of the evaluation can be phrased in terms of a specific research question "What is the hazard identification conclusion as to whether exposure to traffic-related air pollution is associated with adverse health effects?" This research question serves as a focus of the evaluation to be answered by addressing the key questions in **Table 2**.

Table 4. Key Questions (KQ): Assessed by Systematic Review	
Key Questions (KQ): Assessed by Systematic Review	
KQ1	What is the hazard identification category for an association between exposure to traffic-related air pollution and ~~pregnancy-associated hypertensive disorders~~ on based on integrating levels of evidence from human and experimental animal studies: 1) Known, (2) Presumed, (3) Suspected, (4) Not classifiable, or (5) Not identified to be a hazard to humans? **REVISION: hypertensive disorders of pregnancy**
KQ2[1]	~~What is the hazard identification category for an association between exposure to traffic-related air pollution in utero, during infancy or during childhood and neurological development based on integrating levels of evidence from human and experimental animal studies: 1) Known, (2) Presumed, (3) Suspected, (4) Not classifiable, or (5) Not identified to be a hazard to humans?~~

METHODS

Step 1. Problem Formulation

Problem Formulation Activities

Nomination History

OHAT received a nomination to evaluate emerging children's health issues associated with ambient air pollution (Federal Register 77 FR 41406, 13 July 2012; http://ntp.niehs.nih.gov/go/37853). The original nomination was broad, suggesting that OHAT consider a wide range of types of ambient air pollution with a focus on traffic-derived air pollutants and that OHAT avoid reviewing the effects of well-characterized exposures (e.g., tobacco smoke, mercury, lead, arsenic). The health effects nominated for evaluation focused on emerging children's health outcomes, including: adverse birth outcomes; incidence of asthma and allergic disease; respiratory infections in early life; compromised lung function,

[1] **REVISION**: NTP will not pursue a systematic review on TRAP and neurological development.

development and growth; neurological development; pediatric cancer; and developmental basis of adult disease (i.e., adult diseases associated with exposures in utero or during childhood)).

The NTP Executive Committee[2] was informed about the potential evaluation and solicited for input on agency interest/relevance and for names of agency technical staff that should be involved in the evaluation. An evaluation team was identified to include experts from the NTP, US EPA National Center for Environmental Assessment (NCEA), US EPA Office of Transportation and Air Quality (OTAQ), CDC, and non-federal technical advisors (see "About this Protocol, Contributors"). The initial concept proposal for the evaluation was reviewed and approved by NTP's Board of Scientific Counselors in a public meeting on April 18, 2014 (http://ntp.niehs.nih.gov/go/9741). One public comment was received related to the contribution of vehicle tire wear to levels of PM10 in the environment (http://ntp.niehs.nih.gov/go/41362).

Refining the focus of the nomination

To help refine the focus of the nomination, OHAT conducted a preliminary inventory of the literature using a search strategy designed to capture the types of air pollutants outlined in the original nomination and a variety of health outcomes relevant to children's health. The preliminary search results (data not shown) were used to identify emerging health areas of concern as well as types of air pollution that could be considered together in the systematic review (e.g., traffic-related versus industrial source). OHAT prioritized the list of emerging health outcomes to review based on the extent of the literature and evaluations of health effects conducted by other federal agencies.

OHAT decided to focus on the association of traffic-related air pollution with two health outcomes: 1) pregnancy-related hypertensive disorders and 2) neurological outcomes in children exposed *in utero* or during early postnatal life. Specifically, we decided not to pursue the respiratory health outcomes because they were more commonly studied and reviewed. The CDC recently published a review of traffic-related air pollution associated with pediatric cancer (Boothe et al. 2014), and they are conducting a literature review on adverse pregnancy outcomes (e.g., fetal growth and preterm delivery) (Tegan Boehmer, manuscript in preparation). The majority of the literature on ~~pregnancy-associated hypertensive disorders~~ has been published since 2009, and thus this health effect was not included in the 2010 HEI Special Report 17 on the Health Effects of Traffic-Related Air Pollution (HEI 2010) or the EPA's Integrated Science Assessments of CO (US EPA 2010), NOx (US EPA 2008) or particulate matter (US EPA 2009). Likewise, research on neurological disorders and neurological outcomes of children exposed to traffic-related air pollution has resulted in numerous publications since the extensive review of traffic-related pollution by HEI (2010) and the related EPA Integrated Science Assessments (US EPA 2008, 2009, 2010). **However, because traffic-related air pollution and neurological disorders is currently under review by other organizations, NTP will not pursue this topic.**

REVISION: hypertensive disorders of pregnancy

[2]The NTP Executive Committee provides programmatic and policy oversight to the NTP Director and meets once or twice a year in closed forum. Members of this committee include the heads (or their designees) from the following federal agencies: Consumer Product Safety Commission (CPSC), Department of Defense (DoD), Environmental Protection Agency (EPA), Food and Drug Administration (FDA), National Cancer Institute (NCI), National Center for Environmental Health/Agency for Toxic Substances and Disease Registry (NCEH/ATSDR), National Institute of Environmental Health Sciences (NIEHS), National Institute for Occupational Safety and Health (NIOSH), Occupational Safety and Health Administration (OSHA).

Consideration of key scientific issues

Ability to discriminate air pollution derived from traffic versus other sources

The air pollutants identified for inclusion have been commonly used as exposure surrogates for traffic-related air pollution in other evaluations. Many of the exposure surrogates used to identify traffic-related pollution are also generated by other sources (e.g., industrial pollution, home heating fuel) and this will be acknowledged as a limitation in our ability to attribute associations exclusively to traffic sources.

Traffic-related noise as co-exposure for traffic-related air pollution

Traffic noise is a likely co-exposure for all traffic-related air pollution studies. However, relatively few studies have tried to disentangle the possible effects of traffic noise versus traffic-related pollution (e.g., (van Kempen *et al.* 2012)). Correlation between the two exposures is modest and is not always co-linear (Fecht *et al.* 2016). Therefore, it is unclear whether, or to what extent, that traffic-related air pollution is confounded by road traffic noise.

Heterogeneity in methods used to assess exposure

The major challenge in evaluating the association of traffic-related air pollution and selected health effects will be determining the extent to which conclusions can be integrated across the variety of different exposure models for a given air pollutant. The majority of traffic-related air pollutants have spatial variability (Karner *et al.* 2010) and will be most appropriately assessed in studies that use exposure contrasts in models based on space. Pollutants with variability dominated by temporal variability (e.g., PM2.5) are most accurately represented in time-series epidemiology studies. For example, a time-series compares temporal exposures (e.g., between days), while a cohort or a case-control study measures the exposure between different locations (e.g. spatial), and other studies may be a combination of temporal and spatial.

Heterogeneity in exposure measures will be addressed by conducting a series of stratified analyses that take into consideration the exposure contrast evaluated in the epidemiology studies and the exposure model. For studies where exposure metrics are similarly assessed, we will visually display data in a manner that assesses exposure and its associated dose-response. When studies cannot be directly compared, we will group studies based on relative extent of exposure (low, medium, or high). This grouping would occur after data extraction and would be informed by input from evaluation design team members, technical advisors with topic specific expertise, and the framework to categorize traffic-related air pollutants used in the 2010 HEI report.

Changes in composition of traffic-related air pollution over time or country

Another factor to consider are changes in the composition of traffic-related air pollution in the United States since 1975, when the phase-out of lead began, and during the 2000s when use of ethanol blends became more common. The changes related to use of lead in gas will likely not have a significant impact on the evaluations of ~~pregnancy-associated hypertension~~ ~~or neurological development~~ since the majority of the studies to be evaluated will use exposure data collected during the late 1990s or more recently. Similarly, the synthesis of the international literature on traffic-related air pollution with studies in the United States is complicated by differences in the composition of the vehicle fleet; for example, diesel-fueled vehicles make up a larger portion of the vehicle fleet in Europe than in the United

States. To address these possible influences on the composition of traffic-related air pollution we will conduct sub-analyses of studies grouped by time or locale, as possible.

REVISION: hypertensive disorders of pregnancy

Step 2. Search For and Select Studies for Inclusion

Literature Search Strategy

Literature search strategies were developed to identify all relevant published evidence on the traffic-related air pollution associated with ~~pregnancy-associated hypertension~~ ~~or neurological development~~ through (1) reviewing PubMed's Medical Subject Headings (MeSH) for relevant and appropriate terms, (2) extracting key terminology from relevant reviews and a set of previously identified primary data studies that are known to be relevant to the topic ("test set"), and (3) reviewing search strategies presented in other reviews. The search strategy was run and the results are assessed to ensure that 100% of the previously identified relevant primary studies were retrieved. Three databases will be searched from the beginning of the database entries through November 13, 2014 for ~~pregnancy-associated hypertensive disorders~~ ~~or March 18, 2015 for neurological development~~. The search strategy was customized for each database because of differences in syntax (see Appendix 1 ~~Pregnancy-associated Hypertensive Disorders~~ and Appendix 2 Neurological Development[1]). No publication year limits or language restriction will be imposed. The literature search will be updated for a final time 6 weeks prior to release of the final draft. We developed the literature search in collaboration with a librarian trained in systematic review methodology.

REVISION: Hypertensive disorders of pregnancy

REVISION: Hypertensive disorders of pregnancy

REVISION: Hypertensive disorders of pregnancy

Databases Searched

- PubMed
- Scopus
- Web of Science

Searching Other Resources

We will use the following methods to find additional studies that were not identified through the electronic searches:

- Hand searching the reference lists of all included studies after the full text review.
- Hand searching the reference lists of relevant reviews (e.g., HEI (2010), US EPA Integrated Science Assessments (US EPA 2008, 2009, 2010), commentaries, or other non-research articles identified during the initial search. Commentaries or letters on specific studies are also reviewed to see if they contain content that should be noted during data extraction or risk-of-bias assessment of the original report.

[1] **REVISION**: NTP will not pursue a systematic review on TRAP and neurological development.

- Grey literature: To ensure retrieval of the relevant literature, OHAT may try to identify relevant grey literature, which refers to publications that are not commercially published or are not readily publicly available.
- Studies identified by the public when the initial list of included studies is posted on the OHAT website (anticipated for 60-90 days prior to peer review; studies identified within 30 days of posting will be considered for inclusion) or during the public comment period when the draft Monograph is released for public comment (45-60 days prior to peer review).

Studies will be evaluated using the same inclusion and exclusion criteria as used for screening records retrieved from the electronic search. Relevant studies identified through these steps will be marked as "provided from other sources" in the study selection flow diagram.

Unpublished data

NTP only includes publicly accessible and peer-reviewed information in its evaluations. If a study is identified which may be critical to the evaluation and is not peer reviewed, the NTP's practice is to obtain external peer review if the owners of the data are willing to have the study details and results made publicly accessible. The peer review would include an evaluation of the study similar to that for peer review of a journal publication. The NTP would identify and select 2-3 scientists knowledgeable in scientific disciplines relevant for the topic as potential peer reviewers. Persons invited to serve as peer reviewers would be screened for conflict of interest (COI) prior to confirming their service. In most instances, the peer review would be conducted by letter review. The study authors would be informed of the outcome of the peer review and given an opportunity to clarify issues or provide missing details. OHAT would consider the peer-review comments regarding the scientific and technical evaluation of the unpublished study in determining whether to include the study in its evaluation. The study and its related information, if used in the OHAT evaluation, would be publicly available. OHAT would acknowledge via a note for the report that the document underwent external peer review managed by the NTP and the names of the peer reviewers would be identified. Unpublished data from personal author communication can supplement a peer-reviewed study, as long as the information is made publicly available.

Screening Process

DistillerSR®, a web-based, systematic review software program with structured forms and procedures will be used to screen articles for relevance and eligibility to ensure standardization of process[3]. Initially, results of the literature search are assembled in EndNote software and exact article duplicates removed prior to uploading the references into the systematic review software program.

Evidence Selection Criteria

In order to be eligible for inclusion, studies must comply with the criteria specified by the PECO statement (**Table 1**). Studies that do not meet the PECO statement will be excluded. Some articles may be categorized as possible supportive material if they appear inappropriate for inclusion, but appear to contain relevant background information. Those studies would not provide evidence of health effects, or lack of a health effect; however, the background information could provide context or other information

[3]DistillerSR® (http://systematic-review.net/) is a proprietary project management tool for tracking studies through the screening process and storing data extracted from these studies using user-customized forms.

(e.g., exposure or metabolism data) that would be useful when evaluating confidence in bodies of evidence or integrating evidence across human and animal data from the included studies.

Inclusion and exclusion criteria used to screen articles for relevance and eligibility at both the title-and-abstract and full-text screening stages are detailed in **Table 5** (~~pregnancy-associated hypertensive disorders~~) and **Table 6** (neurological development[1]).

REVISION: hypertensive disorders of pregnancy

[1] **REVISION**: NTP will not pursue a systematic review on TRAP and neurological development.

Table 5. Criteria to Determine Study Eligibility for ~~Pregnancy-Associated Hypertensive Disorders~~ REVISION: Hypertensive Disorders of Pregnancy	
Inclusion Criteria	**Exclusion Criteria**
Participants/Population (human studies or experimental model systems)	
• Female humans or non-human animals who were pregnant at exposure and outcome assessment	• Non-pregnant subjects • Other than non-human mammals (e.g., fish, invertebrates, frogs, birds)
Exposure	
• Exposure to direct traffic measurements (*e.g.*, distance to main road, length of main streets with a buffer zone around homes or schools, traffic volume) • Exposure to major air pollutants[a]: CO, NO2, black carbon, elemental carbon, ultrafine PM, coarse PM, and PM2.5 • Exposure to mobile source air toxics[a]: benzene, diesel exhaust, PAH (including urine or blood biomarkers for exposure) • Inhalation exposure (humans and animals) or aspiration route of exposure (animals)	• CO poisoning • Non-inhalation route exposure to relevant air pollutants (*e.g.*, oral or dermal exposure to crude oil spills, gasoline ingestion) • Metals, heavy (*e.g.*, arsenic, lead, mercury) or transitional (*e.g.*, platinum, manganese) • Formaldehyde or other volatile organic compounds relevant to vehicle exhaust
Comparators	
• Humans exposed to lower levels (or no exposure/exposure below detection levels) of relevant air pollutants • **For experimental model studies**: study must include vehicle or untreated control group • **For wildlife or observational studies**: animals exposed to lower levels (or no exposure /exposure below detection levels) of relevant air pollutants	• None
Outcomes	
• Gestational hypertension, preeclampsia, or eclampsia • Blood pressure measurements during pregnancy, which may not meet the definition of gestational hypertension • Comparable mammalian non-human animal model for gestational hypertension, preeclampsia or eclampsia • **REVISION: Chronic hypertension that persists through pregnancy**	• ~~Chronic hypertension prior to pregnancy under evaluation~~ • **REVISION: None**
Publications (e.g., Language Restrictions, Use of Conference Abstracts)	
• Report must contain original data	• Articles with no original data, e.g., editorials, reviews[b] • Conference abstracts or other studies published in abstract form only, grant awards, and theses/dissertations • Retracted articles
[a] Air pollutants must be measured with modelling (e.g., dispersion of traffic exposure or land-use regression models, *etc*.) or with exposure measurements that estimate the pollutant-exposure surrogate that could reasonably be traffic-related (e.g., roadway-specific monitoring or subjects lived within short distances of fixed monitors). [b] Relevant reviews can be used as background and for reference scanning.	

Table 6. Criteria to Determine Study Eligibility for Neurological Development[1]	
Inclusion	**Exclusion**
Participants/population (human studies or experimental model systems)	
• Humans or animals (experimental models or wildlife) restricted to exposure *in utero,* during infancy and/or during childhood, and not restricted by sex or lifestage at assessment	• Humans or animals with exposure assessed only in adulthood • Other than non-human mammals (e.g., fish, frogs, birds)
Exposure	
• Exposure to direct traffic measurements (*e.g.*, distance to main road, length of main streets with a buffer zone around homes or schools, traffic volume) • Exposure to major air pollutants[a]: CO, NO2, black carbon, elemental carbon, ultrafine PM, coarse PM, PM2.5 • Exposure to mobile source air toxics[a]: benzene, diesel exhaust, PAH (including urine or blood biomarkers for exposure) • Inhalation exposure (humans and animals) or aspiration route of exposure (animals)	• CO poisoning • Non-inhalation route exposure to relevant air pollutants (*e.g.*, oral or dermal exposure to crude oil spills, gasoline ingestion) • Metals, heavy (*e.g.*, arsenic, lead, mercury) or transitional (*e.g.*, platinum, manganese) • Formaldehyde or other volatile organic compounds relevant to vehicle exhaust
Comparators	
• Humans exposed to lower levels (or no exposure/exposure below detection levels) of relevant air pollutants • **For experimental animal studies**: study must include vehicle or untreated control group • **For wildlife studies**: animals exposed to lower levels (or no exposure /exposure below detection levels) of relevant air pollutants	• None
Outcomes	
• Any health effect related to neurological development (*e.g.*, learning, memory, motor, *etc.*) and neurobehavioral dysfunction (*e.g.*, attention deficit disorder, autism spectrum disorder)	• Sexual behaviors (*e.g.*, sexual receptivity)
Publications (e.g., Language Restrictions, Use of Conference Abstracts)	
• Report must contain original data	• Articles with no original data (e.g., editorials, reviews[b]) • Conference abstracts or other studies published in abstract form only, grant awards, and theses/dissertations • Retracted articles
[a] Air pollutants must be measured with modelling (e.g., dispersion of traffic exposure or land-use regression models) or with exposure measurements that estimate the pollutant-exposure surrogate that could reasonably be traffic-related (e.g., roadway-specific monitoring or subjects lived within short distances of fixed monitors). [b] Relevant reviews can be used as background and for reference scanning.	

