EPA/635/R-19/243
IRIS Assessment Protocol
www.epa.gov/iris

Systematic Review Protocol for the Methylmercury IRIS Assessment (Preliminary Assessment Materials)

[CASRN 22967-92-6]

May 2020

Integrated Risk Information System
Center for Public Health and Environmental Assessment
Office of Research and Development
U.S. Environmental Protection Agency
Washington, DC

DISCLAIMER

This document is a public comment draft for review purposes only. This information is distributed solely for the purpose of public comment. It has not been formally disseminated by EPA. It does not represent and should not be construed to represent any Agency determination or policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

CPAD	Chemical and Pollutant Assessment Division
CPHEA	Center for Public Health and Environmental Assessment
DNT	developmental neurotoxicity
EPA	U.S. Environmental Protection Agency
HERO	Health and Environmental Research Online
IAP	IRIS Assessment Plan
IRIS	Integrated Risk Information System
NAS	National Academy of Sciences
NRC	National Research Council
OLEM	Office of Land and Emergency Management
PBPK	physiologically based pharmacokinetic
PECO	populations, exposures, comparators, and outcomes
PK	pharmacokinetic
POD	point of departure
RfD	reference dose
SWIFT	Sciome Workbench for Interactive computer-Facilitated Text-mining
UF	uncertainty factor

AUTHORS | CONTRIBUTORS | REVIEWERS

Assessment Team

Leonid Kopylev, Ph.D. (co-Assessment Manager)	U.S. EPA/ORD/CPHEA/CPAD
Deborah Segal, M.H. S. (co-Assessment Manager)	
Johanna Congleton, Ph.D.	
Yu-Sheng Lin, Ph.D.	
Rebecca Nachman, Ph.D.	
Elizabeth Radke-Farabaugh, Ph.D.	

Executive Direction

Wayne Cascio, M.D. (CPHEA Director)

Samantha Jones, Ph.D. (CPHEA Associate Director)

Kristina Thayer, Ph.D. (CPAD Director)

Emma Lavoie, Ph.D. (CPHEA Senior Science Advisor for Assessments)

Andrew Kraft, Ph.D. (CPAD Senior Science Advisor)

Paul White, Ph.D. (CPAD Senior Science Advisor)

Belinda Hawkins, Ph.D. (CPAD Senior Science Advisor)

Ravi Subramaniam, Ph.D. (CPAD Branch Chief)

Production Team

Hillary Hollinger	HERO Librarian
Ryan Jones	HERO Director
Samuel Thacker	HERO Technical Information Specialist
Erin Vining	HERO Data Specialist/ORAU Student Contractor
Vicki Soto	Project Management Team
Dahnish Shams	Project Management Team
Maureen Johnson	NCEA Webmaster

1. INTRODUCTION

A draft assessment plan for methylmercury (MeHg) was presented at a public science meeting on May 15, 2019 (https://www.epa.gov/iris/iris-public-science-meeting-may-2019) to seek input on the problem formulation components of the assessment plan. The assessment plan summarizes the Integrated Risk Information System (IRIS) Program's scoping and problem formulation conclusions, specifies the objectives and specific aims of the assessment, provides draft PECO (populations, exposures, comparators, and outcomes) criteria, and identifies key areas of scientific complexity.

This protocol document presents the methods for conducting the systematic review and dose-response analysis, including any adjustments made to the specific aims and PECO criteria for the assessment in response to public input on the assessment plan. While the IRIS Assessment Plan (IAP) describes *what* the assessment will cover, chemical-specific protocols describe *how* the assessment will be conducted (see Figure 1 for specific aims of the MeHg assessment). The IRIS Program posts assessment protocols on its website and in repositories such as Zenodo (https://zenodo.org/). Public comments will be considered as part of developing the draft assessment. Literature search results will also be posted in HERO (Health and Environmental Research Online) when they are available.

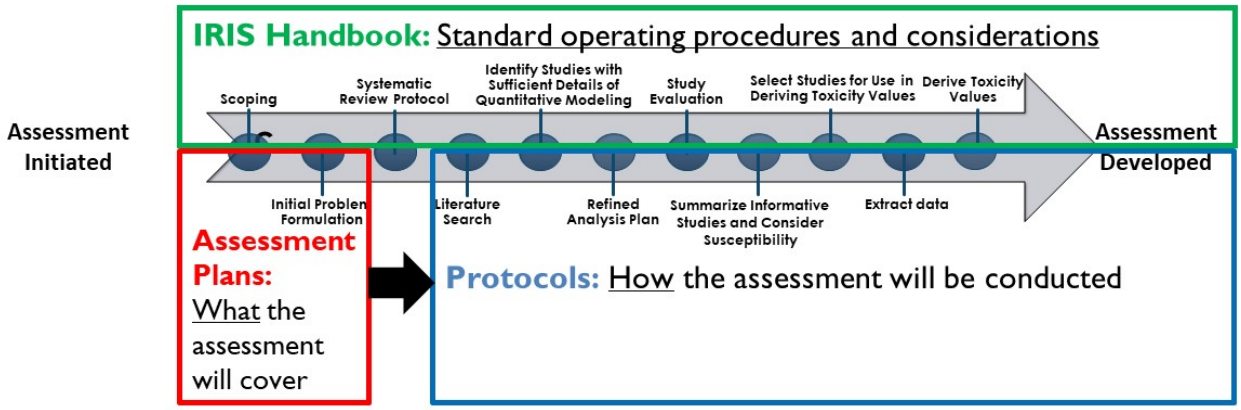


Figure 1. Methylmercury IRIS systematic review problem formulation and method documents.

2. SCOPING AND INITIAL PROBLEM FORMULATION SUMMARY

2.1. BACKGROUND

Multiple health agencies (Health Canada, 2007; UNEP, 2002; U.S. EPA, 2001b; ATSDR, 1999; U.S. EPA, 1997) and the National Academy of Sciences' (NAS) National Research Council (NRC, 2000) have established that prenatal exposure to methylmercury in humans causes developmental neurotoxicity (DNT). An existing IRIS reference dose (RfD) for methylmercury was published in 2001 (U.S. EPA, 2001b) and was based on an NAS assessment from 2000 (NRC, 2000). The outcomes described by NAS included impaired cognitive function, motor function, visuospatial performance, and abnormal (increased or decreased) muscle tone following in utero methylmercury exposure (NRC, 2000). The RfD of 0.1 µg/kg-day[1] was derived from maternal daily intakes of methylmercury of 0.86–1.47 µg/kg-day, estimated to result in cord blood concentrations of 46–79 µg/L associated with multiple DNT measures (specifically, developmental neuropsychological[2] impairment) in a Faroe Island cohort described by Grandjean et al. (1997). This epidemiology study found impaired cognitive function in 7-year-old children from the Faroe Islands who were prenatally exposed to methylmercury (Budtz-Jørgensen et al., 1999; Grandjean et al., 1997). IRIS's previous 1995 RfD for methylmercury was the same as the 2001 RfD and was also based on DNT outcomes from in utero exposure using data from a 1971 Iraqi poisoning incident [derivation described in U.S. EPA (1997)]. In both previous IRIS assessments, following comprehensive literature searches and evaluations in each case, DNT outcomes were concluded to be the most sensitive (other outcomes are discussed in Section 2.3).

Methylmercury is formed when inorganic mercury is methylated by biota in water and soil. Gaseous elemental mercury is released into the atmosphere from natural (e.g., volcanoes) and anthropogenic (e.g., fossil-fuel combustion) sources. Elemental mercury can be converted to inorganic mercury, which then can be transported to land or water through wet or dry deposition processes. Combustion processes can also release inorganic ionic mercury, which can adsorb to particulate matter (Srivastava et al., 2006). Inorganic divalent mercury adsorbed to particulates can deposit after traveling relatively short distances, compared to elemental mercury vapor that

[1]Expressed as a concentration in whole maternal blood, the RfD is approximately 3.5 µg/L (Mahaffey et al., 2009).

[2]In the 2001 IRIS assessment of methylmercury, the term *developmental neuropsychological impairment* was used to describe the adverse effects on the nervous system identified in humans following exposures to methylmercury during developmental life stages. Developmental neuropsychological impairment is a type of DNT, and is the term used in many epidemiological studies.

can travel long distances. Once deposited, microorganisms convert inorganic mercury to methylmercury, which then bioaccumulates in fish tissue. Concentrations of methylmercury in fish tissue, particularly predatory fish higher on the food chain (e.g., swordfish), can be much greater than methylmercury concentrations found in ambient water (U.S. EPA, 2010).

Consumption of contaminated fish and other seafood is the major pathway for exposure to methylmercury in humans (NRC, 2000); however, other foods, such as rice, can also expose humans to methylmercury (Wells et al., 2020; Cui et al., 2017; Rothenberg et al., 2017; Rothenberg et al., 2016). Between 2011 and 2014, average blood methylmercury levels in the U.S. population ranged from 0.434 to 0.498 µg/L (CDC, 2017). During this same period, average total blood mercury levels, which often are used as a basis for determining methylmercury blood levels, ranged from 0.678 to 0.703 µg/L between 2011 and 2016 (CDC, 2018). Males had slightly higher methylmercury blood levels than females. For example, the average methylmercury blood level in 2013–2014 was 0.448 µg/L and 0.422 µg/L for males and females, respectively. Blood methylmercury levels were also found to increase with age. In 2011 and 2012, the most recent years that methylmercury blood levels were available for several age groups, the average for children 6 to 11 years of age was 0.209 µg/L; for 12 to 19 year-olds, it was 0.276 µg/L; and for adults over 19, it was 0.624 µg/L (CDC, 2018). The estimated mean daily intake of total mercury for women older than 20 years in the United States is approximately 1 µg/day[3] (CDC, 2016a; Birch et al., 2014).

Methylmercury readily crosses the placenta and concentrates in cord blood at approximately 1.7 times the levels in maternal blood (Straka et al., 2016; Stern and Smith, 2003; Yang et al., 1997). It is also transferred from mothers to children via breastmilk (CDC, 2009; ATSDR, 1999). As noted earlier, the developing nervous system is particularly sensitive to methylmercury, so these gestational, lactational, and other postnatal exposures are of great concern. Methylmercury exposures to women of childbearing age who could become pregnant might be harmful as well, as studies have reported an average half-life of methylmercury in the body of 50 days, which might then result in fetal exposure early in pregnancy (CDC, 2016b). A one-compartment toxicokinetic model estimated a longer half-life for methylmercury, 80 days, on the basis of blood samples from an adult population (Jo et al., 2015). The half-life of methylmercury varies among individuals, as some individuals have longer clearance times than others. For example, EPA's 2001 assessment reported half-lives for methylmercury ranging from 32 to 189 days after evaluating data from 5 studies (Smith et al., 1994; Sherlock et al., 1984; Kershaw et al., 1980; Al-Shahristani and Shihab, 1974; Miettinen et al., 1971).

Subsistence fishing communities and other populations with high dietary intakes of predatory fish species could be exposed to higher-than-average levels of methylmercury. Therefore, women of childbearing age and children in these communities could have high

[3]Based on the calculated average monthly mercury intake using 2009–2010 NHANES (National Health and Nutrition Examination Survey) data reported by Birch et al. and CDC's anthropometric reference values for 2011–2014 (CDC, 2016a; Birch et al., 2014).

methylmercury exposures during susceptible life stages. People who consume fish from habitats with high methylmercury concentrations due to large microbial populations that convert inorganic mercury to methylmercury also might have particularly high exposures. This includes people eating fish from certain types of wetlands, rivers with a high proportion of wetlands in their watersheds, dilute and low-pH lakes in the Northeast and Northcentral United States, parts of the Florida Everglades, newly flooded reservoirs, and coastal wetlands particularly along the Gulf of Mexico, Atlantic Ocean, and San Francisco Bay (U.S. Department of the Interior, 2000). In some regions of the world, consumption of fish from waters polluted by mercury from small-scale and artisanal gold mining also might result in high methylmercury exposures. Contaminated rice and rice-based food products, such as infant cereals, also can be a source of methylmercury exposure (Cui et al., 2017; Rothenberg et al., 2017; Rothenberg et al., 2016).

2.2. SCOPING SUMMARY

During the scoping process, the IRIS Program met with EPA program and regional offices that had an interest in an IRIS reassessment of methylmercury to discuss specific needs. Table 1 provides a summary of input from this outreach.

Table 1. EPA program and regional office interest in a methylmercury assessment

EPA program or regional office	Oral	Inhalation	Statute/ Regulation	Anticipated uses/interest
Office of Land and Emergency Management (OLEM) EPA Regions 1–10	✓	✓[a]	Comprehensive Environmental Response, Compensation and Liability Act (CERCLA)	CERCLA authorizes EPA to conduct short- or long-term cleanups at Superfund sites and later recover cleanup costs from potentially responsible parties under section 107. Methylmercury toxicological information may be used to make risk determinations for such response actions (e.g., short-term removals, long-term remedial response actions).
			Resource Conservation and Recovery Act (RCRA)[a]	Mercury is listed under RCRA as a characteristic (40 CFR 261.24) and hazardous waste (40 CFR 261.33). Methylmercury toxicological information may be used to evaluate mercury toxicity from releases of elemental mercury and mercury compounds as environmental sources of methylmercury.
			Clean Water Act (CWA)	CWA requires EPA to develop water quality criteria for states and tribes to use in developing water quality standards, requires states and tribes to adopt water quality criteria that protect designated uses such as fish consumption, and requires states and authorized tribes to review water quality standards every three years and modify them on the basis of updated health effects studies derived by EPA.

[a]Although OLEM initially expressed the need for an inhalation reference concentration (RfC) for MeHg, this was de-prioritized during subsequent scoping and problem formulation discussions on the basis of a lack of significant inhalation exposure.

2.3. PROBLEM FORMULATION

Based on a preliminary survey of the methylmercury literature, including review of assessments conducted by other agencies, potential health outcomes identified other than DNT include the following:

- Nervous system outcomes (non-developmental)
- Developmental outcomes (other than nervous system effects)
- Cardiovascular outcomes
- Immune system outcomes
- Reproductive outcomes

DNT resulting from oral exposure was selected as the focus of this first assessment module because it is a well-established hazard and the two previous RfDs for methylmercury were derived for oral exposure DNT outcomes. Some, but not all, recent epidemiology studies have reported DNT adverse outcomes at exposure levels lower than exposure levels found in studies used to derive the current reference dose. Many of these recent studies provide exposure-response information, which justifies and enables reevaluation of the 2001 RfD (U.S. EPA, 2001b; NRC, 2000). Several studies investigated cognitive function [e.g., Golding et al. (2016); Jacobson et al. (2015); Orenstein et al. (2014); Sagiv et al. (2012); Lederman et al. (2008); Oken et al. (2008); Oken et al. (2005)] and motor function [e.g., Prpić et al. (2017); Golding et al. (2016); Suzuki (2016); Lederman et al. (2008); Després et al. (2005); Daniels et al. (2004)] at various ages following prenatal or postnatal exposures to methylmercury. Other DNT outcomes (e.g., behavioral, structural, and electrophysiological) following methylmercury exposures also have been evaluated [e.g., Jin et al. (2016); Ng et al. (2015); Boucher et al. (2010)]. This assessment will only reassess and update the existing dose response for DNT outcomes. It will not reevaluate whether methylmercury causes DNT outcomes because DNT is a well-established human hazard (as discussed in Section 2.1, Background). Also, this assessment will not assess the potential for methylmercury exposure to cause the other possible health outcomes of interest described above (see Section 2.4).

Once completed, the DNT dose-response assessment for oral exposure will undergo public comment/peer review and finalization. After the assessment of DNT outcomes, EPA plans to assess cardiovascular endpoints in a second assessment module and will also consider the need for assessment of other endpoints, such as adult nervous system and reproductive effects. Cardiovascular outcomes were identified as a specific priority during the public comment period and science webinar on the methylmercury IAP (https://www.regulations.gov/docket?D=EPA-HQ-ORD-2018-0655). The decision to evaluate other outcomes aside from DNT and cardiovascular effects will be based on examining whether there is sufficient evidence to assess hazard, derive

reference values, and consider whether the additional analyses are likely to impact EPA decision making beyond what the DNT and cardiovascular assessments provide.

Because ingestion is the primary route of exposure for methylmercury (NRC, 2000), inhalation and dermal routes of exposure are not addressed in this assessment. Although OLEM initially expressed the need for an inhalation reference concentration (RfC) for MeHg, the need was de-prioritized during subsequent scoping and problem formulation discussions on the basis of lack of significant inhalation exposure.

The reassessment of DNT dose response will focus on human studies because the availability of a large epidemiology database on methylmercury exposure and DNT outcomes [e.g., review by Karagas et al. (2012)] eliminates uncertainties associated with interspecies extrapolation. During this reassessment, IRIS will evaluate epidemiology evidence for all types of DNT outcomes resulting from exposure to the fetus, infants, children, or adolescents because during development the brain is more vulnerable to the neurotoxicity of methylmercury. Targeted literature searches might be conducted for animal toxicological or mechanistic studies to address data gaps (e.g., susceptibility) or to replace default uncertainty factors (UFs) with data-derived factors.

Public comments on the MeHg IAP also suggested the IRIS Program should conduct an assessment to examine the adverse effects of developmental MeHg exposure and the beneficial effects related to seafood consumption during pregnancy. Consideration of the health benefits of fish consumption falls outside the traditional scope of the IRIS Program and was not identified as a current EPA National Program priority. However, the IRIS Program understands the importance of both types of effects, in particular, for providing fish consumption advice. In addition, as outlined in the Specific Aims (Section 3.1), the assessment will seek to determine if the available data would also support the derivation of dose-response relationships for DNT outcomes that would be useful for analyses conducted by others to quantify the health impacts of actions to reduce exposures to MeHg. The IRIS Program is also communicating with staff at the U.S. Department of Agriculture supporting the 2020 Dietary Guidelines Advisory Committee (Dietary Fats and Seafood Subcommittee) regarding their review of "What is the relationship between types of dietary fat consumed and neurocognitive development (birth to 18 years) or neurocognitive health (for those 18 years and older)?" (https://www.dietaryguidelines.gov/dietary-fats-and-neurocognitive-health). Although stakeholders indicated a need for evaluation of other forms of mercury (e.g., elemental and inorganic salts), these mercury forms would need to be evaluated through separate assessments. Currently, IRIS is developing an assessment of inorganic mercury salts; however, no other forms of mercury were identified as an Agency priority.

Other public comments on the methylmercury IAP focused on the following topics: an agreement that DNT should be the focus of the first module, the need to consider confounding by fish nutrients and the need to consider biomarker imprecision (see Key Science issues), and the need to consider individual variation (see Specific Aims).

2.4. ASSESSMENT APPROACH

As described above, this assessment will use a modular approach whereby EPA will first evaluate DNT outcomes associated with oral exposure in the first assessment module. DNT was selected for the first module because it is a well-established, sensitive hazard and the two previous RfDs for methylmercury were derived for oral exposure DNT outcomes (see Section 2.1). In addition to DNT, EPA plans to assess cardiovascular outcomes in a second assessment module and will also consider the need for assessing other endpoints, such as the adult nervous system and reproductive effects in additional module(s). Once completed, the draft assessment addressing the DNT dose-response relationship for oral exposure will undergo public comment/peer-review and finalization, rather than waiting for the cardiovascular module to be completed. The decision to evaluate other outcomes aside from DNT and cardiovascular toxicity will be based on examining whether evidence is sufficient to assess hazard, derive reference values, and consider whether the additional analyses are likely to impact EPA decision making beyond what the DNT and cardiovascular assessments provide.

While completing the DNT module, EPA will survey the available hazard information for cardiovascular and other adverse health outcomes (see Section 2.3 for list), primarily by reviewing methylmercury assessments by other agencies and organizations, and recent epidemiology studies. For health effects for which hazard has not been established on the basis of epidemiology evidence, animal and mechanistic studies will also be surveyed. The cardiovascular and any other assessment modules will have their own IAPs that will be released separately. Because inhalation exposure to MeHg is not a significant route of exposure, only oral exposure studies will be evaluated.

2.5. KEY SCIENCE ISSUES

Based on the preliminary literature survey and public comments, the following key scientific issues were identified and will be addressed in this assessment as indicated below. The assessment will consider:

- The accuracy and reliability of measures of the different types of biomarkers (e.g., hair, maternal blood, cord blood) to quantify methylmercury exposure.

  EPA plans to use analytical chemistry criteria to evaluate the accuracy and reliability of measures of the different types of biomarkers (see Appendix C).

- How to best use the different biomarkers measured in PECO-relevant epidemiology studies to inform estimates of the relationship between methylmercury exposure and neurodevelopmental effects. For example, some epidemiology studies measure methylmercury directly in human blood, hair, or nails, while other studies rely on measures of total mercury to estimate methylmercury exposure.

EPA plans to consider several approaches for utilizing studies that rely on total mercury measures to estimate methylmercury exposure: Regression modeling of the methylmercury based on total mercury and, possibly, covariates; accepting that total mercury is an adequate proxy for methylmercury (e.g., for hair); or deriving RfDs for both total mercury and methylmercury.

