EPA/635/R-20/239
IRIS Assessment Protocol
www.epa.gov/iris

Systematic Review Protocol for the Inorganic Mercury Salts IRIS Assessment

CASRN 7487-94-7 (Mercuric Chloride)
CASRN 1344-48-5 (Mercuric Sulfide)
CASRN 10112-91-1 (Mercurous Chloride)

February 2021

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 preliminary draft for review purposes only. This information is distributed solely for review purposes under applicable information quality guidelines. It has not been formally disseminated by the Environmental Protection Agency. It does not represent and should not be construed to represent any Agency determination or policy. It is being circulated for comments on its technical clarity and science policy implications. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

ACToR	Aggregated Computational Toxicology Resource
ADME	absorption, distribution, metabolism, and excretion
ADJ	adjusted
ALT	alanine aminotransferase
AST	aspartate transaminase
ATSDR	Agency for Toxic Substances and Disease Registry
BMDS	Benchmark Dose Software
BMI	body mass index
BMR	benchmark response
BW	body weight
CASRN	Chemical Abstracts Service registry number
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CICAD	Concise International Chemical Assessment Documents
CMA	Continuous modeling average
CPAD	Chemical and Pollutant Assessment Division
CPHEA	Center for Public Health and Environmental Assessment
CTDB	Comparative Toxicogenomics Database
CTL	cytotoxic T lymphocyte
Cu	copper
DEC	Department of Environmental Conservation
DEP	Division of Environmental Protection
DMA	dichotomous modeling average
DNA	deoxyribonucleic acid
DNT	developmental neurotoxicity
DTH	delayed-type hypersensitivity
DWEL	drinking water equivalent level
ECHA	European Chemicals Agency
EGLE	Environment, Great Lakes, and Energy
EPA	Environmental Protection Agency
EPCRA	Emergency Planning and Community Right-to-Know Act
EU JRC	European Union Joint Research Centre
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FOB	functional operational battery
GEO	Gene Expression Omnibus
GI	gastrointestinal
GLP	good laboratory practices
GRADE	Grading of Recommendations Assessment, Development, and Evaluation
HAWC	Health Assessment Workplace Collaborative
HED	human equivalent dose
HEEAD	Health and Environmental Effects Assessment Division
HERO	Health and Environmental Research Online
Hg	mercury
$HgCl_2$	mercuric chloride
Hg_2Cl_2	mercurous chloride
HgS	mercuric sulfide
HPV	High Production Volume
HSDB	Hazardous Substances Data Bank
IAP	IRIS Assessment Plan
IARC	International Agency for Research on Cancer
IRIS	Integrated Risk Information System
LD_{50}	median lethal dose
LOAEL	lowest-observed-adverse-effect level
MeSH	medical subject heading
MOA	mode of action
MLE	maximum likelihood estimation
MLR	mixed leukocyte reaction
MRL	minimal risk level
NAS	National Academy of Sciences
NLM	National Library of Medicine
NMD	normalized mean difference
NOAEL	no-observed-adverse-effect level
NTP	National Toxicology Program
OECD	Organization for Economic Co-operation and Development
OEHHA	Office of Environmental Health Hazard Assessment
OLEM	Office of Land and Emergency Management
OR	odds ratio
ORD	Office of Research and Development
OSF	oral slope factor
OW	Office of Water
PBPK	physiologically based pharmacokinetic
PECO	populations, exposures, comparators, and outcomes
PHG	Public Health Goal
PK	pharmacokinetic
POD	point of departure
PPRTV	Provisional Peer-Reviewed Toxicity Value
RCRA	Resource Conservation Recovery Act
REACH	Registration, Evaluation, Authorization and Restriction of Chemicals
REL	reference exposure level

RfC	inhalation reference concentration
RfD	oral reference dose
RIVM	Dutch National Institute for Public Health and the Environment
ROBINS-I	Risk of Bias in Non-Randomized Studies of Interventions
ROS	reactive oxygen species
Se	selenium
SIDS	Screening Information Dataset
SWIFT	Sciome Workbench for Interactive Computer-Facilitated Text-mining
TDI	tolerable daily intake
TEST	Toxicity Estimation Software Tool
TK	toxicokinetic
TRI	Toxics Release Inventory
TSCA	Toxic Substances Control Act
UF	uncertainty factor
WHO	World Health Organization
WOE	weight of evidence
WOS	Web of Science
Zn	zinc

AUTHORS|CONTRIBUTORS|REVIEWERS

Assessment Team (U.S. EPA/Center for Public Health and Environmental Assessment [CPHEA])

Nagu Keshava	Assessment Manager
Amanda S. Persad	Team Member
Suryanarayana Vulimiri	Team Member
Yu-Sheng Lin	Team Member

Contributors (U.S. EPA/CPHEA)

George Woodall

Santhini Ramasamy

Brittany Schulz (Oak Ridge Associated Universities [ORAU] contractor)

Carolyn Gigot (former ORAU contractor)

Andrew Greenhalgh (former ORAU contractor)

Executive Direction (U.S. EPA/CPHEA)

Wayne Cascio, M.D.	CPHEA Director
Samantha Jones, Ph.D.	CPHEA Associate Director
Kristina Thayer, Ph.D.	Chemical and Pollutant Assessment Division (CPAD) Director
Andrew Kraft, Ph.D.	CPAD Senior Science Advisor
Janice Lee, Ph.D.	CPAD Toxic Effects Assessment Branch – NC Branch Chief
Viktor Morozov, Ph.D.	CPAD Quantitative Assessment Branch – Branch Chief
John Vandenberg, Ph.D.	Health and Environmental Effects Assessment Division (HEEAD) Director
Andrew Hotchkiss, Ph.D.	HEEAD Hazardous Pollutants Assessment and Systems Branch Chief

Production Team and Review (U.S. EPA/CPHEA)

Vicki Soto

Dahnish Shams

Ryan Jones

Maureen Johnson

Erin Vining (ORAU contractor)

1. INTRODUCTION

A draft assessment plan for oral exposures to inorganic mercury salts (mercuric chloride, $HgCl_2$; mercuric sulfide, HgS, also called cinnabar; and mercurous chloride, Hg_2Cl_2, also called calomel) was presented at a public science meeting on December 5, 2019 (https://www.epa.gov/iris/iris-public-science-meeting-dec-2019). 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 assessment is being developed at the request of EPA's Office of Land and Emergency Management (OLEM, see section 1.2.1 for further details), although the assessment may also be used to support actions in other EPA programs and regions and can inform efforts to address inorganic mercury salts by tribes, states, and international health agencies.