[1] **REVISION**: NTP will not pursue a systematic review on TRAP and neurological development.

Title/Abstract Review

Screeners will be trained using project-specific written instructions that reflect the criteria outlined in (**Table 5**; **Table 6**) with an initial pilot phase undertaken to improve clarity of the inclusion and exclusion instructions and to improve accuracy and consistency among screeners. If changes to the inclusion criteria are made based on the pilot phase, they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes. Trained screeners from the evaluation design team will then conduct a title and abstract screen of the search results to determine whether a reference meets the inclusion or exclusion criteria. All references will be independently screened by two screeners (one of which will be the project lead, who will screen all references). Studies that are not excluded based on the title and abstract will be screened through a full-text review. In case of screening conflicts, screeners will independently review their screening results to confirm the inclusion/exclusion decision and, if needed, discuss discrepancies with the other screeners. If a true disagreement exists between screeners, the study passes to the full-text review.

Full-Text Review

After completion of the title/abstract screen, full-text articles are retrieved[4] for those studies that either clearly meet the inclusion criteria or where eligibility to meet the inclusion criteria is unclear. Full-text review will be independently conducted by two screeners that participated in the title/abstract screening (again, one of which will be the project lead, who will screen all references). True disagreements will be resolved by discussion involving another member(s) of the team or, if necessary, through consultation with technical advisors.

Multiple publications of same data

Multiple publications with overlapping data for the same study (e.g., publications reporting subgroups, additional outcomes or exposures outside the scope of an evaluation, or longer follow-up) are identified by examining author affiliations, study designs, cohort name, enrollment criteria, and enrollment dates. If necessary, study authors will be contacted to clarify any uncertainty about the independence of two or more articles. OHAT will include all publications on the study, select one study to use as the primary, and consider the others as secondary publications with annotation as being related to the primary record during data extraction. The primary study will generally be the publication with the longest follow-up, or for studies with equivalent follow-up periods, the study with the largest number of cases or the most recent publication date. OHAT will include relevant data from all publications of the study, although if the same outcome is reported in more than one report, OHAT will exclude the duplicate data.

Tracking study eligibility and reporting the flow of information

The reason for exclusion at the full-text-review stage will be annotated and reported in a study flow diagram in the final report**.** Commonly used categories for exclusion include: (1) is a review, commentary, or editorial with no original data; (2) lacks relevant exposure information; (3) lacks relevant health outcome information; and (4) is a conference abstract, thesis/dissertation.

[4]OHAT will initially attempt to retrieve a full-text copy of the study using an automated program, such as QUOSA, when possible, and NIH library services (NIH subscriptions and interlibrary loans). For publications not available through NIH, OHAT will search the Internet and/or may attempt to contact the corresponding author. Studies not retrieved through these mechanisms are excluded and notated as "not available."

Release of the list of included and excluded studies

The list of included and excluded studies will be posted on the OHAT website (http://ntp.niehs.nih.gov/go/evals) once screening has been completed and prior to completion of the draft OHAT monograph.

Step 3. Data Extraction

Data Extraction Process and Data Warehousing

Data extraction will be managed with structured forms and stored in a database format using ICF International's proprietary Dose Response Analytical Generator and Organizational Network (DRAGON) software.[5]. Data extraction elements are listed separately for human and animal studies in Appendix 3. Study information collected during data extraction will be visualized and made publicly available in Excel format upon publication of the finalized report using Health Assessment Workspace Collaborative (HAWC), an open source, web-based interface.[6]

The extracted data will be used to help summarize study designs and findings, facilitate assessment of risk of bias and/or conduct statistical analyses during evidence synthesis. The content of the data extraction may be revised following the identification of the studies included in the review. Data extraction will be performed by one member of the evaluation team and checked by a second member for completeness and accuracy. Data extractors from the evaluation team will be trained using project-specific written instructions in an initial pilot phase using a subset of studies. Any discrepancies in data extraction will be resolved by discussion or consultation with a third member of the evaluation team. Information that is inferred, converted, or estimated during data extraction will be annotated and marked with brackets.

OHAT will attempt to contact authors of included studies to obtain missing data considered important for evaluating key study findings (e.g., level of data required to conduct a meta-analysis). The evaluation report will note that an attempt to contact study authors was unsuccessful if study researchers do not respond to an email or phone request within one month of the attempt to contact.

Step 4. Quality Assessment of Individual Studies

Internal validity, or risk of bias, will be assessed for individual human and animal studies using a tool developed by OHAT that takes a parallel approach to evaluating risk of bias from human and animal to facilitate consideration of risk of bias across evidence streams with common terms and categories. Instructions for the risk-of-bias evaluation are provided in a guidance document tailored to the specific evidence stream and type of human study design in the detailed guide for using the tool (see "Risk-of-Bias Tool" at http://ntp.niehs.nih.gov/go/38673). The risk-of-bias tool is comprised of a common set of 11 questions that are answered based on the specific details of individual studies to develop risk-of-bias ratings (using one of four response options in **Table 7** for each question). Study design determines the subset of questions that should be used to assess risk of bias for an individual study (**Table 8**). For example, the subset of risk-of-bias questions applicable to all of the experimental study designs includes

[5] DRAGON (Dose Response Analytical Generator and Organizational Network) developed by ICF International (Fairfax, VA; http://www.icfi.com/insights/products-and-tools/dragon-dose-response).

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

a question on randomization of exposure that would not be applicable to observational study designs. Therefore, a similar set of questions are used across experimental study designs (experimental animal and human controlled trials).

Studies are independently assessed by two assessors who answer all applicable risk-of-bias questions with one of four options in **Table 8** (answers from (CLARITY 2013)) following pre-specified criteria detailed in Appendix 4. The criteria describe aspects of study design, conduct, and reporting required to reach risk-of-bias ratings for each question and specify factors that can distinguish among ratings (e.g., what separates "definitely low" from "probably low" risk of bias). The instructions and detailed criteria are tailored to the specific evidence stream and type of human study designs. Risk of bias will be assessed at the outcome level because study design or method specifics may increase the risk of bias for some outcomes and not others within the same study.

Table 7: Answers to the Risk-of-Bias Questions Result in One of Four Risk-of-Bias Ratings	
++	***Definitely Low*** *risk of bias*: There is direct evidence of low risk-of-bias practices
+	***Probably Low*** *risk of bias*: There is indirect evidence of low risk-of-bias practices **OR** it is deemed that deviations from low risk-of-bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias
– NR	***Probably High*** *risk of bias*: There is indirect evidence of high risk-of-bias practices (indicated with "-") **OR** there is insufficient information provided about relevant risk-of-bias practices (indicated with "NR" for not reported). Both symbols indicate probably low risk of bias.
––	***Definitely High*** *risk of bias*: There is direct evidence of high risk-of-bias practices

Table 8: OHAT Risk-of-Bias Questions and Applicability by Study Design

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

[a]Experimental animal studies are controlled exposure studies. Non-human animal observational studies can be evaluated using the design.
[b]Human Controlled Trials are studies in humans with controlled exposure (e.g., Randomized Controlled Trials, non-randomized experimental.
[c]Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).studies) features of observational human studies such as cross-sectional study design.

Risk-of-Bias Assessment Process

Assessors will be trained using the criteria in Appendix 4 with an initial pilot testing phase undertaken to improve clarify of criteria that distinguish between adjacent ratings and to improve consistency among assessors. All team members involved in the assessment will be trained on the same set of studies and asked to identify potential ambiguities in the criteria used to assign ratings for each question. Any ambiguities and rating conflicts will be discussed relative to opportunities to refine the criteria to more clearly distinguish between adjacent ratings. If major changes to the risk-of-bias criteria are made based on the pilot phase (i.e., those that would likely result in revision of response), they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes. It is also expected that information about confounding, exposure characterization, outcome assessment, and other important issues may be identified during or after data extraction, which can lead to further refinement of the risk-of-bias criteria (Sterne *et al.* 2014).

After assessors have independently made risk-of-bias determinations for a study across all risk-of-bias questions, the two assessors will compare their results to identify discrepancies and attempt to resolve them. Any remaining discrepancies will be assessed by the project lead and, if needed, other members of the evaluation design team and/or technical advisors. The final risk-of-bias rating for each question will be recorded along with a statement of the basis for that rating. The risk-of-bias assessment of included studies will be part of the study summaries released in materials for the draft OHAT monograph that will be posted for public comment prior to peer review (anticipated for 45-60 days prior to peer review). Peer review will provide an opportunity for investigators and the public to comment on risk-of-bias.

Missing Information for Risk-of-Bias Assessment

OHAT will attempt to contact authors of included studies by email to obtain missing information considered critical for evaluating risk of bias that cannot be inferred from the study. If additional information or data are received from study authors, risk-of-bias judgments will be modified to reflect the updated study information. If OHAT does not receive a response from the authors by one month of the contact attempt, a risk-of-bias response of "NR" for "not reported; probably high risk of bias" will be used and a note made in the data extraction files that an attempt to contact the authors was unsuccessful.

Exposure Assessment Factors for Risk-of-Bias Assessment

The quality of exposure assessment for individual studies will be addressed during Step 4 assessment of internal validity/risk of bias under question #8 "Can we be confident in the exposure characterization?" Risk-of-bias criteria address issues from purity and stability of compounds in experimental animal studies to exposure variability and misclassification in human studies. In addition, the risk-of-bias assessment will consider the quality of the input data for all exposure models, including spatial and temporal variability, degree of geographic resolution, and consideration of temperature, and meteorology (e.g., precipitation and humidity). Finally, risk-of-bias assessment will consider whether modeling results or other indirect measures have been compared with measurements (a) for the pollutants of interest, (b) for the average time of interest, and (c) for the locations of interest. See Appendix 4, Question 8 for specific risk-of-bias criteria.

Step 5. Organizing and Rating Confidence in the Bodies of Evidence

~~OHAT will consider the collection of studies on the same or closely related pregnancy-associated hypertensive disorders or neurological development, respectively, as bodies of evidence and develop overall confidence ratings in these bodies of evidence using a modification of the GRADE framework~~. OHAT will also consider the availability of information to determine whether peak or average exposure levels are the most important for each health outcome of interest. Procedures for grouping the ~~pregnancy-associated hypertensive disorders and neurological development, respectively~~, considering quantitative or narrative synthesis and developing confidence ratings for this evaluation are described below.

REVISION: OHAT will consider the collection of studies on the components of TRAP as bodies of evidence and develop confidence ratings in these bodies of evidence using a modification of the GRADE framework.

REVISION: hypertensive disorders of pregnancy

Health Outcome and Endpoint Grouping

~~Separate reports will be prepared for the evaluations of (1) pregnancy-related hypertensive disorders and (2) neurological development~~. ~~Pregnancy-associated hypertensive disorders~~ may be grouped in 4 categories based on severity of health condition (**Table 9**). ~~In the evaluation of neurological development, endpoints will be broadly categorized by neurological disorder (human) or neurological domain (human and animal), and will be evaluated as a group under each broad heading. Neurological domains, as well as examples of endpoints, are detailed in **Table 10** and **Table 11**~~.

REVISION: Hypertensive disorders of pregnancy

Table 9: ~~Pregnancy-associated Hypertensive Disorders~~ Outcome Grouping for Human and Animal Studies **REVISION: Hypertensive Disorders of Pregnancy**	
Endpoint	**Description**
Blood pressure change	An increase or decrease related to the reference pregnant population (human or animal)
Gestational hypertension	Systolic blood pressure (SBP) ≥ 140 mmHg or diastolic blood pressure (DBP) ≥ 90 mmHg during the second half of pregnancy
Preeclampsia	Gestational hypertension accompanied by proteinuria after 20 weeks of gestation
Eclampsia and HELLP	Eclampsia: Gestational hypertension, with or without proteinuria, plus at least one observed seizure in a woman with no prior history of a seizure disorder Hemolysis Elevated Liver Enzyme Low Platelet count (HELLP) syndrome: hemolysis (abnormal peripheral smear, bilirubin >1.2 mg/dl, or lactose dehydrogenase >600 IU/L), elevated liver enzymes (aspartate amino transferase or alanine aminotransferase >70 IU/L) and low platelets (<100,000 mm^3)

Table 10: Neurological Developmental Outcome and Neurological Disorder Grouping for Human Studies[1]	
General Domain	**Example test and endpoints**
Attention Deficit/Hyperactive Disorder	Criteria prescribed in the Diagnostic and Statistical Manual of Conduct Disorders of American Pediatric Association, Parent Rating Scale
Autism Spectrum Disorder	Children with developmental records documenting characteristics and behaviors that met a standardized definition for autism spectrum disorders, including medical diagnosis and parental response to questionnaire
Mental development	Bayley Scales of Infant Development (*e.g.*, mental and psychomotor development), Gesell Developmental Schedules (*e.g.*, adaptive, social, language), Wechsler Preschool and Primary Scale of Intelligence (*e.g.*, verbal, language), Wide Range Assessment of Memory and Learning (*e.g.*, memory, learning), Kaufman Brief Intelligence Test (*e.g.*, vocabulary), McCarthy Scales of Children's Abilities (*e.g.*, verbal, quantitative, memory)
Motor/Sensory Development	Bayley Scales of Infant Development (*e.g.*, mental and psychomotor development), McCarthy Scales of Children's Abilities (*e.g.*, gross and fine motor skills)
Anxiety/Emotional	Child Behavior Checklist (*e.g.*, anxious/depressed, attention problems, anxiety problems)
Other	Adjudication data (e.g., delinquent behavior), sleep duration in infants

[1] **REVISION**: NTP will not pursue a systematic review on TRAP and neurological development.

Table 11: Neurological Outcome Grouping for Experimental Animal Studies[1]	
General Domain	**Example test and endpoints**
Memory	Morris water maze (probe phase, path length), fear conditioning (freezing, active avoidance conditioning test, passive avoidance test), delays (early, short, long – term memory)
Learning	Morris water maze (latency – acquisition phase, total time), operant battery test (incremental repeated acquisition, conditioned position responding, accuracy, percent complete, response rate), T-maze, Y-maze, novel object recognition
Motor/Sensory Development	open field (spontaneous activity, distance, locomotion mean), Morris water maze (swim speed), inclined board test, rota-rod, geotaxis, acoustic startle reflex
Anxiety/Emotional	sociability test, elevated plus maze, open field (rearing), forced swimming test, avoidance learning, tail suspension test (depression)
Other	vocalization, territorial aggressive (resident-intruder test)

Considerations for pursuing a narrative or quantitative evidence synthesis

Heterogeneity within the available evidence will determine the type of evidence integration that is appropriate: either a quantitative synthesis (meta-analysis) or narrative approach for evidence integration. When appropriate, we will perform a meta-analysis. Summaries of main characteristics for each included study will be compiled and reviewed by two evaluation team members to determine comparability between studies, identify data transformations necessary to ensure comparability, and determine whether heterogeneity is a concern. The main characteristics evaluated across all eligible studies include the following:

Human Studies

- Study design (e.g., cross-sectional, cohort)
- Details on how participants were classified into exposure groups, if any (e.g., quartiles of exposure concentration)
- Details on source of exposure data (e.g., questionnaire, area monitoring, biomonitoring)
- Concentrations of the traffic-related air pollutants or measurements of the direct traffic measures for each exposure group
- Health outcome(s) reported
- Conditioning variables in the analysis (e.g., variables considered confounders)
- Type of data (e.g., continuous or dichotomous), statistics presented in paper, ability to access raw data
- Variation in degree of risk of bias at individual study level

Animal Studies

- Experimental design (randomized or not, acute or chronic, multigenerational, etc.)
- Animal model used (species, strain, sex, and genetic background)
- Age of animals (at start of treatment, mating, and/or pregnancy status)
- Developmental stage of animals at treatment and outcome assessment

[1] **REVISION**: NTP will not pursue a systematic review on TRAP and neurological development.