- How potential confounding [e.g., Budtz-Jorgensen et al. (2007)] in studies will be accounted for in the analyses. For example, many fish species that contain methylmercury also have beneficial nutrients, such as selenium and polyunsaturated fatty acids, which are important to brain development. In addition, fish could contain other contaminants that might be harmful to brain development, such as polychlorinated biphenyls. Accounting for confounders will be assessed during study evaluation.

EPA plans to assess whether confounders were appropriately accounted for as part of study evaluation.

- The differences in DNT evaluation methods and how their results could be used in this assessment. For example, developmental scores are consistently higher for both term and preterm infants when using the Bayley III test versus the Bayley II test, and some suggest using an adjustment factor to compare the two scores (Lowe et al., 2012).

EPA plans to use criteria (currently in development in collaboration with NTP using contractors who are experts in the field) for evaluating DNT tests and their appropriate use in epidemiology studies evaluating DNT effects of methylmercury exposure.

3. OVERALL OBJECTIVES, SPECIFIC AIMS, AND POPULATIONS, EXPOSURES, COMPARATORS, AND OUTCOMES (PECO) CRITERIA

The overall objective of this assessment is to characterize the dose-response relationship between methylmercury exposure and DNT outcomes and then use this information to update the existing RfD. Because the current RfD for methylmercury was posted by IRIS in 2001 and was based on an NAS (NRC, 2000) assessment, evaluation of studies since 1998 is expected to capture literature that was not considered in the earlier assessments. Any new health risk assessments for methylmercury identified in the search for newer literature will be reviewed as secondary literature sources. The relevant dose-response analyses included in these previous assessments will also be considered in this reassessment. Studies that evaluated the relationships between methylmercury exposures to women of childbearing age and the developing child and DNT outcomes that become apparent at any life stage (infancy through the elderly) will be considered. A conceptual model is presented below to illustrate the focus of the planned assessment (Figure 2). A critical effect will be selected for derivation of a RfD that will be protective against all DNT effects that occur at any age following prenatal to adolescent methylmercury exposure.

Systematic review methods will be used to evaluate the epidemiology literature on DNT outcomes, and the analysis conducted will be consistent with all relevant EPA guidance.[4] As part of this systematic review, potentially susceptible populations, for example, populations with certain genetic polymorphisms, and life stages will be considered. This Systematic Review Protocol reflects scoping and problem formulation changes made to the specific aims and PECO criteria in response to public input received on the MeHg IAP.

[4]EPA guidance documents: http://www.epa.gov/iris/basic-information-about-integrated-risk-information-system#guidance/.

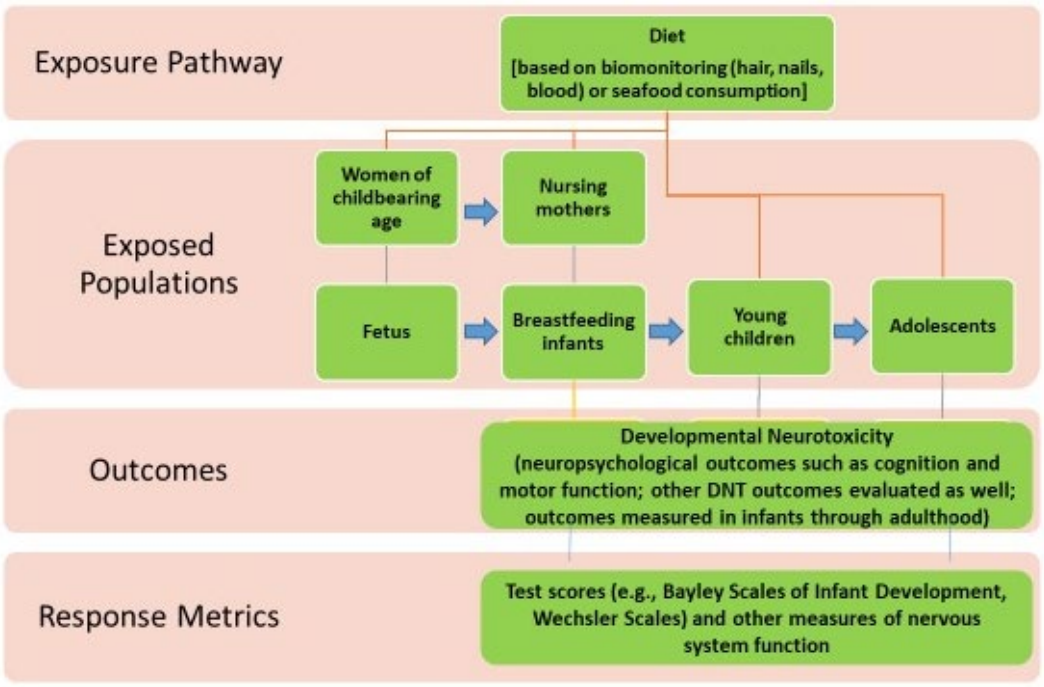


Figure 2. Simplified conceptual model of the reassessment of DNT resulting from exposure to MeHg.

3.1. SPECIFIC AIMS

- Identify epidemiology literature examining effects of exposure to methylmercury as outlined in the PECO criteria (see Section 3.2, Table 2). Develop and execute a literature search strategy to broadly capture data from MeHg epidemiology studies published since 1998, and screen results for relevance.

- Use predefined criteria to identify epidemiology studies from the screened results that provide exposure-response information for DNT outcomes.

- Conduct study evaluations (risk of bias and sensitivity) for identified epidemiology studies. This will include an assessment of the proper consideration of confounders (e.g., fish nutrients such as polyunsaturated fatty acids). Studies with critical deficiencies generally will be considered uninformative and not considered further.

- Summarize study methods and results from epidemiology studies on DNT outcomes, including explicit identification and discussion of issues concerning susceptible populations and life stages, including potentially important genetic polymorphisms.

- Evaluate whether dose conversion [i.e., physiologically based pharmacokinetic (PBPK) modeling] is needed. Depending on the biomarker (e.g., cord blood), conduct a search and review of the relevant literature as needed to determine if calculations used in the previous assessment (to convert from cord blood to oral exposure) need to be updated. If necessary,

individual PBPK models will be evaluated using predefined criteria, and their strengths and uncertainties will be summarized.

- Characterize uncertainties, including individual variability in MeHg toxicokinetics where data are available to do so, thereby reducing reliance on default UFs. Identify key data gaps and research needs, such as limitations of the evidence base and the systematic review.

- Derive a toxicity value (e.g., RfD) for DNT outcomes as supported by the available data.

- Assemble the available data to support analyses conducted by others to quantify the health impacts of actions to reduce exposures to MeHg.

3.2. POPULATIONS, EXPOSURES, COMPARATORS, AND OUTCOMES (PECO) CRITERIA

The PECO criteria are used to identify the evidence that addresses the specific aims of the assessment and to focus the search terms and inclusion/exclusion criteria in a systematic review. The draft PECO criteria for this MeHg assessment (Tables 2 and 3) were based on (1) basis for the chemical's prioritization for assessment, (2) discussions with scientists in EPA program and regional offices to determine the scope of the assessment that will best meet Agency needs, and (3) preliminary review of the DNT literature for MeHg (primarily reviews and authoritative health assessment documents).

Table 2. PECO criteria for epidemiology data

PECO element	Evidence
Populations	Human populations exposed during life stages ranging from the fetus through adolescence.
Exposures	Any quantitative exposure to MeHg based on biomonitoring data (e.g., hair, nails, blood). Measurements must be either direct MeHg measurements or measurements of total mercury (not other forms of mercury, e.g., mercury salts).
Comparators	Referent populations exposed to lower (within the study) levels of MeHg will be used to examine specific effects. The results of the comparisons must be presented with sufficient detail of quantitative modeling (e.g., regression coefficients presented with statistical measure of variation).
Outcomes	DNT outcomes measured at any age, including—but not limited to—tests or measures of cognition, motor function, behavior, vision, and hearing.

Table 3. PECO criteria for PBPK studies

PECO element	Evidence
Populations	Human populations exposed during life stages ranging from the fetus through adolescence.
Exposures	Implemented for relevant information of exposure (defined by route, time of exposure, intensity, and frequency) that informs toxicokinetic modeling to improve estimation procedures for dietary intake of MeHg using biomonitoring data.
Comparators	Any comparison that helps improve the estimation of body burden of MeHg in humans including absorption, distribution, metabolism, and elimination (ADME) processes.
Outcomes	Any examination of MeHg deposition (ADME) and dose metrics (e.g., peak concentration of blood MeHg) that inform the evaluation of MeHg DNT outcomes.

4. LITERATURE SEARCH AND SCREENING STRATEGIES

4.1. LITERATURE SEARCH STRATEGIES

Literature search strategies were developed for epidemiology studies published since 1998 using key terms and words related to the PECO criteria. Of note, the 2001 RfD for MeHg was derived using the toxicokinetic model [one-compartment pharmacokinetic (PK) model] recommended by the NRC (2000). As PK/PBPK studies published prior to 2001 had been extensively evaluated in the 2001 EPA MeHg assessment (U.S. EPA, 2001b), a literature search was conducted for PK/PBPK studies published since 2001. The search strategy involved identifying relevant search terms through the following approaches: (1) extracting key terminology from relevant reviews and (2) consulting with the HERO librarian. Relevant subject headings and text-words were crafted to maximize the sensitivity and specificity of the search results. No language restrictions were applied. The four databases listed below were searched (PK/PBPK search did not include Toxline). Because each database has its own search architecture, the resulting search strategy was tailored to account for each database's unique search functionality (the detailed search strategies are presented in Appendix A, Tables A-1 and A-3).

- PubMed (National Library of Medicine)
- Web of Science (Thomson Reuters)
- Toxline (National Library of Medicine)
- Science Direct (Elsevier)

Literature searches were conducted using EPA's HERO database.[5] Because PECO criteria focused only on studies that report data amenable to dose-response modeling, the literature search for epidemiology studies was organized as follows. First, all MeHg literature was searched in the four databases listed above and duplicates were removed by HERO. After deduplication in HERO, these studies were imported into SWIFT Review software (Howard et al., 2016) to identify epidemiology studies most likely to be suitable for dose-response analysis. In brief, SWIFT Review has preset literature search strategies ("filters") developed by information specialists that can be applied (or modified) by the user to identify PECO-relevant studies. The filters function like a typical search strategy in which studies are tagged as belonging to a certain filter if the terms in the

[5]Health and Environmental Research Online: https://hero.epa.gov/hero/.

filter literature search strategy appear in title, abstract, keyword, or medical subject headings (MeSH) fields content. The SWIFT Review filter for human evidence was modified slightly to identify epidemiology studies (see Appendix A, Table A-2). In addition, a new search filter containing dose-response terms (e.g., regression, *p*-value) was developed and applied to identify those epidemiology studies most likely to be relevant to the MeHg PECO. Studies that included one or more of the search terms in the title, abstract, keyword, or MeSH fields for epidemiology and dose-response search strings were exported as an RIS (Research Information Systems) file for screening in DistillerSR, as described below. Application of the SWIFT Review filters reduced the number of studies for title and abstract screening from 15,277 to 2,905. In addition, because some articles do not contain abstracts and sometimes abstracts are not imported from the HERO database to SWIFT Review, epidemiology articles without abstracts were identified from the SWIFT Review and screened for relevance by one person.

Because the PBPK literature search resulted in relatively few papers (284 unique references after deduplication), the screening was performed by one person using Endnote. The search strategies developed by SWIFT Review for PBPK studies were used to guide the screening.

The literature search will be updated throughout draft development to identify literature published during the course of review. The last full literature search update will be conducted less than 1 year before the planned release of the draft document for public comment. The results returned (i.e., the number of "hits" from each electronic database or other literature source), including the results of any literature search updates, are documented in the literature flow diagrams (see Section 4.3.2), which also reflect the literature screening decisions (see Section 4.4).

The IRIS Program takes extra steps to ensure identification of pertinent studies by encouraging the scientific community and the public to identify additional studies and ongoing research and by considering late-breaking studies that would affect the credibility of the conclusions, even during the review process. Studies identified after peer review begins will only be considered for inclusion if they meet the PECO criteria and are expected to fundamentally alter the assessment's conclusions. Release of the PECO-screened literature in parallel with release of the protocol for public comment provides an opportunity for stakeholders to identify any missing studies, which if identified, will be screened as outlined above for adherence to the PECO criteria.

4.2. NON-PEER-REVIEWED DATA

IRIS assessments rely mainly on publicly accessible, peer-reviewed studies. However, it is possible that unpublished data directly relevant to the PECO may be identified during assessment development. Depending on the potential impact of the study on assessment conclusions, EPA might obtain external peer review if the owners of the data are willing to have the study details and results made publicly accessible (U.S. EPA, 2015). This independent, contractor-driven, peer review would include an evaluation of the study similar to that for peer review of a journal publication. The contractor would identify and select two or three scientists knowledgeable in

scientific disciplines relevant to the topic as potential peer reviewers. Persons invited to serve as peer reviewers would be screened for conflict of interest. In most instances, the peer review would be conducted by letter review. The study authors would be informed of the outcome of the peer review and given an opportunity to clarify issues or provide missing details. The study and its related information, if used in the IRIS assessment, would become publicly available. In the assessment, EPA would acknowledge that the document underwent external peer review managed by EPA, and the names of the peer reviewers would be identified. In certain cases, IRIS will conduct an assessment for utility and data analysis based on having access to a description of study methods and raw data that has undergone rigorous quality assurance/quality control review (e.g., ToxCast/Tox21 data, results of National Toxicology Program studies) but that have not yet undergone external peer review.

Unpublished data from personal author communication can supplement a peer-reviewed study provided the information is made publicly available (typically through documentation in HERO).

4.3. LITERATURE SCREENING STRATEGY

The PECO criteria were used to determine inclusion or exclusion of a reference as a primary source of health effects data or a published PBPK model. Targeted literature searches might be conducted for animal, mechanistic, or ADME studies to address data gaps (e.g., susceptibility) or to replace default UFs with data-derived factors.

Title and abstract-level screening (epidemiology studies). Following a pilot phase to calibrate screening guidance, two screeners independently conducted a title and abstract screen of the search results to identify records that appear to meet the PECO criteria using a structured form in DistillerSR (Evidence Partners; https://www.evidencepartners.com/products/distillersr-systematic-review-software/).

Screening conflicts were resolved by discussion among the primary screeners with consultation by a third reviewer to resolve any remaining disagreements. Eligibility status of non-English studies was assessed using the same approach, and online translation tools were used to assess eligibility at the title and abstract levels.

Studies not meeting the PECO criteria but identified as "potentially relevant supplemental material" were tagged during the title and abstract screening process (see Figures 3 and 4). Conflict resolution was not required during the screening process to identify supplemental information (i.e., tagging by a single screener is sufficient to identify the study as potentially relevant supplemental material that might be considered during draft development).

Full-text level screening (epidemiology studies). Records not excluded on the basis of the title and abstract were advanced to full-text review. Full-text copies of these potentially relevant records were retrieved, exported from the HERO database to Distiller, and independently assessed by two screeners to confirm eligibility according to the PECO criteria. Screening conflicts were

resolved by discussion between the primary screeners with consultation by a third reviewer (as needed to resolve any remaining disagreements). Studies that advanced to full-text review could also be tagged as "potentially relevant supplemental material."

For the PBPK studies PECO, one person conducted title and abstract and full-text screening using Endnote, because relatively few studies were identified.

The results of this screening process were posted on the project page for this assessment in the HERO database https://hero.epa.gov/hero/index.cfm/project/page/project_id/2589. These studies will be "tagged" with appropriate category descriptors (e.g., studies eligible for study evaluation, potentially relevant supplemental material, excluded). Results are also annotated and reported in a literature flow diagram (see Figures 3 and 4).

It is important to emphasize that being tagged as supplemental material does not mean the study would necessarily be excluded from consideration in the assessment. The initial screening level distinctions between a study that meets the PECO criteria and a supplemental study are designed to ensure the supplemental studies are categorized for easy retrieval while conducting the assessment. The impact on the assessment conclusions of individual studies tagged as supplemental material is often difficult to assess during the screening phase of the assessment. These studies might be critical to the assessment, and if so, they will be summarized at the individual study level. Alternatively, they could be cited because they provide context or they might not be cited at all in the assessment (e.g., individual studies that contribute to a well-established scientific conclusion). In addition, studies might be tagged as supplemental material during either title and abstract or full-text screening.

Release of the PECO-screened literature in the protocol (or protocol update) for public comment provides an opportunity for stakeholders to identify any missing studies. If identified, those studies will be screened as outlined above for adherence to the PECO criteria.

4.3.1. Multiple Publications of the Same Cohort

When a cohort is the subject of multiple publications, all publications focused on the cohort will be included. For each cohort, several primary publications could be selected.

4.3.2. Literature Flow Diagram

Flow diagrams for literature searches for epidemiology and PBPK studies are presented in Figures 3 and 4.

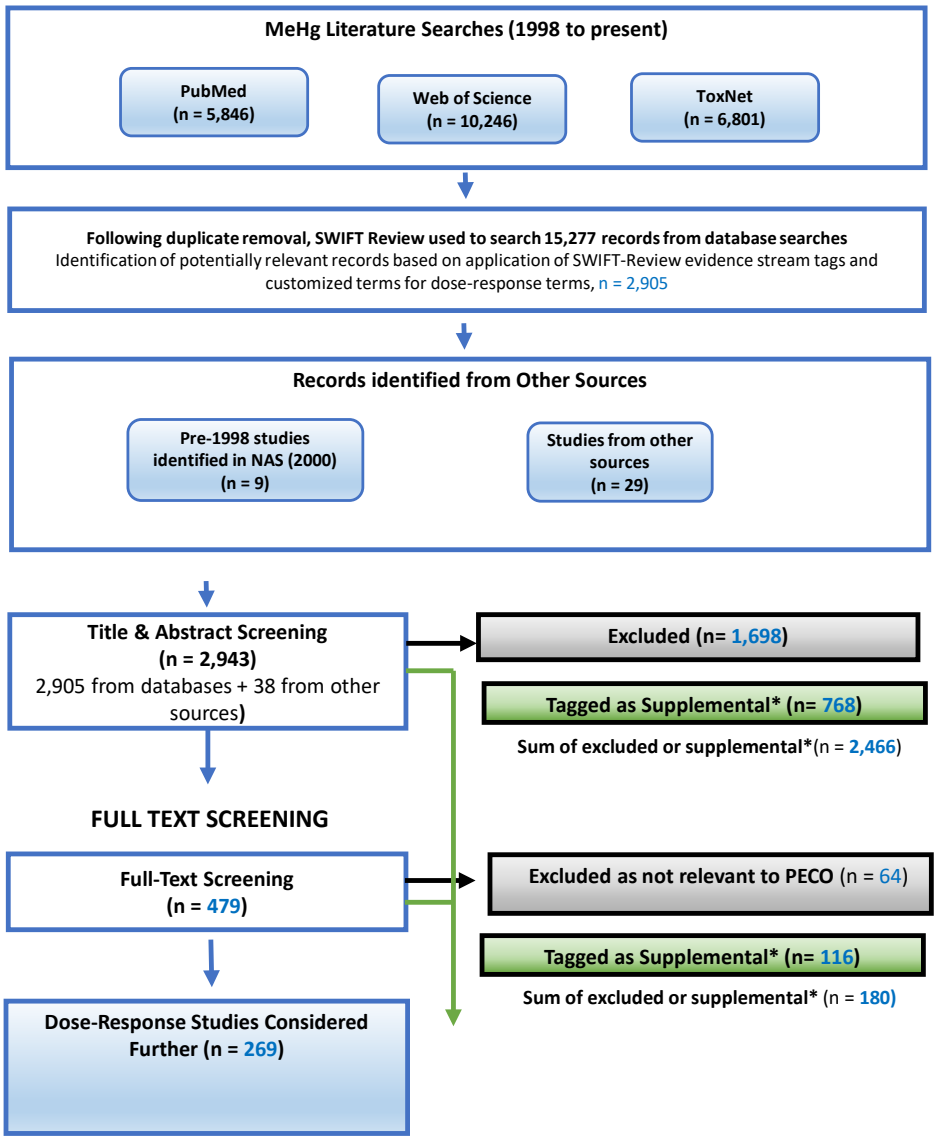


Figure 3. Literature search for MeHg DNT dose-response studies.

[a]Because the literature search was first performed for dose-response epidemiology studies and only then screened, the supplemental literature is not comprehensive and thus not categorized further.