This protocol presents methods for conducting the systematic reviews and dose-response analyses for the inorganic mercury salts (mercuric chloride, mercuric sulfide, and mercurous chloride) assessment, including any adjustments made to the specific aims and draft PECO criteria for the assessment in response to public input on the assessment plan. Whereas the IRIS Assessment Plan (IAP) describes what the assessment will cover, chemical-specific protocols describe how the assessments will be conducted (Figure 1-1). This assessment protocol will be posted on the IRIS website (https://cfpub.epa.gov/ncea/iris2/atoz.cfm) and released for a 30-day public comment period. Public input received on the protocol will be considered during preparation of the draft assessment, and any adjustments made to the protocol will be reflected in an updated version released in conjunction with the draft assessment. The preliminary literature search results for these three inorganic mercury salts are stored in the Health and Environmental Research Online (HERO) database (https://heronet.epa.gov/heronet/index.cfm/project/page/project_id/2357)[1] under inorganic mercury salts project page[2] and available upon public release of the protocol. Furthermore, the literature search results will be regularly updated during draft development and the subsequent stages of assessment review.

[1]EPA's HERO database provides access to the scientific literature behind EPA science assessments. The database includes more than 3 million scientific references and data from the peer-reviewed literature EPA uses to develop its health assessment documents.
[2]Inorganic mercury salts: https://heronet.epa.gov/heronet/index.cfm/project/page/project_id/2697.

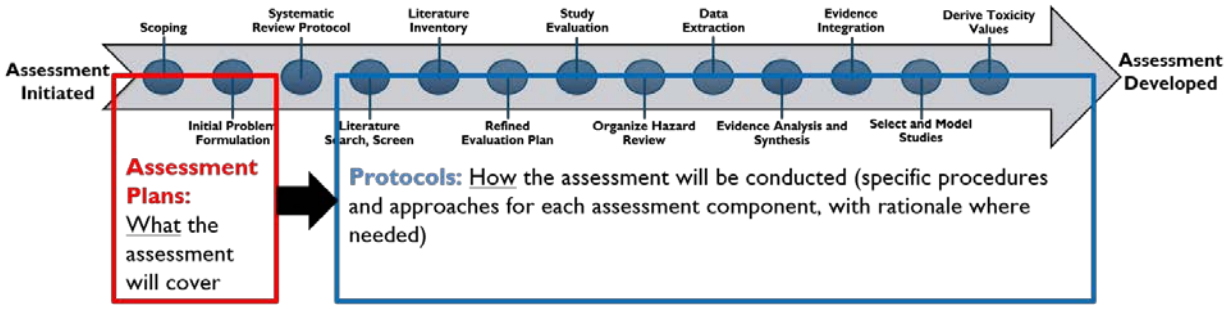


Figure 1-1. IRIS systematic review problem formulation and method documents.

1.1. BACKGROUND INFORMATION ON INORGANIC MERCURY SALTS

1.1.1. Physical and Chemical Properties

Mercury occurs naturally in the environment and can exist as elemental, organic, or inorganic mercury. The IRIS assessment will evaluate the potential human health effects of oral exposures to the three most commonly occurring inorganic mercury salts: mercuric chloride ($HgCl_2$; CASRN 7487-94-7), mercuric sulfide (HgS; CASRN 1344-48-5), and mercurous chloride (Hg_2Cl_2, CASRN 10112-91-1). The chemical structures of mercuric chloride, mercuric sulfide, and mercurous chloride are presented in Figure 1-2 (along with their CASRNs), and their physical and chemical properties are provided in Table 1-1.

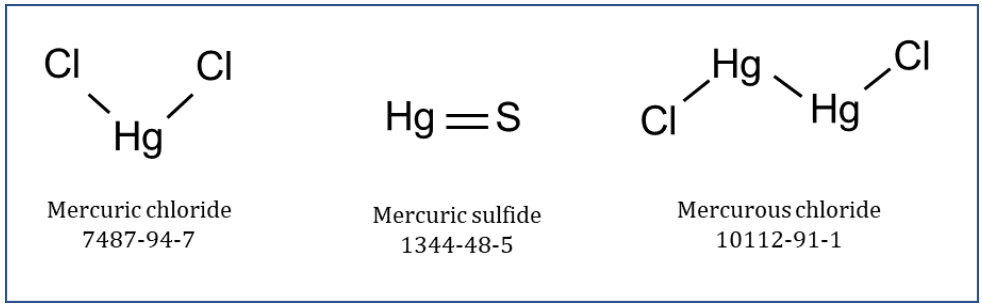


Figure 1-2. Chemical structures of three inorganic mercury salts.

Table 1-1. Physical and chemical properties of inorganic mercury salts

Property (unit)	Inorganic mercury salts		
	Mercuric chloride	**Mercuric sulfide (Cinnabar)**	**Mercurous chloride (Calomel)**
CASRN	7487-94-7[a]	1344-48-5[b]	10112-91-1[c]
DTXSID	5020811[a]	0047747[b]	6044351[c]
Synonyms	Mercury (II) chloride, mercury dichloride[a]	Mercury (II) sulfide, vermilion, (sulfido)mercury[b]	Mercury (I) chloride, dimercury dichloride, mercury subchloride, mercury monochloride[c]
Molecular formula	$HgCl_2$[d]	HgS[b]	Hg_2Cl_2[d]
Molecular wt. (g/mol)	271.49[a]	232.65[b]	472.08[c]
Oxidation state	+2[a]	+2[b]	+1[e]
Melting pt. (°C)	280[a]	583.5[f]	302[d]
Boiling pt. (°C)	302[a]	--	384[d]
Density (g/cm^3)	5.4 at 25°C[d]	8.1[f]	7.15[g]
Vapor pressure (mmHg)	1 at 136.2°C[d]	--	--
Henry's law constant (atm-dm^3/mol)	5×10^{-7} at 293 K[h]	--	--
Water solubility (mol/L)	0.25 at 20°C[i]	Insoluble[d]	--
pKa	--	--	--
LogP	0.22[h]	--	--
Bioconcentration factor	1.03×10^{4}[a]	--	--
Biotransformation	Hg^{2+} to Hg^0	HgS to Hg^{2+} and Hg^{2+} to Hg^0	Hg^{+} to Hg^2

[a]CASRN: 7487-94-7. (U.S. EPA, 2019b) Chemicals Dashboard (https://comptox.epa.gov/dashboard, search = mercuric chloride, accessed April 1, 2020). Values are experimental averages.
[b]CASRN: 1344-48-5. (U.S. EPA, 2019b) Chemicals Dashboard (search = mercuric sulfide, accessed April 1, 2020). Values are experimental averages.
[c]CASRN: 10112-91-1. (U.S. EPA, 2019b) Chemicals Dashboard (search = mercurous chloride, accessed April 1, 2020). Values are experimental averages.
[d](ATSDR, 1999).
[e](Dantith, 2008).
[f](Spectrum, 2016).
[g]Mercurous Chloride. (PubChem, 2018) PubChem Compound Summary (accessed April 1, 2020).
[h](Sommar et al., 2000).
[i]Mercuric Chloride. (PubChem, 2018) PubChem Compound Summary (https://pubchem.ncbi.nlm.nih.gov/, accessed April 1, 2020). Water solubility was converted from g/L to mol/L by dividing by the molar mass.
-- = Not reported. Predicted values might be available from the Environmental Protection Agency (EPA) Chemicals Dashboard.