- Dose levels, frequency of treatment, timing, duration, and exposure route
- Health outcome(s) reported
- Type of data (e.g., continuous or dichotomous), statistics presented in paper, ability to access raw data
- Variation in degree of risk of bias at individual study level

More detailed guidance on evaluating heterogeneity, transforming or normalizing data to ensure comparability, and the process for determining whether a meta-analysis will be pursued is provided in the OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673, see STEP 5). We expect to require input from topic-specific experts to help assess whether studies are too heterogeneous for meta-analysis to be appropriate. Situations where it may not be appropriate to include a study are (1) data on exposure or outcome are too different to be combined, (2) there are concerns about high risk of bias, or (3) other circumstances may indicate that averaging study results would not produce meaningful results. When it is inappropriate or not feasible to quantitatively combine results, OHAT will narratively describe or visually present findings.

Stratified Analyses, Meta-Regression, and Publication Bias

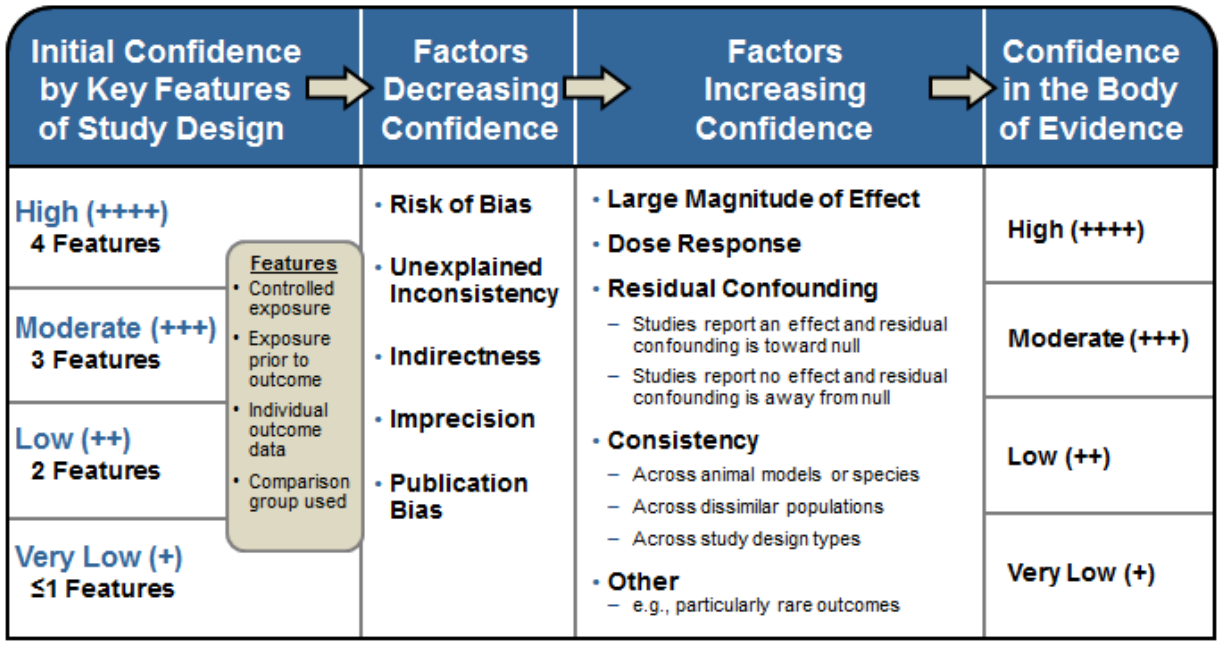
If there is significant study-level heterogeneity, then OHAT may conduct stratified analyses or multivariate meta-regression in an attempt to determine how much heterogeneity can be explained by taking into account both within- and between-study variance (Vesterinen *et al.* 2014). Multivariate meta-regression approaches are especially useful for assessing the significance of associations between study design characteristics. These approaches are considered most suitable if there are at least six to ten studies for a continuous variable and at least four studies for a categorical variable (Fu *et al.* 2011).

If possible, i.e., if there are enough studies; we will assess potential publication bias by developing funnels and performing Egger regression on the estimates of effect size. In addition, if these methods suggest that publication bias is present, we will use trim and fill methods to predict the impact of the hypothetical "missing" studies (Vesterinen *et al.* 2014).

Confidence Rating: Assessment of Body of Evidence

The quality of evidence for each outcome will be graded using the GRADE system for rating the confidence in the body of evidence (Guyatt *et al.* 2011) as adapted by OHAT for observational human and animal studies (Rooney *et al.* 2014) (**Figure 1**). More detailed guidance on reaching confidence ratings in the body of evidence as "high", "moderate", "low" or "very low" is provided in the OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673, see Step 5). In brief, available studies on a particular outcome are initially grouped by key study design features, and each grouping of studies is given an initial confidence rating by those features. This initial rating (column 1 of **Figure 1**) is downgraded for factors that decrease confidence in the results (column 2 of **Figure 1** [risk of bias, unexplained inconsistency, indirectness or lack of applicability, imprecision, and publication bias]) and upgraded for factors that increase confidence in the results (column 3 of **Figure 1** [large magnitude of effect, dose response, consistency across study designs/populations/animal models or species, consideration of residual confounding, and other factors that increase our confidence in the association or effect]).

Figure 1. Assessing Confidence in the Body of Evidence



The reasons for downgrading (or upgrading) confidence may not be due to a single domain of the body of evidence. If a decision to downgrade is borderline for two domains, the body of evidence is downgraded once in a single domain to account for both partial concerns based on considering the key drivers of the strengths or weaknesses. Similarly, the body of evidence is not downgraded twice for what is essentially the same limitation (or upgraded twice for the same asset) that could be considered applicable to more than one domain of the body of evidence. Confidence ratings are independently assessed by members of the evaluation review team, and discrepancies are resolved by consensus and consultation with technical advisors as needed. Confidence ratings are summarized in evidence profile tables.

Relevance of Animal Model to Human Health

- *Rats, mice, and other mammalian model systems:* No limitations of model systems for mammals have been identified *a priori*. Thus, studies conducted in mammalian model systems will be assumed to be relevant for humans (i.e., not downgraded for indirectness) unless compelling data to the contrary is identified during the course of the evaluation.

Relevance of Other than Inhalation Exposure in Animal Studies

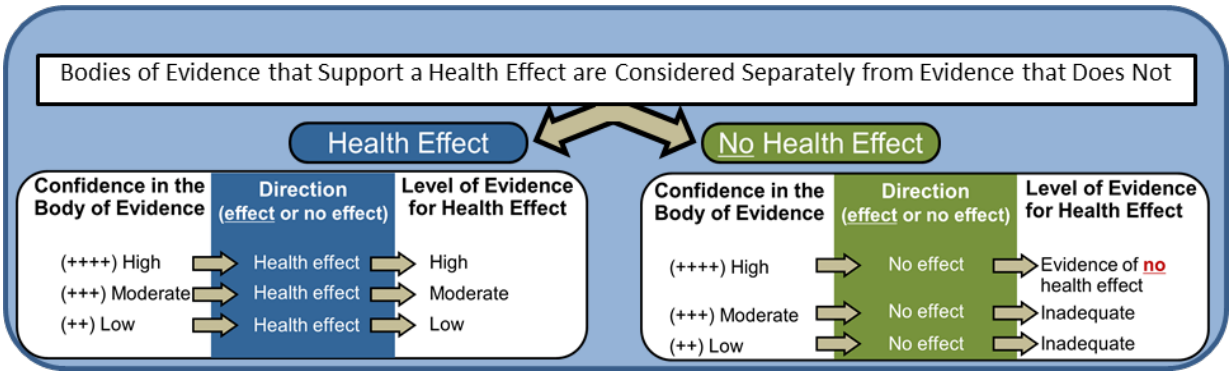
Intratracheal instillation and oropharyngeal aspiration will be downgraded for indirectness, since these exposures result in an unnatural bolus dose that is not well distributed throughout the respiratory tract relative to whole body or nose-only inhalation exposures (Driscoll *et al.* 2000). Oral gavage of particulate matter collected from air will be downgraded for the same reason as intratracheal instillation and oropharyngeal aspiration.

Step 6: Preparation of Draft Level of Evidence Statement

The confidence ratings will be translated into draft level of evidence of health effects for human and animal studies (separately, by evidence stream and health outcome) according to one of four

statements: (1) High, (2) Moderate, (3) Low, or (4) Inadequate (**Figure 2**). The descriptor "evidence of no health effect" is used to indicate confidence that the substance is not associated with a health effect. Because of the inherent difficulty in proving a negative, the conclusion "evidence of no health effect " is only reached when there is high confidence in the body of evidence.

Figure 2. Translation of Confidence Ratings into Evidence of Health Effect Conclusions



Evidence Descriptors	Definition
High Level of Evidence	There is high confidence in the body of evidence for an association between occupational exposure to cancer chemotherapy agents and the health outcome(s).
Moderate Level of Evidence	There is moderate confidence in the body of evidence for an association between occupational exposure to cancer chemotherapy agents and the health outcome(s).
Low Level of Evidence	There is low confidence in the body of evidence for an association between occupational exposure to cancer chemotherapy agents and the health outcome(s), or no data are available.
Inadequate Evidence	There is insufficient evidence available to assess if occupational exposure to cancer chemotherapy agents is associated with the health outcome(s).
Evidence of No Health Effect	There is high confidence in the body of evidence that occupational exposure to cancer chemotherapy agents is not associated with the health outcome(s).

Step 7: Integrate Evidence to Develop Hazard Identification Conclusions.

Finally, the levels of evidence ratings for human and animal data are integrated to reach one of five possible hazard identification categories: (1) Known, (2) Presumed, (3) Suspected, (4) Not classifiable, or (5) Not identified to be a hazard to humans (**Figure 3**).

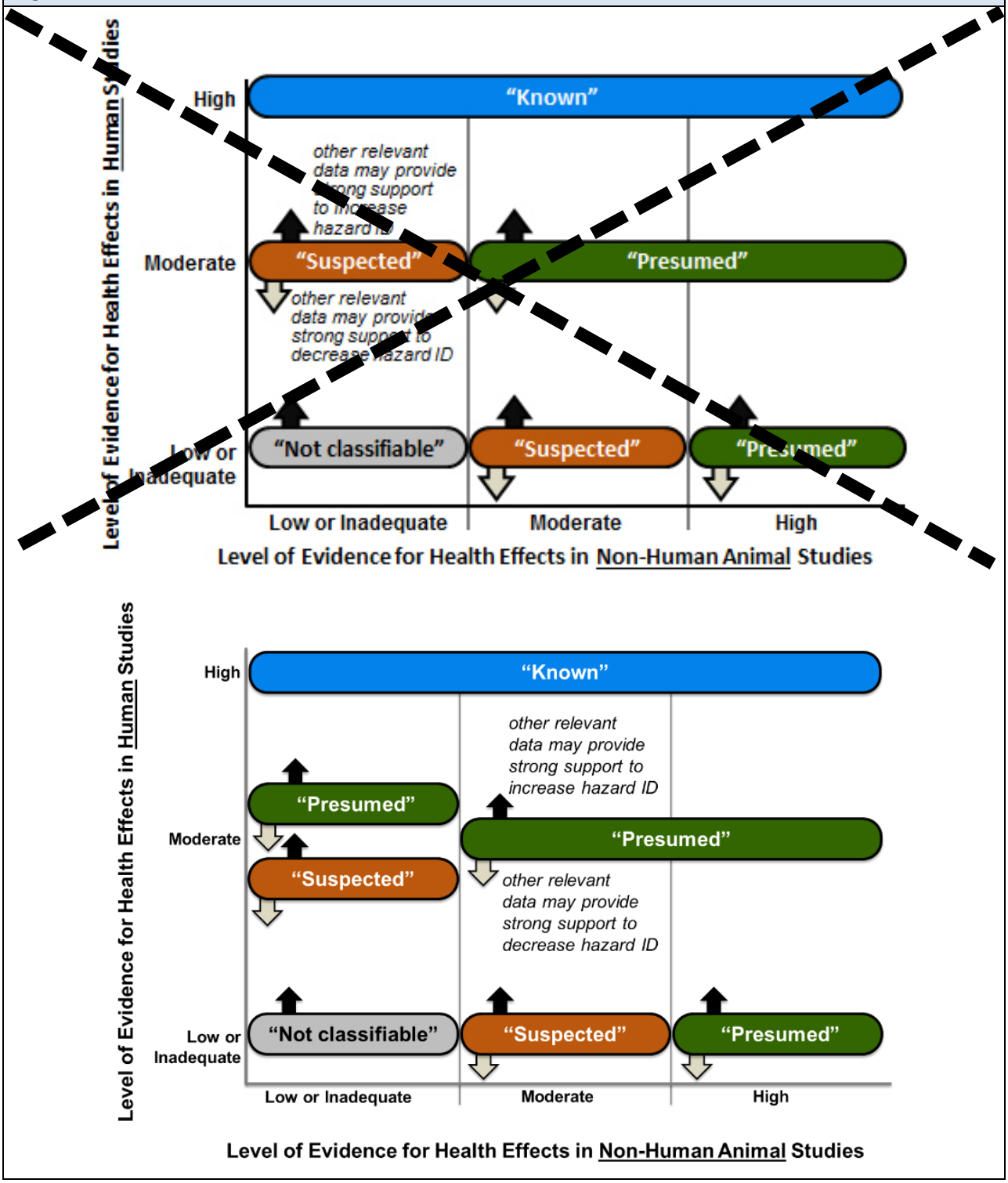
Consideration of Human and Animal Data

Initial hazard identification conclusions will be reached by integrating the highest level-of-evidence conclusion on an outcome basis for the human and the animal evidence streams. Hazard identification conclusions may be reached on the groups of biologically related outcomes (using outcome groups identified in **Table 9** (~~pregnancy-associated hypertensive disorders~~) ~~or **Table 10** and **Table 11** (neurological development)~~ as well as more specific endpoints if data are available to make more specific conclusions. If the data support a health effect, the level-of-evidence conclusion for human data from Step 6 for that health outcome will be considered together with the level of evidence for non-human animal data to reach one of four initial hazard identification conclusions: Known, Presumed, Suspected, or Not classifiable. If either the human or animal evidence stream is characterized as "Inadequate Evidence," then conclusions are based on the remaining evidence stream alone (which is equivalent to treating the missing evidence stream as "Low" in **Figure 3**).

If the human level of evidence rating of "Evidence of no health effect" from Step 6 is supported by a similar level of evidence rating for animal evidence for no health effect, the hazard identification conclusion would be "Not identified to be a hazard to humans."

REVISION: hypertensive disorders of pregnancy

Figure 3. Hazard Identification Scheme



NTP MONOGRAPH FORMAT

The NTP Monograph will include the following information:

Introduction

This section will provide a brief background on the topic.

Methodology

This section will provide a brief overview of the methodologies used in the review process, including:

- the research question;
- the search strategy used to identify and retrieve studies;
- the process for selecting the included studies;
- the methods of data extraction;
- the methods of quality assessment of included studies;
- the methods used to synthesize the data of included studies;
- the methods used to evaluate confidence in the bodies of evidence; and
- the methods used to reach hazard identification conclusions

Results

This section will include the results from the systematic review. Results will be presented in tables or figures as appropriate using HAWC. The results from the included studies will be discussed by outcome. This will include a description of:

- the number of studies identified that reported the outcome;
- a full list of excluded studies, with reasons for exclusion documented for studies excluded at the full text review stage;
- a summary of the results and quality assessment for each individual included study (including files in downloadable format);
- a description of results across studies and analysis of confidence in the body of evidence using OHAT adaptation of GRADE
- a GRADE evidence profile for each health outcome; and
- the presentation of level of evidence and draft hazard identification conclusions for major health outcomes related to ~~pregnancy-associated hypertensive disorders~~ ~~or neurological development~~ for which there are traffic-related air pollution data.

  REVISION: hypertensive disorders of pregnancy

Discussion

The discussion will provide a summary of the review findings, including a discussion of any gaps identified in the evidence and any suggestions of areas for further research. Any important limitations of the review will be described and their impact on the available evidence will be discussed.

Conclusion

This section will present the conclusion of the review.

REFERENCES

AHRQ (Agency for Healthcare Research and Quality). 2014. *AHRQ Training Modules for the Systematic Review Methods Guide*. Rockville, MD. United States Department of Health & Human Services. Available: http://www.effectivehealthcare.ahrq.gov/index.cfm/tools-and-resources/slide-library/ [accessed 15 October 2015].

Bilenko N, Brunekreef B, Beelen R, Eeftens M, de Hoogh K, Hoek G, Koppelman GH, Wang M, van Rossem L, Gehring U. 2015. Associations between particulate matter composition and childhood blood pressure - The PIAMA study. *Environ Int* 84: 1-6.

Boothe VL, Boehmer TK, Wendel AM, Yip FY. 2014. Residential traffic exposure and childhood leukemia: a systematic review and meta-analysis. *Am J Prev Med* 46(4): 413-422.

Calderon-Garciduenas L, Franco-Lira M, D'Angiulli A, Rodriguez-Diaz J, Blaurock-Busch E, Busch Y, Chao CK, Thompson C, Mukherjee PS, Torres-Jardon R, Perry G. 2015. Mexico City normal weight children exposed to high concentrations of ambient PM2.5 show high blood leptin and endothelin-1, vitamin D deficiency, and food reward hormone dysregulation versus low pollution controls. Relevance for obesity and Alzheimer disease. *Environ Res* 140: 579-592.