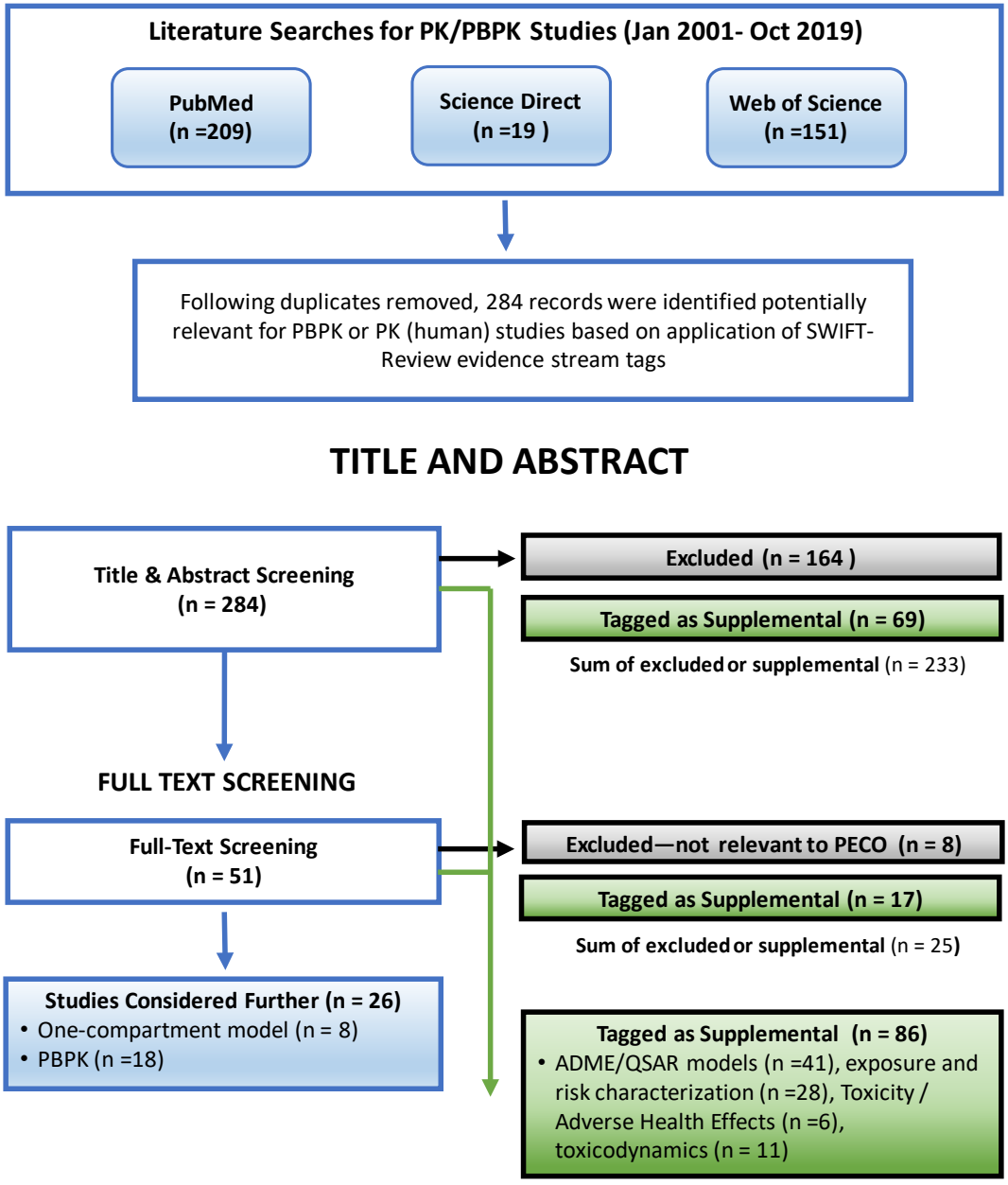


Figure 4. Literature search for MeHg PBPK models.

5. REFINED EVALUATION PLAN

Public comments[6] on the assessment plan for the DNT module did not suggest a change was warranted to the specific aims or PECO; thus, no refined evaluation plan was pursued (i.e., all DNT outcomes in all the studies that met the PECO criteria will be evaluated in this assessment module).

[6] Public comments can be found at https://www.regulations.gov/docket?D=EPA-HQ-ORD-2018-0655.

6. STUDY EVALUATION (REPORTING, RISK OF BIAS, AND SENSITIVITY) STRATEGY

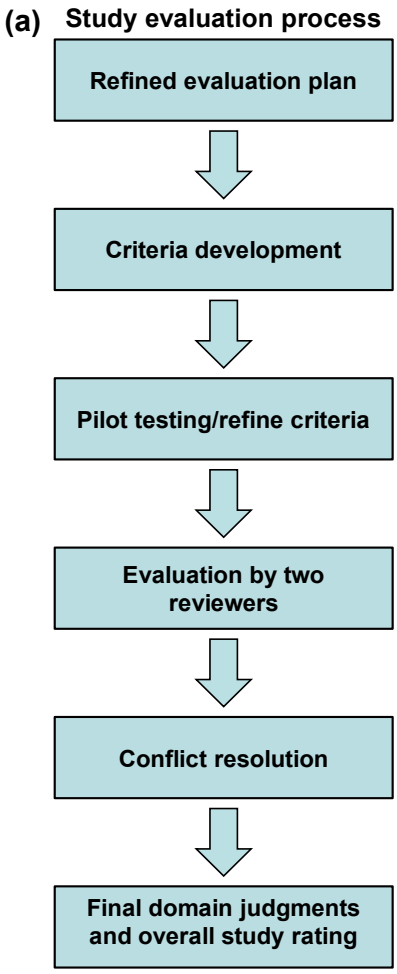
6.1. STUDY EVALUATION OVERVIEW FOR EPIDEMIOLOGY HEALTH EFFECT STUDIES

Evaluation of epidemiology studies of health effects to assess risk of bias and study sensitivity will be conducted for the following domains: exposure measurement, outcome ascertainment, participant selection, potential confounding, analysis, study sensitivity, and selective reporting. Bias can result in false positives and false negatives, while study sensitivity is typically concerned with identifying the latter.

A key concern for the review of epidemiology studies is risk of bias, which is the assessment of internal validity (factors that affect the magnitude or direction of an effect in either direction) and insensitivity (factors that limit the ability of a study to detect a true effect; low sensitivity is a bias toward the null when an effect exists). Reporting quality is evaluated to determine the extent the available information allows for evaluating these concerns. The study evaluations are aimed at discerning the expected magnitude of any identified limitations (focusing on limitations that could substantively change a result), considering also the expected direction of the bias. The study evaluation considerations described below can be refined to address a range of study designs, health effects, and chemicals. The general approach for reaching an overall judgment for the study (or a specific analysis in a study) regarding confidence in the reliability of the results is illustrated in Figure 5.

At least two reviewers will independently evaluate the studies to identify characteristics that bear on the informativeness (i.e., validity and sensitivity) of the results and provide additional chemical- or outcome-specific knowledge or methodological concerns.

Considerations for evaluating studies will be developed in consultation with topic-specific technical experts and existing guidance documents when available, including EPA guidance for neurotoxicity, reproductive toxicity, and developmental toxicity (U.S. EPA, 1998, 1996, 1991). The independent evaluations include a pilot phase to assess and refine the evaluation process. During this phase, decisions will be compared and a consensus reached between reviewers, and when necessary, differences will be resolved by discussion between the reviewers, the chemical assessment team, or technical experts. As reviewers examine a group of studies, additional chemical-specific knowledge or methodological concerns could emerge, and a second pass might become necessary. Refinements to the study evaluation process made during the pilot phase and subsequent implementation will be acknowledged as updates to the protocol.



(b) Individual evaluation Domains

Epidemiology
Participant selection
Confounding
Exposure misclassification
Outcome ascertainment
Analysis
Other sensitivity

Domain judgments

Judgment	Interpretation
Good	Appropriate study conduct relating to the domain and minor deficiencies not expected to influence results.
Adequate	A study that may have some limitations relating to the domain, but they are not likely to be severe or to have a notable impact on results.
Deficient	Identified biases or deficiencies interpreted as likely to have had a notable impact on the results or prevent reliable interpretation of study findings.
Critically Deficient	A serious flaw identified that makes the observed effect(s) uninterpretable. Studies with a critical deficiency will almost always be considered "uninformative" overall.

Overall study rating for an outcome

Rating	Interpretation
High	No notable deficiencies or concerns identified; potential for bias unlikely or minimal; sensitive methodology.
Medium	Possible deficiencies or concerns noted, but resulting bias or lack of sensitivity is unlikely to be of a notable degree.
Low	Deficiencies or concerns were noted, and the potential for substantive bias or inadequate sensitivity could have a significant impact on the study results or their interpretation.
Uninformative	Serious flaw(s) makes study results unusable for hazard identification or dose response.

Figure 5. Overview of IRIS study evaluation process for epidemiology studies: (a) an overview of the evaluation process; (b) the evaluation domains and definitions for ratings (i.e., domain and overall judgments, performed on an outcome-specific basis).

For studies that examine more than one outcome, the evaluation process will be performed separately for each outcome because the utility of a study can vary for different outcomes. If a study examines multiple endpoints for the same outcome,[7] evaluations might be performed at a more granular level if appropriate, but these measures could still be grouped for evidence synthesis.

Authors might be queried to obtain missing critical information, particularly when reporting quality information or data are missing (e.g., content that would be required to conduct a meta-analysis or other quantitative integration) or to provide additional analyses that could address potential limitations. The decision on whether to seek missing information includes considering what additional information would be useful, specifically with respect to any information that could result in a reevaluation of the overall study confidence. Outreach to study authors should be documented and considered unsuccessful if researchers do not respond to an email or phone request within 1 month of the attempt to contact.

For each outcome in a study,[8] reviewers will reach a consensus judgment of *Good*, *Adequate*, *Deficient*, *Not reported*, or *Critically deficient* for each evaluation domain. If a consensus is not reached, a third reviewer will perform conflict resolution. That these evaluations are performed in the context of the study's utility for dose-response analysis is important to stress. These categories are applied to each evaluation domain for each study as follows:

- *Good* represents a judgment that the study was conducted appropriately in relation to the evaluation domain, and any deficiencies, if present, are minor and would not be expected to influence the study results.

- *Adequate* indicates a judgment that there are methodological limitations relating to the evaluation domain, but that those limitations are not likely to be severe or to have a notable impact on the results.

- *Deficient* denotes identified biases or deficiencies that are interpreted as likely to have had a notable impact on the results or that may prevent reliable interpretation of the study findings.

- *Not reported* indicates that the information necessary to evaluate the domain in question was not available in the study. Generally, this term carries the same functional interpretation as *Deficient* for the purposes of the study confidence classification (described below). Depending on the number and severity of other limitations identified in the study, it may or may not be worth reaching out to the study authors to obtain this information (see discussion above).

[7]"Outcome" will be used throughout these methods; the same methods also apply to an endpoint within a larger outcome.

[8]"Study" is used instead of a more accurate term (e.g., "experiment") throughout these sections owing to an established familiarity within the field for discussing a study's risk of bias or sensitivity, etc. However, all evaluations discussed herein are explicitly conducted at the level of an individual outcome within an (un)exposed group of animals or humans, or to a sample of the population within a study.

- *Critically deficient* reflects a judgment that the study conduct introduced a serious flaw that makes the study uninterpretable. Studies with a determination of critically deficient in an evaluation domain will almost always be considered overall "uninformative". For example, in assessing MeHg DNT studies, the studies using DNT tests that are considered deficient or using inappropriate analytical chemistry methods will be considered critically deficient.

Once the evaluation domains have been rated, the identified strengths and limitations will be considered to reach a study confidence classification of *high*, *medium*, or *low* confidence, or *uninformative* for each specific health outcome. This classification is based on the reviewer judgments across the evaluation domains and includes consideration of the likely impact the noted deficiencies in bias and sensitivity or inadequate reporting have on the results. The classifications, which reflect a consensus judgment between reviewers, are defined as follows:

- *High* confidence: A well-conducted study with no notable deficiencies or concerns identified; the potential for bias is unlikely or minimal, and the study used sensitive methodology. *High* confidence studies generally reflect judgments of *good* across all or most evaluation domains.

- *Medium* confidence: A satisfactory (acceptable) study where deficiencies or concerns are noted, but the limitations are unlikely to be of a notable degree. Generally, *medium* confidence studies include *adequate* or *good* judgments across most domains, with the impact of any identified limitation not being judged as severe.

- *Low* confidence: A substandard study where deficiencies or concerns are noted, and the potential for bias or inadequate sensitivity could have a significant impact on the study results or their interpretation. Typically, *low*-confidence studies have a *deficient* evaluation for one or more domains, although some *medium*-confidence studies may have a *deficient* rating in domain(s) considered to have less influence on the magnitude or direction of effect estimates. Generally, *low*-confidence results are given less weight compared to *high*- or *medium*- confidence results during evidence synthesis and integration, and are generally not used as the primary sources of information for derivation of toxicity values unless they are the only studies available. Studies rated as *low* confidence only because of sensitivity concerns about bias towards the null would require additional consideration during evidence synthesis. Observing an effect in these studies may increase confidence, assuming the study is otherwise well conducted (see Section 9).

- *Uninformative*: An unacceptable study where serious flaw(s) make the study results unusable for informing dose response. Studies with *critically deficient* judgments in any evaluation domain are almost always classified as *uninformative* (see explanation above). Studies with multiple *deficient* judgments across domains may also be considered *uninformative*. *Uninformative* studies will not be considered further in the dose-response analysis, but may be used to highlight possible research gaps.

Study evaluation determinations reached by each reviewer and the consensus judgment between reviewers will be documented, and final study evaluations will be made available when the draft is publicly released. The study confidence classifications and their rationales will be

carried forward and considered as part of selecting studies for dose-response, to aid in the interpretation of results across studies.

The principles and framework used for evaluating epidemiology studies are adapted from the principles in the Cochrane Risk of Bias in Nonrandomized Studies of Interventions [ROBINS-I; (Sterne et al., 2016)], modified to address environmental and occupational exposures. The underlying philosophy of ROBINS-I is to describe attributes of an "ideal" study with respect to each of the evaluation domains (e.g., exposure measurement, outcome classification). Core and prompting questions are used to collect information to guide evaluation of each domain.

Core and prompting questions, as well as additional considerations that apply to most outcomes for each domain are presented in Table 4. Core questions represent key concepts, while the prompting questions help the reviewer focus on relevant details under each key domain. Exposure- and outcome-specific criteria to use during evaluation of studies will be developed using the core and prompting questions and refined during a pilot phase with engagement from topic-specific experts. The types of information that might be the focus of those criteria are listed in Table 5.

Table 4. Questions to guide the development of criteria for each domain in epidemiology studies

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
Exposure measurement Does the exposure measure reliably distinguish between levels of exposure in a time window considered most relevant for a causal effect with respect to the development of the outcome?	For all: • Does the exposure measure capture the variability in exposure among the participants, considering intensity, frequency, and duration of exposure? • Does the exposure measure reflect a relevant time window? If not, can the relationship between measures in this time and the relevant time window be estimated reliably? • Was the exposure measurement likely to be affected by a knowledge of the outcome? • Was the exposure measurement likely to be affected by the presence of the outcome (i.e., reverse causality)? For case-control studies of occupational exposures: • Is exposure based on a comprehensive job history describing tasks, setting, time period, and use of specific materials? For biomarkers of exposure, general population: • Is a standard assay used? What are the intra- and inter-assay coefficients of variation? Is the assay likely to be affected by contamination? Are values less than the limit of detection dealt with adequately? • What exposure time period is reflected by the biomarker? If the half-life is short, what is the correlation between serial measurements of exposure?	Is the degree of exposure misclassification likely to vary by exposure level? If the correlation between exposure measurements is moderate, is there an adequate statistical approach to ameliorate variability in measurements? If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations require customization to the exposure and outcome (relevant timing of exposure) **Good** • Valid exposure assessment methods used, which represent the etiologically relevant time period of interest. • Exposure misclassification is expected to be minimal. **Adequate** • Valid exposure assessment methods used, which represent the etiologically relevant time period of interest. • Exposure misclassification may exist but is not expected to greatly change the effect estimate. **Deficient** • Valid exposure assessment methods used, which represent the etiologically relevant time period of interest. Specific knowledge about the exposure and outcome raise concerns about reverse causality, but there is uncertainty whether it is influencing the effect estimate. • Exposed groups are expected to contain a notable proportion of unexposed or minimally exposed individuals, the method did not capture important temporal or spatial variation, or there is other evidence of exposure misclassification that would be expected to notably change the effect estimate. **Critically deficient** • Exposure measurement does not characterize the etiologically relevant time period of exposure or is not valid. • There is evidence that reverse causality is very likely to account for the observed association. • Exposure measurement was not independent of outcome status.

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
Outcome ascertainment Does the outcome measure reliably distinguish the presence or absence (or degree of severity) of the outcome?	For all: • Is outcome ascertainment likely to be affected by knowledge of, or presence of, exposure (e.g., consider access to health care, if based on self-reported history of diagnosis)? For case-control studies: • Is the comparison group without the outcome (e.g., controls in a case-control study) based on objective criteria with little or no likelihood of inclusion of people with the disease? For mortality measures: • How well does cause of death data reflect occurrence of the disease in an individual? How well do mortality data reflect incidence of the disease? For diagnosis of disease measures: • Is the diagnosis based on standard clinical criteria? If it is based on self-report of the diagnosis, what is the validity of this measure? For laboratory-based measures (e.g., hormone levels): • Is a standard assay used? Does the assay have an acceptable level of inter-assay variability? Is the sensitivity of the assay appropriate for the outcome measure in this study population?	Is there a concern that any outcome misclassification is nondifferential, differential, or both? What is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations require customization to the outcome **Good** • High certainty in the outcome definition (i.e., specificity and sensitivity), minimal concerns with respect to misclassification. • Assessment instrument was validated in a population comparable to the one from which the study group was selected. **Adequate** • Moderate confidence that outcome definition was specific and sensitive, some uncertainty with respect to misclassification but not expected to greatly change the effect estimate. • Assessment instrument was validated but not necessarily in a population comparable to the study group. **Deficient** • Outcome definition was not specific or sensitive. • Uncertainty regarding validity of assessment instrument. **Critically deficient** • Invalid/insensitive marker of outcome. • Outcome ascertainment is very likely to be affected by knowledge of, or presence of, exposure. Note: Lack of blinding should not be automatically construed to be critically deficient.

Participant selection Is there evidence that selection into or out of the study (or analysis sample) was jointly related to exposure and to outcome?	For longitudinal cohort: • Did participants volunteer for the cohort based on knowledge of exposure and/or preclinical disease symptoms? Was entry into the cohort or continuation in the cohort related to exposure and outcome? For occupational cohort: • Did entry into the cohort begin with the start of the exposure? • Was follow-up or outcome assessment incomplete, and if so, was follow-up related to both exposure and outcome status? • Could exposure produce symptoms that would result in a change in work assignment/work status ("healthy worker survivor effect")? For case-control study: • Were controls representative of population and time periods from which cases were drawn? • Are hospital controls selected from a group whose reason for admission is independent of exposure? • Could recruitment strategies, eligibility criteria, or participation rates result in differential participation relating to both disease and exposure? For population based- survey: • Was recruitment based on advertisement to people with knowledge of exposure, outcome, and hypothesis?	Were differences in participant enrollment and follow-up evaluated to assess bias? If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)? Were appropriate analyses performed to address changing exposures over time in relation to symptoms? Is there a comparison of participants and nonparticipants to address whether differential selection is likely?	These considerations may require customization to the outcome. This could include determining what study designs effectively allow analyses of associations appropriate to the outcome measures (e.g., design to capture incident vs. prevalent cases, design to capture early pregnancy loss). **Good** • Minimal concern for selection bias based on description of recruitment process (e.g., selection of comparison population, population-based random sample selection, recruitment from sampling frame including current and previous employees). • Exclusion and inclusion criteria specified and would not induce bias. • Participation rate is reported at all steps of study (e.g., initial enrollment, follow-up, selection into analysis sample). If rate is not high, there is appropriate rationale for why it is unlikely to be related to exposure (e.g., comparison between participants and nonparticipants or other available information indicates differential selection is not likely). **Adequate** • Enough of a description of the recruitment process to be comfortable that there is no serious risk of bias. • Inclusion and exclusion criteria specified and would not induce bias. • Participation rate is incompletely reported but available information indicates participation is unlikely to be related to exposure. **Deficient** • Little information on recruitment process, selection strategy, sampling framework and/or participation OR aspects of these processes raises the potential for bias (e.g., healthy worker effect, survivor bias). **Critically deficient** • Aspects of the processes for recruitment, selection strategy, sampling framework, or participation result in concern that selection bias resulted in a large impact on effect estimates (e.g., convenience sample with no information about recruitment and selection, cases and controls are recruited from different sources with different likelihood of exposure, recruitment materials stated outcome of interest and potential participants are aware of or are concerned about specific exposures).

