Applying the Navigation Guide Systematic Review Methodology
Case Study #7

Association between Prenatal Exposures to Ambient Air Pollution and Birthweight

A Systematic Review of the Evidence
Protocol
March 2017

PROTOCOL INFORMATION

Authors

Weihsueh Chiu, PhD
Professor
Department of Veterinary Medicine and Biomedical Sciences
Texas A&M University

Natalie Johnson, PhD
Assistant Professor
Department of Environmental and Occupational Health
Texas A&M University School of Public Health

Megan Moriarty, MPH
Graduate Assistant
Department of Environmental and Occupational Health
Texas A&M University School of Public Health

Jairus Pulczinski, MPH
Graduate Assistant
Department of Environmental and Occupational Health
Texas A&M University School of Public Health

Inyang Uwak, MPH, DrPH
Graduate Assistant
Department of Environmental and Occupational Health
Texas A&M University School of Public Health

Samuel Taiwo, MPH
Environmental Health and Safety Specialist
Environmental Health and Safety
Texas A&M University

Juleen Lam, PhD, MHS, MS
Associate Research Scientist
Department of Obstetrics, Gynecology & Reproductive Sciences
Program on Reproductive Health and the Environment
University of California, San Francisco

Brandie Taylor, PhD, MPH
Assistant Professor
Department of Epidemiology
Director of Program on Reproductive and Child Health
Texas A&M University School of Public Health

Xiaohui Xu, PhD
Associate Professor
Department of Epidemiology
Texas A&M University School of Public Health

TABLE OF CONTENTS

BACKGROUND

The Navigation Guide Systematic Review Methodology

Robust methods to evaluate available scientific evidence to reach conclusions regarding the strength of evidence are fundamental to speeding the translation of the science into policies and decisions to improve health outcomes. In the clinical sciences, methods of evidence integration have played a transformative role in the timely incorporation of science into therapeutic, preventive and cost effective action at the individual and societal level (Fox 2010). Beginning in 2009, researchers began to explore the application of systematic and robust methods of evidence integration in environmental health sciences (Woodruff et al. 2011, Rooney et al. 2014). In 2014, two reports by the National Academy of Sciences (NAS) strongly endorsed the adoption of such improved methods of evidence integration in environmental health sciences, and specifically encouraged their use by the US Environmental Protection Agency (EPA) in determinations of whether environmental chemicals are harmful to human health (National Research Council 2014, National Research Council 2014). Currently, the US EPA is incorporating principles of systematic review into its IRIS process (National Research Council 2014, US Environmental Protection Agency 2014), while the National Institute for Environmental Health Sciences' (NIEHS) National Toxicology Program (NTP) has been developing the tools, expertise, case studies, and other infrastructure that will facilitate increased utilization of systematic review methodologies (Rooney et al. 2014).

The Navigation Guide systematic review methodology (Navigation Guide) was developed in 2011 as part of an interdisciplinary collaboration between clinicians, academicians, and practitioners in an attempt to harmonize the approaches for assessing evidence in the clinical sciences with environmental health sciences (Woodruff et al. 2011). The Navigation Guide is a systematic and transparent methodology that proceeds from best practices in the clinical arena but accounts for the differences in evidence and decision context involved in environmental health risk assessments, such as the reliance on animal toxicology and human observational studies in the absence of randomized controlled trials (RCTs). To date, the Navigation Guide systematic review methodology has been applied in five proof-of-concept studies:

1. To evaluate the human and non-human evidence of perfluorooctanoic acid (PFOA) on fetal growth (Johnson et al. 2014, Koustas et al. 2014, Lam et al. 2014). From this application of the Navigation Guide systematic review methodology, review authors concluded that PFOA was "known to be toxic" to human reproduction and development, based on a finding of "moderate" quality and "sufficient" strength of both the human and non-human mammalian evidence.

2. To evaluate the human and non-human evidence of the association between fetal growth and glomerular filtration rate (GFR) in order to assess the strength of the evidence of a 'reverse causality' hypothesis: that the size of a developing fetus may affect maternal GFR such that a small fetus leads to reduced plasma volume expansion, reduced GFR, and subsequently higher

concentrations of biomarkers in maternal serum. This had been proposed as a potential alternate explanation for observational studies documenting an inverse association between prenatal exposure to chemicals cleared renally and fetal growth (Savitz , Whitworth et al. 2012, Loccisano et al. 2013). The authors of this review found insufficient evidence to support the plausibility of the reverse causality hypothesis and recommended further high quality research (Vesterinen et al. 2014).

3. To evaluate the human and non-human evidence of triclosan on reproductive and/or developmental toxicity. Authors concluded that there was "sufficient" non-human evidence and "inadequate" human evidence of an association between triclosan exposure and thyroxine concentrations, and as a result, triclosan is "possibly toxic" to reproductive and developmental health (Johnson et al. 2016).

4. To evaluate the human evidence of the relationship between polybrominated diphenyl ethers (PBDEs) on human neurodevelopment and quantitative measures of intelligence or ADHD and attention-related behavioral conditions. This review has been completed and the manuscript is in preparation.

5. To evaluate the human evidence for effects of exposure to airborne environmental contaminants on the diagnosis of Autism Spectrum Disorder (ASD). Authors concluded that there was "limited evidence of toxicity" for the association between early life exposure to air pollution as a whole and diagnosis of ASD. However, there was strong evidence supporting the link between prenatal exposure to particulate matter and ASD (Lam et al., 2016).

6. To evaluate the human evidence for effects of exposure to formaldehyde on asthma outcomes in children and adults. This review has been completed and the manuscript is in preparation.

The results of these case studies demonstrate that the methods under development by the USEPA and the NTP are fully achievable (Johnson PI et al. 2014, Koustas et al. 2014, Lam et al. 2014, Woodruff and Sutton 2014).

This 7th case study of the Navigation Guide systematic review method in environmental health will assess the human evidence for effects of exposure to airborne environmental contaminants on the effects of birth weight. The human health rationale for this review is based on the pervasiveness of human exposure to airborne environmental chemicals and human evidence of developmental health impacts of such exposures, as described below.

Rationale for Review: Air Pollution and Birthweight outcomes

Exposure to Air Pollution and Birth weight

We selected particulate air pollution for evaluation based on pervasive human exposure and the considerable amount of evidence supporting associations with low birth weight (LBW). Numerous case-control and cohort studies have demonstrated a positive association between developmental exposures to ambient air pollution and birth weight; however, there have also

been inconsistencies in the conclusions about the association and magnitude of effect. Most reviews published in the past have been based on a small number of studies and were not able to draw conclusions on effect size (Bonzini et al., 2010; Bosetti et al., 2010; Ghosh et al., 2007). More recent systematic reviews that included meta-analysis of a larger number of studies showed pooled estimates of effect size for LBW for a 10 μg/m^3 increase in $PM_{2.5}$ exposure during entire pregnancy ranged from -15.9 g [-26.8, -5.0] (Sun) to -22.17 g [-37.93, -6.41] (Lamichhane) to -23.4 [95% CI, -45.5, -1.4] (Steib). Reviewers cited they were not able to *rule out the consequences of specific biases which may arise from differences in the methodological quality and the study settings,* including consistent strategies for controlling for consistency of characteristics across study populations, ensuring against exposure misclassification, and controlling for confounding. Based on the significant heterogeneity across studies and potential biases in individual studies, we decided to apply the Navigation Guide systematic review methodology to assess the impact of particulate air pollution on birth weight and identify how risk of bias may impact estimated effect size. The accumulating evidence published to date makes this topic an ideal case study for applying the Navigation Guide methodology.

AIM

Study Question

"Does developmental exposure to ambient particulate air pollution affect Birth Weight?"

Study Objectives

- Identify studies or experiments conducted in humans concerning the association of developmental exposure to ambient particulate air pollution with birth weight;
- Assess the risk of bias of individual studies and, where appropriate, assess their impact (including direction) on measures of estimated effect size;
- Conduct a meta-analysis of the size of the effect of exposure to ambient air pollution on birth weight and assess for potential sources of heterogeneity;
- Rate the strength of the human evidence on the effect of ambient air pollution on birth weight according to one of the following four statements: 1. Sufficient; 2. Limited; 3. Inadequate; or 4. Evidence of lack of toxicity.

METHODS

Review Team

Review Team Co-Authors

At the beginning of the case study, Texas A&M will assemble a review team consisting of experts and graduate students from a variety of research fields relevant to the study question at hand (i.e., epidemiology, air pollution/exposure assessment, biostatistics, and/or systematic review methodology). **Every member of our review team** will actively participate in the critical steps of the case study—i.e., developing/approving the protocol, evaluating the quality of evidence, and rating the strength of the evidence. However, in the event that a member of the review team was a coauthor of a study under review, that member must recuse themselves from the evaluating the quality and strength of that study.

Criteria for Selecting Studies

We will select studies in which exposure to ambient air pollution was documented, measured, or estimated, and birth weight was evaluated.

Studies that are eligible for review will address the study question and the characteristics as outlined in the following "PECO" aid.

PECO Statement

"PECO" is an aid used to formulate an answerable question in a systematic review of health studies. The acronym stands for "**P**opulation/Participants", "**E**xposure," "**C**omparator" and "**O**utcomes."

Population: Humans

Exposure: Developmental exposure to ambient particulate air pollution that occurred prior to Birth.

"Developmental exposure" is defined as maternal or paternal exposure incurred during the perinatal period.
"Particulate air pollution" is defined as an outdoor source of an inhaled airborne environmental chemical classified as PM 2.5 or PM10, EXCLUDING active and passive smoking.

Comparator: Humans exposed to lower levels of air pollution than the more highly exposed humans.

Outcome: Birth weight measured as a continuous variable.

Search Methods

We will collaborate with a Systematic Review Coordinator who has training, expertise, and familiarity with developing and performing systematic review literature searches (see Appendix I for MF' biosketch and CV). We will employ a variety of methods to identify relevant data, as outlined below. Our search will not be limited by publication date.

We will perform electronic searches of online databases (Ovid Medline, Embase, and Global Health) using the search terms outlined in Appendix II. Our search strategy and search terms will be developed by a trained librarian who will implement the search for relevant studies.

To assist in the development of a list of terms relevant to our search strategy we will use the Medical Subject Headings (MeSH) database to compile synonyms for ambient air pollution and birthweight-related outcomes.

Ovid MEDLINE

For the exposure, we separated the search into three categories using the terms Particulate Matter, pm 2* or pm 10 and particulate matter or pollut*. These terms were combines using the "OR" statement to create a collection of exposure search terms.

For the outcome, we combined Birth Weight along with low birth weight and its synonyms. We searched using the ti,ab. Function.

We combined the exposure terms and outcome terms using the search terms as well as the "AND" statement to implement the search for papers.

Ovid will be considered our primary online database. Records from subsequent database searches will be first compared to the Ovid set then to other databases already searched to identify and remove duplicates. We will document the number of records retrieved with each search and the total number of duplicates removed, as well as the database where the duplicate being removed originally occurred. This process will be completed using RefWorks and EndNote.

Embase

We will develop our Embase search filter modifying the Ovid MEDLINE search terms. We will use the "ti,ab." function to limit the search to titles and abstracts.

Global Health

We will develop our Global Health search filter using a similar method as described for Ovid Medline and Embase previews modifying the search terms as acceptable to the engine.

Study Selection Criteria

All search results will be imported or manually entered into RefWorks reference management software. We will use RefWorks to eliminate any duplicate references before we begin evaluating the eligibility of the studies identified.

Title and abstract screening

Each reference will be screened in duplicate and independently. Two reviewers (MM, JP) will independently conduct a title and abstract review of each reference from the literature search results to determine whether it meets the selection criteria for inclusion. Each author will screen the entirety of the references, to ensure that all references are screened in duplicate.