Cheng Y, Thomas A, Mardini F, Bianchi SL, Tang JX, Peng J, Wei H, Eckenhoff MF, Eckenhoff RG, Levy RJ. 2012. Neurodevelopmental consequences of sub-clinical carbon monoxide exposure in newborn mice. *PLoS One* 7(2): e32029.

CLARITY (Clinical Advances through Research and Information Translation Research Group). 2013. *Tools to assess risk of bias in cohort studies, case control studies, randomized controlled trials, and longitudinal symptom research studies aimed at the general population*. McMaster University. Available: http://www.evidencepartners.com/resources/ [accessed 15 January 2013].

Coogan PF, White LF, Jerrett M, Brook RD, Su JG, Seto E, Burnett R, Palmer JR, Rosenberg L. 2012. Air pollution and incidence of hypertension and diabetes mellitus in black women living in Los Angeles. *Circulation* 125(6): 767-772.

Dadvand P, Ostro B, Amato F, Figueras F, Minguillon MC, Martinez D, Basagana X, Querol X, Nieuwenhuijsen M. 2014. Particulate air pollution and preeclampsia: a source-based analysis. *Occup Environ Med* 71(8): 570-577.

Driscoll KE, Costa DL, Hatch G, Henderson R, Oberdorster G, Salem H, Schlesinger RB. 2000. Intratracheal instillation as an exposure technique for the evaluation of respiratory tract toxicity: uses and limitations. *Toxicol Sci* 55(1): 24-35.

Duley L. 2009. The global impact of pre-eclampsia and eclampsia. *Semin Perinatol* 33(3): 130-137.

Farhi A, Boyko V, Almagor J, Benenson I, Segre E, Rudich Y, Stern E, Lerner-Geva L. 2014. The possible association between exposure to air pollution and the risk for congenital malformations. *Environ Res* 135: 173-180.

Fecht D, Hansell AL, Morley D, Dajnak D, Vienneau D, Beevers S, Toledano MB, Kelly FJ, Anderson HR, Gulliver J. 2016. Spatial and temporal associations of road traffic noise and air pollution in London: Implications for epidemiological studies. *Environ Int* 88: 235-242.

Fu R, Gartlehner G, Grant M, Shamliyan T, Sedrakyan A, Wilt TJ, Griffith L, Oremus M, Raina P, Ismaila A, Santaguida P, Lau J, Trikalinos TA. 2011. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *J Clin Epidemiol* 64(11): 1187-1197.

Genc S, Zadeoglulari Z, Fuss SH, Genc K. 2012. The adverse effects of air pollution on the nervous system. *J Toxicol* 2012: 782462.

Guo YM, Tong S, Zhang YS, Barnett AG, Jia Y, Pan XC. 2010. The relationship between particulate air pollution and emergency hospital visits for hypertension in Beijing, China. *Sci Total Environ* 408(20): 4446-4450.

Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schunemann HJ. 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 64(4): 383-394.

HEI (Health Effects Institute). 2010. *HEI Panel on the Health Effects of Traffic-Related Air Pollution. Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*. HEI Special Report 17. Boston, MA: Institute HE. Available: http://pubs.healtheffects.org/view.php?id=334 [accessed August 26, 2015].

Higgins J, Green S. 2011. *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration. Available: http://handbook.cochrane.org [accessed 3 February 2013].

Hu H, Ha S, Roth J, Kearney G, Talbott EO, Xu X. 2014. Ambient air pollution and hypertensive disorders of pregnancy: a systematic review and meta-analysis. *Atmos Environ* 97: 336-345.

Kalkbrenner AE, Windham GC, Serre ML, Akita Y, Wang X, Hoffman K, Thayer BP, Daniels JL. 2015. Particulate matter exposure, prenatal and postnatal windows of susceptibility, and autism spectrum disorders. *Epidemiology* 26(1): 30-42.

Karner AA, Eisinger DS, Niemeier DA. 2010. Near-roadway air quality: synthesizing the findings from real-world data. *Environ Sci Technol* 44(14): 5334-5344.

Kicinski M, Vermeir G, Van Larebeke N, Den Hond E, Schoeters G, Bruckers L, Sioen I, Bijnens E, Roels HA, Baeyens W, Viaene MK, Nawrot TS. 2015. Neurobehavioral performance in adolescents is inversely associated with traffic exposure. *Environ Int* 75: 136-143.

Krzyzanowski M, Vandenberg J, Stieb D. 2005. Perspectives on air quality policy issues in Europe and North America. *J Toxicol Environ Health A* 68(13-14): 1057-1061.

Li Z, Chadalapaka G, Ramesh A, Khoshbouei H, Maguire M, Safe S, Rhoades RE, Clark R, Jules G, McCallister M, Aschner M, Hood DB. 2012. PAH particles perturb prenatal processes and phenotypes: protection from deficits in object discrimination afforded by dampening of brain oxidoreductase following in utero exposure to inhaled benzo(a)pyrene. *Toxicol Sci* 125(1): 233-247.

Pedersen M, Stayner L, Slama R, Sorensen M, Figueras F, Nieuwenhuijsen MJ, Raaschou-Nielsen O, Dadvand P. 2014. Ambient air pollution and pregnancy-induced hypertensive disorders: a systematic review and meta-analysis. *Hypertension* 64(3): 494-500.

Pereira G, Nassar N, Bower C, Weinstein P, Cook A. 2010. Residential exposure to traffic emissions and adverse pregnancy outcomes. *Sapiens* 3(1).

Pereira G, Nassar N, Cook A, Bower C. 2011. Traffic emissions are associated with reduced fetal growth in areas of Perth, Western Australia: an application of the AusRoads dispersion model. *Aust N Z J Public Health* 35(5): 451-458.

Pereira G, Cook AG, Haggar F, Bower C, Nassar N. 2012. Locally derived traffic-related air pollution and fetal growth restriction: a retrospective cohort study. *Occup Environ Med* 69(11): 815-822.

Raz R, Roberts AL, Lyall K, Hart JE, Just AC, Laden F, Weisskopf MG. 2015. Autism spectrum disorder and particulate matter air pollution before, during, and after pregnancy: a nested case-control analysis within the Nurses' Health Study II cohort. *Environ Health Perspect* 123(3): 264-270.

Roberts CL, Algert CS, Morris JM, Ford JB, Henderson-Smart DJ. 2005. Hypertensive disorders in pregnancy: a population-based study. *Med J Aust* 182(7): 332-335.

Roberts CL, Ford JB, Algert CS, Antonsen S, Chalmers J, Cnattingius S, Gokhale M, Kotelchuck M, Melve KK, Langridge A, Morris C, Morris JM, Nassar N, Norman JE, Norrie J, Sorensen HT, Walker R, Weir CJ. 2011. Population-based trends in pregnancy hypertension and pre-eclampsia: an international comparative study. *BMJ Open* 1(1): e000101.

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

Shamliyan T, Kane RL, Dickinson S. 2010. A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *J Clin Epidemiol* 63(10): 1061-1070.

Sorensen M, Hoffmann B, Hvidberg M, Ketzel M, Jensen SS, Andersen ZJ, Tjonneland A, Overvad K, Raaschou-Nielsen O. 2012. Long-term exposure to traffic-related air pollution associated with blood pressure and self-reported hypertension in a Danish cohort. *Environ Health Perspect* 120(3): 418-424.

Sterne J, Higgins J, Reeves B (on behalf of the development group for ACROBAT-NRSI). 2014. *A Cochrane Risk of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI). Version 1.0.0*. Available: https://sites.google.com/site/riskofbiastool/ [accessed 28 September 2014].

Sunyer J, Esnaola M, Alvarez-Pedrerol M, Forns J, Rivas I, Lopez-Vicente M, Suades-Gonzalez E, Foraster M, Garcia-Esteban R, Basagana X, Viana M, Cirach M, Moreno T, Alastuey A, Sebastian-Galles N, Nieuwenhuijsen M, Querol X. 2015. Association between traffic-related air pollution in schools and cognitive development in primary school children: a prospective cohort study. *PLoS Med* 12(3): e1001792.

Thirtamara Rajamani K, Doherty-Lyons S, Bolden C, Willis D, Hoffman C, Zelikoff J, Chen LC, Gu H. 2013. Prenatal and early-life exposure to high-level diesel exhaust particles leads to increased locomotor activity and repetitive behaviors in mice. *Autism Res* 6(4): 248-257.

US EPA (United States Environmental Protection Agency). 2008. *Final Report: Integrated Science Assessment for Oxides of Nitrogen – Health Criteria*. EPA/600/R-08/071. Washington, DC. United States Environmental Protection Agency. Available: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=194645 [accessed September 3, 2015].

US EPA (United States Environmental Protection Agency). 2009. *Final Report: Integrated Science Assessment for Particulate Matter*. EPA/600/R-08/139F. United States Environmental Protection Agency. Available: http://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=216546 [accessed September 3, 2015].

US EPA (United States Environmental Protection Agency). 2010. *Final Assessment: Integrated Science Assessment for Carbon Monoxide*. EPA/600/R-09/019F. Washington, DC. United States Environmental Protection Agency. Available: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=218686 [accessed September 3, 2015].

van Kempen E, Fischer P, Janssen N, Houthuijs D, van Kamp I, Stansfeld S, Cassee F. 2012. Neurobehavioral effects of exposure to traffic-related air pollution and transportation noise in primary schoolchildren. *Environ Res* 115: 18-25.

Vesterinen HM, Sena ES, Egan KJ, Hirst TC, Churolov L, Currie GL, Antonic A, Howells DW, Macleod MR. 2014. Meta-analysis of data from animal studies: a practical guide. *J Neurosci Methods* 221: 92-102.

Wang S, Zhang J, Zeng X, Zeng Y, Chen S. 2009. Association of traffic-related air pollution with children's neurobehavioral functions in Quanzhou, China. *Environ Health Perspect* 117(10): 1612-1618.

Wu J, Wilhelm M, Chung J, Ritz B. 2011. Comparing exposure assessment methods for traffic-related air pollution in an adverse pregnancy outcome study. *Environ Res* 111(5): 685-692.

Yokota S, Moriya N, Iwata M, Umezawa M, Oshio S, Takeda K. 2013. Exposure to diesel exhaust during fetal period affects behavior and neurotransmitters in male offspring mice. *J Toxicol Sci* 38(1): 13-23.

ABOUT THE PROTOCOL

Contributors

Evaluation Team

Evaluation teams are composed of federal staff and contractor staff. Contractor staff members are screened for potential conflicts of interest. Federal staff members should do a self-evaluation. Epidemiologists and toxicologists on OHAT evaluation teams should have at least three years' experience and/or training in reviewing studies, including summarizing studies and critical review (e.g., assessing study quality and interpreting findings). Experience in evaluating occupational or environmental studies is preferred. Team members should have at least a master's degree or equivalent experience in epidemiology, toxicology, environmental health sciences, or a related field.

Name	Affiliation
REVISION: Brandiese Beverly, Ph.D.	**NIEHS/NTP, Project Co-Lead**
REVISION: Kembra Howdeshell, PhD	NIEHS/NTP, Project **Co-Lead**
Chad Bailey, PhD	US EPA, Office of Transportation and Air Quality (OTAQ)
Tegan Boehmer, PhD	CDC
Erin Hines, PhD	US EPA, Office of Research and Development (ORD)/National Center for Environmental Assessment (NCEA)
Stephanie Holmgren	NIEHS/Office of Science Information Management
Tom Luben, PhD	US EPA, ORD/NCEA
Cynthia Rider, PhD	NIEHS/NTP
Matt Stout, PhD	NIEHS/NTP
Kristina Thayer, PhD	NIEHS/NTP

Contract support: Will assist in data extraction and risk-of-bias assessment

Name	Affiliation
Ali Goldstone, MS	ICF
Chris Baggett, PhD	Social & Scientific Systems, Inc.
Robin Blain, PhD	ICF
Kate Hoffman, PhD	Social & Scientific Systems, Inc.
Anna Jones, MS	Social & Scientific Systems, Inc.
Kristin Nancy, MS	Social & Scientific Systems, Inc.
Johanna Rochester, PhD	ICF
Pamela Schwingl, PhD	Social & Scientific Systems, Inc.
Fikri Yucel, MS	Social & Scientific Systems, Inc.

Technical Advisors

Technical advisors are outside experts retained on an as-needed basis to provide individual advice to the NTP for a specific topic. The technical advisors were selected for their experience with air pollution, its associated health effects, and systematic review procedures. Technical advisors were screened for conflict of interest prior to their service and did not report any conflicts of interest. Service as a technical advisor does not necessarily indicate that an advisor has read the entire protocol or endorses the final document.

Name	Affiliation
Johanna Boogaard, PhD	Heath Effects Institute
Heather Brumberg, MD, MPH	New York Medical College
Catherine Karr, MD, PhD	University of Washington
Melanie Marty, PhD	State of California/ Office of Environmental Health Hazard Assessment

Sources of Support

National Institute of Environmental Health Sciences/Division of the National Toxicology Program
National Institute of Environmental Health Sciences/Global Environmental Health (2012-2014)

Protocol History and Revisions

Date	Activity or revision
September 5, 2015	Draft protocol reviewed: sent to peer reviewers for comment/review
June 24, 2016	Evaluation protocol to be posted on OHAT website
August 7, 2019	Revised protocol posted on OHAT website to reflect principal updates made during evaluation with justifications and date implemented noted: 1. Added Brandiese Beverly, Ph.D., as the Project Co-Lead (February 2017); 2. Protocol was originally developed to outline two OHAT systematic reviews, however, only TRAP and Hypertensive Disorders of Pregnancy was pursued (January 2018); 3. Systematic review on neurological disorders and TRAP was not pursued because this topic is currently under review by other organizations; here and throughout, text related to neurological disorders has a strike-through notation or is noted with a footnote that NTP will not pursue this topic (January 2018); 4. Here and throughout, the term pregnancy-induced hypertension or pregnancy-associated hypertension has been replaced with hypertensive disorders of pregnancy to reflect the more accurate term to describe the collection of disorders (February 2017). 5. The American College of Obstetrics and Gynecology recognizes 4 categories of hypertensive disorders of pregnancy, which includes chronic hypertension and chronic hypertension with superimposed preeclampsia, and studies were not excluded solely based on the inclusion of pregnant women with chronic hypertension (August 2018); 6. Updated Figure 3 to reflect a clarification and update to the OHAT approach for systematic review and evidence integration which clarified how hazard conclusion are reached when there is moderate level-of-evidence for human data with low or inadequate level-of-evidence for the animal evidence stream (March 2019); 7. Brandiese Beverly, Ph.D., added as Project Co-Lead in affiliation section, along with updates to contributor affiliations (July 2019); and 8. Paradoxically, smoking is a negative risk factor for preeclampsia and is not a confounder. Body mass index (BMI) is a risk factor for preeclampsia and was added as a confounder (August 2018).

APPENDICES

Appendix 1. Literature Search Strategy for ~~Pregnancy-Associated Hypertension~~

REVISION: Hypertensive Disorders of Pregnancy

~~The strategy for this search is broad for the consideration of pregnancy-associated hypertensive disorders and comprehensive for traffic-related air pollution as an exposure or treatment in order to ensure inclusion of relevant papers~~.

REVISION: The strategy for this search is broad for the consideration of hypertensive disorders of pregnancy and comprehensive for traffic-related air pollution as an exposure or treatment in order to ensure inclusion of relevant papers.