Confounding
Is confounding of the effect of the exposure likely?
	Is confounding adequately addressed by considerations in:
	• Participant selection (matching or restriction)?
	• Accurate information on potential confounders and statistical adjustment procedures?
	• Lack of association between confounder and outcome, or confounder and exposure in the study?
	• Information from other sources?
	Is the assessment of confounders based on a thoughtful review of published literature, potential relationships (e.g., as can be gained through directed acyclic graphing), and minimizing potential overcontrol (e.g., inclusion of a variable on the pathway between exposure and outcome)?
		If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations require customization to the exposure and outcome, but this may be limited to identifying key covariates.
			Good
			• Conveys strategy for identifying key confounders. This may include: a priori biological considerations, published literature, causal diagrams, or statistical analyses; with recognition that not all "risk factors" are confounders.
			• Inclusion of potential confounders in statistical models not based solely on statistical significance criteria (e.g., $p < 0.05$ from stepwise regression).
			• Does not include variables in the models that are likely to be influential colliders or intermediates on the causal pathway.
			• Key confounders are evaluated appropriately and considered to be unlikely sources of substantial confounding. This often will include:
			  ○ Presenting the distribution of potential confounders by levels of the exposure of interest and/or the outcomes of interest (with amount of missing data noted);
			  ○ Consideration that potential confounders were rare among the study population, or were expected to be poorly correlated with exposure of interest;
			  ○ Consideration of the most relevant functional forms of potential confounders;
			  ○ Examination of the potential impact of measurement error or missing data on confounder adjustment.
			Adequate
			Similar to Good but may not have included all key confounders, or less detail may be available on the evaluation of confounders (e.g., sub-bullets in Good). It is possible that residual confounding could explain part of the observed effect, but concern is minimal.
			Deficient
			• Does not include variables in the models that are likely to be influential colliders or intermediates on the causal pathway.
			And any of the following:
			• The potential for bias to explain some of the results is high based on an inability to rule out residual confounding, such as a lack of demonstration

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
			that key confounders of the exposure-outcome relationships were considered; • Descriptive information on key confounders (e.g., their relationship relative to the outcomes and exposure levels) are not presented; or • Strategy of evaluating confounding is unclear or is not recommended (e.g., only based on statistical significance criteria or stepwise regression [forward or backward elimination]). **Critically deficient** • Includes variables in the models that are colliders and/or intermediates in the causal pathway, indicating that substantial bias is likely from this adjustment; or • Confounding is likely present and not accounted for, indicating that all of the results were most likely due to bias. o Presenting a progression of model results with adjustments for different potential confounders, if warranted.

Analysis Does the analysis strategy and presentation convey the necessary familiarity with the data and assumptions?	• Are missing outcome, exposure, and covariate data recognized, and if necessary, accounted for in the analysis? • Does the analysis appropriately consider variable distributions and modeling assumptions? • Does the analysis appropriately consider subgroups of interest (e.g., based on variability in exposure level or duration or susceptibility)? • Is an appropriate analysis used for the study design? • Is effect modification considered, based on considerations developed a priori? • Does the study include additional analyses addressing potential biases or limitations (i.e., sensitivity analyses)?	If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations may require customization to the outcome. This could include the optimal characterization of the outcome variable and ideal statistical test (e.g., Cox regression). **Good** • Use of an optimal characterization of the outcome variable. • Quantitative results presented (effect estimates and confidence limits or variability in estimates) (i.e., not presented only as a *p*-value or "significant"/"not significant"). • Descriptive information about outcome and exposure provided (where applicable). • Amount of missing data noted and addressed appropriately (discussion of selection issues—missing at random vs. differential). • Where applicable, for exposure, includes LOD (and percentage below the LOD), and decision to use log transformation. • Includes analyses that address robustness of findings, e.g., examination of exposure-response (explicit consideration of nonlinear possibilities, quadratic, spline, or threshold/ceiling effects included, when feasible); relevant sensitivity analyses; effect modification examined based only on a priori rationale with sufficient numbers. • No deficiencies in analysis evident. Discussion of some details may be absent (e.g., examination of outliers). **Adequate** Same as Good, except: • Descriptive information about exposure provided (where applicable), but may be incomplete; might not have discussed missing data, cutpoints, or shape of distribution. • Includes analyses that address robustness of findings (examples in Good), but some important analyses are not performed. **Deficient** • Does not conduct analysis using optimal characterization of the outcome variable.

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
			• Descriptive information about exposure levels not provided (where applicable). • Effect estimate and *p*-value presented, without standard error or confidence interval. • Results presented as statistically "significant"/"not significant." **Critically deficient** • Results of analyses of effect modification examined without clear a priori rationale and without providing main/principal effects (e.g., presentation only of statistically significant interactions that were not hypothesis driven). • Analysis methods are not appropriate for design or data of the study.
Selective reporting Is there reason to be concerned about selective reporting?	• Were results provided for all the primary analyses described in the methods section? • Is there appropriate justification for restricting the amount and type of results that are shown? • Are only statistically significant results presented?	If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations generally do not require customization and may have fewer than four levels. **Good** • The results reported by study authors are consistent with the primary and secondary analyses described in a registered protocol or methods paper. **Adequate** • The authors described their primary (and secondary) analyses in the methods section and results were reported for all primary analyses. **Deficient** • Concerns were raised based on previous publications, a methods paper, or a registered protocol indicating that analyses were planned or conducted that were not reported, or that hypotheses originally considered to be secondary were represented as primary in the reviewed paper. • Only subgroup analyses were reported suggesting that results for the entire group were omitted. • Only statistically significant results were reported.

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
Sensitivity Is there a concern that sensitivity of the study is not adequate to detect an effect?	• Is the exposure range adequate to detect associations and exposure-response relationships? • Was the appropriate population included? • Was the length of follow-up adequate? Is the time/age of outcome ascertainment optimal given the interval of exposure and the health outcome? • Are there other aspects related to risk of bias or otherwise that raise concerns about sensitivity?		These considerations may require customization to the exposure and outcome, and may have fewer than four levels. Some study features that affect study sensitivity may have already been included in the other evaluation domains. Other features that have not been addressed should be included here. Some examples include: **Adequate** • The range of exposure levels provides adequate variability to evaluate the relevant associations. • The population was exposed to levels expected to have an impact on response. • The study population was sensitive to the development of the outcomes of interest (e.g., ages, life stage, sex). • The timing of outcome ascertainment was appropriate given expected latency for outcome development (i.e., adequate follow-up interval). • The study was adequately powered to observe an association based on underlying population sensitivity and exposure contrasts. • No other concerns raised regarding study sensitivity. **Deficient** • Concerns were raised about the issues described for *adequate* that are expected to notably decrease the sensitivity of the study to detect associations for the outcome.

Table 5. Information relevant to evaluation domains for epidemiology studies

Domain	Types of information that might need to be collected or are important for evaluating the domain
Exposure measurement	Source(s) of exposure (e.g., consumer products, occupational, an industrial accident) and source(s) of exposure data, blinding to outcome, level of detail for job history data, when measurements were taken, type of biomarker(s), assay information, reliability data from repeat measures studies, validation studies.
Outcome ascertainment	Source of outcome (effect) measure, blinding to exposure status or level, how measured/classified, incident vs. prevalent disease, evidence from validation studies, prevalence (or distribution summary statistics for continuous measures).
Participant selection	Study design, where and when was the study conducted, and who was included? Recruitment process, exclusion and inclusion criteria, type of controls, total eligible, comparison between participants and nonparticipants (or followed and not followed), and final analysis group. Does the study include potential susceptible populations or life stages (see discussion in Section 9)?
Confounding	Background research on key confounders for specific populations or settings; participant characteristic data, by group; strategy/approach for consideration of potential confounding; strength of associations between exposure and potential confounders and between potential confounders and outcome; and degree of exposure to the confounder in the population.
Analysis	Extent (and if applicable, treatment) of missing data for exposure, outcome, and confounders; approach to modeling; classification of exposure and outcome variables (continuous vs. categorical); testing of assumptions; sample size for specific analyses; and relevant sensitivity analyses.
Sensitivity	What are the ages of participants (e.g., not too young in studies of pubertal development)? What is the length of follow-up (for outcomes with long latency periods)? Choice of referent group, the exposure range, and the level of exposure contrast between groups (i.e., the extent to which the "unexposed group" is truly unexposed, and the prevalence of exposure in the group designated as "exposed").
Selective reporting	Are results presented with adequate detail for all endpoints and exposure measures reported in the methods section, and are they relevant to the PECO? Are results presented for the full sample and for specified subgroups? Were stratified analyses (effect modification) motivated by a specific hypothesis?

Evaluation of MeHg epidemiology studies includes evaluation of the analytical chemistry methods and the associated QA/QC procedures that were employed. For this purpose, criteria were developed for assessing the analytical chemistry methods used for analysis of mercury/MeHg in blood and in hair. These criteria are presented in Appendix C.

6.2. PHYSIOLOGICALLY BASED PHARMACOKINETIC (PBPK) MODEL DESCRIPTIVE SUMMARY AND EVALUATION

PBPK (and/or classic PK) models should be used in an assessment when an applicable one exists and no equal or better alternative for dosimetric extrapolation is available. Any models used should represent current scientific knowledge and accurately translate the science into computational code in a reproducible, transparent manner. For a specific target organ/tissue, using

or adapting an existing PK/PBPK model or developing a new model or an alternative quantitative approach might be possible following the EPA Quality Assurance Plan for PBPK models (U.S. EPA, 2018). Data for PK/PBPK models might come from studies across various species and could be from in vitro or in vivo model systems.

Because the aim of this assessment is to update the existing RfD by reevaluating the DNT effects associated with MeHg exposure as assessed by MeHg biomarkers (e.g., cord blood), applying pharmacokinetic models in conjunction with biomarker data to estimate MeHg intake doses is essential. In general, two major types of pharmacokinetic models are available for studying MeHg toxicokinetics: a one-compartment (classic PK) model and a multicompartment PBPK model, hereinafter referred to as the PBPK model (U.S. EPA, 2001a). The one-compartment model is the simplest form of a pharmacokinetic model in that the entire body is assumed to act like a single, uniform entity that uses only one volume term, the apparent volume of distribution. As shown in controlled human studies, the absorption rate of MeHg generally is much faster than its elimination rate. In general, the one-compartment model describes existing MeHg data reasonably well and therefore has been the model most often used for estimating MeHg intake doses since 2001.

In 2001, EPA employed a one-compartment (classic PK) model, adapted from the NRC (2000) model, to derive the existing RfD for MeHg by estimating ingestion doses (mg/kg-day) with the use of measured hair mercury from a female population with a consistently high consumption of fish and whale meat in the Faroe Islands (Grandjean et al., 1997). For better characterization of uncertainty and variability in estimating MeHg intake, the 2001 EPA assessment adopted a range of 46–79 ppb of total Hg in maternal blood instead of using a fixed cord blood level of 58 µg/L used by the NRC (2000) model. The Agency for Toxic Substances and Disease Registry (ATSDR, 1999), on the basis of a study of the population with constant fish consumption in the Seychelles Islands, likewise adopted a one-compartment model with an overall UF of 3 to estimate a chronic oral Minimal Risk Level for MeHg. That MeHg levels could reach a steady-state after approximately five elimination half-lives for those who have constant exposure to MeHg has been estimated, as is the case for the populations in the Seychelles and Faroes Islands (NRC, 2000).

In comparison, the PBPK model typically includes several pharmacokinetic parameters (e.g., tissue volumes, partition coefficients, rate constants for metabolism and elimination) that can vary from one individual to another within the subpopulation of interest. These compartments represent organs and tissues that are interconnected in the body via blood flow. PBPK models are conceptually more accurate in predicting body burden of MeHg (e.g., internal dose) as compared to one-compartment models. PBPK modeling is typically considered to be more complex and data intensive than a PK model as it requires more comprehensive ADME data for model development and validation (e.g., for characterization of uncertainty and variability associated with the model parameters and model outputs).

As the pharmacokinetic modeling needs to reflect a balance between the principles of model parsimony and plausibility (i.e., reflective of physiological reality), both the one-compartment

(classic PK) model and the PBPK model will be considered in this assessment. To ensure consistency, both types of toxicokinetic models will be evaluated against measured biomarker data (e.g., hair and maternal blood) following the EPA guidelines, as articulated in Section 6.3.2.

6.2.1. Pharmacokinetic/Physiologically Based Pharmacokinetic (PK/PBPK) Model Descriptive Summary

Classic one-compartment PK and PBPK models for analyzing MeHg in humans that have been published since 2001 are summarized in Table 6.

Table 6. Classic one-compartment PK and PBPK models for MeHg in humans since 2001

Reference	Notes
Classic (one-compartment) PK model	
(Stern, 2005; Stern and Smith, 2003; Stern et al., 2002)	This series of papers extends the 2001 EPA MeHg assessment work to improve the uncertainty and variability analysis. The refinements include the adoption of the Bayesian approach and the use of a range of ratios for cord blood to maternal blood for the maternal dose reconstruction.
(Sirot et al., 2008)	This analysis estimates MeHg intake dose using an EPA one-compartment model and compared it with the intake estimated by the food frequency questionnaire in French frequent seafood consumers.
(Albert et al., 2010)	This study describes a modified one-compartment model based on the WHO (1990) model to estimate integrated variability in dietary MeHg intake and MeHg half-life in blood and to predict mercury level in hair among pregnant women who consumed seafood.
(Yaginuma-Sakurai et al., 2012)	These authors used a one-compartment model described previously by NRC (2000)[a] and a two-way analysis of variance approach to estimate time-dependent hair-to-blood ratio and half-life of blood mercury in a controlled human study.
(Jo et al., 2015)	This study estimates the between-person variability of the MeHg half-life in Korean adults using the same model as described by Swartout and Rice (2000).[a]
(Li et al., 2015)	This study uses the same one-compartment model of Stern (2005) to evaluate the relationship between MeHg intake and blood and hair MeHg levels in a rice-consuming population in China.
PBPK model	
(Byczkowski and Lipscomb, 2001)	This study describes an extension and a refinement of Clewell et al. (1999) model for gestational transfer, along with lactation transfer of MeHg from the exposed mother to the fetus. The results from model simulation were compared with experimental data obtained from rodents for which the model parameters were scaled to humans using allometric procedures.

Reference	Notes
(Carrier et al., 2001b; Carrier et al., 2001a)	These authors first developed a biologically based dynamic multicompartment model for predicting MeHg using animal data Carrier et al. (2001b) and then reparametrized it for humans. Although the model prediction is generally comparable to that of Clewell et al. (1999), this model is considered less informative due to lack of sensitivity and uncertainty analysis.
(Young et al., 2001)	This analysis describes the deposition of mercury (both MeHg and inorganic Hg) in humans using empirical animal data (hamster, rat, guinea pig, cat, rabbit, monkey, sheep, pig, goat, cow). An allometric approach was used to estimate key kinetic parameters (e.g., metabolism rate constants) in the development of the human model, followed by model validation using human autopsy data.
(Leggett et al., 2001)	This study is a literature review of biokinetic models for the deposition of inhaled mercury vapor in the respiratory tract and different patterns of absorption in blood in animals and humans. As the focus of the current assessment is focused on oral exposure, this study is considered irrelevant.
(Mcnally and Loizou, 2015; Lu et al., 2012; Ruiz et al., 2010; Kirman et al., 2003; Pierrehumbert et al., 2002)	These studies propose a generic PBPK model platform with no specific focus on MeHg. They do not add value to the MeHg PBPK database.
(Noisel et al., 2011; Gosselin et al., 2006)	These authors used the model of Carrier et al. (2001b) and data sets on measured total Hg in hair and blood for reconstruction of the likely monthly MeHg intakes among fish consumption populations in Brazil and Canada. No pregnancy and lactation compartments are included in the analysis.
(Allen et al., 2007)	This study uses the same model structure as Clewell et al. (1999) but incorporates a Bayesian approach, implemented using Markov Chain Monte Carlo analysis to characterize interindividual variation in exposure to MeHg and in the pharmacokinetics (distribution, clearance) of MeHg.
(Berthet et al., 2010)	This study presents a generic two-compartment model for 14 chemicals (e.g., mercury, arsenic, cadmium) to characterize biological monitoring variability.
(Lee et al., 2017)	This study uses the Clewell et al. (1999) model as the template but removes the gestation-related compartments (e.g., uterus, fetus) from the model.
(Abass et al., 2018)	This work consists of several linear toxicokinetic equations for depicting toxicokinetics for MeHg, inorganic Hg, and metallic Hg. Compared to the estimated intake of Hg using a food frequency questionnaire, the predicted intake using toxicokinetic modeling based on total mercury levels in the blood tended to be higher.
(Ou et al., 2018)	This study examines a model derived on the basis of a reparameterization of the Clewell model (1999) and organized into three sub-models: (1) the pregnancy model, (2) the lactation model (for lactating mothers), and (3) the infant model (for suckling infants) with the data on repeated measurements of MeHg in children's hair up to 1 year of age.

[a]Both NRC (2000) and Swartout and Rice (2000) models were derived from the WHO (1990) model.

6.2.2. Pharmacokinetic/Physiologically Based Pharmacokinetic (PK/PBPK) Model Evaluation

Once available PBPK models are summarized, the assessment team will evaluate the models in accordance with criteria outlined in U.S. EPA (2018). Judgments on the suitability of a model are separated into two categories: scientific and technical (Table 7). The scientific criteria focus on whether the biology, chemistry, and other information available for chemical mode(s) of action are justified (i.e., preferably with citations to support use) and represented by the model structure and equations. The scientific criteria are judged on the basis of information presented in the publication or report that describes the model and do not require evaluation of the computer code. Initial technical criteria include availability of the computer code and completeness of parameter listing and documentation. Studies that meet the preliminary scientific and technical criteria are then subjected to an in-depth technical evaluation, which includes a thorough review and testing of the computational code. The in-depth technical and scientific analyses focus on the accurate implementation of the conceptual model in the computational code, use of scientifically supported and biologically consistent parameters in the model, and reproducibility of model results reported in journal publications and other documents. This approach stresses (1) clarity in the documentation of model purpose, structure, and biological characterization; (2) validation of mathematical descriptions, parameter values, and computer implementation; and (3) evaluation of each plausible dose metric. The in-depth analysis is used to evaluate the potential value and cost of developing a new model or substantially revising an existing one. PBPK models EPA develops during the course of the assessment will be peer reviewed, either as a component of the draft assessment or by publication in a journal article.

Table 7. Criteria for evaluating PBPK models

Category	Specific criteria
Scientific	Biological basis for the model is accurate. • Consistent with mechanisms that significantly impact dosimetry. • Predicts dose metric(s) expected to be relevant. • Applicable for relevant route(s) of exposure.
	Consideration of model fidelity to the biological system strengthens the scientific basis of the assessment relative to standard exposure-based extrapolation (default) approaches. • Ability of model to describe critical behavior, such as nonlinear kinetics in a relevant dose range, better than the default (i.e., $BW^{3/4}$ scaling). • Model parameterization for critical life stages or windows of susceptibility. Evaluation of these criteria should also consider the model's fidelity vs. default approaches and possible use of an intraspecies uncertainty factor in conjunction with the model to account for variations in sensitivity between life stages.

Category	Specific criteria

	• Predictive power of model-based dose metric vs. default approach, based on exposure
	  ○ Specifically, model-based metrics may correlate better than the applied doses with animal/human dose-response data.
	  ○ The degree of certainty in model predictions vs. default is also a factor. For example, while target tissue metrics are generally considered better than blood concentration metrics, lack of data to validate tissue predictions when blood data are available may lead to choosing the latter.

	Principle of Parsimony • Model complexity or biological scale, including number and parameterization of (sub)compartments (e.g., tissue or subcellular levels) should be commensurate with data available to identify parameters.
	Model describes existing PK data reasonably well, both in "shape" (matches curvature, inflection points, peak concentration time, etc.) and quantitatively (e.g., within factor of 2−3).
	Model equations are consistent with biochemical understanding and biological plausibility.
Initial technical	Well-documented model code is readily available to EPA and public.
	Set of published parameters is clearly identified, including origin/derivation.
	Parameters do not vary unpredictably with dose (e.g., any dose dependence in absorption constants is predictable across the dose ranges relevant for animal and human modeling).
	Sensitivity and uncertainty analysis have been conducted for relevant exposure levels (local sensitivity analysis is sufficient, but global analysis provides more information). • If a sensitivity analysis was not conducted, EPA may decide to independently conduct this additional work before using the model in the assessment. • A sound explanation should be provided when sensitivity of the dose metric to model parameters differs from what is reasonably expected based on experience.

7. DATA EXTRACTION OF STUDY METHODS AND RESULTS

Data extraction and content management will be carried out using HAWC (Health Assessment Workspace Collaborative). Data extraction elements that may be collected from epidemiology studies are listed in Appendix B. The content of the data extraction may be revised following the identification of the studies included in the review as part of a pilot phase to assess the data extraction workflow. Not all studies that meet the PECO criteria go through data extraction. Studies evaluated as being *uninformative* are not considered further and would, therefore, not undergo data extraction. The same may be true for *low*-confidence studies if sufficient *medium*- and *high*-confidence studies are available. All findings are considered for extraction, regardless of statistical significance, although the level of extraction for specific outcomes within a study may differ (i.e., ranging from a narrative to full extraction of dose-response effect size information). Similarly, decisions about data extraction for low-confidence studies are typically made during implementation of the protocol based on consideration of the quality and extent of the available evidence. The version of the protocol released with the draft assessment will outline how low-confidence studies were treated for extraction and evidence synthesis.