References which are included at the title/abstract screening level will be subject to a full text review by two authors (the same authors above MM and JP).

In the event that there is a discrepancy between reviewers, the default will be to push the reference forward to the next step in the process (i.e., if the two reviewers disagree on whether the study is relevant at the title and abstract screening level, the reference will be included by default for full-text screening).

To ensure quality control, each reviewer was blinded from the others response to the studies. These determinations will then be compared to the other reviewers' determinations for these studies.

Reports in English, from any year, will be eligible for inclusion. All reports that compare humans exposed to ambient air pollution to appropriate comparators and evaluate them for Birth weight, as described in the PECO statements above, will be eligible for inclusion.

The title/abstract screening form will be used first to initially screen references. Studies will be EXCLUDED if one or more of the following criteria are met:

1. Article did not report birth weight;
2. Article did not report ambient air pollution exposure;
3. Article contains no original data (e.g., editorial, review etc.)
4. Article did not involve human subjects (i.e., animal evidence only);
5. Other reason (explanation required).

The following instructions were provided to review authors conducting the title and abstract screening:

"When excluding a reference, please select only ONE (1) exclusion reason. Please review the exclusion reasons in order and select the FIRST exclusion reason relevant to the reference being screened. Please add in any additional notes in the comment box to explain your selection if necessary."

The following types of records will be INCLUDED at the title/abstract level:

- Any study conducted in humans on ambient air pollution exposures and birth weight and
- Studies with humans "exposed" to ambient air pollution, even if ambient air pollution exposure levels are not quantified.

For citations where the database contains no abstract, authors will attempt to obtain the abstracts from an Internet search. Articles for which the abstract remains unavailable will be screened based on titles and headings. In cases where titles and headings do not provide sufficient information based on above criteria the study will be included for full-text review.

Full-Text Screening

References included at the title/abstract screening level will then be subject to a full text review by two authors (the same authors involved in title and abstract screening MM and JP). Each reference will be screened independently as all duplicates were removed in the abstract screening.

The authors discussed discrepancies between the reviewers resulting from each step of the review process if necessary. In the event that the discrepancy cannot be resolved, the default will be to push the reference forward to the next step in the process (i.e., if the two reviewers disagree on whether the study is relevant at the full text screening level, the reference will be included by default).

To ensure quality control, the authors blind screened the articles during different times.

Citations eligible for full text review will be screened and EXCLUDED if one or more of the following criteria are met:

1. Article did not report continuous birth weight outcome;
2. Article did not report ambient air pollution exposure;
3. Article contains no original data (e.g., editorial, review etc.)
4. Article did not involve human subjects (i.e., animal evidence only);
5. Other reason (explanation required).

The following instructions were provided to review authors to following full text screening:

"When excluding a reference, please select only ONE (1) exclusion reason. Please review the exclusion reasons in order and select the FIRST exclusion reason relevant to the reference being screened. Please add in any additional notes in the comment box to explain your selection if necessary."

Citations will be INCLUDED if they meet the PECO statement criteria, that is, their subjects represent humans, they include exposure comparisons for relevant periods, and they report the outcome of birth weight.

For articles that are not available in the database, we will attempt to obtain articles from a broad Internet search or using the available articles in the TAMU "get it for me" database.

Data Collection

WA Chiu, I Uwak, and S Taiwo will be involved in data extraction. For each included article, data relevant to the outcomes assessed will be extracted into the Health Assessment Workspace Collaborative (HAWC) database (see Appendix III for the data collection form). The data extracted by each author will be independently reviewed by a second author for quality assurance/quality control. Under the direction of a third author, authors will resolve any discrepancies. The extracted data will be used to develop visual or tabular displays of study characteristics and/or results, to evaluate reporting quality and risk of bias and to conduct statistical analyses. For every study that does not report all the data needed for data analysis, we will request this data from the study contact author by email. If study authors do not respond to requests after being contacted through 2 email messages over the course of 1 month, review authors will note that attempts to contact study researchers were unsuccessful.

Risk of Bias Determination

Risk of bias (ROB) will be assessed for human studies using domains from the Cochrane Collaboration's "Risk of Bias" tool and the Agency for Healthcare Research and Quality's (AHRQ) criteria (Higgins and Deeks 2011, Viswanathan et al. 2012). These tools have been modified to make them appropriate for human observational studies, and include domains that address recruitment strategy, blinding, confounding, exposure assessment, incomplete outcome data, selective outcome reporting, and conflict of interest (Appendix VI). These tools have been modified and applied to evaluate risk of bias in previous case studies applying the Navigation Guide systematic review methodology (Johnson et al. 2014, Johnson et al. 2014).

Informed by empirical data from meta-analyses conducted on pharmacological treatments and studies of risk of bias and sponsorship (Roseman et al. 2011, Lundh et al. 2012, Krauth et al. 2013), we additionally assess funding source and declared conflicts of interest as potential sources of bias. We will also search for each study in PubMed and note if there has been a retraction of the published article in order to determine if the study may be fraudulent or if any corrections have been published.

Two review authors from our review team will independently make risk of bias determinations for each study. Each review author will be assigned a set of studies and will rate these across all ROB domains. An additional QA/QC author will be matched with each study. Any discrepancies will be reviewed by the QA/QC author and discussed among all three reviewers. Any remaining discrepancies will then be reviewed by all other review authors. If, upon further discussion the review authors cannot reach agreement on an appropriate risk of bias determination for a particular domain, the rating judgment will be selected as follows: if one reviewer makes a judgment of 'high' risk of bias and the other makes a judgment of 'probably high' risk of bias, the 'probably high' risk of bias judgment will be used, etc. If additional data or information is acquired from study authors, risk of bias judgments will be modified to reflect the updated study information.

Risk of bias criteria regarding exposure assessment methodology for evaluating air pollution studies.

Empirical evidence exists that risk of bias varies depending on how the exposure was measured. Thus, there is a need to transparently distinguish among those potential biases across different studies. Therefore, we consulted with Dr. K. Koehler to design more specific considerations to evaluate the potential risk of bias for exposure assessment, outlined in Appendix VI. While there is no other empirically-based method to address this aspect of the data, the Review Team includes a recognized experts in the field of air pollution (X Xu, see Appendix I).

Publication Bias.

We will attempt to minimize the impact of publication bias by: (1) implementing a comprehensive search of the published and unpublished literature using multiple sources and methods in order to identify all published studies that meet the eligibility criteria; and (2) if possible, using funnel plot analysis and/or other statistical analyses (e.g. Egger regression (Light and Pillemer 1984) and "trim and fill" (Duval and Tweedie 2000) of the published studies included in the systematic

review), as appropriate. These statistical approaches have been recommended only when the number of studies included in the meta-analysis is sufficiently large (Sterne et al. 2011), and so these analyses will only be performed when >10 studies are included in the meta-analysis. In the event that the number of studies included in the meta-analysis is too small these analytical approaches will not be pursued. Furthermore, in the event of substantial between-study heterogeneity, these methods are known to perform poorly and so we will test for between-study heterogeneity as well to make the determination of whether this method would be appropriate for the collection of included studies (Higgins 2011).

Data Analysis

Where appropriate we (W Chiu and J Lam) will perform a meta-analysis to summarize the effects of exposure to ambient particulate air pollution on birth weight outcome and to assess the impact of study design characteristics on findings. Characteristics from each study will be compiled and reviewed to establish comparability between studies or to identify data transformations necessary to ensure such comparability. Key characteristics include:

- Study design
- Details on how participants were classified into exposure groups, if any (e.g. quartiles of exposure concentrations)
- Details on source of exposure data (questionnaire, air monitoring, biomonitoring, etc.)
- Exposure levels, method of measurement, timing of measurement
- Type of data/summary statistic available

Summaries of these characteristics for each included study will be assessed by two or more review authors to determine comparability between studies and to identify any heterogeneity concerns. Where appropriate, studies with sufficient methodological homogeneity with respect to population, study design, study duration, exposure level and health outcome among other considerations will be combined in a meta-analysis.

If a meta-analysis is deemed appropriate, W Chiu and J Lam will identify appropriate statistical methods to analyze the data, and to determine whether further modifications are required prior to performing the meta-analysis. Our proposed approach is to calculate a pooled estimates of risk estimates (and their 95% confidence intervals) associated with a common exposure unit (e.g., grams change in birth weight per 10 microg/m^3 increase in PM10 or PM2.5) from the included studies using random-effects models, which incorporate both within- and between-study variation, using a restricted maximum likelihood estimator as implemented in the R "metafor" package. We will present the pooled effect estimates and study-specific estimates in a forest plot where the size of the marker corresponds to the inverse of the variance of the effect estimate from each study, and a diamond indicates the overall pooled effect estimate.

Heterogeneity will be assessed using the I-squared statistic as well as tau, the measure of the standard deviation of the random effect. Interpretation of I-squared will be based on the Cochrane Handbook: 0% to 40% (might not be important); 30% to 60% (may represent moderate heterogeneity); 50% to 90% (may represent substantial heterogeneity); 75% to 100% (considerable heterogeneity). Additionally, as described in the Cochrane Handbook, for the last three categories, the importance of the I-squared will be interpreted considering not only the

magnitude of effects but also the strength of the evidence (90% two-tailed confidence interval). The value of tau will be considered relative to the estimated overall effect size.

The review team will also perform sensitivity analyses on the following aspects:

- Sensitivity to exclusion of individual studies in succession,
- Sensitivity to alternative exposure metrics (if available), and
- Sensitivity to alternative outcome metrics (if available).

We will also examine the variation in effect estimates among the studies by carrying out subgroup analyses stratified by key characteristics described above. Assessment of publication bias will be performed as described above in the previous section.

In the event that these proposed methods for data analysis are altered to tailor to the evidence base from included studies, the protocol will be amended accordingly and the reasons for change will be justified in the documentation.

Quality and Strength of Evidence Ratings

Upon completion of the data collection, risk of bias determination, and data analysis, each of the co-authors will compare the results of the systematic review to the criteria in the Navigation Guide systematic review methodology for rating the quality and strength of the evidence. All co-authors will be given explicit directions before rating (see Appendix V, "Instructions for Rating the Quality and Strength of Evidence").

The initial quality level of human observational data will be considered moderate, as has been assigned in prior case studies of applying the Navigation Guide methodology (Woodruff and Sutton 2014).

Factors that may decrease the quality level of the body of evidence include:

1. Risk of Bias Across Studies: Study limitations – a substantial risk of bias across body of evidence;
2. Indirectness: Evidence was not directly comparable to the question of interest (i.e., population, exposure, comparator, and outcome).
3. Inconsistency: Widely different estimates of effect (heterogeneity or variability in results);
4. Imprecision: Studies had few participants and few events (wide confidence intervals); and
5. Publication Bias: Studies missing from body of evidence, resulting in an overestimate or underestimate of true effects from exposure.

Factors that may increase the quality level of the body of evidence include:

1. Large magnitude of effect: Upgraded if modeling suggested confounding alone unlikely to explain associations with large magnitude of effect.
2. Dose-response: Upgraded if consistent dose response gradient in one or multiple studies, and/or dose response across studies.

3. Confounding minimizes demonstrated effect: Upgraded if consideration of all plausible residual confounders or biases would underestimate the effect or suggest a spurious effect when results show no effect.

Possible ratings for quality of evidence are "high," "moderate," or "low." Possible downgrades or upgrades are: 0 (no change), -1 (1 level downgrade), – 2 (2 level downgrade), +1 (1 level upgrade) or +2 (2 level upgrade). The ratings of the separate factors are not added together into a score, e.g. a -1 downgrade for inconsistency and a -1 downgrade for imprecision does not automatically dictate an overall -2 downgrade for the body of evidence. Judgment is exorcised to determine if the rationale behind each downgrade warrants an overall downgrade of 1 or 2 levels. The same applies to upgrading the overall body of evidence. Likewise, a -1 downgrade for one factor and a +1 upgrade for another factor do not automatically cancel out and determine no downgrades or upgrades for the overall body of evidence.