Database	Search term
Pubmed	"air pollution"[mh:noexp] OR "air pollutants"[mh:noexp] OR (("air pollution"[tiab] OR "air pollutant"[tiab] OR "air pollutants"[tiab]) NOT (indoor*[tiab] OR household[tiab])) OR "particulate matter"[mh:noexp] OR smog[mh] OR soot[mh] OR "particulate matter"[tiab] OR PM2.5[tiab] OR "PM(2.5)"[tiab] OR PM10[tiab] OR "PM(10)"[tiab] OR smog[tiab] OR soot[tiab] OR "carbon black"[tiab] OR "black carbon"[tiab] OR "elemental carbon"[tiab] OR ((air[tiab] OR airborne[tiab] OR coarse[tiab] OR ultrafine[tiab] OR fine[tiab]) AND (particle*[tiab] OR particulate*[tiab])) OR "vehicle emissions"[mh] OR motor vehicles[mh] OR ((vehicle[tiab] OR vehicles[tiab] OR vehicular[tiab] OR auto[tiab] OR automobile[tiab] OR bus[tiab] OR buses[tiab] OR car[tiab] OR truck*[tiab] OR engine[tiab] OR transport[tiab]) AND (emissions[tiab] OR exhaust[tiab] OR fume*[tiab])) OR (traffic[tiab] NOT (safety[tiab] OR accident*[tiab] OR injur*[tiab] OR collision*[tiab] OR crash*[tiab])) OR ((proximity OR near) AND (road*[tiab] OR highway*[tiab] OR freeway*[tiab] OR motorway*[tiab])) OR ((air[tiab] OR outdoor[tiab] OR ambient[tiab] OR pollut*[tiab] OR emissions[tiab] OR exhaust*[tiab]) AND ("sulfur dioxide"[mh] OR "sulfur dioxide"[tiab] OR S02[tiab] OR ozone[mh] OR ozone[tiab] OR O3[tiab] OR "hydrogen sulfide"[mh] OR "hydrogen sulfide"[tiab] OR H2S[tiab] OR "carbon monoxide"[mh] OR "carbon monoxide"[tiab] OR "nitric oxide"[tiab] OR "nitrogen oxide"[tiab] OR "nitrogen oxides"[tiab] OR "nitrogen dioxide"[mh] OR "nitrogen dioxide"[tiab] OR NOx[tiab] OR "NO(x)"[tiab] OR NO2[tiab])) OR "volatile organic compounds"[mh] OR "fossil fuels"[mh] OR "volatile organic compounds"[tiab] OR gasoline*[tiab] OR diesel[tiab] OR petrol*[tiab] OR Polycyclic hydrocarbons, aromatic[mh:noexp] OR "benzo(a)pyrene"[mh] OR benzene[mh] OR "polycyclic aromatic hydrocarbons"[tiab] OR "benzopyrene"[tiab] OR "benzo a pyrene"[tiab] OR "3,4-benzopyrene"[tiab] OR benzene[tiab] AND "Hypertension, Pregnancy-Induced"[Mesh] OR "Gestational Hypertension"[tiab] OR "pregnancy-induced hypertension"[tiab] OR (pregnan*[tiab] AND hypertens*[tiab]) OR pre-eclampsia[tiab] OR preeclampsia[tiab] OR (pregnan*[tiab] AND toxemia*[tiab])
Scopus	(("air pollution" OR "air pollutant" OR "air pollutants") AND NOT (indoor* OR household)) OR "particulate matter" OR PM2.5 OR "PM(2.5)" OR PM10 OR "PM(10)" OR smog OR soot OR "carbon black" OR "black carbon" OR "elemental carbon" OR ((air OR airborne OR coarse OR ultrafine OR fine) AND (particle* OR particulate*)) OR ((vehicle OR vehicles OR vehicular OR auto OR automobile OR bus OR buses OR car OR truck* OR engine OR transport) AND (emissions OR exhaust OR fume*)) OR (traffic AND NOT (safety OR accident* OR injur* OR collision* OR crash*)) OR ((proximity OR near) AND (road* OR highway* OR freeway* OR motorway*)) OR ((air OR outdoor OR ambient OR pollut* OR emissions OR exhaust*) AND ("sulfur dioxide" OR S02 OR ozone OR O3 OR "hydrogen sulfide" OR H2S OR "carbon monoxide" OR "nitric oxide" OR "nitrogen oxide"

	OR "nitrogen oxides" OR "nitrogen dioxide" OR NOx OR "NO(x)" OR NO2)) OR "volatile organic" OR "fossil fuel" OR "fossil fuels" OR gasoline* OR diesel OR petrol* OR "polycyclic aromatic hydrocarbons" OR "benzopyrene" OR "benzo a pyrene" OR "3,4-benzopyrene" OR benzene AND "Gestational Hypertension" OR "pregnancy-induced hypertension" OR (pregnan* w/8 hypertens*) OR pre-eclampsia OR preeclampsia OR (pregnan* w/8 toxemia*)
Web of Science	(("air pollution" OR "air pollutant" OR "air pollutants") NOT (indoor* OR household)) OR "particulate matter" OR PM2.5 OR "PM(2.5)" OR PM10 OR "PM(10)" OR smog OR soot OR "carbon black" OR "black carbon" OR "elemental carbon" OR ((air OR airborne OR coarse OR ultrafine OR fine) AND (particle* OR particulate*)) OR ((vehicle OR vehicles OR vehicular OR auto OR automobile OR bus OR buses OR car OR truck* OR engine OR transport) AND (emissions OR exhaust OR fume*)) OR (traffic NOT (safety OR accident* OR injur* OR collision* OR crash*)) OR ((proximity OR nearby) AND (road* OR highway* OR freeway* OR motorway*)) OR ((air OR outdoor OR ambient OR pollut* OR emissions OR exhaust*) AND ("sulfur dioxide" OR S02 OR ozone OR O3 OR "hydrogen sulfide" OR H2S OR "carbon monoxide" OR "nitric oxide" OR "nitrogen oxide" OR "nitrogen oxides" OR "nitrogen dioxide" OR NOx OR NO2)) OR "volatile organic" OR "fossil fuel" OR "fossil fuels" OR gasoline* OR diesel OR petrol* OR "polycyclic aromatic hydrocarbons" OR "benzopyrene" OR "benzo a pyrene" OR "3,4-benzopyrene" OR benzene AND "Gestational Hypertension" OR "pregnancy-induced hypertension" OR (pregnan* near/8 hypertens*) OR pre-eclampsia OR preeclampsia OR (pregnan* near/8 toxemia*)
*For Web of Science search, we cannot use 'near' (as in roadway) as a textword since it is a proximity operator	

Appendix 2. Literature Search Strategy for Neurological Outcomes[1]

The strategy for this search is broad for the consideration of neurological development and comprehensive for traffic-related air pollution as an exposure or treatment in order to ensure inclusion of relevant papers.

Database	Search terms
Pubmed	"air pollution"[mh:noexp] OR "air pollutants"[mh:noexp] OR (("air pollution"[tiab] OR "air pollutant"[tiab] OR "air pollutants"[tiab]) NOT (indoor*[tiab] OR household[tiab])) OR ambient-air[tiab] OR "particulate matter"[mh:noexp] OR "particulate matter"[tiab] OR PM2.5[tiab] OR "PM(2.5)"[tiab] OR PM10[tiab] OR "PM(10)"[tiab] OR smog[mh] OR smog[tiab] OR soot[mh] OR soot[tiab] OR "carbon black"[tiab] OR "black carbon"[tiab] OR "elemental carbon"[tiab] OR ((air[tiab] OR airborne[tiab] OR coarse[tiab] OR ultrafine[tiab] OR fine[tiab]) AND (particle*[tiab] OR particulate*[tiab])) OR "vehicle emissions"[mh] OR ((motor vehicles[mh] OR vehicle*[tiab] OR vehicular[tiab] OR auto[tiab] OR autos[tiab] OR automobile*[tiab] OR bus[tiab] OR buses[tiab] OR car[tiab] OR cars[tiab] OR truck*[tiab] OR engine*[tiab] OR transport*[tiab]) AND (emissions[tiab] OR exhaust[tiab] OR fume[tiab] OR fumes[tiab])) OR (traffic[tiab] NOT (membrane[tiab] OR signaling[tiab] OR safety[tiab] OR accident*[tiab] OR crash*[tiab] OR injur*[tiab] OR collision*[tiab] OR crash*[tiab] OR offence*[tiab] OR violation*[tiab])) OR ((proximity[tiab] OR near[tiab] OR closeness[tiab]) AND (road*[tiab] OR highway*[tiab] OR freeway*[tiab] OR motorway*[tiab])) OR ((air[tiab] OR outdoor[tiab] OR ambient[tiab] OR pollut*[tiab] OR emissions[tiab] OR exhaust[tiab] OR exposure[tiab] OR exposed[tiab]) AND ("sulfur dioxide"[mh] OR "sulfur dioxide"[tiab] OR S02[tiab] OR ozone[mh] OR ozone[tiab] OR O3[tiab] OR "hydrogen sulfide"[mh] OR "hydrogen sulfide"[tiab] OR H2S[tiab] OR "carbon monoxide"[mh] OR "carbon monoxide"[tiab] OR "nitric oxide"[tiab] OR "nitrogen oxide"[tiab] OR "nitrogen oxides"[tiab] OR "nitrogen dioxide"[mh] OR "nitrogen dioxide"[tiab] OR NOx[tiab] OR "NO(x)"[tiab] OR NO2[tiab])) OR "volatile organic compounds"[mh] OR "fossil fuels"[mh] OR volatile-organic-compound*[tiab] OR gasoline*[tiab] OR diesel[tiab] OR petrol*[tiab] OR Polycyclic hydrocarbons, aromatic[mh:noexp] OR "benzo(a)pyrene"[mh] OR benzene[mh] OR polycyclic-aromatic-hydrocarbon*[tiab] OR "benzopyrene"[tiab] OR "benzo a pyrene"[tiab] OR "3,4-benzopyrene"[tiab] OR benzene[tiab] AND in-utero[tiab] OR intrauterine[tiab] OR embryo[tiab] OR embryos[tiab] OR embryonic[tiab] OR fetal[tiab] OR foetal[tiab] OR fetus[tiab] OR foetus[tiab] OR newborn[tiab] OR birth[tiab] OR neonat*[tiab] OR neo-nat*[tiab] OR prenatal[tiab] OR pre-natal[tiab] OR perinatal[tiab] OR peri-natal[tiab] OR postnatal[tiab] OR post-natal[tiab] OR baby[tiab] OR babies[tiab] OR infant*[tiab] OR toddler*[tiab] OR child[tiab] OR children[tiab] OR childhood[tiab] OR adolescen*[tiab] OR youth[tiab] OR early-life[tiab] OR juvenile[tiab] OR preschool*[tiab] OR student*[tiab] OR animals, newborn[mh] OR litter*[tiab] OR Pups[tiab] OR calves[tiab] OR lambs[tiab] OR piglet*[tiab] OR offspring[tiab] OR Maternal exposure[mh] OR prenatal exposure delayed effects[mh] OR maternal-fetal exchange[mh] OR developmental[tiab] OR neurodevelopment*[tiab] AND neurobehavior*[tiab] OR neurodegenerat*[tiab] OR neurodevelop*[tiab] OR

[1] **REVISION**: NTP will not pursue a systematic review on TRAP and neurological development.

	neurodisease*[tiab] OR neuroinflamm*[tiab] OR neuromediat*[tiab] OR neuromodulat*[tiab] OR neuropath*[tiab] OR neuropeptide*[tiab] OR neurophysiol*[tiab] OR neuropsych*[tiab] OR neuroregulat*[tiab] OR neurosecret*[tiab] OR neurotox*[tiab] OR neurotransmitter*[tiab] OR Nervous system[mh] OR nervous-system*[tiab] OR autonomic-nervous[tiab] OR parasympathetic[tiab] OR sympathetic[tiab] OR central-nervous-system[tiab] OR CNS[tiab] OR brain[tiab] OR "gray matter"[tiab] OR "white matter"[tiab] OR amygdala[tiab] OR hippocamp*[tiab] OR hypothalamus[tiab] OR limbic[tiab] OR olfactory-pathway*[tiab] OR perforant-pathway*[tiab] OR meninges[tiab] OR spinal-cord[tiab] OR ganglia[tiab] OR nerve-tissue[tiab] OR nerve[tiab] OR nerves[tiab] OR neural[tiab] OR neurite*[tiab] OR cranial-nerve[tiab] OR neuroblast*[tiab] OR neuroglia*[tiab] OR astrocyte*[tiab] OR microglia*[tiab] OR schwann-cell*[tiab] OR neuron*[tiab] OR axon*[tiab] OR dendrit*[tiab] OR interneuron*[tiab] OR lewy-bod*[tiab] OR nerve-fiber*[tiab] OR purkinje-cell*[tiab] OR pyramidal-cell*[tiab] OR synapse*[tiab] OR synaptic[tiab] OR presynaptic[tiab] OR pre-synaptic[tiab] OR postsynaptic[tiab] OR post-synaptic[tiab] OR demyelinat*[tiab] OR myelin*[tiab] OR receptors, neurotransmitter[mh] OR (receptor*[tiab] AND (bradykinin[tiab] OR calcitonin[tiab] OR catecholamine[tiab] OR corticotropin[tiab] OR GABA[tiab] OR galanin[tiab] OR glutamate[tiab] OR glycine[tiab] OR muscarinic[tiab] OR neurotensin[tiab] OR nicotinic[tiab] OR oxytocin[tiab] OR prolactin[tiab] OR purinergic[tiab] OR somatostatin[tiab] OR serotonin[tiab] OR tachykinin[tiab] OR thyrotropin*[tiab] OR vasoactive[tiab] OR vasopressin[tiab])) OR cerebellar[tiab] OR cerebrum[tiab] OR cerebellum[tiab] OR cerebrospinal[tiab] OR cerebrovascular[tiab] OR intracranial[tiab] OR neural crest[mh] OR neural-crest[tiab] OR neural plate[mh] OR neural-plate[tiab] OR neural tube[mh] OR neural-tube[tiab] OR "long term potentiation"[tiab] OR "long term synaptic depression"[tiab] OR plasticity[tiab] OR gliogenesis[tiab] OR synaptogenesis[tiab] OR Nervous system diseases[mh] OR neurologic*[tiab] OR "multiple sclerosis"[tiab] OR myelitis[tiab] OR "myasthenia gravis" OR myelopathy[tiab] OR polyradiculoneuropath*[tiab] OR guillain-barre[tiab] OR dysautonomia*[tiab] OR ataxia*[tiab] OR stroke[tiab] OR brain-infarction*[tiab] OR dementia[tiab] OR alzheimer*[tiab] OR aphasia[tiab] OR huntington*[tiab] OR encephalitis[tiab] OR encephalomalacia[tiab] OR headache*[tiab] OR migraine*[tiab] OR cerebral-palsy[tiab] OR epilepsy[tiab] OR epileptic[tiab] OR hydrocephal*[tiab] OR dyskinesia*[tiab] OR angelman*[tiab] OR dyston*[tiab] OR essential-tremor*[tiab] OR parkinson*[tiab] OR tourette*[tiab] OR schizophrenia[tiab] OR amyotrophic lateral sclerosis[tiab] OR epilep*[tiab] OR seizure*[tiab] OR hydrocephalus[tiab] OR movement-disorder*[tiab] OR dyskinesia*[tiab] OR dystonia[tiab] OR tic-disorder*[tiab] OR tourette*[tiab] OR nervous system malformations[mh] OR neural-tube-defect*[tiab] OR heavy metal poisoning, nervous system[mh] OR human development[mh] OR ((embryo*[tiab] OR fetal[tiab] OR foetal[tiab] OR fetus[tiab] OR foetus[tiab] OR baby[tiab] OR infant[tiab] OR toddler[tiab] OR child*[tiab]) AND development*[tiab]) OR developmental-delay[tiab] OR delayed-development[tiab] OR developmental-disabilit*[tiab] OR developmental-disorder*[tiab] OR developmental-impair*[tiab] OR Mental disorders[mh:noexp] OR mental-disorder*[tiab] OR mental-development[tiab] OR mental-illness*[tiab] OR Mental disorders diagnosed in childhood[mh] OR anxiety[tiab] OR anxious[tiab] OR obsessive-compulsive[tiab] OR phobia*[tiab] OR phobic[tiab] OR attention-deficit[tiab] OR hyperactiv*[tiab] OR impulse-control*[tiab] OR impulsivity[tiab] OR disruptive-behavior*[tiab] OR aggression[tiab] OR aggressive[tiab] OR asperger*[tiab] OR autism[tiab] OR autistic[tiab] OR communication-disorder*[tiab] OR language[tiab] OR agraphia[tiab] OR dyslexi*[tiab] OR dyscalculia[tiab] OR speech[tiab] OR aphasia[tiab] OR echolalia[tiab] OR mutism[tiab] OR stutter*[tiab] OR down-syndrome[tiab] OR prader-willi[tiab] OR mental-retard*[tiab] OR memory[tiab] OR perception[tiab] OR perceptual[tiab] OR hallucination*[tiab] OR psychomotor*[tiab] OR stereotypic-movement*[tiab] OR mood disorders[mh] OR mood*[tiab] OR bipolar[tiab] OR sleep disorders[mh] OR sleep[mh] OR sleep[tiab] OR motor activity[mh] OR motor-activit*[tiab] OR locomotor[tiab] OR motor-performance[tiab] OR motor-skill*[tiab] OR visual-motor[tiab] OR Nervous system physiological processes[mh] OR arousal[tiab] OR