The data extraction results for included studies will be presented in the assessment and made available for download from EPA HAWC in Excel format when the draft is publicly released. Data extraction will be performed by one member of the evaluation team and checked by one or two other members. Discrepancies in data extraction will be resolved by discussion or consultation with a third member of the evaluation team if needed. Once the data have been verified, they will be "locked" to prevent accidental changes. Digital rulers, such as WebPlotDigitizer (https://automeris.io/WebPlotDigitizer), are used to extract numerical information from figures. Use of digital rulers is documented during extraction.

As previously described, routine attempts will be made to obtain information missing from human health effect studies, if it is considered influential during study evaluations (see Section 6) or when it can provide information required to conduct a meta-analysis (e.g., missing group size or variance descriptors such as standard deviation or confidence interval). Missing data from individual mechanistic (e.g., in vitro) studies will generally not be sought. Outreach to study authors should be documented and considered unsuccessful if researchers do not respond to an email or phone request within 1 month of the attempt to contact.

8. DOSE-RESPONSE ASSESSMENT: STUDY SELECTION AND QUANTITATIVE ANALYSIS

This section of the protocol provides an overview of considerations for conducting the dose-response assessment, particularly statistical considerations specific to dose-response analysis that support quantitative risk assessment. Importantly, these considerations do not supersede existing EPA guidance.

A MeHg oral RfD will be derived. An RfD is an estimate, with uncertainty spanning perhaps an order of magnitude, of an exposure to the human population (including susceptible subgroups) that is likely to be without an appreciable risk of deleterious health effects over a lifetime (U.S. EPA, 2002, §4.2). Reference values are not predictive risk values; that is, they provide no information about risks at higher or lower exposure levels.

8.1. SELECTING STUDIES FOR DOSE-RESPONSE ASSESSMENT

The dose-response assessment begins with a review of the DNT effects, particularly among the studies of highest quality and that exemplify the study attributes summarized in Table 7. This review also considers whether there are opportunities for quantitative evidence integration. Examples of quantitative integration, from simplest to more complex, include (1) characterizing overall toxicity, as in combining effects that comprise a syndrome, or occur on a continuum (e.g., precursors and eventual overt toxicity) and (2) conducting a meta-analysis or meta-regression of all studies addressing a category of important health effects.

Studies of low sensitivity may be less useful if they fail to detect a true effect or yield points of departure with wide confidence limits, but such studies would be considered for inclusion in a meta-analysis.

Studies most useful for dose-response analysis generally have at least one exposure level in the region of the dose-response curve near the benchmark response (the response level to be used for deriving toxicity values), to minimize low-dose extrapolation, and more exposure levels and larger sample sizes overall (U.S. EPA, 2012). These attributes support a more complete characterization of the shape of the exposure-response curve and decrease the uncertainty in the associated exposure-response metric (RfD) by reducing statistical uncertainty in the point of departure and minimizing the need for low-dose extrapolation. In addition to these more general considerations, specific issues that may impact the feasibility of dose-response modeling for individual data sets are described in more detail in the *Benchmark Dose Technical Guidance* (U.S. EPA, 2012).

Table 8. Attributes used to evaluate studies for derivation of toxicity values

Study attributes		Considerations for human studies
Study confidence		*High*- or *medium*-confidence studies are highly preferred over *low*-confidence studies. The available high and medium confidence studies are further differentiated based on the study attributes below as well as a reconsideration of the specific limitations identified and their potential impact on dose-response analyses.
Rationale for choice of species		Human data are preferred over animal data to eliminate interspecies extrapolation uncertainties (e.g., in toxicodynamics, relevance of specific health outcomes to humans).
Relevance of exposure paradigm	Exposure durations	Studies involving human environmental exposures (oral, inhalation).
	Exposure levels	When developing a chronic toxicity value, chronic or subchronic studies are preferred over studies of acute exposure durations. Exceptions exist, such as when a susceptible population or life stage is more sensitive in a particular time window (e.g., developmental exposure).
	Exposure route	Exposures near the range of typical environmental human exposures are preferred. Studies with a broad exposure range and multiple exposure levels are preferred to the extent that they can provide information about the shape of the exposure-response relationship [see the EPA *Benchmark Dose Technical Guidance* (U.S. EPA, 2012, §2.1.1)] and facilitate extrapolation to more relevant (generally lower) exposures.
Subject selection		Studies that provide risk estimates in the most susceptible groups are preferred.
Controls for possible confounding[a]		Studies with a design or analysis (e.g., covariates or other procedures for statistical adjustment) that adequately address the relevant sources of potential critical confounding for a given outcome are preferred.
Measurement of exposure		Studies that can reliably distinguish between levels of exposure in a time window considered most relevant for development of a causal effect are preferred. Exposure assessment methods that provide measurements at the level of the individual and that reduce measurement error are preferred. Measurements of exposure should not be influenced by knowledge of health outcome status.
Measurement of health outcome(s)		Studies that can reliably distinguish the presence or absence (or degree of severity) of the outcome are preferred. Outcome ascertainment methods using generally accepted or standardized approaches are preferred.
		Studies with individual data are preferred in general. Examples include: to characterize experimental variability more realistically, to characterize overall incidence of individuals affected by related outcomes (e.g., phthalate syndrome).
		Among several relevant health outcomes, preference is generally given to those with greater biological significance.
Study size and design		Preference is given to studies using designs reasonably expected to have power to detect responses of suitable magnitude.[b] This does not mean that studies with substantial responses but low power would be ignored, but that they should be interpreted in light of a confidence interval or variance for the response. Studies that address changes in the number at risk (through decreased survival, loss to follow-up) are preferred.

[a]An exposure or other variable that is associated with both exposure and outcome, but is not an intermediary between the two.

[b]Power is an attribute of the design and population parameters, based on a concept of repeatedly sampling a population; it cannot be inferred post hoc using data from one experiment (Hoenig and Heisey, 2001).

8.2. CONDUCTING DOSE-RESPONSE ASSESSMENTS

EPA uses a two-step approach for dose-response assessment that distinguishes analysis of the dose-response data in the range of observation from any inferences about responses at lower environmentally relevant exposure levels (U.S. EPA, 2012; 2005, §3):

1) Within the observed dose range, the preferred approach is to use dose-response modeling to incorporate as much of the data set as possible into the analysis. This modeling yields a POD, an exposure level ideally near the lower end of the range of observation, without significant extrapolation to lower exposure levels. See Section 8.2.1 for more details.

2) Derivation of reference values nearly always involves extrapolation to exposures lower than the POD and is described in more detail in Section 8.2.3.

For reference values, IRIS assessments typically derive a candidate value from each suitable data set. Evaluating these candidate values grouped within a particular DNT domain yields a single DNT value for each domain under consideration. Next, evaluation of these domain values results in the selection of a single overall reference value to cover all DNT domains. While this overall reference value is the focus of the assessment, the domain values can be useful for subsequent cumulative risk assessments that consider the combined effect of multiple agents acting on a common domain, regardless of the domain selected as the basis for an RfD.

8.2.1. Dose-response Analysis in the Range of Observation

For conducting a dose-response assessment, toxicodynamic ("biologically based") modeling can be used when there are sufficient data to ascertain the mode of action and quantitatively support model parameters that represent rates and other quantities associated with the key precursor events of the mode of action. Toxicodynamic modeling is potentially the most comprehensive way to account for the biological processes involved in a response. Such models seek to reflect the sequence of key precursor events that lead to a response. Toxicodynamic models can contribute to dose-response assessment by revealing and describing nonlinear relationships between internal dose and response. Such models may provide a useful approach for analysis in the range of observation, provided the purpose of the assessment justifies the effort involved.

When a toxicodynamic model is not available for dose-response assessment or when the purpose of the assessment does not warrant developing such a model, empirical modeling should be used to fit the data (on the apical outcome or a key precursor event) in the range of observation. For this purpose, EPA has developed a standard set of models (http://www.epa.gov/bmds) that can be applied to typical data sets, including those that are nonlinear. In situations where there are alternative models with significant biological support, the decision maker can be informed by the presentation of these alternatives along with the models' strengths and uncertainties. The EPA has developed guidance on modeling dose-response data, assessing model fit, selecting suitable models,

and reporting modeling results [see the EPA *Benchmark Dose Technical Guidance* (U.S. EPA, 2012)]. Additional judgment or alternative analyses are used if the procedure fails to yield reliable results, for example, if the fit is poor, modeling may be restricted to the lower doses, especially if there is competing toxicity at higher doses.

For each modeled response, a POD from the observed data should be estimated to mark the beginning of extrapolation to lower doses. The POD is an estimated dose (expressed in human-equivalent terms) near the lower end of the observed range without significant extrapolation to lower doses. The POD is used as the starting point for subsequent extrapolations and analyses.

The response level at which the POD is calculated is guided by the severity of the endpoint. For dichotomous data, a response level of 10% extra risk is generally used for minimally adverse effects, 5% or lower for more severe effects. For continuous data, a response level is ideally based on an established definition of biological significance. In the absence of such definition, one control standard deviation from the control mean is often used for minimally adverse effects, one-half standard deviation for more severe effects. The POD is the 95% lower bound on the dose associated with the selected response level.

8.2.2. Extrapolation: Slope Factors and Unit Risks

A cancer assessment is not included in the scope of this assessment for MeHg. Accordingly, this assessment will not derive an oral slope factor or inhalation unit risk.

8.2.3. Extrapolation: Reference Values

Reference value derivation is EPA's most frequently used type of nonlinear extrapolation method and is most commonly used for noncancer effects.

For each data set selected for reference value derivation, reference values are estimated by applying relevant adjustments to the PODs to account for the conditions of the reference value definition—for human variation, extrapolation from animals to humans (not necessary for this assessment as only human studies are being evaluated), extrapolation to chronic exposure duration, and extrapolation to a minimal level of risk (if not observed in the data set). Increasingly, data-based adjustments (U.S. EPA, 2014) and Bayesian methods for characterizing population variability (NRC, 2014) are feasible and can be distinguished from the UF considerations outlined below. The assessment will discuss the scientific bases for estimating these data-based adjustments and UFs:

- *Human variation*: The assessment accounts for variation in susceptibility across the human population and the possibility that the available data may not represent individuals who are most susceptible to the effect, by using a data-based adjustment or UF or a combination of the two. Where appropriate data or models for the effect or for characterizing the internal dose are available, the potential for data-based adjustments for toxicodynamics or

toxicokinetics is considered (U.S. EPA, 2014, 2002).[9] [10] When sufficient data are available, an intraspecies UF either less than or greater than 10-fold may be justified (U.S. EPA, 2002). This factor may be reduced if the POD is derived from or adjusted specifically for susceptible individuals [not for a general population that includes both susceptible and nonsusceptible individuals; (U.S. EPA, 2002, §4.4.5; 1998, §4.2; 1996, §4; 1994, §4.3.9.1; 1991, §3.4)]. When the use of such data or modeling is not supported, a UF with a default value of 10 is considered.

- *LOAEL to NOAEL*: If a POD is based on a LOAEL (lowest-observed-adverse-effect level), the assessment includes an adjustment to an exposure level where such effects are not expected. This can be a matter of great uncertainty if no evidence is available at lower exposures. A factor of 3 or 10 is generally applied to extrapolate to a lower exposure expected to be without appreciable effects (NOAEL, or no-observed-adverse-effect level). A factor other than 10 may be used, depending on the magnitude and nature of the response and the shape of the dose-response curve (U.S. EPA, 2002, 1998, 1996, 1994, 1991).

- *Subchronic-to-chronic exposure*: When using subchronic studies to make inferences about chronic/lifetime exposure, the assessment considers whether lifetime exposure could have effects at lower levels of exposure. A factor of up to 10 may be applied to the POD, depending on the duration of the studies and the nature of the response (U.S. EPA, 2002, 1998, 1994).

- *Database deficiencies*: In addition to the adjustments above, if database deficiencies raise concern that further studies might identify a more sensitive effect, organ system, or life stage, the assessment may apply a database UF (U.S. EPA, 2002, 1998, 1996, 1994, 1991). The size of the factor depends on the nature of the database deficiency. For example, the EPA typically follows the recommendation that a factor of 10 be applied if both a prenatal toxicity study and a two-generation reproduction study are missing and a factor of $10^{1/2}$ (i.e., 3) if either one or the other is missing (U.S. EPA, 2002, §4.4.5).

The POD that is used for an RfD is divided by the product of these factors. U.S. EPA (2002, §4.4.5) recommends that any composite factor that exceeds 3,000 represents excessive uncertainty, and recommends against relying on the associated RfD.

The derivation of an RfD for DNT health effects for MeHg conducted as part of the current module will be performed consistent with EPA guidance summarized above.

[9]Examples of adjusting the toxicokinetic portion of interhuman variability include the IRIS boron assessment's use of nonchemical-specific kinetic data [e.g., glomerular filtration rate in pregnant humans as a surrogate for boron clearance (U.S. EPA, 2004)] and the IRIS trichloroethylene assessment's use of population variability in trichloroethylene metabolism, via a PBPK model, to estimate the lower 1st percentile of the dose metric distribution for each POD (U.S. EPA, 2011).

[10]Note that when a PBPK model is available for relating human internal dose to environmental exposure, relevant portions of this UF may be more usefully applied prior to animal-to-human extrapolation, depending on the correspondence of any nonlinearities (e.g., saturation levels) between species.

9. PROTOCOL HISTORY

Release date:

Revisions history:

APPENDICES

APPENDIX A. ELECTRONIC DATABASE SEARCH STRATEGIES

Table A-1. Overall database search strategy

Search	Search strategy	Date and results
PubMed		
Chemical terms	((((((("methylmercury"[All Fields] OR "methyl mercury"[All Fields]) OR "methyl-mercury"[All Fields]) OR "MeHg"[All Fields]) OR "monomethylmercury"[All Fields]) OR "22967-92-6"[EC/RN Number]) OR "Methylmercure"[All Fields]) OR " Methylquecksilber"[All Fields]) AND ("1998"[PDAT]: "3000"[PDAT])	1998–March 2017: 5,028 May 2019 update: 818
Web of Science		
Chemical terms	(TS="methylmercury" OR TS="methyl mercury" OR TS="methyl-mercury" OR TS="Methylmercury (MeHg)" OR TS="monomethylmercury" OR TS="22967-92-6" OR TS="Méthylmercure" OR TS="Methylquecksilber" OR TS="methylmercury ii" OR TS="MeHg") AND PY=(1998-2017)	1998–March 2017: 8,962 May 2019 update: 1,284
Toxline		
Chemical terms	@SYN1+@AND+@OR+(methylmercury+"methyl+mercury"+monomethylmercury+Methylmercure+Methylquecksilber+MeHg+@TERM+@rn+22967-92-6)+@RANGE+yr+1998+2017+@NOT+@org+"nih+reporter"	1998–March 2017: 5,714 May 2019 update: 1,087
Science Direct		
Chemical terms	(methylmercury OR "methyl mercury" OR methyl-mercury OR "Methylmercury (MeHg)" OR monomethylmercury OR "22967-92-6" OR Méthylmercure OR Methylquecksilber OR "methylmercury ii" OR MeHg)	1998–March 2017: 5,330 May 2019 update: 0 (HERO could not search Science Direct)
Total	Total unique records from database searches with duplicates removed: 15,277	

Table A-2. SWIFT Review search of titles and abstracts to identify epidemiology dose-response articles

Target	Search string
Epidemiology study[a]	(mesh_mh:(humans OR "human development") OR tiab: (human* OR person* OR people) OR mesh_mh:(age groups) OR tiab: (pediatric* OR paediatric* OR baby OR babies OR toddler* OR child* OR youth* OR youngster* OR tween* OR teen OR teens OR teenager*) OR (tiab:("in utero" OR prenat* OR perinat* OR neonat* OR postnat*) AND NOT tiab: (mice OR mouse OR rat OR rats)) OR tiab:(preschool* OR "pre-school*" OR kindergarten* OR schoolchild* OR student*) OR tiab:("middle age*" OR elder* OR "senior citizen*" OR seniors OR retiree* OR septuagenarian* OR octagenarian* OR sexagenarian* OR nonagenarian* OR centenarian*) OR mesh_mh:("nuclear family") OR tiab:(famil* OR parent* OR father* OR mother* OR sibling* OR brother* OR sister* OR twin OR twins OR "stepfather*" OR "step father*" OR "stepmother*" OR "step mother*" OR "stepdaughter*" OR "step daughter*" OR "stepson*" OR "step son*" OR aunt* OR uncle* OR niece* OR nephew* OR grandparent* OR grandfather* OR "grand father*" OR grandmother* OR "grand mother*" OR grandchild* OR granddaughter* OR grandson* OR spouse* OR partner* OR husband* OR wife OR wives OR guardian* OR caregiver* OR "care giver*") OR mesh_mh:(men OR women) OR tiab:(men OR man OR boy OR boys OR boyhood OR women OR woman OR girl OR girls OR girlhood) OR mesh_mh:("population groups" OR "vulnerable populations") OR tiab:("african american*" OR "asian american*" OR hispanic* OR latina* OR latino* OR "mexican american*" OR underserved OR disadvantaged) OR mesh_mh:("epidemiologic studies" OR "double-blind method" OR "single-blind method") OR mesh_sh:(epidemiology) OR tiab:("case control*" OR cohort OR "cross sectional" OR "follow-up study" OR longitudinal OR prospective OR retrospective) OR mesh_pubtype:("case reports" OR "clinical trial" OR "observational study" OR "randomized control trial" OR "twin study") OR tiab:("clinical trial*" OR observational OR "randomized control trial*") OR mesh_mh:("research subjects" OR "human experimentation" OR patients OR "Patient Participation") OR tiab:("human subject*" OR "research subject*" OR client* OR patient* OR inpatient* OR outpatient* OR participant* OR volunteer*) OR mesh_mh:("occupational groups" OR "occupational exposure") OR tiab:(occupation* OR workplace OR "work place" OR "work-related" OR administrator* OR aides OR assistant* OR crew OR crews OR employee* OR personnel OR professional OR staff OR technician* OR worker* OR educator* OR instructor* OR teacher* OR clinician* OR doctor* OR physician* OR pharmacist* OR nurs* OR residents OR veterinarian*))
Dose-response	("meta-analysis" OR "Systematic review" OR tiab_punct:"P <" OR tiab_punct:"p <" OR tiab_punct:"P <=" OR tiab_punct:"p <=" OR tiab_punct:"P >*" OR tiab_punct:"p >*" OR tiab_punct:"p=*" OR tiab_punct:"P =" OR tiab_punct:"p =" OR tiab_punct:"p>*" OR tiab_punct:"p<*" OR tiab:"significan*" OR tiab:"nonsignificant" OR RR OR RRs OR SMR OR SMRs OR "rate ratio*" OR "prevalence ratio*" OR "hazard ratio*" OR "odds ratio*" OR "risk ratio*" OR "relative risk*" OR "prevalence ratio*" OR tiab:"covariate*" OR tiab:"adjust*" OR tiab:"control* for" OR tiab:"associat*" OR tiab:"confound*" OR CI OR "confidence interval*" OR "credible interval" OR regression* OR "explanatory variable*" OR tiab:"dose-response")

[a]The search strategy from epidemiology studies is adapted from standard SWIFT Review search strategies for humans.