Authors who decide to rate quality down or up need to specify the 1 or 2 criteria most responsible for their decision while documenting all factors that contributed to the final decision. After independently evaluating the quality of the evidence, co-authors will compare their evaluations and any discrepancies between the reviewers' decisions will be resolved through discussion until consensus on the overall quality of the body of evidence is reached. The rationale for each decision on each of the five factors will be recorded.

Subsequent to consensus on the quality of the evidence, the review authors will rate the strength of evidence. The overall strength of the body of human evidence is based on a combination of four criteria: (1) Quality of body of evidence (i.e., the rating from the previous step); (2) Direction of effect; (3) Confidence in effect; and (4) Other compelling attributes of the data that may influence certainty. The results of these four criteria are summarized according to one of the following four concluding statements: 1. Sufficient; 2. Limited; 3. Inadequate; or 4. Evidence of lack of toxicity (Appendix VII, Table 1) (Woodruff et al. 2011). Any discrepancies between the reviewers' decisions will be resolved through discussion. The senior authors (W Chiu, N Johnson) will be the ultimate arbiter of the discrepancies that cannot be resolved through consensus among the review authors. The results of the review, including implications for public health, will be compiled in a manuscript for submission to the peer-review literature.

SUPPLEMENTARY INFORMATION

Appendix I. Coauthor/Liberian Biosketches and Conflict of Interest Statements

WEIHSUEH A. CHIU

Weihsueh A. Chiu, Ph.D., is a professor in the Department of Veterinary Integrative Biosciences in the College of Veterinary Medicine and Biomedical Sciences at Texas A&M University. His academic background includes an AB in Physics from Harvard University, a PhD in Physics from Princeton University, and a Certificate in Science, Technology, and Environmental Policy from the Woodrow Wilson School of Public and International Affairs at Princeton University. Dr. Chiu's previous experience includes 2 years at the U.S. Government Accountability Office and over 14 years at the U.S. Environmental Protection Agency. His research has focused on human health risk assessment, particularly with respect to toxicokinetics, mechanisms of toxicity, physiologically-based pharmacokinetic modeling, dose-response assessment, and characterizing uncertainty and variability. He has served as an expert peer reviewer for multiple academic journals, including as Associate Editor for Environmental Health Perspectives, as well as for the National Toxicology Program, the Agency for Toxic Substances and Disease Registry, and the California Environmental Protection Agency. He has also served on multiple committees and workgroups for the World Health Organization (WHO) International Agency for Research on Cancer, the WHO International Program on Chemical Safety, as well as for the U.S. National Academy of Sciences. As a member of these workgroups, he has developed and/or implemented systematic review-based approaches for evaluating carcinogenicity and endocrine-related health effects from exposures to environmental chemicals. He chaired a peer review committee for the National Toxicology Program reviewing their systematic review of the potential immunotoxicity effects of PFOA and PFOS. Dr. Chiu declares no conflicts of interest.

NATALIE M. JOHNSON

Natalie M. Johnson, Ph.D., is an assistant professor in the Department of Environmental and Occupational Health in the Texas A&M University School of Public Health. She received her PhD in toxicology from Texas A&M University and completed a postdoctoral fellowship at the Johns Hopkins School of Public Health. She has extensive training in animal modeling and conducting exposure assessment applying chemical-specific biomarkers in underserved communities. Her current research is focused on investigating the link between prenatal exposure to air pollution

and adverse health outcomes in later in life, mainly the development of childhood asthma. This work includes experimental models, exploration of underlying mechanisms, and exposure assessment in vulnerable communities along the Texas-Mexico border and Eastern China. Dr. Johnson declares no conflicts of interest.

MEGAN A. MORIARTY

Megan A. Moriarty is a master's student in Environmental Health in the School of Public Health at Texas A&M University. Her academic background includes a BS in Biology from Texas A&M with a minor in business. She has previously worked in Dr. Weston Porter's lab assisting in research on the Sim2 gene and its relationship to breast cancer. Additionally, Megan completed her practicum at HollyFrontier where she assessed safety measures and completed an in-house audit at the Navajo Refinery. She serves as the Treasurer of the Environmental Sustainability Group where they educate on sustainable practices and environmental disparities. Ms. Moriarty declares no conflicts of interest.

INYANG UWAK

Inyang Uwak, MBBS, MPH, is a doctoral student and research assistant in Epidemiology and Environmental Health at Texas A&M University School of Public Health. Her academic background includes a Medical degree from the University of Calabar, Nigeria, a Certificate in Tropical Medicine & Public Health and a Master of Public Health degree from Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland. Inyang's previous experience includes 4 years in clinical practice as a primary care physician in Nigeria. She has also worked at the Harris County Public Health and Environmental Services (HCPHES) in Pasadena, Texas and was involved with data analysis and conducting Health Impact Assessments (HIA) in communities in Houston. Her research interests include cancer prevention, human health risk assessment and environmental toxicology. She is a member of EpiAssist a student volunteer group that provides health departments with assistance during community assessments, outbreak investigation and public health emergencies. Ms. Uwak declares no conflicts of interest.

SAMUEL O. TAIWO

Samuel O. Taiwo is an Environmental Health and Safety Specialist, providing services to faculty, staff and students of Texas A&M University, College Station, Texas. He is a part of the Industrial Hygiene group of the Environmental Health and Safety Department and specializes in ergonomics, asbestos inspection and management planning, indoor air quality consulting, respirator fit-testing, CPR training, audiometric testing, campus AED maintenance, emergency response, among other health and safety functions. His academic background includes a Bachelor of Science in Biochemistry from Olabisi Onabanjo University, Nigeria, and a Master of Public Health in Environmental Health in Texas A&M School of Public Health, College Station, Texas. Mr. Taiwo briefly worked with the Toxicology Division of the Texas Commission on Environmental Quality where he co-authored a Development Support Document for Diethylamine. Mr. Taiwo's research interest lies in environmental/inhalation/developmental toxicology. Mr. Taiwo declares no conflicts of interest.

JULEEN LAM

Juleen Lam is an Associate Research Scientist at the University of California at San Francisco (UCSF). She has been involved with the Navigation Guide since 2011 while employed at the Environmental Protection Agency's Office of Policy at the National Center for Environmental Economics as an Oak Ridge Institute for Science and Education (ORISE) postdoctoral fellow and as a researcher at the Johns Hopkins University Bloomberg School of Public Health in the Department of Health, Policy and Management. She has been involved in two case studies to date applying the Navigation Guide to address problems in the field of environmental health. Juleen received her PhD from Johns Hopkins University in Environmental Health Policy, MHS from Johns Hopkins University in Biostatistics, MS from George Washington University in Environmental Engineering Management, and two BS degrees from the University of California at Davis in Math and Environmental Toxicology. She has over a decade of experience in environmental health research and policy, holding positions at state and federal government agencies, academic institutions, and in the consulting and nonprofit sectors. She specializes in analysis of environmental health data and focuses her research on the translation of scientific findings into making informed decisions and policies. Dr. Lam declares no conflicts of interest.

BRANDIE D. TAYLOR

My training has provided me with a strong foundation in epidemiologic methods and biostatistics with particular emphasis on immune system alterations in reproductive and pregnancy complications. As a NIH T32 predoctoral fellow (2009-2011) in Reproductive, Perinatal, and Pediatric Epidemiology, I focused on identifying immunological and genetic markers involved in *C. trachomatis* pathogenesis. In 2011, I received a NIH T32 postdoctoral fellowship in Perinatal Epidemiology at Michigan State University. During this time, I utilized a large population based cohort to identify inflammatory markers for preterm birth prediction. In 2012, I returned to the University of Pittsburgh as a postdoctoral associate and examined associations between immune system biomarkers and preeclampsia. Currently, I am an Assistant Professor of Epidemiology at Texas A&M University. My current projects are multidisciplinary and continue to focus on the immune system in reproductive and pregnancy health. Overall, my work has led to 22 publications in the field of reproductive and perinatal epidemiology. Dr. Taylor declares no conflicts of interest.

XIAOHUI XU

I am currently Associate Professor in the Department of Epidemiology and Biostatistics at Texas A&M University, School of Public Health and have received excellent training in epidemiology, biostatistics, toxicology and medicine. As environmental epidemiologist, I have strong methodological and analytical skills in epidemiological research including study design, GIS analysis, and analyzing different format data with complex statistical models such as longitudinal data, spatial data, data from the complex sampling surveys and administrative resources. My epidemiological research primarily focuses on studying health effects of environmental exposures, particularly ambient air pollution exposure on chronic diseases and maternal and child health

outcomes. Over the past decade, I have conducted many epidemiological studies on assessing health effects of ambient air pollution exposure in both China and USA. I am currently the recipient, as Principal Investigator, of a career development grant from the NIEHS to investigate the health effects of air pollution on adverse pregnancy outcomes in the state of Florida. I have also conducted projects funded by AHA, the CDC environmental public health tracking program, and NASA to study health effects of ambient air pollution in USA. In summary, I have demonstrated track record of accomplished and productive research projects in the area of environmental epidemiology, particularly in air pollution epidemiologic research. Dr. Xu declares no conflicts of interest.

LIBRARIAN BIOSKETCH AND RESUME

MARGARET FOSTER

Title: Systematic Reviews and Research Services Coordinator
Degrees: MPH from the University of Texas Health Science Center at Houston (2009); M.S. from the University of North Texas (2003)
Biosketch:

Margaret Foster

Systematic Reviews and Research Services Coordinator

Education

MPH, University of Texas Health Science Center at Houston, Houston, TX, 8/2005-5/2009.
Major: Behavioral Sciences and Health Promotion
Thesis: A systematic evaluation of the search strategies, methods, and selection of systematic reviews of interventions to prevent obesity

M.S., University of North Texas, Denton, TX, 8/2003.
Major: Library and Information Sciences
Specialty: Health Informatics

B.S. Texas A&M University, College Station, TX, 12/1998.
Major: Psychology

No degree. University of Texas at San Antonio, San Antonio, TX 8/1991- 5/1996.

Professional and/or Library Experience

Systematic Reviews and Research Services Coordinator, Medical Sciences Library, Texas

A&M
University, January 2012

- Guide the establishment of an MSL Center for Systematic Reviews (SR)
- Define, develop, and market CSR services to all MSL constituencies
- Instruct faculty and students on conducting SRs and meta-analyses
- Perform searches according to SR protocol
- Report results in format suitable for publications and grant proposals
- Collaborate with appropriate Evans Library personnel to establish a campus-wide SR service
- Lead campus Refworks support services by maintaining a RefWorks guide, providing training for library staff and faculty, and supporting user needs
- Provide liaison services in support of faculty and student research, promote relevant library services and resources, and communicate information needs of researchers to MSL
- Provide liaison services to the Department of Kinesiology and Health Education
- Work as part of the Client Services team, delivering clinical and reference services to library clients both onsite and as part of the embedded librarian team

Joint Assistant Professor, Department of Health Promotion and Community Health Sciences, School of Rural Public Health, Texas A&M Health Science Center, September 2012

Social Sciences and Education Librarian, Sterling C. Evans Library, Texas A&M University, August 2007 - December 2011

- Provide reference assistance in person, telephone, virtual, and email
- Liaison with the Health and Kinesiology Department (HLKN) and Women's and Gender Studies
- Department (WGST)
- Provide instructional sessions for the HLKN and WGST departments and assist in other in
- instructional sessions
- Create brochures and other educational materials
- Conduct collection development in subject areas relevant to the HLKN and WGST departments

Reference Librarian, School of Public Health Library, University of Texas Health Science Center at Houston, September 2005 - August 2007

- Provide reference assistance in person, phone, and email
- Educate users on library resources through instruction sessions
- Create brochures and other educational materials
- Assess needs of patrons and participate in planning of library activities
- Provide outreach and develop collection for health promotion and behavioral sciences division

- Promote utilization of library resources
- Collaborate faculty to integrate information literacy competencies into the classroom

Adjunct Faculty, School of Library and Information Sciences, University of North Texas May 2005 – May 2008

Graduate Library Assistant, University of Texas School of Public Health, August 2004 - September 2005

Library Associate II, Medical Sciences Library, Texas A&M University, March 1999 - July 2004

Documentation of Publications, Research, and Presentations

A. Publications:

1. Refereed Publications

Shim, M., Gimeno, D., Pruitt, S.; McLeod, C., Foster, M. J., and Amick III, B (2012). "A systematic review of retirement as a risk factor for mortality". Applied Demography and Public Health. Ed. Hoque, N. Springer.