	awake[tiab] OR reflex*[tiab] OR proprioception[tiab] OR gait[tiab] OR auditory[tiab] OR hearing[tiab] OR behavior[mh:noexp] OR adolescent behavior[mh] OR behavioral symptoms[mh] OR child behavior[mh] OR impulsive behavior[mh] OR impulsive[tiab] OR risk-taking[mh] OR risk-tak*[tiab] OR sexual-behavior[tiab] OR social-behavior[tiab] OR sociable[tiab] OR spatial behavior[mh] OR emotions[mh] OR emotion*[tiab] OR personality[mh] OR personality[tiab] OR aggression[tiab] OR aggressive[tiab] OR Mental processes[mh] OR cognition[tiab] OR cognitive[tiab] OR awareness[tiab] OR comprehension[tiab] OR learning[tiab] OR memory[tiab] OR mental-recall[tiab] OR perception[tiab] OR intellect*[tiab] OR intelligen*[tiab] OR behavior, animal[mh] OR morris-water[tiab] OR morris-maze[tiab] OR water-maze[tiab] OR fear-conditioning[tiab] OR active-avoidance[tiab] OR passive-avoidance[tiab] OR t-maze[tiab] OR y-maze[tiab] OR novel-object*[tiab] OR rota-rod[tiab] OR geotaxis[tiab] OR startle[tiab] OR open-field-test[tiab] OR sociability[tiab] OR tail-suspension[tiab] OR vocaliz*[tiab] OR territorial[tiab]
Scopus	(("air pollution" OR "air pollutant" OR "air pollutants") AND NOT (indoor* OR household)) OR ambient-air OR "particulate matter" OR "particulate matter" OR PM2.5 OR "PM(2.5)" OR PM10 OR "PM(10)" OR smog OR smog OR soot OR soot OR "carbon black" OR "black carbon" OR "elemental carbon" OR ((air OR airborne OR coarse OR ultrafine OR fine) w/4 (particle* OR particulate*)) OR "vehicle emissions" OR (("motor vehicles" OR vehicle* OR vehicular OR auto OR autos OR automobile* OR bus OR buses OR car OR cars OR truck* OR engine* OR transport*) w/4 (emissions OR exhaust OR fume OR fumes)) OR (traffic AND NOT (membrane OR signal* OR safety OR accident* OR crash* OR injur* OR collision* OR crash* OR offence* OR violation*)) OR ((proximity OR near OR closeness) pre/3 (road* OR highway* OR freeway* OR motorway*)) OR ((air OR outdoor OR ambient OR pollut* OR emissions OR exhaust OR exposed OR exposure) w/4 ("sulfur dioxide" OR "sulfur dioxide" OR S02 OR ozone OR ozone OR O3 OR "hydrogen sulfide" OR "hydrogen sulfide" OR H2S OR "carbon monoxide" OR "carbon monoxide" OR "nitric oxide" OR "nitrogen oxide" OR "nitrogen oxides" OR "nitrogen dioxide" OR "nitrogen dioxides" OR NOx OR "NO(x)" OR NO2)) OR "volatile organic compounds" OR "fossil fuels" OR volatile-organic-compound* OR gasoline* OR diesel OR petrol* OR polycyclic-aromatic-hydrocarbon* OR "benzo(a)pyrene" OR "benzopyrene" OR "benzo a pyrene" OR "3,4-benzopyrene" OR benzene AND in-utero OR intrauterine OR embryo OR embryos OR embryonic OR fetal OR foetal OR fetus OR foetus OR newborn OR birth OR neonat* OR neo-nat* OR prenatal OR pre-natal OR perinatal OR peri-natal OR postnatal OR post-natal OR baby OR babies OR infant* OR toddler* OR child OR children OR childhood OR adolescen* OR youth OR early-life OR juvenile OR preschool* OR litter* OR Pups OR calves OR lambs OR piglet* OR offspring OR "Maternal exposure" OR "prenatal exposure" OR maternal-fetal-exchange AND neurobehavior* OR neurodegenerat* OR neurodevelop* OR neurodisease* OR neuropath* OR neurophysiol* OR neuropsych* OR neurotox* OR "central nervous system" OR CNS OR brain OR neurologic* OR "multiple sclerosis" OR dementia OR alzheimer* OR aphasia OR huntington* OR "cerebral palsy" OR epilepsy OR epileptic OR hydrocephal* OR dyskinesia* OR "essential tremor" OR parkinson* OR tourette* OR schizophrenia OR "movement disorder" OR "movement disorders" OR dyskinesia* OR dystonia OR tourette* OR "neural tube defect" OR "neural tube defects" OR ((baby OR infant OR toddler OR child*) w/3 development*) OR "developmental delay" OR "delayed development" OR "developmental disability" OR "developmental disabilities" OR "developmental disorder" OR "developmental

	disorders" OR "developmental impairment" OR "mental disorder" OR "Mental disorders" OR "mental development" OR "mental illness" OR "Mental illnesses" OR anxiety OR anxious OR "obsessive compulsive" OR phobia* OR phobic OR "attention deficit" OR hyperactiv* OR "impulse control" OR impulsiv* OR "disruptive behavior" OR aggression OR aggressive OR asperger* OR autism OR autistic OR "communication disorder" OR language OR agraphia OR dyslexi* OR dyscalculia OR speech OR aphasia OR stutter* OR "down syndrome" OR psychomotor* OR "stereotypic movement" OR "stereotypic movements" OR bipolar OR sleep OR "motor activity" OR "Motor activities" OR locomotor OR "motor performance" OR "motor skills" OR "visual motor" OR reflex* OR proprioception OR gait OR auditory OR hearing OR "risk taking" OR "sexual behavior" OR "social behavior" OR sociable OR sociability OR personality OR "mental processes" OR cognition OR cognitive OR comprehension OR learning OR memory OR "mental recall" OR intellect* OR intelligen* OR "problem solving" OR ((child OR rat OR rats OR mice OR mouse) pre/3 (behavior OR behaviour)) OR (("morris water" OR "morris maze" OR "water maze" OR "fear conditioning" OR freezing OR "active avoidance" OR "passive avoidance" OR "operant battery" OR "t maze" OR "y maze" OR "novel object" OR "novel objects" OR "rota rod" OR geotaxis OR startle OR "open field test" OR "tail suspension" OR vocaliz* OR territorial) AND (rat OR rats OR mice OR mouse))
Web of Science*	(("air pollution" OR "air pollutant" OR "air pollutants") NOT (indoor* OR household)) OR ambient-air OR "particulate matter" OR "particulate matter" OR PM2.5 OR "PM(2.5)" OR PM10 OR "PM(10)" OR smog OR smog OR soot OR soot OR "carbon black" OR "black carbon" OR "elemental carbon" OR ((air OR airborne OR coarse OR ultrafine OR fine) near/4 (particle* OR particulate*)) OR "vehicle emissions" OR (("motor vehicles" OR vehicle* OR vehicular OR auto OR autos OR automobile* OR bus OR buses OR car OR cars OR truck* OR engine* OR transport*) near/4 (emissions OR exhaust OR fume OR fumes)) OR (traffic NOT (trafficked OR trafficking OR membrane OR signal* OR safety OR accident* OR crash* OR injur* OR collision* OR crash* OR offence* OR violation*)) OR ((proximity OR closeness) near/4 (road* OR highway* OR freeway* OR motorway*)) OR ((air OR outdoor OR ambient OR pollut* OR emissions OR exhaust OR exposed OR exposure) near/4 ("sulfur dioxide" OR "sulfur dioxide" OR S02 OR ozone OR ozone OR O3 OR "hydrogen sulfide" OR "hydrogen sulfide" OR H2S OR "carbon monoxide" OR "carbon monoxide" OR "nitric oxide" OR "nitrogen oxide" OR "nitrogen oxides" OR "nitrogen dioxide" OR "nitrogen dioxides" OR NOx OR "NO(x)" OR NO2)) OR "volatile organic compounds" OR "fossil fuels" OR volatile-organic-compound* OR gasoline* OR diesel OR petrol* OR polycyclic-aromatic-hydrocarbon* OR "benzo(a)pyrene" OR "benzopyrene" OR "benzo a pyrene" OR "3,4-benzopyrene" OR benzene AND "in utero" OR intrauterine OR embryo OR embryos OR embryonic OR fetal OR foetal OR fetus OR foetus OR newborn OR birth OR neonat* OR neo-nat* OR prenatal OR pre-natal OR perinatal OR peri-natal OR postnatal OR post-natal OR baby OR babies OR infant* OR toddler* OR child OR children OR childhood OR adolescen* OR youth OR "early life" OR juvenile OR preschool* OR litter* OR Pups OR calves OR lambs OR piglet* OR offspring OR "Maternal exposure" OR "prenatal exposure" OR maternal-fetal-exchange AND neurobehavior* OR neurodegenerat* OR neurodevelop* OR neurodisease* OR neuropath* OR neurophysiol* OR neuropsych* OR neurotox* OR "central nervous system" OR CNS OR brain OR neurologic* OR "multiple sclerosis" OR dementia OR alzheimer* OR aphasia OR huntington* OR "cerebral palsy" OR epilepsy OR epileptic OR hydrocephal* OR dyskinesia* OR "essential tremor" OR parkinson* OR tourette* OR schizophrenia OR "movement

	disorder" OR "movement disorders" OR dyskinesia* OR dystonia OR tourette* OR "neural tube defect" OR "neural tube defects" OR ((baby OR infant OR toddler OR child*) near/3 development*) OR "developmental delay" OR "delayed development" OR "developmental disability" OR "developmental disabilities" OR "developmental disorder" OR "developmental disorders" OR "developmental impairment" OR "mental disorder" OR "Mental disorders" OR "mental development" OR "mental illness" OR "Mental illnesses" OR anxiety OR anxious OR "obsessive compulsive" OR phobia* OR phobic OR "attention deficit" OR hyperactiv* OR "impulse control" OR impulsiv* OR "disruptive behavior" OR aggression OR aggressive OR asperger* OR autism OR autistic OR "communication disorder" OR language OR agraphia OR dyslexi* OR dyscalculia OR speech OR aphasia OR stutter* OR "down syndrome" OR psychomotor* OR "stereotypic movement" OR "stereotypic movements" OR bipolar OR sleep OR "motor activity" OR "Motor activities" OR locomotor OR "motor performance" OR "motor skills" OR "visual motor" OR reflex* OR proprioception OR gait OR auditory OR hearing OR "risk taking" OR "sexual behavior" OR "social behavior" OR sociable OR sociability OR personality OR "mental processes" OR cognition OR cognitive OR comprehension OR learning OR memory OR "mental recall" OR intellect* OR intelligen* OR "problem solving" OR ((child OR rat OR rats OR mice OR mouse) near/3 (behavior OR behaviour)) OR (("morris water" OR "morris maze" OR "water maze" OR "fear conditioning" OR freezing OR "active avoidance" OR "passive avoidance" OR "operant battery" OR "t maze" OR "y maze" OR "novel object" OR "novel objects" OR "rota rod" OR geotaxis OR startle OR "open field test" OR "tail suspension" OR vocaliz* OR territorial) AND (rat OR rats OR mice OR mouse))
*For Web of Science search, we cannot use 'near' (as in roadway) as a textword since it is a proximity operator	

Appendix 3. Data Extraction Elements Captured in DRAGON for Human and Animal Studies

Data Extraction Elements for Human Studies	
Funding	Funding source(s)
	Reporting of conflict of interest (COI) by authors (*reporting bias)
Subjects	Study population name/description
	Dates of study and sampling time frame
	Geography (country, region, state, etc.)
	Demographics (sex, race/ethnicity, age or lifestage at exposure and at outcome assessment)
	Number of subjects (target, enrolled, n per group in analysis, and participation/follow-up rates) (*missing data bias)
	Inclusion/exclusion criteria/recruitment strategy (*selection bias)
	Description of reference group (*selection bias)
Methods	Study design (e.g., prospective or retrospective cohort, nested case-control study, cross-sectional, population-based case-control study, intervention, case report, etc.)
	Length of follow-up (*information bias)
	Health outcome category, e.g., cardiovascular
	Health outcome, e.g., blood pressure (*reporting bias)
	Diagnostic or methods used to measure health outcome (*information bias)
	Confounders or modifying factors and how considered in analysis (e.g., included in final model, considered for inclusion but determined not needed (*confounding bias)
	Substance name and CAS number
	Exposure assessment (e.g., blood, urine, hair, air, drinking water, job classification, residence, administered treatment in controlled study, etc.) (*information bias)
	Methodological details for exposure assessment (e.g., HPLC-MS/MS, limit of detection) (*information bias)
	Statistical methods (*information bias)
Results	Exposure levels (e.g., mean, median, measures of variance as presented in paper, such as SD, SEM, 75th/90th/95th percentile, minimum/maximum); range of exposure levels, number of exposed cases
	Statistical findings (e.g., adjusted β, standardized mean difference, adjusted odds ratio, standardized mortality ratio, relative risk, etc.) or description of qualitative results. When possible, OHAT will convert measures of effect to a common metric with associated 95% confidence intervals (CI). Most often, measures of effect for continuous data are expressed as mean difference, standardized mean difference, and percent control response. Categorical data are typically expressed as odds ratio, relative risk (RR, also called risk ratio), or β values, depending on what metric is most commonly reported in the included studies and on OHAT's ability to obtain information for effect conversions from the study or through author query.

Data Extraction Elements for Human Studies	
	If not presented in the study, statistical power can be assessed during data extraction using an approach that can detect a 10% to 20% change from response by control or referent group for continuous data, or a relative risk or odds ratio of 1.5 to 2 for categorical data, using the prevalence of exposure or prevalence of outcome in the control or referent group to determine sample size. For categorical data where the sample sizes of exposed and control or referent groups differ, the sample size of the exposed group will be used to determine the relative power category. Recommended sample sizes to achieve 80% power for a given effect size, i.e., 10% or 20% change from control, will be compared to sample sizes used in the study to categorize statistical power as "appears to be adequately powered" (sample size for 80% power met), somewhat underpowered (sample size is 75% to < 100% of number required for 80% power), "underpowered" (sample size is 50% to < 75% of number required for 80% power), or "severely underpowered" (sample size is < 50% of number required for 80% power).
	Observations on dose response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, non-monotonic)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

Data Extraction Elements for Animal Studies	
Funding	Funding source(s)
	Reporting of COI by authors (*reporting bias)
Animal Model	Sex
	Species
	Strain
	Source of animals
	Age or lifestage at start of dosing and at health outcome assessment
	Diet and husbandry information (e.g., diet name/source)
Treatment	Chemical name and CAS number
	Source of chemical
	Purity of chemical (*information bias)
	Dose levels or concentration (as presented and converted to mg/kg bw/d when possible)
	Other dose-related details, such as whether administered dose level was verified by measurement, information on internal dosimetry (*information bias)
	Vehicle used for exposed animals
	Route of administration (e.g., oral, inhalation, dermal, injection)
	Duration and frequency of dosing (e.g., hours, days, weeks when administration was ended, days per week)
Methods	Study design (e.g., single treatment, acute, subchronic (e.g., 90 days in a rodent), chronic, multigenerational, developmental, other)
	Guideline compliance (i.e., use of EPA, OECD, NTP or another guideline for study design, conducted under GLP guideline conditions, non-GLP but consistent with guideline study, non-guideline peer-reviewed publication)
	Number of animals per group (and dams per group in developmental studies) (*missing data bias)
	Randomization procedure, allocation concealment, blinding during outcome assessment (*selection bias)
	Method to control for litter effects in developmental studies (*information bias)
	Use of negative controls and whether controls were untreated, vehicle-treated, or both
	Report on data from positive controls – was expected response observed? (*information bias)
	Endpoint health category (e.g., reproductive)
	Endpoint (e.g., infertility)
	Diagnostic or method to measure endpoint (*information bias)
	Statistical methods (*information bias)
Results	Measures of effect at each dose or concentration level (e.g., mean, median, frequency, and measures of precision or variance) or description of qualitative results. When possible, OHAT will convert measures of effect to a common metric with associated 95% confidence intervals (CI). Most often, measures of effect for continuous data will be expressed as mean difference, standardized mean difference, and percent control response. Categorical data will be expressed as relative risk (RR, also called risk ratio).

Data Extraction Elements for Animal Studies	
	No Observed Effect Level (NOEL), Lowest Observed Effect Level (LOEL), benchmark dose (BMD) analysis, statistical significance of other dose levels, or other estimates of effect presented in paper. **Note:** The NOEL and LOEL are highly influenced by study design, do not give any quantitative information about the relationship between dose and response, and can be subject to author's interpretation (e.g., a statistically significant effect may not be considered biologically important). Also, a NOEL does not necessarily mean zero response. Ideally, the response rate at specific dose levels is used as the primary measure to characterize the response.
	If not presented in the study, statistical power can be assessed during data extraction using an approach that assesses the ability to detect a 10% to 20% change from control group's response for continuous data, or a relative risk or odds ratio of 1.5 to 2 for categorical data, using the outcome frequency in the control group to determine sample size. Recommended sample sizes to achieve 80% power for a given effect size, i.e., 10% or 20% change from control, will be compared to sample sizes used in the study to categorize statistical power as "appears to be adequately powered" (sample size for 80% power met), "somewhat underpowered" (sample size is 75% to < 100% of number required for 80% power), "underpowered" (sample size is 50% to < 75% of number required for 80% power), or "severely underpowered" (sample size is < 50% of number required for 80% power).
	Observations on dose response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, non-monotonic)
	Data on internal concentration, toxicokinetics, or toxicodynamics (when reported)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

Appendix 4. Risk-of-Bias Criteria

The OHAT risk-of-bias tool for human and animal studies (version date January 2015 and available at http://ntp.niehs.nih.gov/go/38673) reflects OHAT's current best practices and provides the detailed discussion and instructions for the risk-of-bias practices used in this evaluation. The OHAT tool uses a single set of questions (also called "elements" or "domains") to assess risk of bias across various study types to facilitate consideration of conceptually similar potential sources of bias across the human and animal evidence streams with a common terminology. Individual risk-of-bias questions are designated as only applicable to certain study designs (e.g., cohort studies or experimental animal studies), and a subset of the questions apply to each study design (**Table 6**).