Table A-3. Database search strategies (PBPK studies)

Search	Search string	Date and results
PubMed	(pbpk[tiab] OR "pb-pk"[tiab] OR pk[tiab] OR tk[tiab] OR pbtk[tiab] OR "pb-tk"[tiab] OR httk[tiab] OR pk-model*[tiab] OR tk-model*[tiab] OR (pharmacokinetic*[tiab] OR pharmacokinetics[mh:noexp] OR pharmacokinetics[sh] OR toxicokinetic*[tiab] OR toxicokinetics[mh:noexp] OR "physiologically based"[tiab] OR "biologically based"[tiab])) AND (model*[tiab] OR models[tiab] OR modeling[tiab]) AND (methylmercury[tiab] OR "methyl mercury"[tiab] OR mercury[tiab])	Jan 2001–Nov 2019: 209
Science Direct	(pbpk OR pb-pk OR pk OR tk OR pbtk OR pb-tk OR httk) OR (pharmacokinetic* OR toxicokinetic* OR physiologically OR "biologically)) AND (model* OR models OR modeling) AND (methylmercury OR "methyl mercury" OR mercury)	Jan 2001–Nov 2019: 19
Web of Science	(pbpk OR "pb-pk" OR pk OR tk OR pbtk OR "pb-tk" OR httk OR pk-model* OR tk-model* OR (pharmacokinetic* OR pharmacokinetics[mh:noexp] OR pharmacokinetics[sh] OR toxicokinetic* OR toxicokinetics[mh:noexp] OR "physiologically based" OR "biologically based")) AND (model* OR models OR modeling) AND (methylmercury OR "methyl mercury" OR mercury)	Jan 2001–Nov 2019: 151

APPENDIX B. DATA EXTRACTION FIELDS

Table B-1. Key data extraction elements to summarize study design, experimental model, methodology, and results

Field label	Possible data extraction elements
Epidemiology studies	
Funding	Funding source(s)
	Reporting of conflict of interest by authors
Subjects	Study population name/description
	Dates of study and sampling time frame
	Geography (country, region, state, etc.)
	Demographics (sex, race/ethnicity, age or life stage at exposure, and at outcome assessment)
	Number of subjects (target, enrolled, *n* per group in analysis, and participation/follow-up rates)
	Inclusion/exclusion criteria/recruitment strategy
	Description of reference group
Methods	Study design (e.g., prospective or retrospective cohort, nested case-control study, cross-sectional, population-based case-control study, intervention, case report)
	Length of follow-up
	Health outcome category (e.g., cardiovascular)
	Health outcome (e.g., blood pressure)
	Diagnostic or methods used to measure health outcome
	Confounders or modifying factors and how considered in analysis (e.g., included in final model, considered for inclusion but determined not needed)
	Chemical name
	Exposure assessment (e.g., blood, hair, placenta)
	Methodological details for exposure assessment (e.g., analysis method, limit of detection)
	Statistical methods
Results	Exposure levels (e.g., mean, median, measures of variance as presented in paper, such as standard deviation, standard error of the mean, 75th/90th/95th percentile, minimum/maximum); range of exposure levels, number of exposed cases
	Statistical findings (e.g., adjusted β, standardized mean difference, adjusted odds ratio, standardized mortality ratio, relative risk) or description of qualitative results. When possible, convert measures of effect to a common metric with associated 95% confidence intervals. Most often, measures of effect for continuous data are expressed as mean difference, standardized mean difference, and percentage control response. Categorical data are typically expressed as odds ratio, relative risk (also called risk ratio), or β values, depending on what metric is most commonly reported in the included studies and ability to obtain information for effect conversions from the study or through author query.
	Observations on dose-response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, nonmonotonic)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

APPENDIX C. CRITERIA FOR EVALUATION OF ANALYTICAL CHEMISTRY METHODS USED FOR ANALYSIS OF MERCURY/METHYLMERCURY IN BLOOD AND HAIR

Table C-1. Evaluation of analytical methods for blood total and methylmercury in epidemiology studies

Level		Criteria
Good		**Acceptable proportion of samples (>50%) above the limit of detection (LOD) OR LOD less than median blood mercury concentration in the study population** **Method** All papers must include a description of methodological factors: • All papers must indicate the type of sample analyzed; for Good quality, should be liquid whole blood (collected in tubes with ethylenediaminetetraacetic acid [EDTA] anticoagulant, stored in the refrigerator within 1 year) or a freeze-dried (within 2 years) blood sample. NOTE: If samples have been stored for longer than noted above, recovery of storage stability spiked samples must be reported but results will not be evaluated. Additional information regarding storage stability samples is provided in "Useful Terms Defined." • For laboratories using a standard method: The paper will include the method number and a citation for a standard method for measuring mercury levels in blood (e.g., Centers for Disease Control and Prevention method ITB003A or National Health and Nutrition Examination Survey methods 3001.1 or 3016.8 for total mercury, method DLS-3020.5 for methylmercury) OR a description of the method discussed as follows. • For laboratories using a nonstandard, validated method: The paper will provide a citation for a peer-reviewed report of the validation in the main body of the paper or in supplementary information (refer to "Useful Terms Defined" for more information on validation) OR a description of the method discussed, as follows. • For laboratories using nonstandard, nonvalidated methods: Nonstandard methods may perform acceptably if sufficient evidence of data quality (quality control [QC] results) is provided (as follows). For nonstandard methods, the paper must provide a description of the method, including the following: o Analytical limits including LOD and/or the limit of quantitation (LOQ) must be reported for nonstandard, nonvalidated methods to obtain a Good rating. o Combustion Atomic Absorption Spectroscopy (CAAS) refers to direct mercury analysis and may use several instruments, such as the Milestone Direct Mercury Analyzer (DMA) or the Nippon Mercury Analyzer (MA). ▪ For studies using CAAS for total mercury analysis, critical method variables that should be noted are time and

Level		Criteria
		temperature of sample drying step, charring/decomposition step of analysis, and amalgamator flash time and temperature. ○ Sample stabilization measures (e.g., addition of gold or thiols; refer to "Useful Terms Defined" for additional details). ○ Sample preparation method including, but not limited to, the following: ▪ Derivatization steps for gas chromatography-inductively coupled plasma-mass spectrometry (GC-ICP-MS) analysis or another speciation technique for methylmercury. ▪ Oxidizing agents for cold vapor atomic fluorescence spectroscopy (CVAFS) and cold vapor atomic absorption spectroscopy (CVAAS) (e.g., potassium permanganate [$KMnO_4$], nitric acid, or sulfuric acid for total mercury analysis). ▪ Reducing agent for cold vapor atomic fluorescence spectroscopy (CVAFS) and cold vapor atomic absorption spectroscopy (CVAAS) (e.g., tin chloride [$SnCl_2$], sodium borohydride [$NaBH_4$], sodium hypochlorite [NaClO]). ▪ Digestion reagents, temperatures (≤65°C), and times. ▪ Extraction solvent composition. • <u>For speciation analysis (i.e., methylmercury):</u> Separation steps before detection. Examples include the following: ○ <u>For liquid chromatography (LC; e.g., high-performance liquid chromatography [HPLC] or ultra-high-performance liquid chromatography [UHPLC]):</u> Column type, eluent composition (e.g., mercaptoethanol, EDTA, cysteine), eluent gradient (or indicate whether an isocratic program [where the same eluent content was used through the run] was used), and injection volume. ○ <u>For gas chromatography:</u> Transfer line temperature, column temperature program, and column identity. ○ <u>For distillation:</u> Solvent and temperature. • Use of an appropriate internal standard (e.g., bismuth [Bi], praseodymium [Pr], holmium [Ho]; ICP-MS only; other organomercury compound [e.g., propylmercury, butylmercury; speciation by LC- or GC-ICP-MS only]): ○ NOTE: Internal standard is not generally included in CVAAS/CVAFS/CAAS methods. *NOTE: The items listed earlier should be included, but because of the great variability in nonstandard methods, it is not necessary to evaluate the quality of all methodological factors, only to ensure that they are included in the method description.* **<u>Quality Control</u>** All papers (standard methods, validated methods, and nonstandard methods) must include a description of QC procedures and results performed to verify method performance and data quality for the study:

Level		Criteria
		• At least three laboratory QC procedures and results in the methods section, supplementary materials, cited papers, or standardized laboratory protocol, including, but not limited to, the following: o Blood-based standard reference materials (SRMs) (955C, levels 1 and 3 only for total mercury, level 3 for methylmercury only; 966 if the study was done before December 2015, level 2 for methylmercury only): % recovery information, 90%–110%. If the study was done before 2000 (the year SRM 966 was issued), acceptable reference materials include "Control Blood for Metals" by the Behring Institute, OSSD 20/21. o Duplicate sample preparation and relative percent difference (RPD) results, <15% RPD. o Representative blank sample analyses or chromatograms to demonstrate the absence of interferences. o Recovery of spiked study samples, 90%–110%. o Post-extraction spiked sample recovery or method blank spiked (MBS) samples, 90%–110%. o Replicate QC precision (also called uncertainty, repeatability, or reproducibility; or percent relative standard deviation [%RSD] or coefficient of variation [%CV], <15%, or correlation coefficient, >0.90). o Participation in an interlaboratory testing program with documented results for mercury in blood samples and "satisfactory" results. NOTE: A wide range of interlaboratory testing programs is available for trace metals, but only ones that monitor mercury in whole blood, serum, or plasma, or that use standard methods, are relevant for assessing data quality. o Incurred sample reanalysis (ISR) to demonstrate reproducibility on different days, and RPD results, <15% RPD; refer to "Useful Terms Defined" for more information. o Control charts (e.g., Bland-Altman, Levey-Jennings, Harrell-Davis, Shewhart) for QC samples of the same sample type, prepared by the cited method, that show method performance over time (large sample populations only). NOTE: If any of the above method parameters or quality control measures exhibit values that fall outside the range of value noted above, then the entire study should be categorized at the lowest quality level of the individual method parameter(s) or QC measure(s).

Level		Criteria
Adequate		**Method** Same as Good, with the following exceptions: • All papers must indicate the type of sample analyzed; for Adequate quality, may be liquid whole blood (collected in tubes with ethylenediaminetetraacetic acid [EDTA] anticoagulant, stored in the refrigerator within 1 year) or a freeze-dried (within 2 years) blood sample. NOTE: If samples have been stored for longer than noted above, recovery of storage stability spiked samples must be reported but results will not be evaluated. • For laboratories using standard methods: Samples analyzed by Environmental Protection Agency (EPA) method 7473 (total mercury only). • For nonstandard, nonvalidated methods: The paper will comment on two or three methodological details listed earlier such as extraction times, cleaning times between sample analyses, instrumental technique, and so on. • Analytical limits including LOD and/or the limit of quantitation (LOQ) must be reported for nonstandard, nonvalidated methods to obtain an Acceptable rating. • For measurements of methylmercury: Determined methylmercury by subtracting inorganic mercury from total mercury. • For ICP-MS only: Use of less common or less appropriate internal standards such as terbium (Tb), rhodium (Rh), gallium (Ga), thallium (Tl), indium (In), yttrium (Y), or scandium (Sc). **Quality Control** Any Adequate analysis must include the following: • For standard methods and validated methods (with literature citation): Discussion of one or two laboratory QC procedures and results in the methods section, supplementary materials, cited papers, or standardized laboratory protocol, including either of the following: o Reference material percentage recovery information (may be the National Institute of Standards and Technology [NIST] SRM mentioned earlier or another such as NIST 3133 or 3177 for total mercury only, or commercial blood SRM [e.g., ClinCheck or Seronorm for total mercury]); recovery will fall in the range 85%–90% or 110%–115%. o Replicate sample preparation information (e.g., %CV or RPD, for duplicate preparation, of 10%–15%). • For nonstandard, nonvalidated studies: Discussion of at least three laboratory QC procedures and results in the methods section, supplementary materials, cited papers, or standardized laboratory protocol, including a combination of the procedures mentioned under "Good" and "Adequate" levels.

Level		Criteria
Deficient		**Method** Same as Adequate, with the following exceptions: • All papers must indicate the type of sample analyzed; for Deficient quality, may be liquid whole blood (collected in vacutainer tubes with ethylenediaminetetraacetic acid [EDTA] anticoagulant, 1 year storage in the refrigerator, or lithium heparin anticoagulant due to the lower stability of heparin) or a freeze-dried (2 years) blood sample. NOTE: If samples have been stored for longer than noted above, recovery of storage stability spiked samples must be reported but results will not be evaluated. Additional information regarding storage stability samples is provided in "Useful Terms Defined". • Minimal methodological details provided in the paper, supplemental information, or cited papers; only one or two of the items from the aforementioned method detail list. o For example, mentioning only the instrumental technique used for analysis, omitting all collection and storage procedures, analytical limits, or references for the sample preparation method. NOTE: The analytical limits for nonstandard, nonvalidated methods must be provided for papers to be considered Good or Adequate, as it is necessary to determine whether a method was used with sensitivity levels appropriate to the matrix and that the method has been appropriately optimized by the analytical laboratory. If analytical limits for nonstandard, nonvalidated methods are not provided, the maximum rating possible is Deficient. **Quality Control** Studies will be evaluated as Deficient if they include the following: • Use of NIST standard reference materials or commercial certified reference materials (CRMs) in non–blood-based matrices (e.g., NIST 2976, 3668, 1641e; several European Union Joint Research Centre [JRC] CRMs). • QC results fall outside of "Adequate" acceptance ranges but are not severe enough to warrant exclusion of the study (e.g., recovery of QC samples in the range 80%–85% or 115%–120%; variability of replicate samples 15%–20% RPD). OR • QC procedures and results not discussed in the paper or supplemental information. NOTE: Where QC procedures are not described, requesting additional information from the corresponding author will be necessary because QC results are instrumental in gauging reliability of reported sample data.
Critically Deficient		Low proportion of samples (<50%) above the LOD OR LOD greater than median blood mercury concentration in the study population.

Level		Criteria
		Method If any of the following are true: • Analysis of dried blood spots (DBS). NOTE: To date, there have been virtually no studies that have explored the analysis of dried blood spot samples for mercury or methylmercury. Additionally, the field of dried blood spot analysis of metals is still in development, especially for quantitative applications. • Use of inappropriate standard method (e.g., methods EPA 245.2 or 7471B). NOTE: The above methods describe the analysis of bulk volumes of water for mercury content, so those and similar methods cannot be directly used for analysis of mercury in blood. It may be possible to adapt such methods for use with blood samples, but they should be treated as nonstandard, nonvalidated methods and detailed descriptions of the preparation and analysis methods must be provided. • No collection, preparation, or analysis method details described in the paper, supplementary information, or cited reports. NOTE: If no description of the preparation and analysis methods are provided, it is not possible to confirm that appropriate measures were taken to ensure the accuracy of results. Quality control results alone may not account for all essential method parameters in study samples (e.g., collection and storage measures). • For ICP-MS or GC-ICP-MS studies: No internal standard reported. NOTE: The use of internal standards is an essential aspect of ICP-MS analysis to appropriately account for instrument drift and matrix impact on analyte signals. **Quality Control** • QC results reveal concerns about reliability of measurements (e.g., recovery of QC samples outside 80%–120%, variability of replicate samples >20% RPD). • If after inquiry, it is found that no QC samples were analyzed. NOTE: Analysis of quality control samples is essential to demonstrate the accuracy of chemical analyses and provide confidence in data quality and is generally considered a standard practice in analytical laboratories. QC samples can demonstrate that analytical accuracy is maintained even in complex sample matrices, such as blood. If no quality control samples were prepared by the same method and analyzed alongside study samples, then it is not possible to have confidence in the quality of analytical data generated in support of an epidemiology study.

Table C-2. Evaluation of analytical methods for hair total and methylmercury in epidemiology studies

Level		Criteria
Good		**Acceptable proportion of samples (>50%) above the limit of detection (LOD) OR LOD less than median hair mercury concentration in study population** **Method** All papers must include a description of methodological factors: • All papers must indicate the sample collection considerations including: age and sex of study participants, hair segment characteristics (e.g., length from scalp, length of hair analyzed). Studies should clearly indicate the area of the body from which hair was collected (e.g., scalp, underarm, pubis), which is important for interpreting results and comparing across studies. • Sample storage and shipment procedures must be described including sample containers/bags, tying procedures, etc. • Cleaning procedures for hair samples must be described. Acceptable washing procedures include deionized water, ionic or nonionic detergent solution, acetone, methanol, etc. Multiple washes or heat, or both, may be used below 65°C. An example standard cleaning technique is presented in IAEA Report 50 (IAEA/RL/50), "Activation analysis of hair as an indicator of contamination of man by environmental trace element pollutants."[a] • For laboratories using a standard method: The paper will include the method number and a citation for a standard method for measuring mercury levels in hair (e.g., United States Department of Agriculture (USDA) Food Safety Inspection Service [FSIS] method MER for total mercury, Environmental Protection Agency (EPA) SW-846 Method 3200 for extraction only, EPA SW-846 method 6800 for Hg speciation only, European Union Consortium to Perform Human Biomonitoring on a European Scale [COPHES] D3.6 "Mercury: Determination in Scalp Hair" for total mercury only) OR a method description discussed as follows. • For laboratories using a nonstandard, validated method: The paper will provide a citation for a peer-reviewed report of the validation in the main body of the paper or in supplementary information (refer to "Useful Terms Defined" for more information on validation) OR a description of the method discussed, as follows. • For laboratories using nonstandard, nonvalidated methods: Nonstandard methods may perform acceptably if sufficient evidence of data quality (quality control [QC] results) is provided (as follows). For nonstandard methods, the report must provide a description of the method including the following: o Analytical limits including LOD and/or the limit of quantitation (LOQ) must be reported for nonstandard, nonvalidated methods to obtain a Good rating. o Combustion Atomic Absorption Spectroscopy (CAAS) refers to direct mercury analysis and may use several instruments, such as

Level		Criteria
		the Milestone Direct Mercury Analyzer (DMA) or the Nippon Mercury Analyzer (MA). ▪ For studies using CAAS for total mercury analysis, critical method variables that should be noted are time and temperature of sample drying step, charring/decomposition step of analysis, and amalgamator flash time and temperature. o Sample stabilization measures (e.g., addition of gold or thiols during or after extraction; refer to "Useful Terms Defined" for additional details). o Sample preparation method including, but not limited to, the following: ▪ Hair sample drying method and conditions (e.g., temperature, time). ▪ Derivatization steps for gas chromatography-inductively coupled plasma-mass spectrometry (GC-ICP-MS) analysis or another speciation technique for MeHg. ▪ Oxidizing agents for cold vapor atomic fluorescence spectroscopy (CVAFS) and cold vapor atomic absorption spectroscopy (CVAAS) (e.g., potassium permanganate [$KMnO_4$], nitric acid, or sulfuric acid for total mercury analysis). ▪ Reducing agent for CVAFS and CVAAS (e.g., tin chloride [$SnCl_2$], sodium borohydride [$NaBH_4$], sodium hypochlorite [$NaClO_2$], hydroxylamine hydrochloride). ▪ For total mercury analysis only: Digestion reagents, temperatures (≤65°C for open vessel digestion or digestion without stabilizers mentioned previously), and times. ▪ Extraction solvent composition. • For speciation analysis (i.e., methylmercury): Separation steps before detection. Examples include the following: o For liquid chromatography (LC; e.g., high-performance liquid chromatography [HPLC] or ultra-high-performance liquid chromatography [UHPLC]): Column type, eluent composition (e.g., mercaptoethanol, ethylenediaminetetraacetic acid [EDTA], cysteine), eluent gradient (or indicate whether an isocratic program [where the same eluent content was used through the run] was used), and injection volume. o For gas chromatography: Transfer line temperature, column temperature program, and column identity. o For distillation: Solvent and temperature. • Use of an appropriate internal standard (e.g., bismuth [Bi], praseodymium [Pr], holmium [Ho]; ICP-MS only; other organomercury compound [e.g., propylmercury, butylmercury; speciation by LC- or GC-ICP-MS only]): o NOTE: Internal standard is not generally included in CVAAS/CVAFS/CAAS methods.

Level		Criteria
		NOTE: the items listed above should be included, but due to the great variability in nonstandard methods, it is not necessary to evaluate the quality of all methodological factors, only ensure that they are included in the method description. **Quality Control** All papers (standard methods, validated methods, and nonstandard methods) must include a description of QC procedures and results performed to verify method performance and data quality for their study: • At least three laboratory QC procedures and results in methods section, supplementary materials, cited papers, or standardized laboratory protocol including, but not limited, to the following: o Hair-based certified reference material (CRM) (NIES CRM 13: Human Hair, IAEA-085 or -086: Human Hair (Methyl Mercury), Joint Research Centre CRM BCR-397: Trace elements in human hair for total mercury, Chinese CRM GBW 07601): % recovery information, 85–115%. NOTE: As of this time, no NIST (National Institute of Standards and Technology) Standard Reference Materials are available for hair mercury. o Duplicate sample preparation and relative percent difference (RPD) results, <20% RPD. o Representative blank sample analyses or chromatograms to demonstrate the absence of interferences. o Recovery of spiked study samples, 85%–115%. o Post-extraction spiked sample recovery or method blank spiked (MBS) samples, 85%–115%. o Replicate QC precision (also called uncertainty, imprecision, repeatability, or reproducibility; or percent relative standard deviation [%RSD] or coefficient of variation [%CV], <20%, or correlation coefficient >0.85). o Participation in an interlaboratory testing program with documented results for mercury in hair samples and "satisfactory" results. NOTE: A wide range of interlaboratory testing programs is available for *trace* metals, but only ones that monitor mercury in hair, or that use standard methods, are relevant for assessing data quality. o Incurred sample reanalysis (ISR) to demonstrate reproducibility on different days, and RPD results, <20% RPD; refer to "Useful Terms Defined" for more information. o Control charts (e.g., Bland-Altman, Levey-Jennings, Harrell-Davis, Shewhart) for QC samples of the same sample type, prepared by the cited method that show method performance over time (large sample populations only).