1.1.2. Sources, Production, and Use

Mercury occurs naturally in geologic materials in the environment and can exist in inorganic form as salts. It can also exist in an elemental form as a liquid or gas or in its organic form (methylmercury), which the IRIS Program is also assessing (https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=73#tab-2). In its inorganic form, mercury occurs abundantly in the environment, primarily as the minerals cinnabar

(HgS) and metacinnabar (cubic form of HgS) and as impurities in other minerals (USGS, 1970). Mercury occurs in Earth's crust at levels averaging 0.5ppm, but the actual concentration varies considerably depending on location (Sidle, 1993; Budavari, 1989). Its geologic associations are with volcanic rocks and hydrothermal systems, where it can readily combine with chlorine, sulfur, and other elements and subsequently weather to form inorganic salts.

Although the use of inorganic mercury salts in many consumer products, such as medicinal products, has been phased out, ammoniated mercuric chloride is still widely used in skin lightening soaps and creams (Park and Zheng, 2012; ATSDR, 1999). Mercuric chloride also is used in photography and as a topical antiseptic and disinfectant, wood preservative, and fungicide. Mercuric sulfide (cinnabar) is used to color paints and is one of the red coloring agents used in tattoo dyes (ATSDR, 1999). Mercurous chloride was used widely in medicinal products, including laxatives, worming medications, and teething powders, but has since been replaced by agents considered to be safer and more effective (ATSDR, 1999).

Historically, mercury (in its various forms, including as inorganic mercury salts) is mined using open pit (10% of production) and underground mining techniques (90%) (Drake, 1981). In 1998, 34 facilities were producing or processing mercury in the United States (U.S. EPA, 1998a). Approximately 58 metric tons of mercury were produced as a by-product from eight mines in 1991, and 64 metric tons were produced as a by-product from nine mines in 1992. Although mercury mines generate most of the world production of mercury, the majority of the mercury produced in the United States comes from secondary production sources (recycling) (U.S. EPA, 1997). It should be noted that environmental concerns have led to numerous regulations that have dramatically decreased the production and use of mercury in the United States. For example, mercury use in the chloralkali sector has declined by 98 percent from 136 metric tons in 1996 to about 0.3 metric tons in 2010 because of increased processing and recycling efficiencies, plant closures, or conversion to other technologies (Wilburn, 2013).

1.1.3. Environmental Fate and Transport

Inorganic mercury salts can be transported in water and occur in soil. Dust containing these salts can enter the air from mining deposits of ores that contain mercury. Emissions of inorganic mercury can occur from coal-fired power plants and the burning of municipal and medical waste, and from factories that use mercury. Inorganic mercury also can enter water or soil from the weathering of rocks that contain inorganic mercury salts and from factories or water treatment facilities that release water contaminated with mercury (ATSDR, 1999).

1.1.4. Environmental Concentrations

Inorganic mercury salts have not been evaluated under the Environmental Protection Agency's (EPA) National Air Toxics Assessment (https://www.epa.gov/national-air-toxics-assessment). The concentration of total mercury in ambient air in the United States in 2014 has been reported to be ~1.7 ng/m^3 (https://www.epa.gov/national-air-toxics-assessment/2014-nata-

assessment-results#emissions). Substantially higher levels (10,000–15,000 ng/m^3) have been detected in ambient air near mercury mines, refineries, and agricultural fields with fungicides containing mercury (WHO, 2003).

1.1.5. Potential for Human Exposure

Human exposure to inorganic mercury salts can occur in both general and occupational settings (ATSDR, 1999). The general population can be exposed to inorganic mercury salts through contact with mercury-containing products and soil, exposure from dental amalgam fillings (Kingman et al., 1998), ingestion of mercury-contaminated food or drinking water, or inhalation of ambient air near mercury refineries, mines, and industrial plants. Humans could also be exposed through application of inorganic mercury-containing skin lightening creams and soaps, topical antiseptics, and disinfectants (Mckelvey et al., 2011; WHO/UNEP, 2008). Another, less well-documented source of exposure to inorganic mercury salts among the general population is from their use in ethnic religious, magical, and ritualistic practices and in herbal remedies (WHO, 2003).

Occupational exposure can occur in gold and silver mining, electrical equipment manufacturing, chemical and metal processing in which mercury is used, construction where building parts contain mercury (such as electrical switches, thermometers), and dental offices. Generally, trends in environmental concentrations and human exposure/intake have been decreasing; thus, estimates in older citations are likely to be overestimates.

1.1.6. Populations and Lifestages with Potentially Greater Exposures

Populations and lifestages that might experience exposures greater than those of the general population include individuals in occupations requiring frequent contact with inorganic mercury salts. Workplace environments presenting the largest potential sources of occupational exposure to mercury include chloralkali production facilities, cinnabar mining and processing operations, and industrial facilities involved in the manufacture or use of instruments containing liquid mercury, although exposure has significantly decreased in recent years due to numerous regulations and a dramatic decrease in production and use of mercury in the United States (Wilburn, 2013).

1.1.7. Previous Assessments of Inorganic Mercury Salts by the Environmental Protection Agency (EPA) and Other Health Agencies

An overview of existing oral values for inorganic mercury salts from different state and federal agencies and international bodies is provided in Figure 1-3 and Table 1-2. The endpoints and the basis for derivation of the oral toxicity values are included in Table 1-2. EPA published an IRIS health effects assessment for mercuric chloride in 1995 (U.S. EPA, 1995) (https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=692), which included a reference dose (RfD) for lifetime oral exposure and an assessment of carcinogenicity. EPA derived

an RfD value of 3×10^{-4} mg/kg-day for mercuric chloride on the basis of autoimmune effects (autoimmune glomerulonephritis) in Brown Norway rats in subchronic-duration feeding and subcutaneous studies (Andres, 1984; Bernaudin et al., 1981; Druet et al., 1978). Based on the qualitative weight-of-evidence characterization, mercuric chloride was classified as group C, a possible human carcinogen; resulting from limited evidence of carcinogenicity in rats and mice, and no available human data. No quantitative cancer values were derived for either oral or inhalation exposures. EPA has not published health effects assessments for mercuric sulfide or mercurous chloride. EPA attempted to derive a provisional RfD value for mercuric sulfide in 2002, but was not successful due to the lack of human data and inadequate subchronic or chronic oral data in animals (https://cfpub.epa.gov/ncea/pprtv/documents/MercuricSulfide.pdf). EPA has derived no inhalation toxicity values (RfCs) for any inorganic mercury salt (mercuric chloride, mercuric sulfide, or mercurous chloride).