The eight questions relevant to experimental or controlled-exposure studies were used as the basis for development of an OHAT *in vitro* risk-of-bias tool. This tool will be applied to mechanistic studies using cells or tissues from humans or other animals with an *in vitro* exposure regime; in contrast to *in vivo* exposures that are already addressed by risk-of-bias tools for experimental animal studies or controlled human exposures. A manuscript detailing the *in vitro* risk-of-bias method to be used in this evaluation is currently under peer review for publication (Andy Rooney, manuscript in preparation). Comments received during the manuscript review process will be considered for potential revisions to the risk-of-bias method used to evaluate *in vitro*/ mechanistic studies.

The specific criteria used to assess risk of bias for this evaluation are outlined below for Human/observational studies, experimental animal studies, and in vitro/mechanistic studies. Based on literature searches done for the case study we do not expect any controlled exposure studies in humans (i.e., human controlled trials) and therefore have not included risk-of-bias criteria for that study design. If relevant human controlled trials of direct traffic measures or traffic-related air pollutants are identified, the criteria from the January 2015 OHAT risk-of-bias tool will be used to evaluate risk of bias.

Observational Studies (Human or wildlife studies)

Cohort studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups?

Risk-of-Bias Criteria for Appropriate Comparison Groups (Cohort Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, • **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables (see question #4).
Probably Low Risk of Bias (+)
• Indirect evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates,

Risk-of-Bias Criteria for Appropriate Comparison Groups (Cohort Studies)
• **OR** differences between groups would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates, • **OR** there is insufficient information provided about the comparison group including a different rate of non-response without an explanation (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates.

4. Did study design or analysis account for important confounding and modifying variables?

Risk-of-Bias Criteria for Confounding and Modifying Variables (Cohort Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that appropriate adjustments or explicit considerations were made for the variables listed below as potential confounders and/or effect measure modifiers in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching, adjustment in multivariate model, stratification, propensity scoring, or other methods that were appropriately justified. Acceptable consideration of appropriate adjustment factors includes cases when the factor is not included in the final adjustment model because the author conducted analyses that indicated it did not need to be included, • **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, • **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. In occupational studies or studies of contaminated sites, other chemical exposures known to be associated with those settings were appropriately considered. • **Note:** Questionnaires that are not explicitly reported as valid/reliable are assumed to be valid/reliable measurements of primary covariates and confounders. ~~• **Note:** General confounders considered *a priori* for pregnancy-associated hypertension are: age, race/ethnicity, smoking, and socioeconomic factors (e.g., education, income, etc).~~ • **REVISION: Note: General confounders considered *a priori* for hypertensive disorders of pregnancy are: age, race/ethnicity, body mass index, and socioeconomic factors (e.g., education, income, etc.)** ~~• **Note**: General confounders considered *a priori* for neurological outcomes are: age, race/ethnicity, smoking, and socioeconomic factors.~~ • **Note:** ~~Pregnancy-associated hypertensive disorder~~ specific confounders or exclusion criteria: past history of chronic hypertension. **REVISION: Hypertensive disorders of pregnancy-** • **Note:** Traffic noise is a co-exposure for all traffic-related air pollution studies, but is unlikely to be measured. Therefore, studies that did not account for traffic noise were not penalized in the risk of bias rating.

Risk-of-Bias Criteria for Confounding and Modifying Variables (Cohort Studies)
Probably Low Risk of Bias (+)
• Indirect evidence that appropriate adjustments were made, • **OR** it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results, • **AND** there is evidence (direct or indirect) that covariates and confounders considered were assessed using valid and reliable measurements, • **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research), • **AND** there is evidence (direct or indirect) that other co-exposures anticipated to bias results were not present or were appropriately adjusted for, • **OR** it is deemed that co-exposures present would not appreciably bias results. **Note:** this includes insufficient information provided on co-exposures in general population studies.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the distribution of important covariates and known confounders differed between the groups and was not appropriately adjusted for in the final analyses, • **OR** there is insufficient information provided about the distribution of known confounders (record "NR" as basis for answer), • **OR** there is indirect evidence that covariates and confounders considered were assessed using measurements of unknown validity, • **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record "NR" as basis for answer), • **OR** there is indirect evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for, • **OR** there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the distribution of important covariates and known confounders differed between the groups, confounding was demonstrated, and was not appropriately adjusted for in the final analyses, • **OR** there is direct evidence that covariates and confounders considered were assessed using non valid measurements, • **OR** there is direct evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Risk-of-Bias Criteria for Data Attrition or Exclusion (Cohort Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study. • **Note**: Acceptable handling of subject attrition includes: very little missing outcome data; reasons for

Risk-of-Bias Criteria for Data Attrition or Exclusion (Cohort Studies)
missing subjects unlikely to be related to outcome (for survival data, censoring unlikely to be introducing bias); missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups, • **OR** missing data have been imputed using appropriate methods and characteristics of subjects lost to follow up or with unavailable records are described in identical way and are not significantly different from those of the study participants.
Probably Low Risk of Bias (+)
• Indirect evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study, • **OR** it is deemed that the proportion lost to follow-up would not appreciably bias results. This would include reports of no statistical differences in characteristics of subjects lost to follow up or with unavailable records from those of the study participants. Generally, the higher the ratio of participants with missing data to participants with events, the greater potential there is for bias. For studies with a long duration of follow-up, some withdrawals for such reasons are inevitable.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed, • **OR** there is insufficient information provided about numbers of subjects lost to follow-up (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed. • **Note**: Unacceptable handling of subject attrition includes: reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across study groups; or potentially inappropriate application of imputation.

8. Can we be confident in the exposure characterization?

Risk-of-Bias Criteria for Exposure Characterization (Cohort Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that more than one traffic-related air pollutant was reported • **AND** exposure was consistently assessed using well-established methods that directly measure exposure, • **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods, • **AND** exposure was assessed in a relevant time-window and reasonably well aligned with the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay, and measured with good accuracy and precision such that different exposure groups can be distinguished. • **Note:** Data on cross-validation R^2 and/or sensitivity/subgroup analyses (e.g., selecting only subjects residing within a specified short distance from a road site monitor) may indicate a study has lower risk of bias, but the absence of such analyses will not penalize a study.
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure was consistently assessed using well-established methods that

Risk-of-Bias Criteria for Exposure Characterization (Cohort Studies)
directly measure exposure, • **OR** exposure was assessed using less-established methods that directly measure exposure, • **AND** exposure was assessed in a relevant time-window and reasonably well aligned with the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay and measured with good accuracy and precision such that different exposure groups can be distinguished.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure • **AND** indirect evidence that exposure assessment does not adequately reflect relevant exposure levels (e.g., poor density of data, poor data quality, many missing values, substantial data misalignment), • **OR** there is evidence that the exposure was assessed using indirect measures that have not been validated or empirically shown to be consistent with methods that directly measure exposure (e.g., questionnaire, self-report without validation), • **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the exposure was assessed using methods with poor validity, • **AND** direct evidence that exposure assessment does not adequately reflect relevant exposure levels (e.g., poor density of data, poor data quality, many missing values, substantial data misalignment), • **OR** evidence of substantial exposure misclassification.

9. Can we be confident in the outcome assessment?

Risk-of-Bias Criteria for Outcome Assessment (Cohort Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that ~~pregnancy-associated hypertensive disorders~~ ~~or neurological development~~ were assessed using well-established methods (e.g., gold standard) **REVISION: hypertensive disorders of pregnancy** • **AND** subjects had been followed for the same length of time in all study groups, • **AND** there is direct evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the study group or exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes.
Probably Low Risk of Bias (+)
• Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard), • **AND** subjects had been followed for the same length of time in all study groups • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes, • **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results, which is more likely to apply to objective outcome measures,

Risk-of-Bias Criteria for Outcome Assessment (Cohort Studies)
• **NOTE:** Acceptable, but not ideal assessment methods will depend on the outcome, but examples of such methods may include proxy reporting of outcomes such as and mining of data collected for other purposes. Proxy reporting (e.g., parental reporting of days sick or doctor-diagnosis) of immune disease, colds, etc. should be considered on a case-by-case basis with consideration of whether or not there is empirical evidence as to the reliability of proxy reporting for that outcome.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome assessment method is an insensitive instrument (e.g., a questionnaire used to assess outcomes with no information on validation), • **OR** the length of follow up differed by study group, • **OR** there is indirect evidence that it was possible for outcome assessors (including study subjects if outcomes were self-reported) to infer the study group prior to reporting outcomes, • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the outcome assessment method is an insensitive instrument, • **OR** the length of follow up differed by study group, • **OR** there is direct evidence for lack of adequate blinding of outcome assessors (including study subjects if outcomes were self-reported), including no blinding or incomplete blinding.

10. Were all measured outcomes reported?

Risk-of-Bias Criteria for Outcome Reporting (Cohort Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).

Risk-of-Bias Criteria for Outcome Reporting (Cohort Studies)
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

Cross Sectional and Case Series Studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups?[NA to Case series]

Risk-of-Bias Criteria for Appropriate Comparison Groups (Cross Sectional Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, • **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables (see question #4).
Probably Low Risk of Bias (+)
• Indirect evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, • **OR** differences between groups would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates, • **OR** there is insufficient information provided about the comparison group including a different rate of non-response without an explanation (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates.

4. Did study design or analysis account for important confounding and modifying variables?

Risk-of-Bias Criteria for Confounding and Modifying Variables (Cross Sectional and Case Series Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that appropriate adjustments or explicit considerations were made for the variables listed below as potential confounders and/or effect measure modifiers in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching, adjustment in multivariate model, stratification, propensity scoring, or other methods that were appropriately justified. Acceptable consideration of appropriate adjustment factors includes cases when the factor is not included in the final adjustment model because the author conducted analyses that indicated it did not need to be included, • **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, • **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. In occupational studies or studies of contaminated sites, other chemical exposures known to be associated with those settings were appropriately considered. • **Note:** Questionnaires that are not explicitly reported as valid/reliable are assumed to be valid/reliable

Risk-of-Bias Criteria for Confounding and Modifying Variables (Cross Sectional and Case Series Studies)
measurements of primary covariates and confounders. • ~~**Note:** General confounders considered *a priori* for pregnancy-associated hypertension are: age, race/ethnicity, smoking, and socioeconomic factors (e.g., education, income, etc).~~ **REVISION: Note: General confounders considered *a priori* for hypertensive disorders of pregnancy are: age, race/ethnicity, body mass index, and socioeconomic factors (e.g., education, income, etc).** • ~~**Note**: General confounders considered *a priori* for neurological outcomes are: age, race/ethnicity, smoking, and socioeconomic factors~~. • **Note:** ~~Pregnancy-associated hypertension~~ specific confounders or exclusion criteria: past history of chronic hypertension. **REVISION: Hypertensive disorders of pregnancy-** • **Note:** Traffic noise is a co-exposure for all traffic-related air pollution studies, but is unlikely to be measured. Therefore, studies that did not account for traffic noise were not penalized in the risk of bias rating.
Probably Low Risk of Bias (+)
• Indirect evidence that appropriate adjustments were made, • **OR** it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results, • **AND** there is evidence (direct or indirect) that covariates and confounders were considered using valid and reliable measurements, • **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research), • **AND** there is evidence (direct or indirect) that other co-exposures anticipated to bias results were not present or were appropriately adjusted for, • **OR** it is deemed that co-exposures present would not appreciably bias results. • **Note:** this includes insufficient information provided on co-exposures in general population studies.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the distribution of important covariates and known confounders differed between the groups and was not appropriately adjusted for in the final analyses, • **OR** there is insufficient information provided about the distribution of known confounders (record "NR" as basis for answer), • **OR** there is indirect evidence that covariates and confounders were considered using measurements of unknown validity, • **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record "NR" as basis for answer), • **OR** there is indirect evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for, • **OR** there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the distribution of important covariates and known confounders differed between the groups, confounding was demonstrated, and was not appropriately adjusted for in the final analyses, • **OR** there is direct evidence that covariates and confounders considered were assessed using non valid

Risk-of-Bias Criteria for Confounding and Modifying Variables (Cross Sectional and Case Series Studies)
measurements, • **OR** there is direct evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Risk-of-Bias Criteria for Data Attrition or Exclusion (Cross-Sectional and Case Series Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably Low Risk of Bias (+)
• Indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that exclusion of subjects from analyses was not adequately addressed, • **OR** there is insufficient information provided about why subjects were removed from the study or excluded from analyses (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that exclusion of subjects from analyses was not adequately addressed. • **Note**: Unacceptable handling of subject exclusion from analyses includes: reason for exclusion likely to be related to true outcome, with either imbalance in numbers or reasons for exclusion across study groups.

8. Can we be confident in the exposure characterization?

Risk-of-Bias Criteria for Exposure Characterization (Cross-Sectional and Case Series Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that more than one traffic-related air pollutant was reported • **AND** exposure was consistently assessed using well-established methods that directly measure exposure, • **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods, • **AND** exposure was assessed in a relevant time-window and reasonably well aligned with the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay, and measured with good accuracy and precision such that different exposure groups can be distinguished. • **Note:** Data on cross-validation R^2 and/or sensitivity/subgroup analyses (e.g., selecting only subjects residing within a specified short distance from a road site monitor) may indicate a study has lower risk of bias, but the absence of such analyses will not penalize a study.
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure was consistently assessed using well-established methods that

Risk-of-Bias Criteria for Exposure Characterization (Cross-Sectional and Case Series Studies)
directly measure exposure, • **OR** exposure was assessed using less-established methods that directly measure exposure, • **AND** exposure was assessed in a relevant time-window and reasonably well aligned with the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay and measured with good accuracy and precision such that different exposure groups can be distinguished.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure • **AND** indirect evidence that exposure assessment does not adequately reflect relevant exposure levels (e.g., poor density of data, poor data quality, many missing values, substantial data misalignment), • **OR** there is evidence that the exposure was assessed using indirect measures that have not been validated or empirically shown to be consistent with methods that directly measure exposure (e.g., questionnaire, self-report without validation), • **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the exposure was assessed using methods with poor validity, • **AND** direct evidence that exposure assessment does not adequately reflect relevant exposure levels (e.g., poor density of data, poor data quality, many missing values, substantial data misalignment), • **OR** evidence of substantial exposure misclassification.

9. Can we be confident in the outcome assessment?

Risk-of-Bias Criteria for Outcome Assessment (Cross Sectional and Case Series Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that ~~pregnancy-associated hypertensive disorders or neurological development~~ were assessed using well-established methods (the gold standard), **REVISION: hypertensive disorders of pregnancy** • **AND** there is direct evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes. • **NOTE** Well-established assessment methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits, commercial laboratories, or standard assays such as ELISAs for IgG with sufficiently low variation and limits of detection to allow discrimination between groups (or evidence that the assay could have detected a difference based on responses to a positive control); doctor diagnosis of asthma or incidence data obtained from medical records; obtained from registries (Shamliyan *et al.* 2010).
Probably Low Risk of Bias (+)
• Indirect evidence that the outcome was assessed using acceptable methods, • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors were adequately blinded to the exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes,

Risk-of-Bias Criteria for Outcome Assessment (Cross Sectional and Case Series Studies)
• **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results (including that subjects self-reporting outcomes were likely not aware of reported links between the exposure and outcome lack of blinding is unlikely to bias a particular outcome). • **NOTE:** Acceptable, but not ideal assessment methods will depend on the outcome, but examples of such methods may include proxy reporting of outcomes such as asthma and mining of data collected for other purposes. Proxy reporting (e.g., parental reporting of days sick or doctor-diagnosis) of immune disease, colds, etc. should be considered on a case-by-case basis with consideration of whether or not there is empirical evidence as to the reliability of proxy reporting for that outcome.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome assessment method is an insensitive instrument, • **OR** there is indirect evidence that it was possible for outcome assessors to infer the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were likely aware of reported links between the exposure and outcome), • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the outcome assessment method is an insensitive instrument, • **OR** there is direct evidence that outcome assessors were aware of the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were aware of reported links between the exposure and outcome).

10. Were all measured outcomes reported?

Risk-of-Bias Criteria for Outcome Reporting (Cross Sectional and Case Series Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis

Risk-of-Bias Criteria for Outcome Reporting (Cross Sectional and Case Series Studies)
for answer).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

Case Control Studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups?