Level		Criteria
		Note: If any of the above method parameters or quality control measures exhibit values that fall outside the range of values noted above, then the entire study should be categorized at the lowest quality level of the individual method parameter(s) or QC measure(s).
Adequate		**Method** Same as Good, with the following exceptions: • All papers may indicate the sample collection considerations including age and sex of study participants, hair segment characteristics (e.g., length from scalp, length of hair analyzed). Studies should clearly indicate the area of the body from which hair was collected (e.g. scalp, underarm, pubis), which is important for interpreting results and comparing across studies. • Sample storage and shipment procedures must be described including sample containers/bags, tying procedures, etc. • Cleaning procedures for hair samples must be described. Acceptable washing procedures include deionized water, ionic or nonionic detergent solution, acetone, methanol, etc. Multiple washes or heat, or both, may be used below 65°C. An example standard cleaning technique is presented in IAEA Report 50 (IAEA/RL/50), "Activation analysis of hair as an indicator of contamination of man by environmental trace element pollutants."[a] • For laboratories using standard methods: Samples analyzed by EPA SW-846 method 6800 (Hg speciation only) or method 7473 (total mercury only). • For nonstandard, nonvalidated methods: The report will comment on two or three methodological details above such as extraction times, cleaning times between sample analyses, instrumental technique, and so on. • Analytical limits including LOD and/or the limit of quantitation (LOQ) must be reported for nonstandard, nonvalidated methods to obtain an Acceptable rating. • For measurements of methylmercury: Determined methylmercury by subtracting inorganic mercury (iHg) from total mercury (THg). • For ICP-MS only: Use of less common or less appropriate internal standards such as terbium (Tb), rhodium (Rh), gallium (Ga), thallium (Tl), indium (In), yttrium (Y), or scandium (Sc). **Quality Control** Any Adequate analysis must include: • For standard methods and validated methods (with literature citation): Discussion of one or two laboratory QC procedures and results in methods section, supplementary materials, cited paper, or standardized laboratory protocol including either: o Reference material percentage recovery information (may be the certified reference materials [CRMs] mentioned earlier or another

Level		Criteria
		such as NIST Standard Reference Material [SRM] 3133 or 3177 for use in spiking total Hg only), recovery will fall in the range 80%–85% or 115%–120%. o Replicate sample preparation information (e.g., %CV or RPD) for duplicate preparation of 20%–25%. • <u>For nonstandard, nonvalidated studies:</u> Discussion of at least three laboratory QC procedures and results in the methods section, supplementary materials, cited papers, or standardized laboratory protocol, including a combination of the procedures mentioned under "Good" and "Adequate" quality levels.
Deficient		**<u>Method</u>** Same as Adequate except: • For Deficient quality: All papers may indicate 1-2 sample collection considerations including age and sex of study participants, hair segment characteristics (e.g., length from scalp, length of hair analyzed). Studies should clearly indicate the area of the body from which hair was collected (e.g. scalp, underarm, pubis), which is important for interpreting results and comparing across studies. • Sample storage and shipment procedures must be described including sample containers/bags, tying procedures, etc. • Cleaning procedures for hair samples must be described. Acceptable washing procedures include deionized water, ionic or nonionic detergent solution, acetone, methanol, etc. Multiple washes or heat, or both, may be used below 65°C. An example standard cleaning technique is presented in IAEA Report 50 (IAEA/RL/50), "Activation analysis of hair as an indicator of contamination of man by environmental trace element pollutants."[a] • Minimal methodological details provided in the paper, supplemental information, or cited papers; only <u>one or two</u> of the items from the aforementioned method detail list. o For example, only mentioning the instrumental technique used for analysis, omitting all collection, storage, and cleaning procedures, analytical limits or references for the sample preparation method. NOTE: The analytical limits for nonstandard, nonvalidated methods must be provided for papers to be considered Good or Adequate, as it is necessary to determine whether a method was used with sensitivity levels appropriate to the matrix and that the method has been appropriately optimized by the analytical laboratory. If analytical limits for nonstandard, nonvalidated methods are not provided, the maximum rating possible is Deficient. **<u>Quality Control</u>** Studies will be evaluated as Deficient if they include the following:

Level		Criteria
		• Use of NIST standard reference materials or commercial certified reference materials in non-hair-based tissue matrices (e.g., NIST 1575a [total mercury only], 1946, 1947, 2976, several European Union Joint Research Centre [JRC] CRMs). • QC results fall outside of "Adequate" acceptance ranges but are not severe enough to warrant exclusion of the study (e.g., recovery of QC samples in the range 75%–80% or 120%–125%; variability of replicate samples 25%–50% RPD). OR • QC Procedures and results not discussed in the paper or supplemental information. NOTE: Where QC procedures are not described, requesting additional information from the corresponding author will be necessary because QC results are instrumental in gauging reliability of reported sample data.
Critically Deficient		Low proportion of samples (<50%) above the LOD OR LOD greater than median blood mercury concentration in the study population. **Method** If any of the below are true: • Use of inappropriate standard method (e.g., methods EPA 245.1, EPA 245.2, EPA 245.7, EPA 7470A, EPA 1630, EPA 1631E). NOTE: The above methods describe the analysis of bulk volumes of water for mercury content, so those and similar methods cannot be directly used for analysis of mercury in hair. It may be possible to adapt such methods for use with hair samples, but they should be treated as nonstandard, nonvalidated methods and detailed descriptions of the preparation and analysis methods must be provided. • No collection, preparation, or analysis method details described in the paper, supplementary information, or cited reports. NOTE: If no description of the preparation and analysis methods are provided, it is not possible to confirm that appropriate measures were taken to ensure the accuracy of results. Quality control results alone may not account for all essential method parameters in study samples (e.g., collection and storage measures). • Cleaning, drying, or open-vessel extraction of samples at any temperature greater than 65°C except sample preparation for speciation analysis by distillation and closed-vessel digestion of hair samples for total mercury analysis in the presence of sulfur-containing stabilizing agents. NOTE: Mercury volatilizes at temperatures above 65°C in the absence of a stabilizer during sample preparation such as thiols or gold. Cleaning and drying procedures are not generally performed in the presence of such stabilizers and open vessel digestions will readily allow release of volatilized mercury.

Level		Criteria
		• For ICP-MS or GC-ICP-MS studies: No internal standard reported. NOTE: The use of internal standards is an essential aspect of ICP-MS analysis to appropriately account for instrument drift and matrix impact on analyte signals. **Quality Control** • QC results reveal concerns about reliability of measurements (e.g., recovery of QC samples outside 75%–120% or variability of replicate samples >20% RPD). • If after inquiry, it is found that no QC samples were analyzed. NOTE: Analysis of quality control samples is essential to demonstrate the accuracy of chemical analyses and provide confidence in data quality and is generally considered a standard practice in analytical laboratories. QC samples can demonstrate that analytical accuracy is maintained even in complex sample matrices, such as blood. If no quality control samples were prepared by the same method and analyzed alongside study samples, then it is not possible to have confidence in the quality of analytical data generated in support of an epidemiology study.

[a]Ryabukhin, Y. S. (1976). Activation analysis of hair as an indicator of contamination of man by environmental trace element pollutants (No. IAEA-RL--50). International Atomic Energy Agency.

In comparison of the quality control procedures between hair and blood, it should be noted that the criteria for each rating category are different for the two matrices, specifically the recovery/relative error (accuracy) of spiked QC samples and percent difference (precision) of duplicate samples. The differences serve as an acknowledgment that analysis of hair is subject to greater variability than analysis of blood as a result of several factors, primarily sample collection, instrumental error when weighing samples, segmenting of hair samples, etc.

Useful Terms Defined

Acceptance criteria: Predetermined values for accuracy or precision that demonstrate the generation of data of acceptable quality or accuracy. If a measurement falls outside predefined acceptance criteria, then the reliability of data generated during analysis or in the same analytical batch may be affected.

Coefficient of variation: A measure of variability of repeated measurement of a value, or precision. Calculated as the ratio of standard deviation of a set of measurements to the average value; often expressed as a percentage:

$$\text{Coefficient of variation (\%CV)} = \frac{\text{Standard deviation (SD,}\sigma\text{)}}{\text{Mean concentration}} \times 100\%$$

Also called relative standard deviation, %RSD, and uncertainty.

Control chart: A graphical representation of quality control results over time, intended to demonstrate reliable method performance for longer analysis projects, usually encompassing anywhere from several days to several months. Many types of control charts can be designed to show instrument or method performance measures. The most relevant charts for assessing confidence in analytical data quality show method performance measures, such as quality control sample recovery (accuracy) or precision over time. Specific examples of types of control charts include Bland–Altman, Levey–Jennings, Harrell–Davis, Shewhart, and others.

Correlation coefficient: Statistical analysis showing agreement between replicate measurements over a range of concentrations. Calculated by plotting the concentration of duplicate samples over several concentrations, then plotting a line of best fit to the pairs of data. The correlation coefficient of the line shows how closely the points fall to the ideal line with a slope of 1.

Derivatization: A sample preparation step involving chemical reaction of the target analyte to enhance a particular chemical property for the purposes of analysis. Examples include reacting mercury and methylmercury with sodium tetraethylborate ($NaBEt_4$). The resulting products are more reactive and thus better suited to analysis, in this case, by gas chromatography. Products can also still be separated and accurately analyzed.

Digestion: A sample preparation method involving degradation of the sample, with acid or base, to break down the matrix and efficiently separate the analyte from the sample matrix. Digestion methods are generally very harsh and are less able to preserve analyte speciation, often making them preferable for total metal analysis (e.g., total mercury analysis).

Dried blood spots: A sample collection technique for analyzing blood samples collected for screening purposes on a filter paper or similar material, often done in newborn screening but also increasingly implemented in public health studies because of ease of collection and lower concerns about sample stability or storage. Sometimes abbreviated as DBS. Also called dried matrix spots (DMS).

Extraction: A sample preparation method involving the separation of a target analyte from a sample by selective dissolution in a liquid solvent (organic or aqueous). Extraction methods are often gentler than digestion methods and are often more suitable for analyzing speciation of samples. Most common extraction techniques are distillation and alkaline extraction (using a basic pH to gently break down tissues or solids).

Incurred sample reanalysis: Reanalysis of select samples on a second analytical day to demonstrate reproducibility of results over time. This quality control measure differs from duplicate analysis, which is generally performed on the same day.

Interference: Any substance besides the target analyte that results in inaccurate measurements of a chemical signal during analysis. There are several types of interferences, including chemical interferences, spectral interferences, and so on. Interferences may result in higher or lower measured concentrations of a chemical depending on the nature of the interference. Instrumental factors are generally not included in the definition of interferences.

Internal standard: Chemical substance added to standards and samples in ICP-MS analysis to allow monitoring of instrument drift or effects on analyte signal due to sample composition. For mercury analysis, usually relevant only to ICP-MS methods.

Laboratory accreditation: A certification for laboratories to demonstrate the application of quality systems and laboratory practices that support the generation of high-quality data. Programs are organized by a range of commercial, government, or academic entities and are often administered annually.

Matrix: A specific type of sample, such as blood, hair, water, or biological tissue.

Matrix blank: A type of quality control sample prepared containing only extraction reagents or solvents and control biological matrix that are carried through the sample preparation process; analyzed to monitor endogenous concentration of analyte in unexposed samples. In toxicology studies, these are expected to be low unless the target analyte is a pervasive environmental contaminant.

Memory effect: A chemical phenomenon specific to mercury and a few other elements where the metal interacts strongly with the surface of plastic materials of an instrument and adsorb to the surface, only to be released slowly over time. This phenomenon results in a gradual decrease of mercury signal to zero between samples and requires special considerations in the design of an experiment including special wash solutions, longer cleaning times, or substitution of alternate materials to prevent the effect from biasing measured results.

Method blank: A type of quality control sample prepared containing only extraction reagents or solvents that are carried through the sample preparation process; analyzed to monitor contribution to a background analyte signal arising from the sample preparation or reagents. Also called reagent blank.

Recovery: A measure of method accuracy; comparison of the measured concentration of an analyte against the expected or "known" concentration of the analyte. Usually calculated as follows:

$$\text{Recovery} = \frac{\text{Measured concentration}}{\text{Expected concentration}} \times 100\%$$

Reference material: A well-characterized standard sample containing known concentrations of one or more analytes, intended to be processed and analyzed alongside study samples to verify accurate measurement of an analyte in samples when using an analytical method.

- Certified reference material (CRM): A reference material that has been repeatedly characterized for use as a quality control sample. Usually produced in smaller batches and sold by commercial entities, with a less rigorous characterization than standard reference materials.

- Standard reference material (SRM): A reference material that has been produced in large single batches and characterized to a very high degree, usually by an official organization such as the National Institute of Standards and Technology, for use as a quality control sample. Considered the "gold standard" of reference materials for use as a quality control sample.

Optimal reference material selection will include the exact combination of analyte and matrix being analyzed, but if such a combination is not available, then similar matrices containing the target analyte may be used.

Relative error: An alternate measure of method accuracy, demonstrating the variance of a measured concentration of an analyte from the expected or "known" concentration of the analyte. Usually calculated as follows:

$$\text{Relative error (\%RE)} = \frac{(\text{Measured concentration} - \text{Expected concentration})}{\text{Expected concentration}} \times 100\%$$

Replicate: Preparation of a quality control or study sample two or more times for repeated determination of the concentration of an analyte, intended to demonstrate precision of analytical measurements, usually performed on the same day to demonstrate reproducibility in an analytical batch.

Speciation: Study of the different chemical forms of an analyte; in the case of mercury, referring to the analysis of the concentration and percentage of inorganic mercury and organic mercury, such as methylmercury, dimethylmercury, ethylmercury, and so on, present in a sample. Important considerations in speciation methods are the sample preparation methods, chemical composition of eluents, temperatures, and column types, all of which control the separation of different forms and the ability to maintain the forms of mercury throughout a speciation analysis.

Spike: Addition of the analyte to a method blank, matrix blank, or study sample, intended to demonstrate accurate measurement of a known concentration of the analyte after processing alongside study samples.

Stabilization: Ensuring that the concentration and species of analytes remain constant over time. This process can apply to sample storage before analysis or sample preparation, to ensure that no mercury is lost or converted before analysis. For analysis of mercury, this is most often performed through addition of gold to interact with mercury, addition of hydrochloric acid (specific to total mercury analysis after digestion), addition of a metal binding agent like EDTA, or addition of sulfur-containing chemicals (thiols). This is usually only necessary for long-term storage of blood samples or during extraction and digestion processes to prevent volatilization. One part of a method validation may be to characterize chemical stability in the sample tissue or in digests or extracts.

Storage Stability Samples: A specific type of quality control sample that is prepared and analyzed to demonstrate the impact of long storage periods on recovery of mercury in samples. Storage stability samples are a set of control samples of the same type as study samples (blood for the purposes of this document) spiked with the analyte of interest (mercury in this case) and stored alongside study samples from time of receipt until time of analysis. The recovery of mercury in storage stability samples provides a measure of potential loss of mercury as a result of long storage times.

Uncertainty: A measure of precision in analytical measurements, encompassing several types of calculations including relative standard deviation, coefficient of variation, reproducibility, and repeatability.

Validation: Systematic performance of the preparation and analysis method, to demonstrate consistent accurate performance of a method under a range of conditions. Includes repeated analysis of replicates of standard samples or spiked matrix samples, often over several days or with multiple analysts. Characterizes linear range, accuracy, precision, analytical limits (LOD and LOQ), specificity (the absence of analyte signal due to interferences), stability of the analyte in samples or extracts. One approach to validation can be found in the FDA publication *Bioanalytical Method Validation: Guidance for Industry* (FDA, 2018), which discusses parameters to be validated. However, this publication is specifically aimed at the design of studies regulated by Good Manufacturing Practices and Good Laboratory Practices; alternative approaches can be used for research purposes that address the critical parameters discussed earlier.

Useful Abbreviations Defined

CAAS	combustion atomic absorption spectroscopy
CRM	certified reference material
CV	coefficient of variation OR cold vapor (depending on context)
CVAAS	cold vapor atomic absorption spectroscopy
CVAFS	cold vapor atomic fluorescence spectroscopy
DBS	dried blood spot
DMS	dried matrix spot
EDTA	ethylenediaminetetraacetic acid
EPA	Environmental Protection Agency
FDA	Food and Drug Administration
GC	gas chromatography
HPLC	high-performance liquid chromatography
ICP-MS	inductively coupled plasma–mass spectrometry
ISR	incurred sample reanalysis
JRC	European Commission Joint Research Centre
LC	liquid chromatography
LOD	limit of detection
LOQ	limit of quantitation
MBS	method blank spiked
NIST	National Institute of Standards and Technology
QC	quality control
RE	relative error
RPD	relative percent difference
RSD	relative standard deviation
SRM	standard reference material
UHPLC	ultra-high-performance liquid chromatography

REFERENCES

Abass, K; Huusko, A; Knutsen, HK; Nieminen, P; Myllynen, P; Meltzer, HM; Vahakangas, K; Rautio, A. (2018). Quantitative estimation of mercury intake by toxicokinetic modelling based on total mercury levels in humans. Environ Int 114: 1-11. http://dx.doi.org/10.1016/j.envint.2018.02.028

Al-Shahristani, H; Shihab, KM. (1974). Variation of biological half-life of methylmercury in man. Arch Environ Occup Health 28: 342-344. http://dx.doi.org/10.1080/00039896.1974.10666505

Albert, I; Villeret, G; Paris, A; Verger, P. (2010). Integrating variability in half-lives and dietary intakes to predict mercury concentration in hair. Regul Toxicol Pharmacol 58: 482-489. http://dx.doi.org/10.1016/j.yrtph.2010.08.020

Allen, BC; Hack, CE; Clewell, HJ. (2007). Use of Markov Chain Monte Carlo analysis with a physiologically-based pharmacokinetic model of methylmercury to estimate exposures in US women of childbearing age. Risk Anal 27: 947-959. http://dx.doi.org/10.1111/j.1539-6924.2007.00934.x

ATSDR (Agency for Toxic Substances and Disease Registry). (1999). Toxicological profile for mercury [ATSDR Tox Profile]. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service. http://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=115&tid=24

Berthet, A; de Batz, A; Tardif, R; Charest-Tardif, G; Truchon, G; Vernez, D; Droz, PO. (2010). Impact of biological and environmental variabilities on biological monitoring-an approach using toxicokinetic models. J Occup Environ Hyg 7: 177-184. http://dx.doi.org/10.1080/15459620903530052

Birch, RJ; Bigler, J; Rogers, JW; Zhuang, Y; Clickner, RP. (2014). Trends in blood mercury concentrations and fish consumption among U.S. women of reproductive age, NHANES, 1999-2010. Environ Res 133: 431-438. http://dx.doi.org/10.1016/j.envres.2014.02.001

Boucher, O; Bastien, CH; Saint-Amour, D; Dewailly, É; Ayotte, P; Jacobson, JL; Jacobson, SW; Muckle, G. (2010). Prenatal exposure to methylmercury and PCBs affects distinct stages of information processing: An event-related potential study with Inuit children. Neurotoxicology 31: 373-384. http://dx.doi.org/10.1016/j.neuro.2010.04.005

Budtz-Jørgensen, E; Keiding, N; Grandjean, P. (1999). Benchmark modeling of the Faroese methylmercury data: Final report to U.S. EPA. (Research Report 99/5). Copenhagen, Denmark: University of Copenhagen.

Budtz-Jorgensen, E; Keiding, N; Grandjean, P; Weihe, P. (2007). Confounder selection in environmental epidemiology: Assessment of health effects of prenatal mercury exposure. Ann Epidemiol 17: 27-35. http://dx.doi.org/10.1016/j.annepidem.2006.05.007

Byczkowski, JZ; Lipscomb, JC. (2001). Physiologically based pharmacokinetic modeling of the lactational transfer of methylmercury. Risk Anal 21: 869-882. http://dx.doi.org/10.1111/0272-4332.215158

Carrier, G; Bouchard, M; Brunet, RC; Caza, M. (2001a). A toxicokinetic model for predicting the tissue distribution and elimination of organic and inorganic mercury following exposure to methyl mercury in animals and humans. II. Application and validation of the model in humans. Toxicol Appl Pharmacol 171: 50-60. http://dx.doi.org/10.1006/taap.2000.9113

Carrier, G; Brunet, RC; Caza, M; Bouchard, M. (2001b). A toxicokinetic model for predicting the tissue distribution and elimination of organic and inorganic mercury following exposure to methyl mercury in animals and humans. I. Development and validation of the model using experimental data in rats. Toxicol Appl Pharmacol 171: 38-49. http://dx.doi.org/10.1006/taap.2000.9112

CDC (Centers for Disease Control and Prevention). (2009). Mercury fact sheet [Fact Sheet]. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service, Agency for Toxic Substances and Disease Registry.