Role: Developed, updated, and managed all search results; screened portion of articles; assisted in writing methods; assisted in editing entire chapter

Stephens, J., Sare, L., Kimball, R., Foster, M., and Kitchens, J. (2011) "The tenure support mechanisms provided by the Faculty Research Committee at Texas A&M University: A Model for Academic Libraries" Library Management 32(8/9), 531-539.

Role: Collected portion of data; created all charts and tables after synthesizing data; assisted in editing paper

Shurtz, S. and Foster, M. (2011) "Developing and using a rubric for evaluating evidence-based medicine point-of-care tools," JMLA: Journal of the Medical Library Association, 99(3): 247-54.

Role: Created all charts and tables; wrote Methods and Results sections; co-wrote all other sections

Impact: selected as 1 of 8 articles selected from JMLA to be part of the Medical Library Association's Independent Reading Program which allows members to receive 1 MLA continuing education contact hour for each article read and analyzed; JMLA impact factor of .988 (median impact factor in the Information and Library Science category is .641); for 2011 this article had 1108 full text views and 439 pdf downloads

vanDuinkerken, W., Coker, C., & Anderson, M. (2010). PERSPECTIVES ON...: Looking Like Everyone Else: Academic Portfolios for Librarians. Journal of Academic Librarianship, 36(2), 166-172.

Role: Developed guidelines for faculty librarian academic portfolio which was the basis of the article

Yoshii, A., Plaut, D. A., McGraw, K. A., Anderson, M. and Wellik, K. E. (2009) "Analysis of the Reporting of Search Strategies in Cochrane Systematic Reviews" JMLA: Journal of the Medical Library Association 97(1): 21-29.

Role: Analyzed data from previous studies and data from study, created all charts and tables;
wrote methods and results section; participated in editing all parts of manuscript
Impact: Times cited: Scopus- 7; Web of Science-5; Google Scholar-13; JMLA has impact factor
of .988 (median impact factor in the Information and Library Science category is .641)

Anderson, M. and Olmstadt W. (2003) "Providing systematic training for patient support groups." Journal of Hospital Librarianship. 3(3): 13-24.

B. Presentations

1. Juried Presentations

International

Plaut, D., McGraw, K. A., Anderson, M.J., Nguyen, L., Wellik, K. E., and Yoshii, A., "Analysis of the reporting of search strategies in Cochrane Systematic Reviews". Juried poster session with published abstract, North American Conference on Systematic Reviews, Baltimore, MD, 2006

National

Shim, M., Gimeno, D., Pruitt, S.; McLeod, C., Foster, M. J., and Amick III, B. "A systematic review of retirement as a risk factor for mortality". Paper presented at Population Association of America, San Antonio, Texas, 2010

Cleveland, A. Philbrick, J. L., Pipes, T., and Anderson, M. "Training Future Health Information Professionals to Manage Disaster Situations" Paper presented at Medical Library Association Conference, Honolulu, 2009

Plaut, D., McGraw, K. A., Anderson, M.J., Nguyen, L., Wellik, K. E., and Yoshii, A., "Analysis of the reporting of search strategies in Cochrane Systematic Reviews". Poster presented at MLA Conference, Philadelphia, 2007

Olmstadt, William, and Anderson, Margaret J. "Information needs among genetic counselors", Juried poster session with published abstract, 2005 MLA Conference, San Antonio, Texas, 2005

Regional/State

Foster, M. and Shurtz, S. "Making the CASE for EBM: The development and evaluation of the Critical Appraisal for Summaries of Evidence (CASE) worksheet" Juried paper session with published abstract, South Central Chapter of the MLA Conference, Baton Rouge, LA, 2011. Won 1st place for research

Foster, M. and Shurtz, S. "Evidence Based Public Health in Practice: designing an evidence based health intervention" Heart and Stroke Healthy City Liaison Summit, Texas Department

of Health, Austin, TX Aug 5, 2011. Duration: 4 hours

Trumble, J. M., Anderson, M. J., Caldwell, M., Chuang, F., Fulton, S., Howard, A., and Varman, B. "Systematic Evaluation of Evidence Based Medicine Tools for Point of Care", Juried paper session with published abstract 2006, South Central Chapter of the MLA Conference, College Station, Texas, 2006

Anderson, Margaret J. "Feed the need for Public Health News: RSS Feeds, Blogs, and More", Juried poster session with published abstract, 2006 MLA Conference, College Station, Texas, 2006

Anderson, Margaret J. "Systematic training for patient support groups", Juried paper session with published abstract, 2002 South Central Chapter of the MLA Conference, San Antonio, Texas, 2002

2. Invited Presentations

Regional/State

Anderson, Margaret J. "Emergency Management 101" presentation, University of North Texas All School Day, Houston, TX, March 2007 Duration: 30 minutes

3. Other Presentations

University *(not related to professional assignment)*

Anderson, Margaret J. "Scholarly Research: Guide to effective and systematic literature searches" presentation, Texas A&M University, Black Graduate Students' Association, November 2008, Duration: 30 minutes

Anderson, Margaret J. "Virtual Worlds" Instructional Technology Showcase, Oct 31, 2007. Duration: 30 minutes.

Anderson, Margaret J. "A critical evaluation of search filters" TAMU Libraries Research Forum, Nov 2, 2007. Duration: 20 minutes

Community

Foster, Margaret J. Interviewed by Matthew Smith "Health Information" Brazos Valley Health, College Station. KEOS. May 26, 2011. Duration: 1 hour

Foster, Margaret J. Interviewed by Matthew Smith "Health Information and Technology Questions" Brazos Valley Health, College Station. KEOS. August 7, 2008. Duration: 1 hour

Foster, Margaret J. Interviewed by Matthew Smith "Fitness Gadgets" Brazos Valley Health, College Station. KEOS. October 2, 2008. Duration: 1 hour

Documentation of Professional Activities

A. National Committees

American Library Association, American of College and Research Libraries, Women's Studies Section, Assistant Editor, Women's Studies Section- ALA Newsletter, 2008-2010

American Library Association, American of College and Research Libraries, Women's Studies Section, member, Electronic Resources and Access Committee, 2008-2010

American Library Association, American of College and Research Libraries, Women's Studies

Section, ex-officio member, WSS Executive Committee 2008-2010
Medical Library Association, Veterinary Medical Libraries Section, Editor, Newsletter of the Veterinary Medical Libraries Section of the Medical Library Association, 2001-2004, 2012

B. Texas A&M Health Science Center committees

QEP CARE committee

C. Library Committee

TAMU Libraries, CSI, Tech Trends, 2009 (appointed)
TAMU Libraries, CSI, Second Life Group, chair 2009-2010
TAMU Libraries, Academic Portfolio Committee, 2009 (appointed)
TAMU Libraries, SFX Menu Redesign Committee, 2009-2010 (appointed)
TAMU Libraries, Research Committee, 2009-2011 (elected)
TAMU Libraries, SECC Committee, 2010, chair 2011

D. Teaching and Workshops

1. Teaching (for credit courses)

Co-instructor, Evidence Based Medicine and Critical Thinking (M1), Texas A&M Health Science Center, College of Medicine, Fall 2012, Spring 2013 (1 credit hour)
Co-instructor, Evidence Based Medicine and Critical Thinking (M2), Texas A&M Health Science Center, College of Medicine, Fall 2012-Spring 2013 (1 credit hour)
Co-instructor, Public Health Informatics (SRPH 640), Texas A&M Health Science Center, School of Rural Public Health, Spring 2008-2012 (3 credit hour)
Co-instructor, Disaster Management for Information Professionals (SLIS 5670), University of North Texas, School of Library and Information Science, Summer 2006-2008 (3 credit hour)
Co-instructor, Community-Based Health Information (SLIS 5960), University of North Texas, School of Library and Information Science, Summer 2005-2009 (3 credit hour)
Co-instructor, Evidence Based Public Health (PH 1430), University of Texas School of Public Health, Spring 2006 (3 credit hour)
Lead Instructor, Zombies 101: Surviving Freshmen Year (UGST 181-538), Texas A&M University, Fall 2011. Designed to teach incoming freshmen critical thinking skills. (1 credit hour)
Lead Instructor, Avatars and Second Life (UGST 181-513), Texas A&M University, Fall 2010 (1 credit hour)
Lead Instructor, Second Life: Flying into Freshmen Year (UPAS 181-511), Texas A&M University, Fall 2009 (1 credit hour)

2. Workshops

Developer and presenter, Systematic Reviews: Role of the Librarian, South Central Chapter of the MLA Conference, Lubbock, TX 6 hours, October 2012
Developer and presenter, Systematic Reviews: Role of the Librarian, South Central Chapter of the MLA Conference, Baton Rouge, LA, 6 hours, October 2011

Co-developer and presenter, Evidence Based Public Health, Medical Sciences Library, Summer 2010, Fall 2011
Developed and presented, Systematic Review Workshop, Texas A&M University Library, 9 hours over 3 days, offered twice during Fall 2010
Library facilitator, "Learning to Practice and Teach Evidence-based Health: An intensive workshop", UT MD Anderson Cancer Center, Houston, Texas, 2006-2007

E. Honors

Research Award for Contributed Papers (1st place), South Central Chapter of MLA, 2011
Certificate of Appreciation, College of Education & Human Development (TAMU), July 2008
Research Award for Posters (1st place), MLA, 2007
Research Award for Contributed Papers (1st place), South Central Chapter of MLA, 2006
McLemore Educational Opportunity Scholarship, 2003
Research Award for Contributed Papers (2nd place), South Central Chapter of MLA, 2002
Louise David Scholarship, 2001

F. Memberships

Evidence Based Veterinary Medicine Association, 2012

Medical Library Association, 2003-2006, 2010-12
Consumer & Patient Health Information 2011
Veterinary Medical Libraries Section 2012
Public Health/Health Administration 2011-12
South Central Chapter, Medical Library Association, 2004-5, 2010-11

American Libraries Association 2006, 2008-10
American of College and Research Libraries 2006, 2008-10
Educational & Behavioral Sciences Section, 2008-10
Women's Studies Section, 2008-10
Library Research Round Table 2008-10
Library Instruction Round Table 2006

G. Advising

Faculty Advisor for Mixed Martial Arts Student Organization, 2008-2012; group provides free selfdefense for women on campus once a week in addition to 4 MMA training sessions each week.