EPA identified the Agency for Toxic Substances and Disease Registry's (ATSDR) *Toxicological Profile for Mercury* (ATSDR, 1999) as the most recent assessment of mercury salts by a federal health agency. The ATSDR toxicological profile includes information on different forms of mercury including metallic mercury (also known as elemental mercury), inorganic mercury, and organic mercury by all routes of exposure. ATSDR (1999) has derived oral minimal risk levels (MRLs[3]) for acute (0.007 mg/kg-day) and intermediate (0.002 mg/kg-day) durations of exposure to individual inorganic mercury salts based on renal effects reported in a 1993 National Toxicology Program (NTP) study of mercuric chloride (NTP, 1993). The World Health Organization adopted the ATSDR toxicity value of 0.002 mg/kg-day based on renal effects in rats (WHO, 2003). The International Agency for Research (IARC) concluded that evidence in experimental animals for the carcinogenicity for mercuric chloride is limited and it is not classifiable as to its carcinogenicity to humans (Group 3) (IARC, 1993).

[3]The ATSDR, in response to congressional mandate under the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), develops comparison values to help identify chemicals that may be of concern at hazardous waste sites. One type of these values is called minimal risk levels (MRLs). An MRL is an estimate of the amount of a chemical a person can eat, drink, or breathe each day without a detectable risk to health. MRLs are developed for health effects other than cancer.

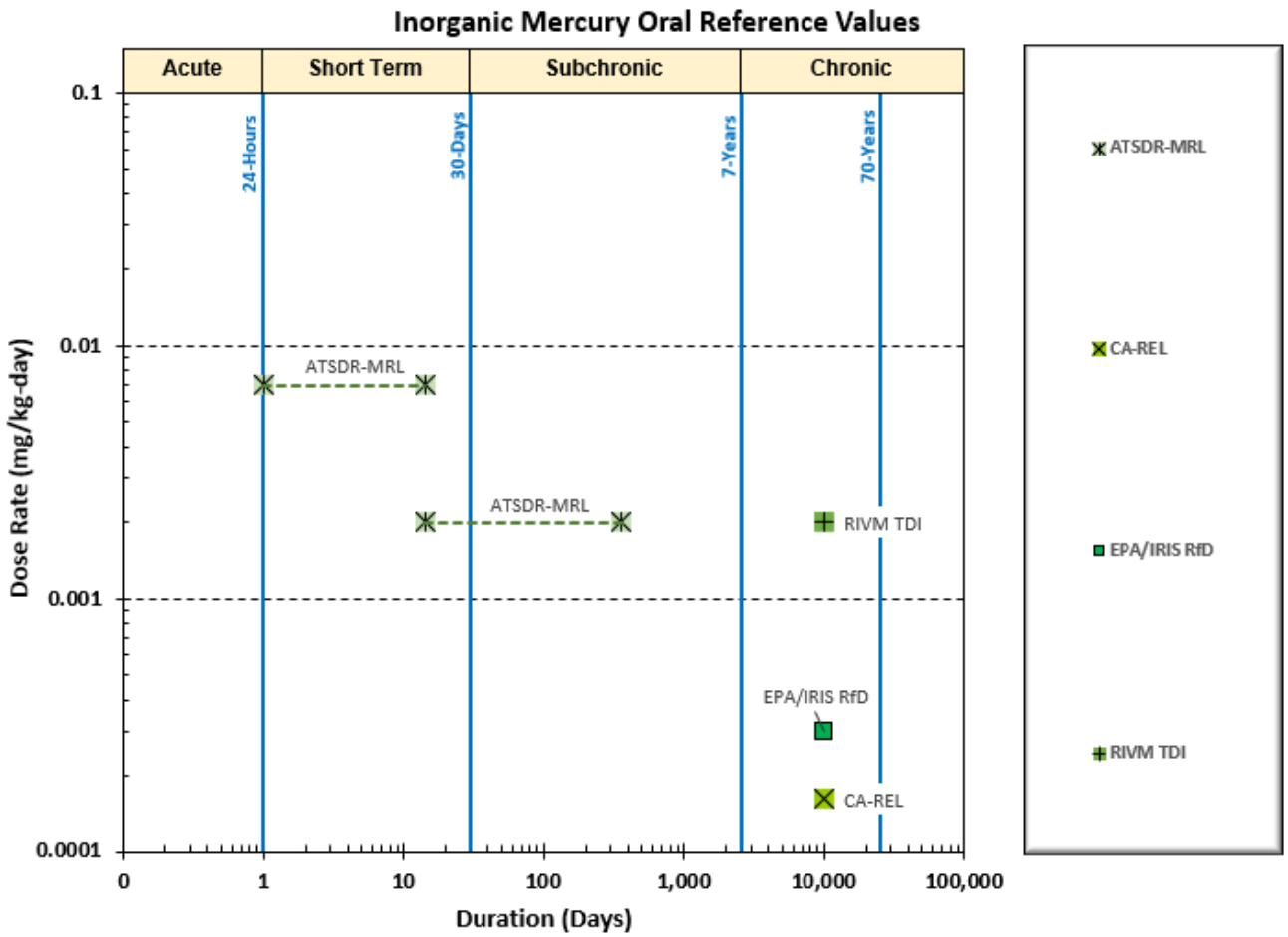


Figure 1-3. Available human health values for oral exposure to inorganic mercury.

ATSDR = Agency for Toxic Substances and Disease Registry; MRL = minimal risk level; RIVM = *Rijksinstituut voor Volksgezondheid en Milieu,* The Netherlands Institute for Public Health and the Environment; TDI = tolerable daily intake; EPA = U.S. Environmental Protection Agency; IRIS = Integrated Risk Information System; RfD = reference dose; CA = California; REL = reference exposure level.

Table 1-2. Details on derivation of the available health effect values for oral exposure to inorganic mercury salts

Value name	Duration	Compound(s)	Oral value (mg/kg-d)	Health effect	Point of departure	Qualifier	Source	Uncertainty factors	Notes on derivation	Review status
ATSDR MRL[a]	Acute (1–14 d)	Inorganic mercury	0.007	Renal effects in rats exposed to Hg in the form of $HgCl_2$ in water for 14 d	0.93 mg Hg/kg-d 0.66 mg Hg/kg-d	NOAEL $NOAEL_{ADJ}$	(NTP, 1993)	Total UF = 100 UF_A = 10 UF_H = 10	Duration adjusted: 5-d/7-d	Final (ATSDR, 1999)
	Intermediate (15–365 d)		0.002	Renal effects in rats exposed to Hg in the form of $HgCl_2$ in water for 26 wk	0.23 mg Hg/kg-d 0.16 mg Hg/kg-d	NOAEL $NOAEL_{ADJ}$				
CA-REL[b]	Chronic	Inorganic mercury	0.00016	Renal effects in rats exposed to $HgCl_2$	0.23 mg Hg/kg-d 0.16 mg Hg/kg-d	NOAEL $NOAEL_{ADJ}$	(NTP, 1993)	Total UF = 1,000 UF_A = 10 UF_H = 10 UF_S = 10	Duration adjusted: 5-d/7-d POD and UFs adopted from earlier PHG assessment	Final (OEHHA, 2008)
EPA RfD (IRIS)[c]	Chronic	Mercuric chloride	0.0003	Autoimmune glomerulonephritis in rats exposed to $HgCl_2$ (Based on OW DWEL for inorganic mercury)	--	--	(U.S. EPA, 1988a)	--	Based on OW DWEL[d]	Final (U.S. EPA, 1995)
RIVM TDI	Chronic	Inorganic mercury	0.002	Renal effects in lab animals exposed to $HgCl_2$	0.23 mg Hg/kg-d	NOAEL	(NTP, 1993)	Total UF = 100 UF_A = 10 UF_H = 10		Final (RIVM, 2001)

[a]In a 2003 CICAD, based on ATSDR Toxicological Profile, WHO established a TDI with the same value and derivation details as the intermediate MRL (WHO, 2003).