Risk-of-Bias Criteria for Appropriate Comparison Groups (Case Control Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that cases and controls were similar (e.g., recruited from the same eligible population including being of similar age, gender, ethnicity, and eligibility criteria other than outcome of interest as appropriate), recruited within the same time frame, and controls are described as having no history of the outcome, • **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables (see question #4).
Probably Low Risk of Bias (+)
• Indirect evidence that cases and controls were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age), recruited within the same time frame, and controls are described as having no history of the outcome, • **OR** it is deemed differences between cases and controls would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that controls were drawn from a very dissimilar population than cases or recruited within very different time frames, • **OR** there is insufficient information provided about the appropriateness of controls including rate of response reported for cases only (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that controls were drawn from a very dissimilar population than cases or recruited within very different time frames.

4. Did study design or analysis account for important confounding and modifying variables?

Risk-of-Bias Criteria for Confounding and Modifying Variables (Case Control Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that appropriate adjustments were made for primary covariates and confounders in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching of cases and controls, adjustment in multivariate model, stratification, propensity scoring, or other methods were appropriately justified, • **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, • **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. • **Note:** Questionnaires that are not explicitly reported as valid/reliable are assumed to be valid/reliable measurements of primary covariates and confounders. ~~• **Note:** General confounders considered *a priori* for pregnancy-associated hypertension are: age, race/ethnicity, smoking, and socioeconomic factors (e.g., education, income, etc).~~ • **REVISION Note: General confounders considered *a priori* for hypertensive disorders of pregnancy are: age race/ethnicity, body mass index, and socioeconomic factors (e.g., education, income,**

Risk-of-Bias Criteria for Confounding and Modifying Variables (Case Control Studies)
etc.) • ~~**Note**: General confounders considered *a priori* for neurological outcomes are: age, race/ethnicity, smoking, and socioeconomic factors~~. • **Note:** ~~Pregnancy-associated hypertension-~~specific confounders or exclusion criteria: past history of chronic hypertension. **REVISION: Hypertensive disorders of pregnancy-**
Probably Low Risk of Bias (+)
• Indirect evidence that appropriate adjustments were made, • **OR** it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results, • **AND** there is evidence (direct or indirect) that primary covariates and confounders were assessed using valid and reliable measurements, • **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research), • **AND** there is evidence (direct or indirect) that other co-exposures anticipated to bias results were not present or were appropriately adjusted for, • **OR** it is deemed that co-exposures present would not appreciably bias results. • **Note:** this includes insufficient information provided on co-exposures in general population studies.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the distribution of primary covariates and known confounders differed between cases and controls and was not investigated further, • **OR** there is insufficient information provided about the distribution of known confounders in cases and controls (record "NR" as basis for answer), • **OR** there is indirect evidence that primary covariates and confounders were assessed using measurements of unknown validity, • **OR** there is insufficient information provided about the measurement techniques used (record "NR" as basis for answer), • **OR** there is indirect evidence that there was an unbalanced provision of additional co-exposures across cases and controls, which were not appropriately adjusted for, • **OR** there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the distribution of primary covariates and known confounders differed between cases and controls, confounding was demonstrated, but was not appropriately adjusted for in the final analyses, • **OR** there is direct evidence that primary covariates and confounders were assessed using non valid measurements, • **OR** there is direct evidence that there was an unbalanced provision of additional co-exposures across cases and controls, which were not appropriately adjusted for.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Risk-of-Bias Criteria for Data Attrition or Exclusion (Case Control Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably Low Risk of Bias (+)
• Indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that exclusion of subjects from analyses was not adequately addressed, • **OR** there is insufficient information provided about why subjects were removed from the study or excluded from analyses (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that exclusion of subjects from analyses was not adequately addressed. • **Note**: Unacceptable handling of subject exclusion from analyses includes: reason for exclusion likely to be related to true outcome, with either imbalance in numbers or reasons for exclusion across study groups.

8. Can we be confident in the exposure characterization?

Risk-of-Bias Criteria for Exposure Characterization (Case Control Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that more than one traffic-related air pollutant was reported • **AND** exposure was consistently assessed using well-established methods that directly measure exposure, • **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods, • **AND** exposure was assessed in a relevant time-window and reasonably well aligned with the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay, and measured with good accuracy and precision such that different exposure groups can be distinguished. • **Note:** Data on cross-validation R^2 and/or sensitivity/subgroup analyses (e.g., selecting only subjects residing within a specified short distance from a road site monitor) may indicate a study has lower risk of bias, but the absence of such analyses will not penalize a study.
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure was consistently assessed using well-established methods that directly measure exposure, • **OR** exposure was assessed using less-established methods that directly measure exposure, • **AND** exposure was assessed in a relevant time-window and reasonably well aligned with the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay and measured with good accuracy and precision such that different exposure groups can be distinguished

Risk-of-Bias Criteria for Exposure Characterization (Case Control Studies)
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure • **AND** indirect evidence that exposure assessment does not adequately reflect relevant exposure levels (e.g., poor density of data, poor data quality, many missing values, substantial data misalignment), • **OR** there is evidence that the exposure was assessed using indirect measures that have not been validated or empirically shown to be consistent with methods that directly measure exposure (e.g., questionnaire, self-report without validation), • **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record "NR" as basis for answer). •
Definitely High Risk of Bias (--)
• Direct evidence that the exposure was assessed using methods with poor validity, • **AND** direct evidence that exposure assessment does not adequately reflect relevant exposure levels (e.g., poor density of data, poor data quality, many missing values, substantial data misalignment), • **OR** evidence of substantial exposure misclassification.

9. Can we be confident in the outcome assessment?

Risk-of-Bias Criteria for Outcome Assessment (Case Control Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that ~~pregnancy-associated hypertensive disorders or neurological development~~ were assessed in cases (i.e., case definition) and controls using well-established methods (the gold standard), **REVISION: hypertensive disorders of pregnancy** • **AND** subjects had been followed for the same length of time in all study groups, • **AND** there is direct evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the exposure level when outcome was assessed in cases (i.e., case definition) and controls. • **NOTE** Well-established methods will depend on the outcome, but examples of such methods may include: doctor diagnosis of asthma or doctor diagnosis obtained from medical records.
Probably Low Risk of Bias (+)
• Indirect evidence that the outcome was assessed in cases (i.e., case definition) and controls using acceptable methods), • **AND** subjects had been followed for the same length of time in all study groups, • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors were adequately blinded to the exposure level when reporting outcomes, • **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results (including that subjects self-reporting outcomes were likely not aware of reported links between the exposure and outcome or lack of blinding is unlikely to bias a particular outcome). • **NOTE** Acceptable, but not ideal assessment methods will depend on the outcome, but examples of such methods may include proxy reporting of outcomes such as asthma and mining of data collected for other purposes. Proxy reporting of immune disease should be considered on a case-by-case basis with consideration of whether or not there is empirical evidence as to the reliability of proxy reporting for that outcome.

Risk-of-Bias Criteria for Outcome Assessment (Case Control Studies)
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome was assessed in cases (i.e., case definition) using an insensitive instrument, • **OR** there is insufficient information provided about how cases were identified (record "NR" as basis for answer). • **OR** there is indirect evidence that it was possible for outcome assessors to infer the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were likely aware of reported links between the exposure and outcome), • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the outcome was assessed in cases (i.e., case definition) using an insensitive instrument, • **OR** there is direct evidence that outcome assessors were aware of the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were aware of reported links between the exposure and outcome).

10. Were all measured outcomes reported?

Risk-of-Bias Criteria for Outcome Reporting (Case Control Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).

Risk-of-Bias Criteria for Outcome Reporting (Case Control Studies)
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

Experimental Animal Studies

1. Was administered dose or exposure level adequately randomized?

Risk-of-Bias Criteria for Dose or Exposure Level Randomization (Experimental Animal Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that animals were allocated to any study group including controls using a method with a random component, • **AND** there is direct evidence that the study used a concurrent control group as an indication that randomization covered all study groups, • **Note:** Acceptable methods of randomization include: referring to a random number table, using a computer random number generator, coin tossing, or shuffling cards (2011). • **Note:** Restricted randomization (e.g., blocked randomization) to ensure particular allocation ratios will be considered low bias. Similarly, stratified randomization approaches that attempt to minimize imbalance between groups on important prognostic factors (e.g., body weight) will be considered acceptable.
Probably Low Risk of Bias (+)
• Indirect evidence that animals were allocated to any study group including controls using a method with a random component (i.e., authors state random allocation, without description of method), • **AND** evidence that the study used a concurrent control group as an indication that randomization covered all study groups, • **OR** it is deemed that allocation without a clearly random component would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that animals were allocated to study groups using a method with a non-random component, • **OR** indirect evidence that there was a lack of a concurrent control group, • **OR** there is insufficient information provided about how cells were allocated to study groups (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that animals were allocated to study groups using a non-random method including judgment of the investigator, the results of a laboratory test or a series of tests, • **OR** direct evidence that there was a lack of a concurrent control group.

2. Was allocation to study groups adequately concealed?

Risk-of-Bias Criteria for Allocation Concealment (Experimental Animal Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that at the time of assigning study groups the research personnel did not know what group animals were allocated to, and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable. • **Note:** Acceptable methods used to ensure allocation concealment include sequentially numbered treatment containers of identical appearance or equivalent methods.
Probably Low Risk of Bias (+)
• Indirect evidence that at the time of assigning study groups the research personnel did not know what group animals were allocated to and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable, • **OR** it is deemed that lack of adequate allocation concealment would not appreciably bias results.

Risk-of-Bias Criteria for Allocation Concealment (Experimental Animal Studies)
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that at the time of assigning study groups it was possible for the research personnel to know what group animals were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable, • **OR** there is *insufficient* information provided about allocation to study groups (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that at the time of assigning study groups it was possible for the research personnel to know what group animals were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable.

3. Did selection of study participants result in the appropriate comparison groups? [NA]

4. Did study design or analysis account for important confounding and modifying variables? [NA]

5. Were experimental conditions identical across study groups?

Risk-of-Bias Criteria for Experimental Conditions (Experimental Animal Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that same vehicle was used in control and experimental animals, • **AND** direct evidence that non-treatment-related experimental conditions were identical across study groups (i.e., the study report explicitly provides this level of detail).
Probably Low Risk of Bias (+)
• Indirect evidence that the same vehicle was used in control and experimental animals, • **OR** it is deemed that the vehicle used would not appreciably bias results, • **AND** identical non-treatment-related experimental conditions are assumed if authors did not report differences in housing or husbandry.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the vehicle differed between control and experimental animals, • **OR** authors did not report the vehicle used (record "NR" as basis for answer), • **OR** there is indirect evidence that non-treatment-related experimental conditions were not comparable between study groups.
Definitely High Risk of Bias (--)
• Direct evidence from the study report that control animals were untreated, or treated with a different vehicle than experimental animals, • **OR** there is direct evidence that non-treatment-related experimental conditions were not comparable between study groups.

6. Were the research personnel blinded to the study group during the study?

Risk-of-Bias Criteria for Blinding of Research Personnel to Study Group (Experimental Animal Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study. Methods used to ensure blinding include central allocation; sequentially numbered treatment containers of identical appearance; sequentially numbered animal cages; or equivalent methods,

Risk-of-Bias Criteria for Blinding of Research Personnel to Study Group (Experimental Animal Studies)
Probably Low Risk of Bias (+)
• Indirect evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study, • **OR** it is deemed that lack of adequate blinding during the study would not appreciably bias results. This would include cases where blinding was not possible but research personnel took steps to minimize potential bias, such as restricting the knowledge of study group to veterinary or supervisory personnel monitoring for overt toxicity, or randomized husbandry or handling practices (e.g., placement in the animal room, necropsy order, etc.).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the research personnel were not adequately blinded to study group, • **OR** there is insufficient information provided about blinding to study group during the study (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the research personnel were not adequately blinded to study group.

7. Were outcome data complete without attrition or exclusion from analysis?

Risk-of-Bias Criteria for Data Completeness (Experimental Animal Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that loss of animals was adequately addressed and reasons were documented when animals were removed from a study. • **Note**: Acceptable handling of attrition includes: very little missing outcome data; reasons for missing animals unlikely to be related to outcome (or for survival data, censoring unlikely to be introducing bias); missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups; missing outcomes is not enough to impact the effect estimate. • **OR** missing data have been imputed using appropriate methods (insuring that characteristics of animals are not significantly different from animals retained in the analysis).
Probably Low Risk of Bias (+)
• Indirect evidence that loss of animals was adequately addressed and reasons were documented when animals were removed from a study, • **OR** it is deemed that the proportion lost would not appreciably bias results. This would include reports of no statistical differences in characteristics of animals removed from the study from those remaining in the study.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that loss of animals was unacceptably large and not adequately addressed, • **OR** there is insufficient information provided about loss of animals (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that loss of animals was unacceptably large and not adequately addressed. • **Note**: Unacceptable handling of attrition or exclusion includes: reason for loss is likely to be related to true outcome, with either imbalance in numbers or reasons for loss across study groups.

8. Can we be confident in the exposure characterization?

Risk-of-Bias Criteria for Exposure Characterization (Experimental Animal Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that exposure to air pollutant(s) were independently characterized and purity

Risk-of-Bias Criteria for Exposure Characterization (Experimental Animal Studies)
confirmed generally as ≥98% (when applicable), • **AND** that exposure was consistently administered (i.e., with the same method and time-frame) across treatment groups, • **AND** for inhalation or aspiration studies that information is provided on consumption or internal dose metrics to confirm expected exposure levels sufficiently to allow discrimination between exposure groups, • **AND** if internal dose metrics are available, there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished, • **AND** if internal dose metrics are available, the study used spiked samples to confirm assay performance. • **Note:** Air pollutant exposure should be characterized in the following manner: 1) concentration of test article in the treated and control chambers was monitored and recorded at a minimum of once per hour; 2) average chamber concentration did not vary by more than ± 10% from the target concentration from exposure period to exposure period, and the daily average chamber concentration did not vary by more than ± 10% relative standard deviation (RSD); 3) it was confirmed by photometric or other appropriate means, that the test atmosphere does not contain aerosolized test article; and 4) particle size distribution was controlled and monitored, and had a sigma g of less than 3. • **Note:** Environmental gases will not have internal dose metrics.
Probably Low Risk of Bias (+)
• Indirect evidence that the air pollutant(s) were appropriately characterized and purity confirmed generally as ≥98% (when applicable), • **OR** direct evidence that purity was independently confirmed as ≥95% and it is deemed that impurities of up to 5% would not appreciably bias results (when applicable), • **AND** that exposure was consistently administered (i.e., with the same method and time-frame) across treatment groups, • **AND** for inhalation or aspiration studies no information is provided on consumption or internal dose metrics, • **AND** if internal dose metrics are available, there is indirect evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure (including purity of the test substance and compliance with the treatment, if applicable) was assessed using poorly validated methods, • **OR** there is insufficient information provided about the validity of the exposure assessment method, but no evidence for concern (record "NR" as basis for answer), • **AND** if internal dose metrics are available, there is indirect evidence that most of the exposure data measurements are below the limit of quantitation for the assay such that different exposure groups cannot be distinguished.
Definitely High Risk of Bias (--)
• Direct evidence that the exposure (including purity of the test substance and compliance with the treatment, if applicable) was assessed using poorly validated methods.

9. Can we be confident in the outcome assessment?

Risk-of-Bias Criteria for Outcome Assessment (Experimental Animal Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that ~~pregnancy-associated hypertensive disorders~~ ~~or neurological development~~ were assessed using well-established methods (e.g., gold standard) **REVISION: hypertensive disorders of pregnancy** • **AND** assessed at the same length of time after initial exposure in all study groups, • **AND** there is direct evidence that the outcome assessors were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes. • **NOTE** Well-established methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits, commercial laboratories with experience in the assay, or standard assays such as ELISAs for IgG and with sufficiently low variation and limits of detection to allow discrimination of responses between treatment groups (or direct evidence that the assay could have detected a difference based on responses to a positive control).
Probably Low Risk of Bias (+)
• Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard), • **AND** assessed at the same length of time after initial exposure in all study groups, • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes, • **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results, which is more likely to apply to objective outcome measures. • **NOTE** For some outcomes, particularly histopathology assessment, outcome assessors are not blind to study group as they require comparison to the control to appropriately judge the outcome, but additional measures such as multiple levels of independent review by trained pathologists can minimize potential bias. • **NOTE** Acceptable assessment methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits with some variation, but ability to discriminate between the high dose treatment and control group (or indirect evidence that the assay could have detected a difference based on responses to a positive control).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome assessment method is an insensitive instrument, • **OR** the length of time after initial exposure differed by study group, • **OR** there is indirect evidence that it was possible for outcome assessors to infer the study group prior to reporting outcomes without sufficient quality control measures, • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the outcome assessment method is an insensitive instrument, • **OR** the length of time after initial exposure differed by study group, • **OR** there is direct evidence for lack of adequate blinding of outcome assessors, including no blinding or incomplete blinding without quality control measures.

10. Were all measured outcomes reported?

Risk-of-Bias Criteria for Outcome Reporting (Experimental Animal Studies)
Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as answer basis).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.