Significant continuing education

Discovery Workshop, NCBI, 2011, Houston, TX, 2 day course (covered genetics related databases)
Professional certification in online teaching, Instructional Technology Services, Texas A&M University, Fall 2012

ARL Project management institute, Association of College & Research Libraries, Mar 18-19, 2008

Disaster preparedness and planning symposium, Medical Library Association, 2006, 6 contact hours

Planning and evaluating health information outreach projects, Medical Library Association, 2006, 6 contact hours

Assessing student learning outcomes, Association of College & Research Libraries, 2005, 3-week online webinar

Appendix II. Search Terms

1. Exp Particulate Matter/
2. (pm 2* or pm 10 or pm2* or pm10).ti,ab.
3. (particulate adj2 (matter* or pollut*)).ti,ab.
4. Or/1-3
5. Exp "Birth Weight"/
6. (lbw or (low adj1 (birthweight* or birthweight*))).ti,ab.
7. Or/5-6
8. 4 and 7
9. limit 8 to English language

Appendix III. HAWC Data Collection

Fields are free- form except where choices (in italics) are shown

SOURCE

Reviewer:

- (identify yourself)

Publication year:

Authors declared conflicts of interest:

- *None declared*
- *Declared*

  If declared, provide details:

Study funding source:

- *Government grant*
- *Industry funded*
- *Nonprofit organization grant*
- *Other*

  Study funding source details:

What are the study objectives?:

Site(s) of data collection (city, state, country):

METHODS

Study duration:

Study design:

- *Cross-sectional*
- *Cohort, prospective*
- *Cohort, retrospective*
- *Case-control*
- *Ecological*

- *Other (list details below)*

Study design details:

Characteristics of study population:

- *Cohort (give description, e.g. NHANES 2004-2006)*
- *Sample size of total cohort*
- *Sample size (each exposure group)*
- *Age (each exposure group)*
- *Co-morbidities*
- *Other relevant details (list below)*

Study subject details:

Exposure period:

- *Pregnancy*
- *Other (details below)*
- *Record when exposure occurred or was measured, in relation to outcome measurement*

Source of exposure data:

- *Air pollution monitoring (list specific methods)*
- Modeling (list specific methods)
- *Questionnaire (list specific proxy used to determine of exposure)*
- *Other (specify)*

Total number of exposed groups:

Total number of non-exposed groups:

Number of subjects in each group:

If a power calculation was done, was the sample size of the study sufficient?

- *Yes*
- *No*

Concentrations of PM measured, and units:

Frequency of PM measurements if more than once:

Number of replicate measurements taken

Chemical name:

- PM 2.5
- PM 10

Chemical name details:

Other chemical details:

Outcomes measured:

Method of fetal growth measurement:

- *Weight*
- *Length*
- *Other (details below)*

Method of fetal growth measurement details:

Gestational age at outcome measurements:

- *At birth*
- *Other (details below)*

Birth outcome measurement details:

Unit of measurement (for weight, etc.):

- *Grams*
- *Milimeters*
- *Other (details below)*

Unit of measurement (for weight, etc.) details:

Sex (where outcome measured):

- *Males only*
- *Females only*
- *Males and females*
- *Other (details below)*

Number subjects analyzed (for exposure and outcome):

Number of missing participants:

RESULTS

Statistical methods:

- Statistical tests employed
- Statistic (odd ratio, adjusted odds ratio, beta estimate, etc.)
- P-values given
- Confidence intervals given
- Confounding adjustments in statistical tests

Were known confounders accounted for by study design?

Were known confounders accounted for by analysis?

How were data reported (mean, median, raw data, etc.)?:

Growth measurement data for each group (ie, outcome):

How growth measurement data were reported (table, figures, etc.):

Summary data for each exposure group

Estimate of effect with confidence interval and p-value

How was precision reported (standard error, CI, etc.):?

- *Standard error*
- *Standard deviation*
- *Confidence intervals*
- *Other (details below)*
- *Not stated*

How precision reported details:

Precision estimates:

How precision estimates were reported (tabe, figure, etc.):

Miscellaneous comments by reviewer regarding data analysis:

Appendix IV. Instructions for Making Risk of Bias Determinations

Human Studies

Please answer LOW RISK, PROBABLY LOW RISK, PROBABLY HIGH RISK, HIGH RISK or NOT APPLICABLE and provide details/justification.

Note: These criteria for judging risk of bias are for human studies only since we are not evaluating animal studies in this case study. These questions have also been modified from previous applications of the Navigation Guide, with edits intended so that answering "No" to each question aligns with a rating of "High risk of bias", "Probably No" → "Probably high risk of bias", "Probably Yes" → "Probably low risk of bias" and "Yes" → "Low risk of bias."

1. Was the strategy for recruiting participants consistent across study groups?

Criteria for a judgment of LOW risk of bias (i.e., answer: "YES"):

Protocols for recruitment and inclusion/exclusion criteria were applied similarly across study groups, and any one of the following:

- Study participants were recruited from the same population at the same time frame; or
- Study participants were not all recruited from the same population, but proportions of participants from each population in each study group are uniform

Criteria for the judgment of PROBABLY LOW risk of bias (i.e., answer: "Probably Yes"):

There is insufficient information about participant selection to permit a judgment of 'YES', but there is indirect evidence that suggests that participant recruitment and inclusion/exclusion criteria was consistent, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of HIGH risk of bias (i.e., answer: "No"):
Any of the following:

- Protocols for recruitment or inclusion/exclusion criteria were applied differently across study groups; or
- Study participants were recruited at different time frames; or
- Study participants were recruited from different populations and proportions of participants from each population in each study group are not uniform
- Differential loss to follow-up between groups
- Reported refusal/non-response is uniform between groups

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e., answer: "Probably No"):

There is insufficient information about participant selection to permit a judgment of 'NO', but there is indirect evidence that suggests that participant recruitment or inclusion/exclusion criteria was inconsistent, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that participant selection is not an element of study design capable of introducing risk of bias in the study.

2. Was knowledge of the exposure adequately prevented during the study?

Criteria for a judgment of LOW risk of bias (i.e., answer: "YES"):

Any of the following:

- No blinding, but the review authors judge that the outcome and the outcome measurement as well as the exposure and exposure measurement are not likely to be influenced by lack of blinding (such as differential outcome assessment where the outcome is assessed using different measurement or estimation metrics across exposure groups, or differential exposure assessment where exposure is assessed using different measurement or estimation metrics across diagnostic or outcome groups); or
- Blinding of key study personnel was ensured, and it is unlikely that the blinding could have been broken; or
- Some key study personnel were not blinded, but exposure and outcome assessment was blinded and the non-blinding of others is unlikely to introduce bias. For example, investigators were effectively blinded to the exposure and/or outcome groups, for example if the exposure was measured by a separate entity and the outcome was obtained from a hospital record.

Criteria for the judgment of PROBABLY LOW risk of bias (i.e., answer: "Probably Yes"):

There is insufficient information about blinding to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was adequately blinded, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of HIGH risk of bias (i.e., answer: "No"):

Any of the following:

- No blinding or incomplete blinding, and the outcome or outcome measurement or exposure and exposure measurement is likely to be influenced by lack of blinding (i.e., differential outcome or exposure assessment); or
- Blinding of key study personnel attempted, but likely that the blinding could have been broken so as to introduce bias; or
- Some key study personnel were not blinded, and the non-blinding of others was likely to introduce bias.

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e., answer: "Probably No"):

There is insufficient information about blinding to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not adequately blinded, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that blinding is not an element of study design capable of introducing risk of bias in the study.

3. Were exposure assessment methods robust?

Note: For this risk of bias domain, we will consider exposure assessment metric for PM. Risk of bias will be assessed for each data set. The risk of bias over the body of evidence will be rated by review authors' review of risk of bias across all datasets (not across all studies). Our rationale for breaking up studies into data sets is that: (1) there is empirical evidence that risk of bias varies depending on which air pollution exposure was measured (i.e., chemical component) and how it was measured (i.e., exposure metric) (US Environmental Protection Agency 2013); (2) there is a need to transparently distinguish among these potential biases within a given study; and (3) co-authors did not identify an empirically-based or otherwise scientifically preferable alternative method to address this aspect of heterogeneity in the data.

The following list of considerations represents a collection of factors proposed by experts in various fields that may potentially influence the internal validity of the exposure assessment in a systematic manner (not those that may randomly affect overall study results). ***These should be interpreted only as suggested considerations, and <u>should not be viewed as scoring or a checklist</u>.***

<u>List of Considerations</u>:

Exposure assessment metric:

1) *Modeling*
2) *Monitoring*

For each, overall considerations include:

1) *What is the quality of the metric being used?*
2) *Has the metric been validated for the scenario for which it is being used?*
3) *Is the exposure measured in the study a surrogate for air pollution (i.e., distance to freeway)?*
4) *What was the temporal coverage (i.e., whole developmental period, or a shorter duration)?*
5) *Did the analysis account for prediction uncertainty?*
6) *How was missing data accounted for, and any data imputations incorporated?*
7) *Were sensitivity analyses performed?*

In particular, for exposure assessment models:

1) *Were the input data in the study suspected to systematically under- or over-estimate exposure?*
2) *What type of model was used (geostatistical interpolation, land-use regression, dispersion models, personal air sampling models, hybrid models, etc.)?*
3) *Were meteorological variables incorporated in the model and justified by authors in their selection?*
4) *Were data on land use, topography, traffic, monitoring data, emission rates, etc. incorporated and justified by authors in their selection?*
5) *What was the spatial variation (e.g., distance from source) and geographic/spatial accuracy (county, census tract, individual residence)?*

6) *What was the temporal specificity and variation (accuracy to level of the day, pregnancy trimester, year, etc.?)*
7) *What was the address completeness (e.g., only home address at one point in time, or more complete address history throughout pregnancy/postnatal life and other locations such as work)?*
8) *What was the space-time coverage of the model?*
9) *Were time-activity patterns accounted for?*
10) *Was mixing height considered as a covariate?*

Criteria for a judgment of LOW risk of bias (i.e., answer: "Yes"):

The reviewers judge that there is low risk of exposure misclassification, i.e.:

- There is high confidence in the accuracy of the exposure assessment methods, such as methods that have been tested for validity and reliability in measuring the targeted exposure; or
- Less-established or less direct exposure measurements are validated against well-established or direct methods; or:

A) Monitoring: direct and personal monitoring devices that were used that have been validated for the chemical and scenario for which it was used and there is sufficient evidence that relevant factors from the List of Considerations above would imply minimal risk of bias in the exposure assessment; or
B) Modeling: the model accounted for the time-activity pattern specific to each research participant, (e.g. includes more than exposure at the residential address) and included air pollution modeling methods that have been validated or shown to have a high degree of spatial accuracy (e.g. point location), and/or methods that are themselves validated with good agreement compared to person-based air data collection; and there is sufficient evidence that relevant factors from the List of Considerations above would imply minimal risk of bias in the exposure assessment.

AND if applicable (e.g. for laboratory measurements), appropriate QA/QC for methods are described and are satisfactory, with at least three of the following items reported, or at least two of the following items reported plus evidence of satisfactory performance in a high quality inter-laboratory comparison:

- Limit of detection or quantification;
- standards recovery;
- measure of repeatability;
- investigation and prevention of blanks contamination.

Criteria for the judgment of PROBABLY LOW risk of bias (i.e., answer: "Probably No"):

There is insufficient information about the exposure assessment methods to permit a judgment of low risk of bias, but there is indirect evidence that suggests that methods were robust, as described by the criteria for a judgment of low risk of bias. Studies only reporting that the QA/QC items above were satisfactory but not reporting all of the actual numbers may receive a judgment of "probably low risk of bias." Additionally:

A) Monitoring: methodologies which directly assess exposure were used, such as personal exposure instruments, but had not been validated for that purpose, or if such instruments were worn for less than 4 hours per day, or there is some evidence that relevant factors from the List of Considerations above would imply minimal risk of bias in the exposure assessment.
B) Modeling: the model used methods that do not meet the criteria of including time-activity patterns AND spatial accuracy, and so may not have the level of validation compared to person-based air measurement, but include measurements that have evidence of quality, such as good-quality data inputs, validation against area-based air measurement, or other establishments of the accuracy of the data inputs and models, or there is some evidence that relevant factors from the List of Considerations above would imply minimal risk of bias in the exposure assessment.