[b]This value has been adopted by New York State DEC (NYSDEC, 2006).

[c]This RfD has been adopted by Michigan EGLE, Nevada DEP, and Health Canada (NDEP, 2017; DEQ, 2015; Health Canada, 2010). A 2002 PPRTV document states that, "based on the limited available pharmacokinetic data for mercuric sulfide, the [IRIS] RfD for mercuric chloride (0.0003 mg/kg-day) could be considered protective for mercuric sulfide" (U.S. EPA, 2002a).

[d]RfD = DWEL × drinking water intake rate ÷ BW = 0.01 mg/L × 2 L/day ÷ 70 kg = 0.0003 mg/kg-day.

ADJ = adjusted; ATSDR = Agency for Toxic Substances and Disease Registry; BW = body weight; CICAD = Concise International Chemical Assessment Document; DEC = Department of Environmental Conservation; DEP = Division of Environmental Protection; DWEL = drinking water equivalent level; EGLE = Environment, Great Lakes & Energy; EPA = Environmental Protection Agency; IRIS = Integrated Risk Information System; MRL = minimal risk level; NOAEL = no-observed-adverse-effect level; NTP = National Toxicology Program; OEHHA = Office of Environmental Health Hazard Assessment; OW = Office of Water; PHG = Public Health Goal; POD = point of departure; PPRTV = Provisional Peer-Reviewed Toxicity Value; REL = reference exposure level; RfD = oral reference dose; RIVM = *Rijksinstituut voor Volksgezondheid en Milieu,* The Netherlands Institute for Public Health and the Environment; TDI = tolerable daily intake; UF_A = animal-to-human variability; UF_H = interhuman variability; UF_S = subchronic-to-chronic adjustment; WHO = World Health Organization.

1.2. SCOPING AND PROBLEM FORMULATION SUMMARY

1.2.1. SCOPING SUMMARY

Table 1-3 summarizes Agency interest for prioritizing the three inorganic mercury salts (mercuric chloride, mercuric sulfide, and mercurous chloride) described in this protocol. Initially, both oral and inhalation routes were identified as priorities; however oral toxicity values were subsequently prioritized on the basis of anticipated human exposure pathways and because very little (or no) evidence is available for the inhalation route of exposure (see Section 4, Appendix C). If, during the assessment development, significant data for the inhalation route of exposure becomes available, inhalation toxicity values may be included in the assessment.

Table 1-3. Environmental Protection Agency (EPA) program and regional office interest in an assessment of inorganic mercury salts

EPA program or regional office	Oral	Inhalation	Statutes/regulations	Anticipated uses/interest
OLEM	✓	✓[a]	CERCLA; EPCRA; RCRA Subtitle I (underground storage tanks)	Toxicological information from inorganic mercury salts may be used to make risk determinations for response actions (e.g., short-term removals, long-term remedial response actions) under CERCLA and RCRA including Subtitle I. For example, 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.

[a]Discussions with OLEM indicated a primary need for oral exposure values. Although there was interest in inhalation exposure values, during the development of the IAP and this protocol, very little (or no) evidence for inhalation route was found, thus inhalation values were deprioritized by OLEM. Dermal exposure information was not indicated as a need.

CERCLA = Comprehensive Environmental Response, Compensation, and Liability Act; EPCRA = Emergency Planning and Community Right-to-Know Act; OLEM = Office of Land and Emergency Management; RCRA = Resource Conservation Recovery Act.

1.2.2. PROBLEM FORMULATION

A preliminary literature survey was conducted to understand the extent of the available evidence for inorganic mercury salts to address the scoping needs. Based on this literature survey, it was determined that there will not be sufficient evidence to conduct an RfC analysis for any of the three mercury salts; therefore, as noted above, the IRIS Program will not evaluate inhalation exposure in the mercury salts assessment. If significant inhalation data become available, an inhalation assessment will be considered. Insufficient evidence is available to develop an RfD for mercuric sulfide and mercurous chloride, hence, an analogue-based, read-across approach will be

considered for analysis. Furthermore, a number of key science issues will be identified based on scientific data.

Further details on both literature inventory and methods and results, see section 3 and 4 and also previously published IAP (https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=346843).

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

The overall objective of the assessment is to identify adverse human health effects and characterize exposure-response relationships potentially resulting from exposure to mercuric chloride, mercuric sulfide, and mercurous chloride to support the development of toxicity values. This assessment will use systematic review methods to evaluate the epidemiological and toxicological literature, including consideration of relevant mechanistic evidence (e.g., to inform key science issues; see Section 6). A preliminary literature search and inventory (described in Sections 3 and 4) was conducted to aid problem formulation by assessing the amount and type of evidence available to conduct the assessment. The assessment methods described in this protocol utilize EPA guidance.[4]

The Agency plans to utilize a standard approach to derive oral toxicity values (RfDs) for mercuric chloride. For the other two salts where the current literature search indicates evidence is lacking, an analogue-based, read-across approach will be considered.

2.1. SPECIFIC AIMS

The aims of this assessment are to:

- Identify epidemiological (i.e., human) and toxicological (i.e., experimental animal) literature reporting the effects of exposure to mercuric chloride, mercuric sulfide, or mercurous chloride, as outlined in the PECO (see Section 2.2, Table 2-1).

- Evaluate mechanistic information (including toxicokinetic evidence) associated with exposure to mercuric chloride, mercuric sulfide, and mercurous chloride, as necessary to inform the interpretation of findings related to potential health effects in studies of humans and animals. The scope of these analyses of mechanistic information will be determined by the complexity and confidence in the phenotypic evidence in humans and animals, the likelihood of the analyses to affect evidence synthesis conclusions for human health (e.g., considering the mechanistic studies available based on literature inventory), and the directness or relevance of the available model systems for understanding potential human

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

health hazards. The mechanistic evaluations will focus primarily on the key science issues identified in Section 6.