CDC (Centers for Disease Control and Prevention). (2016a). Anthropometric reference data for children and adults: United States, 2011–2014. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

CDC (Centers for Disease Control and Prevention). (2016b). Mercury biomonitoring summary. Available online at https://www.cdc.gov/biomonitoring/Mercury_BiomonitoringSummary.html

CDC (Centers for Disease Control and Prevention). (2017). Table blood methyl mercury (2011-2014), fourth national report on human exposure to environmental chemicals, updated tables, January 2017, volume one (pp. 278). Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service, Agency for Toxic Substances and Disease Registry. https://www.cdc.gov/exposurereport/pdf/FourthReport_UpdatedTables_Volume1_Jan2017.pdf

CDC (Centers for Disease Control and Prevention). (2018). Fourth national report on human exposure to environmental chemicals, updated tables, March 2018, volume one. Atlanta, GA: U.S. Department of Health and Human Services. https://www.cdc.gov/exposurereport/pdf/FourthReport_UpdatedTables_Volume1_Mar2018.pdf

Clewell, HJ; Gearhart, JM; Gentry, PR; Covington, TR; Van Landingham, CB; Crump, KS; Shipp, AM. (1999). Evaluation of the uncertainty in an oral reference dose for methylmercury due to interindividual variability in pharmacokinetics. Risk Anal 19: 547-558. http://dx.doi.org/10.1023/a:1007017116171

Cui, W; Liu, G; Bezerra, M; Lagos, DA; Li, Y; Cai, Y. (2017). Occurrence of methylmercury in rice-based infant cereals and estimation of daily dietary intake of methylmercury for infants. J Agric Food Chem 65: 9569-9578. http://dx.doi.org/10.1021/acs.jafc.7b03236

Daniels, JL; Longnecker, MP; Rowland, AS; Golding, J; Team, AS. (2004). Fish intake during pregnancy and early cognitive development of offspring. Epidemiology 15: 394-402. http://dx.doi.org/10.1097/01.ede.0000129514.46451.ce

Després, C; Beuter, A; Richer, F; Poitras, K; Veilleux, A; Ayotte, P; Dewailly, É; Saint-Amour, D; Muckle, G. (2005). Neuromotor functions in Inuit preschool children exposed to Pb, PCBs, and Hg. Neurotoxicol Teratol 27: 245-257. http://dx.doi.org/10.1016/j.ntt.2004.12.001

FDA (U.S. Food and Drug Administration). (2018). Bioanalytical method validation guidance for industry. Rockville, MD: Food and Drug Administration. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/bioanalytical-method-validation-guidance-industry

Golding, J; Gregory, S; Iles-Caven, Y; Hibbeln, J; Emond, A; Taylor, CM. (2016). Associations between prenatal mercury exposure and early child development in the ALSPAC study. Neurotoxicology 53: 215-222. http://dx.doi.org/10.1016/j.neuro.2016.02.006

Gosselin, NH; Brunet, RC; Carrier, G; Bouchard, M; Feeley, M. (2006). Reconstruction of methylmercury intakes in indigenous populations from biomarker data. J Expo Sci Environ Epidemiol 16: 19-29. http://dx.doi.org/10.1038/sj.jea.7500433

Grandjean, P; Weihe, P; White, RF; Debes, F; Araki, S; Yokoyama, K; Murata, K; Sørensen, N; Dahl, R; Jørgensen, PJ. (1997). Cognitive deficit in 7-year-old children with prenatal exposure to methylmercury. Neurotoxicol Teratol 19: 417-428. http://dx.doi.org/10.1016/S0892-0362(97)00097-4

Health Canada. (2007). Mercury and human health. Ottawa, Ontario, Canada. https://www.canada.ca/en/health-canada/services/healthy-living/your-health/environment/mercury-human-health.html

Hoenig, JM; Heisey, DM. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. Am Stat 55: 19-24.

Howard, BE; Phillips, J; Miller, K; Tandon, A; Mav, D; Shah, MR; Holmgren, S; Pelch, KE; Walker, V; Rooney, AA; Macleod, M; Shah, RR; Thayer, K. (2016). SWIFT-Review: a text-mining workbench for systematic review. Syst Rev 5: 87. http://dx.doi.org/10.1186/s13643-016-0263-z

Jacobson, JL; Muckle, G; Ayotte, P; Dewailly, É; Jacobson, SW. (2015). Relation of prenatal methylmercury exposure from environmental sources to childhood IQ. Environ Health Perspect 123: 827-833. http://dx.doi.org/10.1289/ehp.1408554

Jin, L; Liu, M; Zhang, L; Li, Z; Yu, J; Liu, J; Ye, R; Chen, L; Ren, A. (2016). Exposure of methyl mercury in utero and the risk of neural tube defects in a Chinese population. Reprod Toxicol 61: 131-135. http://dx.doi.org/10.1016/j.reprotox.2016.03.040

Jo, S; Woo, HD; Kwon, HJ; Oh, SY; Park, JD; Hong, YS; Pyo, H; Park, KS; Ha, M; Kim, H; Sohn, SJ; Kim, YM; Lim, JA; Lee, SA; Eom, SY; Kim, BG; Lee, KM; Lee, JH; Hwang, MS; Kim, J. (2015). Estimation of the biological half-life of methylmercury using a population toxicokinetic model. Int J Environ Res Public Health 12: 9054-9067. http://dx.doi.org/10.3390/ijerph120809054

Karagas, MR; Choi, AL; Oken, E; Horvat, M; Schoeny, R; Kamai, E; Cowell, W; Grandjean, P; Korrick, S. (2012). Evidence on the human health effects of low-level methylmercury exposure [Review]. Environ Health Perspect 120: 799-806. http://dx.doi.org/10.1289/ehp.1104494

Kershaw, TG; Clarkson, TW; Dhahir, PH. (1980). The relationship between blood levels and dose of methylmercury in man. Arch Environ Occup Health 35: 28-36.

Kirman, CR; Sweeney, LM; Meek, ME; Gargas, ML. (2003). Assessing the dose-dependency of allometric scaling performance using physiologically based pharmacokinetic modeling. Regul Toxicol Pharmacol 38: 345-367. http://dx.doi.org/10.1016/j.yrtph.2003.07.004

Lederman, SA; Jones, RL; Caldwell, KL; Rauh, V; Sheets, SE; Tang, D; Viswanathan, S; Becker, M; Stein, JL; Wang, RY; Perera, FP. (2008). Relation between cord blood mercury levels and early child development in a World Trade Center cohort. Environ Health Perspect 116: 1085-1091. http://dx.doi.org/10.1289/ehp.10831

Lee, S; Tan, YM; Phillips, MB; Sobus, JR; Kim, S. (2017). Estimating methylmercury intake for the general population of South Korea using physiologically based pharmacokinetic modeling. Toxicol Sci 159: 6-15. http://dx.doi.org/10.1093/toxsci/kfx111

Leggett, RW; Munro, NB; Eckerman, KF. (2001). Proposed revision of the ICRP model for inhaled mercury vapor. Health Phys 81: 450-455. http://dx.doi.org/10.1097/00004032-200110000-00010

Li, P; Feng, X; Chan, H; Zhang, X; Du, B. (2015). Human body burden and dietary methylmercury intake: The relationship in a rice-consuming population. Environ Sci Technol 49: 9682-9689. http://dx.doi.org/10.1021/acs.est.5b00195

Lowe, JR; Erickson, SJ; Schrader, R; Duncan, AF. (2012). Comparison of the Bayley II Mental Developmental Index and the Bayley III Cognitive Scale: Are we measuring the same thing? Acta Paediatr 101: e55-e58. http://dx.doi.org/10.1111/j.1651-2227.2011.02517.x

Lu, G; Abduljalil, K; Jamei, M; Johnson, TN; Soltani, H; Rostami-Hodjegan, A. (2012). Physiologically-based pharmacokinetic (PBPK) models for assessing the kinetics of xenobiotics during pregnancy: Achievements and shortcomings [Review]. Curr Drug Metab 13: 695-720. http://dx.doi.org/10.2174/138920012800840374

Mahaffey, KR; Clickner, RP; Jeffries, RA. (2009). Adult women's blood mercury concentrations vary regionally in the United States: Association with patterns of fish consumption (NHANES 1999-2004). Environ Health Perspect 117: 47-53. http://dx.doi.org/10.1289/ehp.11674

Mcnally, K; Loizou, GD. (2015). A probabilistic model of human variability in physiology for future application to dose reconstruction and QIVIVE. Front Pharmacol 6: 213. http://dx.doi.org/10.3389/fphar.2015.00213

Miettinen, JK; Rahola, T; Hattula, T; Rissanen, K; Tillander, M. (1971). Elimination of 203-Hg-methylmercury in man. Ann Clin Res 3: 116-122.

Ng, S; Lin, CC; Jeng, SF; Hwang, YH; Hsieh, WS; Chen, PC. (2015). Mercury, APOE, and child behavior. Chemosphere 120: 123-130. http://dx.doi.org/10.1016/j.chemosphere.2014.06.003

Noisel, N; Bouchard, M; Carrier, G; Plante, M. (2011). Comparison of a toxicokinetic and a questionnaire-based approach to assess methylmercury intake in exposed individuals. J Expo Sci Environ Epidemiol 21: 328-335. http://dx.doi.org/10.1038/jes.2010.33

NRC (National Research Council). (2000). Toxicological effects of methylmercury. Washington, DC: National Academy Press. http://dx.doi.org/10.17226/9899

NRC (National Research Council). (2014). Review of EPA's Integrated Risk Information System (IRIS) process. Washington, DC: The National Academies Press. http://www.nap.edu/catalog.php?record_id=18764

Oken, E; Osterdal, ML; Gillman, MW; Knudsen, VK; Halldorsson, TI; Strom, M; Bellinger, DC; Hadders-Algra, M; Michaelsen, KF; Olsen, SF. (2008). Associations of maternal fish intake during pregnancy and breastfeeding duration with attainment of developmental milestones in early childhood: A study from the Danish National Birth Cohort. Am J Clin Nutr 88: 789-796.

Oken, E; Wright, RO; Kleinman, KP; Bellinger, D; Amarasiriwardena, CJ; Hu, H; Rich-Edwards, JW; Gillman, MW. (2005). Maternal fish consumption, hair mercury, and infant cognition in a U.S. Cohort. Environ Health Perspect 113: 1376-1380. http://dx.doi.org/10.1289/ehp.8041

Orenstein, ST; Thurston, SW; Bellinger, DC; Schwartz, JD; Amarasiriwardena, CJ; Altshul, LM; Korrick, SA. (2014). Prenatal organochlorine and methylmercury exposure and memory and learning in school-age children in communities near the New Bedford Harbor Superfund site, Massachusetts. Environ Health Perspect 122: 1253-1259. http://dx.doi.org/10.1289/ehp.1307804

Ou, L; Wang, H; Chen, C; Chen, L; Zhang, W; Wang, X. (2018). Physiologically based pharmacokinetic (PBPK) modeling of human lactational transfer of methylmercury in China. Environ Int 115: 180-187. http://dx.doi.org/10.1016/j.envint.2018.03.018

Pierrehumbert, G; Droz, PO; Tardif, R; Charest-Tardif, G; Truchon, G. (2002). Impact of human variability on the biological monitoring of exposure to toluene, phenol, lead, and mercury: II. Compartmental based toxicokinetic modelling. Toxicol Lett 134: 165-175. http://dx.doi.org/10.1016/S0378-4274(02)00186-8

Prpić, I; Milardović, A; Vlašić-Cicvarić, I; Špiric, Z; Radić Nišević, J; Vukelić, P; Snoj Tratnik, J; Mazej, D; Horvat, M. (2017). Prenatal exposure to low-level methylmercury alters the child's fine motor skills at the age of 18 months. Environ Res 152: 369-374. http://dx.doi.org/10.1016/j.envres.2016.10.011

Rothenberg, SE; Jackson, BP; Carly McCalla, G; Donohue, A; Emmons, AM. (2017). Co-exposure to methylmercury and inorganic arsenic in baby rice cereals and rice-containing teething biscuits. Environ Res 159: 639-647. http://dx.doi.org/10.1016/j.envres.2017.08.046

Rothenberg, SE; Yu, X; Liu, J; Biasini, FJ; Hong, C; Jiang, X; Nong, Y; Cheng, Y; Korrick, SA. (2016). Maternal methylmercury exposure through rice ingestion and offspring neurodevelopment: A prospective cohort study. Int J Hyg Environ Health 219: 832-842. http://dx.doi.org/10.1016/j.ijheh.2016.07.014

Ruiz, P; Fowler, BA; Osterloh, JD; Fisher, J; Mumtaz, M. (2010). Physiologically based pharmacokinetic (PBPK) tool kit for environmental pollutants–Metals. SAR QSAR Environ Res 21: 603-618. http://dx.doi.org/10.1080/1062936X.2010.528942

Sagiv, SK; Thurston, SW; Bellinger, DC; Amarasiriwardena, C; Korrick, SA. (2012). Prenatal exposure to mercury and fish consumption during pregnancy and attention-deficit/hyperactivity disorder-related behavior in children. Arch Pediatr Adolesc Med 166: 1123-1131. http://dx.doi.org/10.1001/archpediatrics.2012.1286

Sherlock, J; Hislop, J; Newton, D; Topping, G; Whittle, K. (1984). Elevation of mercury in human-blood from controlled chronic ingestion of methylmercury in fish. Hum Toxicol 3: 117-131.

Sirot, V; Guerin, T; Mauras, Y; Garraud, H; Volatier, J; Leblanc, JC. (2008). Methylmercury exposure assessment using dietary and biomarker data among frequent seafood consumers in France - CALIPSO study. Environ Res 107: 30-38. http://dx.doi.org/10.1016/j.envres.2007.12.005

Smith, JC; Allen, PV; Turner, MD; Most, B; Fisher, HL; Hall, LL. (1994). The kinetics of intravenously administered methylmercury in man. Toxicol Appl Pharmacol 128: 251-256. http://dx.doi.org/10.1006/taap.1994.1204

Srivastava, RK; Hutson, N; Martin, B; Princiotta, F; Staudt, J. (2006). Control of mercury emissions from coal-fired electric utility boilers. Environ Sci Technol 40: 1385-1393.

Stern, AH. (2005). A revised probabilistic estimate of the maternal methyl mercury intake dose corresponding to a measured cord blood mercury concentration. Environ Health Perspect 113: 155-163. http://dx.doi.org/10.1289/ehp.7417

Stern, AH; Clewell, HJ; Swartout, J. (2002). An objective uncertainty factor adjustment for methylmercury pharmacokinetic variability. Hum Ecol Risk Assess 8: 885-894.

Stern, AH; Smith, AE. (2003). An assessment of the cord blood:maternal blood methylmercury ratio: implications for risk assessment. Environ Health Perspect 111: 1465-1470. http://dx.doi.org/10.1289/ehp.6187

Sterne, JAC; Hernán, MA; Reeves, BC; Savović, J; Berkman, ND; Viswanathan, M; Henry, D; Altman, DG; Ansari, MT; Boutron, I; Carpenter, JR; Chan, AW; Churchill, R; Deeks, JJ; Hróbjartsson, A; Kirkham, J; Jüni, P; Loke, YK; Pigott, TD; Ramsay, CR; Regidor, D; Rothstein, HR; Sandhu, L; Santaguida, PL; Schünemann, HJ; Shea, B; Shrier, I; Tugwell, P; Turner, L; Valentine, JC; Waddington, H; Waters, E; Wells, GA; Whiting, PF; Higgins, JPT. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. Br Med J 355: i4919.

Straka, E; Ellinger, I; Balthasar, C; Scheinast, M; Schatz, J; Szattler, T; Bleichert, S; Saleh, L; Knöfler, M; Zeisler, H; Hengstschläger, M; Rosner, M; Salzer, H; Gundacker, C. (2016). Mercury toxicokinetics of the healthy human term placenta involve amino acid transporters and ABC transporters. Toxicology 340: 34-42. http://dx.doi.org/10.1016/j.tox.2015.12.005

Suzuki, C. (2016). Assessing change of environmental dynamics by legislation in Japan, using red tide occurrence in Ise Bay as an indicator. Mar Pollut Bull 102: 283-288. http://dx.doi.org/10.1016/j.marpolbul.2015.08.010

Swartout, J; Rice, G. (2000). Uncertainty analysis of the estimated ingestion rates used to derive the methylmercury reference dose. Drug Chem Toxicol 23: 293-306. http://dx.doi.org/10.1081/DCT-100100116

U.S. Department of the Interior. (2000). Mercury in the environment fact sheet 146-00 (October) [Fact Sheet]. Washington, D.C. https://www2.usgs.gov/themes/factsheet/146-00/

U.S. EPA (U.S. Environmental Protection Agency). (1991). Guidelines for developmental toxicity risk assessment (pp. 1-71). (EPA/600/FR-91/001). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=23162

U.S. EPA (U.S. Environmental Protection Agency). (1994). Methods for derivation of inhalation reference concentrations and application of inhalation dosimetry [EPA Report]. (EPA/600/8-90/066F). Research Triangle Park, NC: U.S. Environmental Protection Agency, Office of Research and Development, Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office. https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=71993&CFID=51174829&CFTOKEN=25006317

U.S. EPA (U.S. Environmental Protection Agency). (1996). Guidelines for reproductive toxicity risk assessment (pp. 1-143). (EPA/630/R-96/009). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. https://www.epa.gov/sites/production/files/2014-11/documents/guidelines_repro_toxicity.pdf

U.S. EPA (U.S. Environmental Protection Agency). (1997). Mercury study report to congress. Volume 4. An assessment of exposure to mercury in the United States. (EPA/452/R-97-006). U.S. EPA, Office of Air Quality Planning and Standards and Office of Research and Development.

U.S. EPA (U.S. Environmental Protection Agency). (1998). Guidelines for neurotoxicity risk assessment [EPA Report] (pp. 1-89). (EPA/630/R-95/001F). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. http://www.epa.gov/risk/guidelines-neurotoxicity-risk-assessment

U.S. EPA (U.S. Environmental Protection Agency). (2001a). IRIS summary for methylmercury (MeHg) (CASRN 22967-92-6). Available online at http://www.epa.gov/ncea/iris/subst/0073.htm (accessed September 13, 2010).

U.S. EPA (U.S. Environmental Protection Agency). (2001b). Methylmercury chemical assessment summary. Washington, D.C.: Integrated Risk Information System, National Center for Environmental Assessment.

U.S. EPA (U.S. Environmental Protection Agency). (2002). A review of the reference dose and reference concentration processes. (EPA/630/P-02/002F). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. https://www.epa.gov/sites/production/files/2014-12/documents/rfd-final.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2004). Toxicological review of boron and compounds. In support of summary information on the Integrated Risk Information System (IRIS) [EPA Report]. (EPA/635/04/052). Washington, DC: U.S. Environmental Protection Agency, IRIS. http://nepis.epa.gov/exe/ZyPURL.cgi?Dockey=P1006CK9.txt

U.S. EPA (U.S. Environmental Protection Agency). (2005). Guidelines for carcinogen risk assessment [EPA Report]. (EPA/630/P-03/001F). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. https://www.epa.gov/sites/production/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2010). Guidance for implementing the January 2001 methylmercury water quality criterion, April 2010. (NTIS/11880158).

U.S. EPA (U.S. Environmental Protection Agency). (2011). Toxicological review of trichloroethylene (CASRN 79-01-6) in support of summary information on the Integrated Risk Information System (IRIS) [EPA Report]. (EPA/635/R-09/011F). Washington, DC. https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0199tr/0199tr.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2012). Benchmark dose technical guidance. (EPA/100/R-12/001). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. https://www.epa.gov/risk/benchmark-dose-technical-guidance

U.S. EPA (U.S. Environmental Protection Agency). (2014). Guidance for applying quantitative data to develop data-derived extrapolation factors for interspecies and intraspecies extrapolation [EPA Report]. (EPA/100/R-14/002F). Washington, DC: Risk Assessment Forum, Office of the Science Advisor. https://www.epa.gov/sites/production/files/2015-01/documents/ddef-final.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2015). Science policy council peer review handbook [EPA Report] (4th ed.). (EPA/100/B-15/001). Washington, DC: U.S. Environmental Protection Agency, Science Policy Council. https://www.epa.gov/osa/peer-review-handbook-4th-edition-2015

U.S. EPA (U.S. Environmental Protection Agency). (2018). An umbrella Quality Assurance Project Plan (QAPP) for PBPK models [EPA Report]. (ORD QAPP ID No: B-0030740-QP-1-1). Research Triangle Park, NC.

UNEP (United Nations Environment Programme). (2002). Global mercury assessment. Geneva, Switzerland: UNEP Chemicals. https://www.unenvironment.org/explore-topics/chemicals-waste/what-we-do/mercury/global-mercury-assessment

Wells, EM; Kopylev, L; Nachman, R; Radke, EG; Segal, D. (2020). Seafood, wine, rice, vegetables, and other food items associated with mercury biomarkers among seafood and non-seafood consumers: NHANES 2011-2012. J Expo Sci Environ Epidemiol. http://dx.doi.org/10.1038/s41370-020-0206-6

WHO (World Health Organization). (1990). Methylmercury.

Yaginuma-Sakurai, K; Murata, K; Iwai-Shimada, M; Nakai, K; Kurokawa, N; Tatsuta, N; Satoh, H. (2012). Hair-to-blood ratio and biological half-life of mercury: experimental study of methylmercury exposure through fish consumption in humans. J Toxicol Sci 37: 123-130. http://dx.doi.org/10.2131/jts.37.123

Yang, JM; Jiang, ZZ; Wang, YL; Qureshi, IA; Wu, XD. (1997). Maternal-fetal transfer of metallic mercury via the placenta and milk. Ann Clin Lab Sci 27: 135-141.

Young, JF; Wosilait, WD; Luecke, RH. (2001). Analysis of methylmercury disposition in humans utilizing a PBPK model and animal pharmacokinetic data. J Toxicol Environ Health A 63: 19-52. http://dx.doi.org/10.1080/152873901750128344