Criteria for the judgment of HIGH risk of bias (i.e., answer: "No"):

The reviewers judge that there is high risk of exposure misclassification and any one of the following:

- There is low confidence in the accuracy of the exposure assessment methods; or
- Less-established or less direct exposure measurements are not validated and are suspected to introduce bias that impacts the outcome assessment (example: participants are asked to report exposure status retrospectively, subject to recall bias); or
- Uncertain how exposure information was obtained; or:

A) Monitoring: Information from databases or otherwise was gathered that indirectly assessed exposure without considering variables noted in the List of Considerations above, such as spatial variability, land use regression, etc., or there is sufficient evidence that relevant factors from the List of Considerations above would imply risk of bias in the exposure assessment.
B) Modeling: the air pollution model used has been demonstrated not to pertain to area-based or person-based measures or has otherwise been previously demonstrated to be unable to describe air levels of exposure for assigning exposure in a research situation, or there is sufficient evidence that relevant factors from the List of Considerations above would imply risk of bias in the exposure assessment.

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e., answer: "Probably No"):

There is insufficient information about the exposure assessment methods to permit a judgment of high risk of bias, but there is indirect evidence that suggests that methods were not robust, as described by the criteria for a judgment of high risk of bias. Additionally:

A) Monitoring: measurement of exposures that may not have been validated for use to study air pollution were used, or there is some evidence that relevant factors from the List of Considerations above would imply risk of bias in the exposure assessment.
B) Modeling: air pollution models were used that have not been compared to person-based or area-based air measurements and have suspicion of problems estimating true exposure because, for example, they do not have spatial accuracy (e.g. county-level measures), do not pertain to the correct time frame, are based on limited data, or differ in methodology

between cases and controls in a study, or there is some evidence that relevant factors from the List of Considerations above would imply risk of bias in the exposure assessment.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that exposure assessment methods are not capable of introducing risk of bias in the study.

4. Was confounding adequately addressed?

Prior to the evaluation of studies, coauthors collectively developed the following list of confounders as well as the rationale for inclusion. The top 5 confounders are considered "important" based on strength of association with the outcome as well as exposure variable.

1. Social class.
This is measured differently from study to study, such as by education and income. Note that variables like marital status and insurance can even reflect aspects of social class. Sometimes social class is accounted for by individual-level measurements, and other times by group-level measurements (such as census variables).

Rationale: Where people live (neighborhood) is strongly influenced by social class. Additionally, we know that urban residence is associated with higher air pollution (Williams et al. 2006). And the airborne pollutants that someone is exposed to are influenced by neighborhood, so social class is related to neighborhood.

2. Race/ ethnicity
Race and ethnicity is known to influence studies of air pollution and birthweight.

3. Tobacco use or Environmental tobacco smoke exposure
Tobacco use is known to be a form of air pollution and can indirectly influence studies of air pollution and birthweight.

4. Maternal (and/or paternal) age.
Maternal age is related to social class, because very young mothers tend to be of lower social class and older mothers tend to be of higher social class so maternal age may be correlated with air pollutant levels (i.e., younger women may be of lower social class, and lower social classes may be exposed to higher levels of air pollutants). Note that while paternal age is likely also a risk factor, it is often not adjusted for because this variable has a high degree of missing-ness. Because maternal age and paternal age are highly related, it is often thought that adjusting for maternal age is sufficient, and so may not be necessary to include both.

5. Season of conception/birth (calendar time of conception/birth).
Air pollutant concentrations also vary by season due to sunlight and other factors. Air pollutants will only vary by season if there is temporal refinement in the air pollutant measure, such as monthly or trimester-long values. A study with annual averages or air pollutant levels, or static levels such as distance to a road, will NOT show a correlation structure between season and air pollutants, and so season will not confound in this type of study.

6. Comorbidity
Comorbidity is known to influence studies of air pollution and birthweight.

7. Alcohol use

Alcohol use is known to decrease birthweight and be associated with social class. Those who consume alcohol while pregnant are more likely to develop pregnancy complications and have a lower birthweight.

Criteria for a judgment of 'YES' (i.e. low risk of bias):

The study accounted for (i.e., matched, stratified, multivariate analysis or otherwise statistically controlled for) all 5 important potential confounders, or reported that potential confounders were evaluated and omitted because inclusion did not substantially affect the results. The determination of specific confounders may be informed by the data, including the studies included in the review.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

The study accounted for most, at least 3 of the 5 of the important potential confounders AND this lack of accounting is not expected to introduce substantial bias.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

The study accounted for two or less of our listed potential confounders.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

The study accounted for at least two of the important potential confounders, but included other potential confounders AND this lack of accounting may have introduced substantial bias.

5. Were incomplete outcome data adequately addressed?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

Participants were followed long enough to obtain outcome measurements;
OR any one of the following:

- No missing outcome data; or
- Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to introduce bias); or
- Attrition or missing outcome data balanced in numbers across exposure groups, with similar reasons for missing data across groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a relevant impact on the intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a relevant impact on the observed effect size; or
- Missing data have been imputed using appropriate methods

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information about incomplete outcome data to permit a judgment of low risk of bias, but there is indirect evidence that suggests incomplete outcome data was adequately addressed, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

Participants were not followed long enough to obtain outcome measurements;
OR any one of the following:

- Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across exposure groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce biologically relevant bias in intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce biologically relevant bias in observed effect size; or
- Potentially inappropriate application of imputation.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information about incomplete outcome data to permit a judgment of high risk of bias, but there is indirect evidence that suggests incomplete outcome data was not adequately addressed, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that incomplete outcome data is not capable of introducing risk of bias in the study.

6. Does the study report appear to have been comprehensive in its outcome reporting?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

All of the study's pre-specified (primary and secondary) outcomes outlined in the protocol, methods, abstract, and/or introduction that are of interest in the review have been reported in the pre-specified way.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information about selective outcome reporting to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was free of selective reporting, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

Any one of the following:

- Not all of the study's pre-specified primary outcomes (as outlined in the protocol, methods, abstract, and/or introduction) have been reported; or
- One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; or
- One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected effect); or
- One or more outcomes of interest are reported incompletely

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information about selective outcome reporting to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not free of selective reporting, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that selective outcome reporting is not capable of introducing risk of bias in the study.

7. Is the study free of support from any company, study author, or other entity having a financial interest in any of the exposures studied?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

The study did not receive support from a company, study author, or other entity having a financial interest in the outcome of the study. Examples include the following:

- Funding source is limited to government, non-profit organizations, or academic grants funded by government, foundations and/or non-profit organizations;
- Chemicals or other treatment used in study were purchased from a supplier;
- Company affiliated staff are not mentioned in the acknowledgements section;
- Authors were not employees of a company with a financial interest in the outcome of the study;
- Company with a financial interest in the outcome of the study was not involved in the design, conduct, analysis, or reporting of the study and authors had complete access to the data;
- Study authors make a claim denying conflicts of interest;
- Study authors are unaffiliated with companies with financial interest, and there is no reason to believe a conflict of interest exists;
- All study authors are affiliated with a government agency (are prohibited from involvement in projects for which there is a conflict of interest or an appearance of conflict of interest).

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was free of support from a company, study author, or other

entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of 'NO' (i.e. high risk of bias):
The study received support from a company, study author, or other entity having a financial interest in the outcome of the study. Examples of support include:

- Research funds;
- Chemicals, equipment or testing provided at no cost;
- Writing services;
- Author/staff from study was employee or otherwise affiliated with company with financial interest;
- Company limited author access to the data;
- Company was involved in the design, conduct, analysis, or reporting of the study;
- Study authors claim a conflict of interest

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that conflicts of interest are not capable of introducing risk of bias in the study.

8. Did the study appear to be free of other problems that could put it at a risk of bias?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

The study appears to be free of other sources of bias.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was free of other threats to validity.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

There is at least one important risk of bias. For example, the study:

- Had a potential source of bias related to the specific study design used; or
- Stopped early due to some data-dependent process (including a formal-stopping rule); or
- The conduct of the study is affected by interim results (e.g. recruiting additional participants from a subgroup showing greater or lesser effect); or
- Has been claimed to have been fraudulent; or
- Had some other problem

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not free of other threats to validity, as described by the criteria for a judgment of high risk of bias.

Appendix V. Instructions for Grading the Quality and Strength of Evidence

A. Grading Quality

Each of the categories to consider in downgrading or upgrading the evidence is described in detail below. Please record your results on the chart at the end of each category, including a brief explanation for your ratings.

Downgrade Categories

Category 1. Quality of Study Limitations (Risk of Bias)(Guyatt et al. 2011)

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

The evidence from studies can be rated down if most of the relevant evidence comes from studies that suffer from a high risk of bias. Risk of bias is rated by outcome across studies. Study limitations for each outcome for individual studies and across studies are summarized in the heat maps.

GRADE outlines the following principles for moving from risk of bias in individual studies to rating quality of evidence across studies.

1. In deciding on the overall quality of evidence, one does not average across studies (for instance if some studies have no serious limitations, some serious limitations, and some very serious limitations, one does not automatically rate quality down by one level because of an average rating of serious limitations). Rather, judicious consideration of the contribution of each study, with a general guide to focus on the high-quality studies is warranted.[1]

2. This judicious consideration requires evaluating the extent to which each study contributes toward the estimate of magnitude of effect. The contribution that each study makes will usually reflect study sample size and number of outcome events. Larger studies with many events will contribute more, much larger studies with many more events will contribute much more.

3. One should be conservative in the judgment of rating down. That is, one should be confident that there is substantial risk of bias across most of the body of available evidence before one rates down for risk of bias.

4. The risk of bias should be considered in the context of other limitations. If, for instance,

[1] Note: Limitations to GRADE's risk of bias assessments as stated by GRADE: "First, empirical evidence supporting the criteria is limited. Attempts to show systematic difference between studies that meet and do not meet specific criteria have shown inconsistent results. Second, the relative weight one should put on the criteria remains uncertain. The GRADE approach is less comprehensive than many systems, emphasizing simplicity and parsimony over completeness. GRADE's approach does not provide a quantitative rating of risk of bias. Although such a rating has advantages, we share with the Cochrane Collaboration methodologists a reluctance to provide a risk of bias score that, by its nature, must make questionable assumptions about the relative extent of bias associated with individual items and fails to consider the context of the individual items."

reviewers find themselves in a close-call situation with respect to two quality issues (risk of bias and, say, precision), GRADE suggests rating down for at least one of the two.

5. Notwithstanding the first four principles, reviewers will face close-call situations. You should acknowledge that you are in such a situation, make it explicit why you think this is the case, and make the reasons for your ultimate judgment apparent.

Rating for Risk of Bias (Study Limitations) 0 no change -1 decrease quality 1 level -2 decrease quality 2 levels		**Rationale for your judgment**
Human		

Category 2. Indirectness of Evidence

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

Quality of evidence (your confidence in estimates of effect) may decrease when substantial differences exist between the population, the exposure, or the outcomes measured in research studies under consideration in the review.

Evidence is direct when it directly compares the exposures in which we are interested when applied to the populations in which we are interested and measures outcomes important to the study question (in GRADE the outcomes must be important to patients).

Based on GRADE (Guyatt et al. 2011), evidence can be indirect in one of three ways.[2]

1. The population studied differs from the population of interest (the term applicability is often used for this form of indirectness). GRADE states that in general, one should not rate down for population differences unless one has compelling reason to think that the biology in the population of interest is so different than the population tested that the magnitude of effect will differ substantially. According to GRADE, most often, this will not be the case.