- Conduct study evaluations for individual epidemiological and toxicological studies (evaluating reporting quality, risk of bias, and sensitivity) and physiologically based pharmacokinetic (PBPK) models (scientific and technical review). The evaluation of epidemiological studies, when available, will specifically consider, to the extent possible, the likelihood and impact of potential confounding factors for other mercury salts. Studies considered *uninformative* will not be considered further.

- Extract data on relevant health outcomes from epidemiological and toxicological studies of *high*, *medium*, and *low* confidence based on the study evaluations (full data extraction of low confidence studies may not be performed for poorly studied health effects or for health effects on which extensive *medium* and *high* confidence studies exist in the evidence base).

- Synthesize the evidence across studies, assessing similar health outcomes using a narrative approach.

- For each health outcome, express strength of evidence judgments across studies (or subsets of studies) separately for studies of exposed humans and for animal studies. Based on the focused mechanistic analyses specific to each inorganic mercury salt, if any, the mechanistic evidence will be integrated with the available health effects evidence (or lack thereof).

- For each health outcome (or grouping of outcomes), develop an integrated expert judgment across lines of evidence as to whether and to what extent the evidence supports that exposure to inorganic mercury salts has the potential to be hazardous to humans. The judgment will be directly informed by the evidence syntheses and be based on structured review of an adapted set of considerations for causality first introduced by Austin Bradford Hill (Hill, 1965) (see Sections 10 and 11), including consideration (e.g., based on available mechanistic information) and discussion of biological understanding. As part of the evidence integration narrative, characterize the overall strength of evidence for the available database of studies and its uncertainties, and identify and discuss issues concerning potentially susceptible populations and lifestages.

- Derive toxicity values (e.g., oral reference dose [RfD], cancer risk estimates) as supported by the available data. Apply toxicokinetic and dosimetry modeling (if available) to account for interspecies differences, as appropriate. Given the apparent species and sex differences in the toxicokinetic profile of the different inorganic mercury salts, methods to address these potential differences will be a key consideration.

- If no relevant or adequate data are identified for the point of departure (POD) derivation, an analogue-based approach will be attempted using information gathered during the literature search screening process to compile an initial list of candidate analogues and determine the feasibility of conducting a read-across analysis.

- Characterize uncertainties and identify key data gaps and research needs across each inorganic mercury salts database, such as limitations of the available evidence, limitations of the systematic review, and consideration of dose relevance and pharmacokinetic differences when extrapolating findings from higher dose animal studies to lower levels of human exposure.

2.2. 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 literature screening, including study inclusion/exclusion, in a systematic review. The PECO for inorganic mercury salts (see Table 2-1) was based on discussions with scientists in EPA program and regional offices, a review of health assessment documents from other federal agencies such as ATSDR, and consideration of comments on the IAP which was released on October 8, 2019 (https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=346843). To help with problem formulation, a preliminary literature survey using the PECO below was conducted from January 1997 to February 28, 2019.

Studies were included that met the PECO criteria. In addition to those studies meeting the PECO criteria, studies containing supplemental material that are potentially relevant to addressing the specific aims of the assessment were tracked during the literature screening process (Table 2-2). Although these studies did not meet PECO criteria, they were not necessarily excluded from further consideration, as described in Section 3.2.

Table 2-1. Populations, exposures, comparators, and outcomes (PECO) criteria

PECO element	Evidence
Populations	**Human:** Any population and lifestage (occupational or general population, including children and other sensitive populations). **Animal:** Nonhuman mammalian animal species (whole organism) of any life stage (including preconception, in utero, lactation, peripubertal, and adult stages). Nonmammalian models and in vitro studies will be tracked as "potentially relevant supplemental information."
Exposures	**Chemical Forms** • Mercuric chloride (7487-94-7) and all synonyms including mercuric perchloride, mercury bichloride, mercury chloromercurate (II), mercury dichloride, mercury

PECO element	Evidence
	perchloride, mercury (II) chloride, $HgCl_2$, dichloromercury, calochlor, bichloride of mercury • Mercuric sulfide (1344-48-5) and synonyms including cinnabar, mercury (II) sulfide, mercury (II) sulfide black, mercury (II) sulfide red, mercury sulfide, mercury sulphide, vermilion, Chinese red, ethiops mineral, HgS • Mercurous chloride (10112-91-1) and synonyms including calomel, calogreen, chloromercury, dimercury dichloride, mercury (I) chloride, mercury chloride, mercury monochloride, mercury protochloride, mercury subchloride, mild mercury chloride, Hg_2Cl_2 **Human:** Exposure to the relevant forms of inorganic mercury salts listed above, including occupational exposures via oral and inhalation route. Exposure can be based on administered dose or concentration, biomonitoring data (e.g., urine, blood, or other specimens), environmental or occupational-setting measures (e.g., air, water levels), or job title or residence. **Animal:** Exposure to inorganic mercury salts via the oral and inhalation route. Studies involving exposures to mixtures will be included only if they include exposure to inorganic mercury salts alone. Other exposure routes, including dermal, or injection exposures, will be tracked during screening as "potentially relevant supplemental information."
Comparators	**Human:** A comparison or referent population exposed to lower levels (or no exposure/exposure below detection limits) of inorganic mercury salts, or exposure to inorganic mercury salts for shorter periods of time. Worker surveillance studies are considered to meet PECO criteria, however, even if no referent group is presented. Case reports describing findings in 1–3 people in nonoccupational or occupational settings will be tracked as "potentially relevant supplemental information." **Animal:** A concurrent control group exposed to vehicle-only treatment or untreated control (control could be a baseline measurement, e.g., acute toxicity studies of mortality).
Outcomes	All health outcomes (both cancer and noncancer).
Classical Pharma-cokinetic (PK) or Physiologically based Pharmacokinetic (PBPK) model studies	Classical Pharmacokinetic (PK) or Physiologically based Pharmacokinetic (PBPK) model studies **Classical pharmacokinetic or dosimetry model studies:** Classical PK or dosimetry modeling usually divides the body into just one or two compartments, which are not specified by physiology, where movement of a chemical into, between, and out of the compartments is quantified empirically by fitting model parameters to ADME (absorption, distribution, metabolism, and excretion) data. This category is for papers that provide detailed descriptions of PK models, that are not a PBPK model. **Physiologically based pharmacokinetic or mechanistic dosimetry model studies:** PBPK models represent the body as various compartments (e.g., liver, lung, slowly perfused tissue, richly perfused tissue) to quantify the movement of chemicals or particles into and out of the body (compartments) by defined routes of exposure, metabolism, and elimination, and thereby estimate concentrations in blood or target tissues.