[2] GRADE includes a fourth type of indirectness that occurs when there are no direct (i.e., head-to-head) comparisons between two or more interventions of interest. This criterion is not relevant to our study question; it could be relevant to future case studies.

2. The intervention (exposure) tested may differ from the exposure of interest, i.e., a difference in the chemical, route and/or dose. Decisions regarding indirectness of populations and exposure depend on an understanding of whether biological or social factors are sufficiently different that one might expect substantial differences in the magnitude of effect. GRADE also states, "As with all other aspects of rating quality of evidence, there is a continuum of similarity of the intervention that will require judgment. It is rare, and usually unnecessary, for the intended populations and interventions to be identical to those in the studies, and we should only rate down if the differences are considered sufficient to make a difference in outcome likely."

3. Outcomes may differ from those of primary interest; for instance, surrogate outcomes that are not themselves important, but measured in the presumption that changes in the surrogate reflect changes in an important outcome. The difference between desired and measured outcomes may relate to time frame. When there is a discrepancy between the time frame of measurement and that of interest, whether to rate down by one or two levels will depend on the magnitude of the discrepancy. Another source of indirectness related to measurement of outcomes is the use of substitute or surrogate endpoints in place of the exposed population's important outcome of interest. In general, the use of a surrogate outcome requires rating down the quality of evidence by one, or even two, levels. Consideration of the biology, mechanism, and natural history of the disease can be helpful in making a decision about indirectness. Surrogates that are closer in the putative causal pathway to the adverse outcomes warrant rating down by only one level for indirectness. GRADE states that rarely, surrogates are sufficiently well established that one should choose not to rate down quality of evidence for indirectness. In general, evidence based on surrogate outcomes should usually trigger rating down, whereas the other types of indirectness will require a more considered judgment.

Rating for Indirectness 0 no change -1 decrease quality 1 level -2 decrease quality 2 levels		**Rationale for your judgment**
Human		

Category 3. Inconsistency of Evidence

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

According to Cochrane, "when studies yield widely differing estimates of effect (heterogeneity or variability in results) investigators should look for robust explanations for that heterogeneity. …When heterogeneity exists and affects the interpretation of results, but authors fail to identify a plausible explanation, the quality of the evidence decreases."

Based on GRADE (Guyatt et al. 2011), **a body of evidence is not rated up in quality if studies yield consistent results**, **but may be rated down in quality if inconsistent**. Their stated reason is that a consistent bias will lead to consistent, spurious findings.

GRADE suggests rating down the quality of evidence if large inconsistency (heterogeneity) in study results remains after exploration of a priori hypotheses that might explain heterogeneity. Judgment of the extent of heterogeneity is based on similarity of point estimates, extent of overlap of confidence intervals, and statistical criteria. GRADE's recommendations refer to inconsistencies in effect size, specifically to relative measures (risk ratios and hazard ratios or odds ratios), not absolute measures.

Based on GRADE, reviewers should consider rating down for inconsistency when:

1. Point estimates vary widely across studies;

2. Confidence intervals (CIs) show minimal or no overlap;

3. The statistical test for heterogeneity-which tests the null hypothesis that all studies in a meta-analysis have the same underlying magnitude of effect- shows a low P-value;

4. The I^2 -which quantifies the proportion of the variation in point estimates due to among-study differences-is large. (I.e., the I^2 index quantifies the degree of heterogeneity in a meta-analysis).

GRADE states that inconsistency is important **only when it reduces confidence in results in relation to a particular decision**. Even when inconsistency is large, it may not reduce confidence in results regarding a particular decision. For example, studies that are inconsistent related to the magnitude of a beneficial or harmful effect (but are in the same direction) would not be rated down; in instances when results are inconsistent as to whether there is a benefit or harm of treatment, GRADE would rate down the quality of evidence as a result of variability in results, because the meaning of the inconsistency is so relevant to the decision to treat or not to treat.

Rating for Inconsistency	**Rationale for your judgment**
0 no change -1 decrease quality 1 level -2 decrease quality 2 levels	

Human		

Category 4. Imprecision of Evidence

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

Cochrane states that when studies have few participants and few events, and thus have wide confidence intervals (CIs), authors can lower their rating of the quality of evidence. These ratings of precision are made as judgments by review authors. The ratings are made by looking across studies, or, if available, on the results of a meta-analysis.

GRADE defines evidence quality differently for systematic reviews and guidelines. For systematic reviews, quality refers to confidence in the estimates of effect. For guidelines, quality refers to the extent to which confidence in the effect estimate is adequate to support a particular decision (Guyatt et al. 2011). For the purpose of step 3 of Navigation Guide, we will use the systematic review definition, because the decision phase does not occur until step 4 when recommendations for prevention are made. Thus, when reviewing the data for imprecision, evaluate your confidence in the estimate of the effect.

According to GRADE, to a large extent, CIs inform the impact of random error on evidence quality. Thus, when considering imprecision, the issue is whether the CI around the estimate of exposure effect is sufficiently narrow. If it is not, GRADE rates down the evidence quality by one level (for instance, from high to moderate). If the CI is very wide, GRADE might rate down by two levels.

Rating for Imprecision 0 no change -1 decrease quality 1 level -2 decrease quality 2 levels		**Rationale for your judgment**
Human		

Category 5. Publication Bias

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

GRADE (Guyatt et al. 2011) and Cochrane (Higgins and Green 2011) assess publication bias in a similar manner. Whereas "selective outcome reporting" is assessed for each study included in the review as part of the risk of bias assessment, "publication bias" is assessed on the body of evidence. GRADE states that "when an entire study remains unreported and the results relate to the size of the effect- publication bias- one can assess the likelihood of publication bias only by looking at a group of studies."

Cochrane's definition of publication bias is "the *publication* or *non-publication* of research findings depending on the nature and direction of the results." Cochrane and GRADE are primarily concerned with *overestimates* of true effects of treatments or pharmaceuticals, especially related to "small studies effects", i.e., the tendency for estimates of an intervention to be more beneficial in smaller studies. There is empirical evidence in the clinical sciences that publication and other reporting biases result in over estimating the effects of interventions (Higgins and Green 2011).

In contrast, in environmental health, we are primarily concerned with *underestimating* the true effects of a chemical exposure, since in many cases population wide exposure has already occurred. We are also concerned that studies finding no association are less likely to be published because journals are less likely to publish "negative" findings.

Applying this inverted concern to GRADE's assessment for publication bias, leads to these considerations when rating publication bias:

- Early *negative* studies, particularly if small in size, are suspect. (GRADE is concerned with early *positive* studies).
- Authors of systematic reviews should suspect publication bias when studies are uniformly small, particularly when sponsored by the industry. (Same as GRADE)
- Empirical examination of patterns of results (e.g., funnel plots) may suggest publication bias but should be interpreted with caution. (Same as GRADE)
- More compelling than any of these theoretical exercises is authors' success in obtaining the results of some unpublished studies and demonstrating that the published and unpublished data show different results. (Same as GRADE)
- Comprehensive searches of the literature including unpublished studies, i.e., the grey literature, and a search for research in other languages are important to addressing publication bias. Note that Cochrane also states "comprehensive searching is not sufficient to prevent some substantial potential biases."

Rating for Publication Bias 0 no change -1 decrease quality 1 level -2 decrease quality 2 levels		Rationale for your judgment
Human		

Upgrade Categories

GRADE states that the circumstances for upgrading likely occur infrequently and are primarily relevant to observational and other non-randomized studies. Although it is possible to rate up results from randomized controlled trials, GRADE has yet to find a compelling circumstance for doing so (Guyatt et al. 2011).

GRADE specifies 3 categories for increasing the quality of evidence (Guyatt et al. 2011)

Category 6. Large Magnitude of Effect

Possible ratings: 0=no change; +1 or +2 upgrade 1 or 2 levels

Modeling studies suggests that confounding (from non-random allocation) alone is unlikely to explain associations with a relative risk (RR) greater than 2 (or less than 0.5), and very unlikely to explain associations with an RR greater than 5 (or less than 0.2). Thus, these are the definitions of "large magnitude of effect" used by GRADE to upgrade 1 or 2 levels, respectively. Also, GRADE is more likely to rate up if the effect is rapid and out of keeping with prior trajectory; usually supported by indirect evidence. GRADE presents empirical evidence to support these conclusions, and states that "although further research is warranted, both modeling and empirical work suggest the size of bias from confounding is unpredictable in direction but bounded in size. Hence, the GRADE group has previously suggested guidelines for rating quality of evidence up by one category (typically from low to moderate) for associations greater than 2, and up by two categories for associations greater than 5."

Applying the GRADE definitions of large magnitude of effect i.e., RR greater than 2 or 5 is problematic in environmental health because for dichotomous outcomes RR is a function of the exposure comparator; these definitions also are not applicable to results from continuous variables. At present, we do not have an empirically defined "large magnitude of effect."

Therefore, for the purpose of this case study, co-authors should assess whether the results indicate a large magnitude of effect using their expert judgment of "large effects" in environmental health and state their definition for discussion by the group.

Rating for Large Magnitude of Effect 0 no change +1 increase quality 1 level +2 increase quality 2 levels		**Rationale for your judgment**
Human		

Category 7. Dose-response

Possible ratings: 0=no change; +1 or +2 upgrade 1 or 2 levels

Possible considerations include consistent dose response gradients in one or multiple studies, and/or dose response across studies, depending on the overall relevance to the body of evidence.

Rating for Dose-Response 0 no change +1 increase quality 1 level +2 increase quality 2 levels		**Rationale for your judgment**
Human		

Category 8. Confounding Minimizes Effect

Possible ratings: 0=no change; +1 or +2 upgrade 1 or 2 levels

All plausible residual confounders or biases would reduce a demonstrated effect, or suggest a spurious effect when results show no effect. GRADE provides the following example of grading up evidence when observational studies have failed to demonstrate an association.

Rating for Confounding Minimizes Effect 0 no change +1 increase quality 1 level +2 increase quality 2 levels		**Rationale for your judgment**
Human		

The results of the reviewers' ratings by population will be compiled and discussed leading to a final decision on overall quality of human evidence. The rationale for the decision will be fully documented.

1. **Final decision on overall quality of human evidence:**
(Example: Moderate quality is upgraded 1 step to high for XYZ reason(s))

---- High

---- Moderate

---- Low

---- Very

B. Rate the Strength of Evidence

The evidence quality ratings will be translated into strength of evidence for each population based on a combination of four criteria: (1) Quality of body of evidence; (2) Direction of effect; (3) Confidence in effect; and (4) Other compelling attributes of the data that may influence certainty. The strength of evidence ratings are summarized in Table 1 below, where their meaning is further defined.

Table 1. Strength of evidence definitions for human evidence

Strength Rating	Definition
Sufficient evidence of toxicity	A positive relationship is observed between exposure and outcome where chance, bias, and confounding can be ruled out with reasonable confidence. The available evidence includes results from one or more well-designed, well-conducted studies, and the conclusion is unlikely to be strongly affected by the results of future studies[3].
Limited Evidence of Toxicity	A positive relationship is observed between exposure and outcome where chance, bias, and confounding cannot be ruled out with reasonable confidence. Confidence in the relationship is constrained by such factors as: the number, size, or quality of individual studies, or inconsistency of findings across individual studies[4]. As more information becomes available, the observed effect could change, and this change may be large enough to alter the conclusion.
Inadequate Evidence of Toxicity	The available evidence is insufficient to assess effects of the exposure. Evidence is insufficient because of: the limited number or size of studies, low quality of individual studies, or inconsistency of findings across individual studies. More information may allow an assessment of effects.
Evidence of Lack of Toxicity	No relationship is observed between exposure and outcome, and chance, bias and confounding can be ruled out with reasonable confidence. The available evidence includes consistent results from more than one well-designed, well-conducted study at the full range of exposure levels that humans are known to encounter, and the conclusion is unlikely to be strongly affected by the results of future studies5. The conclusion is limited to the age at exposure and/or other conditions and levels of exposure studied.