Table 2-2. Major categories of "Potentially Relevant Supplemental Material"

Category	Evidence
In vitro, ex vivo, or in silico studies "mechanistic" studies	In vitro, ex vivo, or in silico studies reporting measurements related to a health outcome that inform the biological or chemical events associated with phenotypic effects, in both mammalian and non-mammalian model systems. Studies where the chemical is used as a laboratory reagent generally need not be tagged (e.g., chemical probe used to measure antibody response).
Toxicokinetic (ADME)	Studies designed to capture information regarding absorption, distribution, metabolism, and excretion, including toxicokinetic studies. This category includes studies of bioavailability and solubility because inorganic mercury salts are soluble or insoluble in various media. Such information could be helpful in updating or revising the parameters used in existing PBPK models. Studies describing environmental fate and transport or metabolism in bacteria or model systems not applicable to humans or animals should not be tagged.
Non-mammalian model systems	Studies in nonmammalian model systems (e.g., fish, birds, *C. elegans*).
Exposure characteristics (no health outcome assessment)	Exposure characteristic studies include data unrelated to toxicological endpoints but provide information on exposure sources or measurement properties of the environmental agent (e.g., demonstrate a biomarker of exposure).
Mixture studies	Studies involving exposures to mixtures will be included if the exposure includes exposure to mercuric chloride, mercuric sulfide, or mercurous chloride. This categorization generally does not apply to epidemiological studies where the exposure source might be unclear. Such studies are tracked as PECO relevant when inhalation or oral exposure is plausible.
Routes of exposure not meeting PECO criteria	Studies other than oral and inhalation routes of exposure, (e.g., dermal or injection exposure). This categorization generally does not apply to epidemiological studies where the exposure route might be unclear. Such studies are tracked as PECO relevant when oral exposure is plausible.
Case reports or case series	Case reports describing health outcomes after exposure will be tracked as potentially relevant supplemental information when the number of subjects is ≤3.
Records with no original data	Records that do not contain original data, such as other agency assessments, informative scientific literature reviews, editorials, or commentaries.
Conference abstracts	Records that do not contain sufficient documentation to support study evaluation and data extraction.
Unnamed reports	Records identified that have no citation information (typically identified during gray literature searches).

3. LITERATURE SEARCH AND SCREENING STRATEGIES

The ATSDR *Toxicological Profile for Mercury* (ATSDR, 1999) was selected as the starting point for the literature searches because it is the most recent and comprehensive review of health effects of inorganic mercury salts published by a U.S. federal agency. It should be noted that the World Health Organization (WHO) released a Concise International Chemical Assessment Document (CICAD) on the human health effects of mercury in 2003; however, because WHO adopted the toxicity values from ATSDR, this protocol used the ATSDR toxicological profile as a starting point. All references from the health effects section (Chapter 2) of the 1999 ATSDR *Toxicological Profile for Mercury* were extracted by an EPA information specialist and stored in the Health and Environmental Research Online (HERO) database (https://heronet.epa.gov/heronet/index.cfm/project/page/project_id/2357).

To identify records that had been published since the release of the 1999 ATSDR *Toxicological Profile for Mercury,* database searches were conducted with a start date of January 1997 by an EPA information specialist in three online databases (PubMed, Web of Science, Toxline). The initial literature search was completed in February 2019, and search results are available within HERO project page on inorganic mercury salts (https://heronet.epa.gov/heronet/index.cfm/project/page/project_id/2697).

The literature searches will be updated throughout the development of the assessment and review process to identify newly published literature. Accordingly, the methods for literature search and screening (and some of the approaches to refining the evaluation plan on the basis of identified literature; see Section 4) are described in the protocol using past tense, whereas approaches for the other assessment methods are outlined using future tense. The last full literature search update will be conducted prior to (several months) the planned release of the draft document for public comment. The literature flow diagrams (see Section 3.2.2) will be updated on the basis of the new search results. During assessment development and review, additional literature searches might be performed (e.g., to supplement an analysis of a specific biological mechanism). Any such ancillary searches will be documented in updates to the protocol. Studies identified after peer review begins will be considered for inclusion only if they are directly relevant to the PECO criteria and are expected to alter the assessment's conclusions fundamentally. 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 would be screened as outlined above for adherence to the PECO criteria.

The methods described in this protocol were implemented in accordance with EPA quality assurance policies and procedures (Quality Policy Procedures[5] and CIO 2105.0[6], formerly known as 5360.1 A2).

3.1. LITERATURE SEARCH STRATEGIES

3.1.1. Database Search Term Development

The literature search to identify studies published since the ATSDR *Toxicological Profile for Mercury* focused only on the chemical name (and synonyms or trade names) with no additional limits. Chemical synonyms were identified by first using the "Find Chemical Synonyms" feature in SWIFT Review (Howard et al., 2016). In brief, this feature automatically creates a PubMed-formatted chemical search using (1) the common name for the chemical as presented in the Tox21 chemical inventory list; (2) the Chemical Abstract Services Registry Number (CASRN); (3) synonyms from the ChemIDPlus database, which currently contains chemical names and synonyms for over 400,000 chemicals; and (4) removal of ambiguous or short alphanumeric terms that could lead to false positives. This search was manually reviewed to ensure any synonyms listed in EPA's Chemicals Dashboard (U.S. EPA, 2019b) as "valid" or "good" were included. The PubMed search created from SWIFT Review then was modified as needed for use on other databases (Appendix A).

3.1.2. Database Searches

The databases listed below were searched by an EPA information specialist with no language restrictions applied. Searches included studies from 1997 through February 28, 2019 to overlap at least 2 years with the publication of the ATSDR *Toxicological Profile for Mercury* to help ensure no studies were missed. Full details of the search strategy for each database are presented in Appendix A. All records were stored in the HERO database.

- PubMed (National Library of Medicine)
- Web of Science (WoS, Thomson Reuters)
- Toxline (National Library of Medicine)

Because the number of studies retrieved during the initial literature search was relatively large (~4,000–6,000) for each chemical (Figure 3-1), the studies were imported into SWIFT Review software (https://www.sciome.com/swift-review/; see also (Howard et al., 2016) to identify those

[5]U.S. Environmental Protection Agency Procedures for Quality Policy: https://www.epa.gov/sites/production/files/2015-10/documents/21060.pdf.
[6]Policy and Program Requirements for the Mandatory Agency-Wide Quality System: https://www.epa.gov/sites/production/files/2015-09/documents/epa_order_cio_21050.pdf.

most likely to be applicable to human health assessment. In brief, SWIFT Review has preset literature search filters developed by information specialists that can be applied to separate studies more likely to be useful for identifying human health content than those that likely are not (e.g., analytical methods). The filters function like a typical search strategy where studies are tagged as belonging to a certain filter if the terms in the filter literature search strategy appear in title, abstract, keyword or *m*edical *subject headings (MeSH*) fields content. The SWIFT Review filters applied focused on lines of evidence (human, animal, *in vitro*). The details of the search strategies underlying the filters are available at https://hawcprd.epa.gov/media/attachment/SWIFT-Review_Search_Strategies.pdf. Studies not retrieved using these filters were not considered further in screening. Studies that included one or more of the search terms in the title, abstract, keyword, or *MeSH* fields were exported as a RIS file for screening in DistillerSR[7] as described in section 3.2.1.