[3] The Navigation Guide rates the quality and strength of evidence of human and non-human evidence streams separately as "sufficient", "limited", "inadequate" or "evidence of lack of toxicity" and then these two ratings are combined to produce one of five possible statements about the overall strength of the evidence of a chemical's reproductive/developmental toxicity. The methodology is adapted from the criteria used by the International Agency for Research on Cancer (IARC) to categorize the carcinogenicity of substances International Agency for Research on Cancer (2006). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Preamble (amended January 2006). Lyon, France, World Health Organization. except as noted.

[4]Language for the definitions of the rating categories were adapted from descriptions of levels of certainty provided by the U.S. Preventive Services Task Force Levels of Certainty Regarding Net Benefit. http://www.uspreventiveservicestaskforce.org/uspstf07/methods/benefit.htm

[5] Language for the definitions of the rating categories were adapted from descriptions of levels of certainty provided by the U.S. Preventive Services Task Force Levels of Certainty Regarding Net Benefit Sawaya, G. F., J. Guirguis-Blake, M. LeFevre, R. Harris, D. Petitti and U. S. P. S. T. Force (2007). "Update on the methods of the U.S. Preventive Services Task Force: estimating certainty and magnitude of net benefit." Ann Intern Med **147**(12): 871-875.

ABOUT THE ARTICLE

Declarations of Interest: No competing interests reported

Sources of Support:
Internal: None
External: None

References

Bonzini M, Carugno M, Grillo P, Mensi C, Bertazzi PA, Pesatori AC. Impact of ambient air pollution on birth outcomes: systematic review of the current evidences. Med Lav 2010;101(5):341-363.

Bosetti C, Nieuwenhuijsen MJ, Gallus S, Cipriani S, La Vecchia C, Parazzini F. Ambient particulate matter and preterm birth or BW: a review of the literature. Arch Toxicol 2010;84(6):447-460.

DerSimonian, R. and N. Laird (1986). "Meta-analysis in clinical trials." Control Clin Trials **7**(3): 177-188.

Duval, S. and R. Tweedie (2000). "Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis." Biometrics **56**(2): 455-463.

Fox, D. M. (2010). "Realizing and allocating savings from improving health care quality and efficiency." Prev Chronic Dis **7**(5): A99.

Ghosh R, Rankin J, Pless-Mulloli T, Glinianaia S. Does the effect of air pollution on pregnancy outcomes differ by gender? A systematic review. Environ Res 2007;105(3):400-408.

Guyatt, G. H., A. D. Oxman, R. Kunz, J. Woodcock, J. Brozek, M. Helfand, P. Alonso-Coello, Y. Falck-Ytter, R. Jaeschke, G. Vist, E. A. Akl, P. N. Post, S. Norris, J. Meerpohl, V. K. Shukla, M. Nasser and H. J. Schunemann (2011). "GRADE guidelines: 8. Rating the quality of evidence-indirectness." J Clin Epidemiol **64**(12): 1303-1310.

Guyatt, G. H., A. D. Oxman, R. Kunz, J. Woodcock, J. Brozek, M. Helfand, P. Alonso-Coello, P. Glasziou, R. Jaeschke, E. A. Akl, S. Norris, G. Vist, P. Dahm, V. K. Shukla, J. Higgins, Y. Falck-Ytter and H. J. Schunemann (2011). "GRADE guidelines: 7. Rating the quality of evidence-inconsistency." J Clin Epidemiol **64**(12): 1294-1302.

Guyatt, G. H., A. D. Oxman, S. Sultan, P. Glasziou, E. A. Akl, P. Alonso-Coello, D. Atkins, R. Kunz, J. Brozek, V. Montori, R. Jaeschke, D. Rind, P. Dahm, J. Meerpohl, G. Vist, E. Berliner, S. Norris, Y. Falck-Ytter, M. H. Murad and H. J. Schunemann (2011). "GRADE guidelines: 9. Rating up the quality of evidence." J Clin Epidemiol **64**(12): 1311–1316.

Guyatt, G. H., A. D. Oxman, V. Montori, G. Vist, R. Kunz, J. Brozek, P. Alonso-Coello, B. Djulbegovic, D. Atkins, Y. Falck-Ytter, J. W. Williams, Jr., J. Meerpohl, S. L. Norris, E. A. Akl and H. J. Schunemann (2011). "GRADE guidelines: 5. Rating the quality of evidence-publication bias." J Clin Epidemiol **64**(12): 1277–1282.

Guyatt, G., A. Oxman, R. Kunz, J. Brozek, P. Alonso-Coello, D. Rind, P. Devereaux, V. Montori, B. Freyschuss, G. Vist, R. Jaeschke, J. Williams Jr, M. Murad, D. Sinclair, Y. Falck-Ytter, J. Meerpohl, C. Whittington, K. Thorlund, J. Andrews and H. Schünemann (2011). "GRADE guidelines: 6. Rating the quality of evidence--imprecision." J Clin Epidemiol **64**(12): 1283-1293.

Higgins, J. and J. Deeks (2011). Chapter 7: Selecting studies and collecting data. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. J. Higgins and S. Green, The Cochrane Collaboration.

Higgins, J. P. T. and S. Green (2011). Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [Updated March 2011], The Cochrane Collaboration. Available from www.cochrane-handbook.org.

Johnson PI, Koustas E, Vesterinen HM, Sutton P, Atchley DS, Kim AN, Campbell M, Donald JM, Sen S, Bero L, Zeise L, Woodruff TJ. (2016) Application of the Navigation Guide systematic review methodology to the evidence for developmental and reproductive toxicity of triclosan. Environ Int. 2016 Jul-Aug;92-93:716-28. doi: 10.1016/j.envint.2016.03.009. PMID: 27156197

Johnson, P., P. Sutton, D. S. Atchley, E. Koustas, J. Lam, K. A. Robinson, S. Sen, D. A. Axelrad and T. J. Woodruff (2014). "The Navigation Guide - Evidence-Based Medicine Meets Environmental Health: Systematic Review of Human Evidence for PFOA Effects on Fetal Growth." Environmental Health Perspectives. 122(10): 1040-51.

Koustas, E., J. Lam, P. Sutton, P. I. Johnson, D. S. Atchley, S. Sen, K. A. Robinson, D. A. Axelrad and T. J. Woodruff (2014). "The Navigation Guide - Evidence-Based Medicine Meets Environmental Health:

Systematic Review of Non-Human Evidence for PFOA Effects on Fetal Growth." Environmental Health Perspectives. 122(10): 1015-27.
Krauth, D., T. J. Woodruff and L. Bero (2013). "Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review." Environmental Health Perspectives **121**(9): 985-992.
Lam, J., E. Koustas, P. Sutton, P. I. Johnson, D. S. Atchley, S. Sen, K. A. Robinson, D. A. Axelrad and T. J. Woodruff (2014). "The Navigation Guide-Evidence-Based Medicine Meets Environmental Health: Integration of Animal and Human Evidence for PFOA Effects on Fetal Growth." Environ Health Perspect.
Lam J, Sutton P, Kalkbrenner A, Windham G, Halladay A, Koustas E, Lawler C, Davidson L, Daniels N, Newschaffer C, Woodruff T. (2016). A systematic review and meta-analysis of multiple airpborne pollutants and autism spectrum disorder. PLoS One. 11(9):e0161851.
Lamichhane DK, Leem JH, Lee JY, Kim HC. A meta-analysis of exposure to particulate matter and adverse birth outcomes. Environ Health Toxicol. 2015;30:e2015011.
Light, R. J. and D. B. Pillemer (1984). Summing up: the science of reviewing research, Harvard University Press.
Loccisano, A. E., M. P. Longnecker, J. L. Campbell, Jr., M. E. Andersen and H. J. Clewell, 3rd (2013). "Development of PBPK models for PFOA and PFOS for human pregnancy and lactation life stages." J Toxicol Environ Health A **76**(1): 25-57.
Lundh, A., S. Sismondo, J. Lexchin, O. A. Busuioc and L. Bero (2012). "Industry sponsorship and research outcome." Cochrane database of systematic reviews (Online) **12**.
National Research Council (2014). Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC, National Academies Press.
National Research Council (2014). Review of the Environmental Protection Agency's State-of-the-Science Evaluation of Nonmonotonic Dose–Response Relationships as They Apply to Endocrine Disruptors. Washington, DC, National Academies Press.
Rooney, A. A., A. L. Boyles, M. S. Wolfe, J. R. Bucher and K. A. Thayer (2014). "Systematic review and evidence integration for literature-based environmental health science assessments." Environ Health Perspect **122**(7): 711-718.
Roseman, M., K. Milette, L. A. Bero, J. C. Coyne, J. Lexchin, E. H. Turner and B. D. Thombs (2011). "Reporting of conflicts of interest in meta-analyses of trials of pharmacological treatments." JAMA **305**(10): 1008-1017.
Savitz, D. A. "Guest editorial: Biomarkers of perfluorinated chemicals and birth weight. ." Environ Health Perspect **115**: A528-529.
Sterne, J. A., A. J. Sutton, J. P. Ioannidis, N. Terrin, D. R. Jones, J. Lau, J. Carpenter, G. Rucker, R. M. Harbord, C. H. Schmid, J. Tetzlaff, J. J. Deeks, J. Peters, P. Macaskill, G. Schwarzer, S. Duval, D. G. Altman, D. Moher and J. P. Higgins (2011). "Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials." BMJ **343**: d4002.
Stieb DM, Chen L, Eshoul M, Judek S. Ambient air pollution, BW and preterm birth: a systematic review and meta-analysis. Environ Res 2012;117:100-111.
Sun X, Luo X, Zhao C, Zhang B, Tao J, Yang Z, Ma W, Liu T. The associations between birth weight and exposure to fine particulate matter (PM2.5) and its chemical constituents during pregnancy: A meta-analysis. Environ Pollut 2016;211:38-47.
US Environmental Protection Agency (2014). Draft Developmental Materials for the Integrated Risk Information System (IRIS) Toxicological Review of Inorganic Arsenic. Washington, DC.
Vesterinen, H. M., P. I. Johnson, D. S. Atchley, P. Sutton, J. Lam, M. G. Zlatnik, S. Sen and T. J. Woodruff (2014). "The relationship between fetal growth and maternal glomerular filtration rate: A systematic review." In Press. Journal of Maternal-Fetal & Neonatal Medicine.

Viswanathan, M., M. Ansari, N. Berkman, S. Chang, L. Hartling, L. McPheeters, P. Santaguida, T. Shamliyan, K. Singh, A. Tsertsvadze and J. Treadwell (2012). Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews.

Whitworth, K. W., L. S. Haug, D. D. Baird, G. Becher, J. A. Hoppin, R. Skjaerven, C. Thomsen, M. Eggesbo, G. Travlos, R. Wilson, L. A. Cupul-Uicab, A. L. Brantsaeter and M. P. Longnecker (2012). "Perfluorinated compounds in relation to birth weight in the Norwegian Mother and Child Cohort Study." Am J Epidemiol **175**(12): 1209-1216.

Woodruff, T. J. and P. Sutton (2014). "The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes." Environ Health Perspect **122**(10): 1007-1014.

Woodruff, T. J. and P. Sutton (2014). "The Navigation Guide sytematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes." Environmenal Health Perspectives.

Woodruff, T. J., P. Sutton and T. N. G. W. Group (2011). "An Evidence-Based Medicine Methodology To Bridge The Gap Between Clinical And Environmental Health Sciences." Health Affairs **30**(5): 931-937.