3.1.3. Other Sources Consulted

The literature search strategies are designed to be broad, but like any search strategy, studies could be missed (e.g., cases where the specific chemical is not mentioned in the title, abstract, or keywords; ability to capture "gray" literature not indexed in the databases listed above). Thus, in addition to the database searches, the sources below were or will be used to identify studies that could have been missed based on the database search. Records appearing to meet the PECO criteria will be uploaded into DistillerSR, annotated with source of the record, and screened using the methods described in Section 3.2. Appendix B describes the specific methods for searching each source.

- For information published before 1997, the ATSDR *Toxicological Profile for Mercury* (ATSDR, 1999) will be used as the primary resource for identifying pertinent studies. Specifically, studies cited in the Health Effects chapter (Chapter 2) of the ATSDR document will be screened against the PECO using the methods described in Section 3.2.

- Manual review of the reference lists from final or publicly available draft assessments (e.g., EPA's Provisional Peer-Reviewed Toxicity Values [PPRTV]) or a published journal review specifically focused on human health.

- Manual review of the reference lists of studies screened as PECO-relevant after full-text review for potentially relevant studies.

[7]DistillerSR is a web-based systematic review software used to screen studies available at https://www.evidencepartners.com/products/distillersr-systematic-review-software.

- Retrieving references from the EPA Chemicals Dashboard ToxVal database (U.S. EPA, 2019b)[8] to identify studies or assessments that present point-of-departure (POD) information. In brief, ToxValDB collates publicly available summary values on toxicity dose effects typically used in chemical assessments. Many of the PODs presented in ToxValDB are based on gray literature studies or on assessments not available in databases such as PubMed, WoS, etc. Although many of the resources included in the "Other Sources Consulted" list are represented in ToxValDB, they are also manually searched because most of the ToxValDB entries have not undergone quality assurance to ensure accuracy or completeness and might not include recent studies.

- European Chemicals Agency (ECHA) registration dossiers to identify data submitted by registrants (http://echa.europa.eu/information-on-chemicals/information-from-existing-substances-regulation).

- EPA ChemView database (U.S. EPA, 2019a) to identify unpublished studies, information submitted to EPA under Toxic Substances Control Act (TSCA) section 4 (chemical testing results); section 8(d) (health and safety studies); section 8(e) (substantial risk of injury to health or the environment notices); and FYI (voluntary documents). Other databases accessible via ChemView include EPA's High Production Volume (HPV) Challenge database (https://iaspub.epa.gov/oppthpv/public_search.html_page) and the Toxic Release Inventory (TRI) database.

- National Toxicology Program (NTP) database of study results and research projects (https://ntp.niehs.nih.gov/data/index.html).

- The Organization for Economic Cooperation and Development (OECD) Screening Information DataSet (SIDS) High Production Volume Chemicals (https://www.echemportal.org/echemportal/substancesearch/page.action?pageID=9)

- Review the list of references in the ECOTOX database for the chemical(s) of interest.

- EPA CompTox Chemical Dashboard (U.S. EPA, 2019b) to retrieve a summary of any ToxCast or Tox21 high throughput screening information. These data can be used to generate mechanistic insight, predict outcomes using appropriate models, and potentially inform dose-response modeling. Their importance for outcome prediction and dose-response modeling depends on the context, size, and quality of retrieved results and the lack of availability of other data typically used for these purposes.

[8]ToxValDB is a database designed to store a wide range of public toxicity information while maintaining the linkages to original source information so that users can access available details. In particular, ToxValDB collates publicly available summary values on toxicity dose-effect related typically used in risk assessments. These include point of departure (POD) data collected from data sources within ACToR and ToxRefDB, and no-observed- and lowest-observed (adverse)-effect levels (NOEL, NOAEL, LOEL, LOAEL) data extracted from repeated dose toxicity studies submitted under REACH. Also included are reference dose and concentration values (RfDs and RfCs) from EPA's Integrated Risk Information System (IRIS) and dose descriptors from EPA's Provisional Peer-Reviewed Toxicity Values (PPRTV) documents. Acute toxicity information was extracted from several different sources, including: OECD eChemPortal; ECHA (European Chemicals Agency); NLM (National Library of Medicine) HSDB (Hazardous Substances Data Bank); ChemIDplus via EPA TEST (Toxicity Estimation Software Tool); and the EU JRC (Joint Research Centre) AcutoxBase. Finally, data from the eChemPortal and the EU COSMOS project have also been included in ToxValDB.

- Comparative Toxicogenomics Database (CTD), available at http://ctdbase.org/. The list of genes retrieved from the CTD can be informative, depending on its size and context, and help generate mechanistic inferences or support mechanistic evidence generated by other studies. References that report the chemical-gene interactions retrieved from the CTD can enrich the inventory of mechanistic references.

- Public data from gene expression studies (Gene Expression Omnibus, GEO https://www.ncbi.nlm.nih.gov/geo/ and ArrayExpress https://www.ebi.ac.uk/arrayexpress/). These data can be used to generate mechanistic evidence, predict adverse outcomes and derive toxicity values depending on the design of gene expression studies that produced the data.

- References identified during public comment periods, by technical consultants, during peer review.

3.1.4. Non-Peer-Reviewed Data

IRIS assessments rely mainly on publicly accessible, peer-reviewed studies. Gray literature (i.e., studies not reported in the peer-reviewed literature) directly relevant to the PECO might be identified during assessment development (e.g., Good Laboratory Practice [GLP] studies submitted to EPA, dissertations). In this case, if the data could substantially influence assessment decisions or conclusions (i.e., could affect hazard conclusions or significantly alter dose-response analyses), EPA can obtain external peer-review from experts independent of the Agency if the owners of the data are willing to have the study details and results made publicly accessible. This independent, contractor-driven peer review would include an evaluation of the study, as is done 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 before confirming their service. In most instances, the peer review would be conducted by letter review. The study authors would be informed of the outcome of the peer review and given an opportunity to clarify issues or provide missing details. 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, EPA's IRIS Program will conduct an assessment for utility and data analysis on the basis of having access to a description of study methods and raw data that have undergone rigorous quality assurance/quality control review (e.g., ToxCast/Tox21 data; results of National Toxicology Program [NTP] studies) but that have not yet undergone external peer review.

Unpublished (e.g., raw) data from personal author communication can supplement a peer-reviewed study, provided that information is made publicly available. If such ancillary information is acquired, it will be documented in the Health Assessment Workspace Collaborative

(HAWC)[9] or HERO project page for the inorganic mercury salts being assessed (depending on the nature of the information received).

3.2. SCREENING PROCESS

As described below, PECO criteria were used by two independent reviewers to screen and inventory studies at the title and abstract level and full-text levels.

3.2.1. Application of PECO Criteria

For those studies considered "relevant" or "unclear" at the title and abstract level, the PECO criteria were then used to determine "inclusion" or "exclusion" of a reference based on the full-text review. In addition to including studies that meet PECO criteria (Table 2-1), other studies containing material potentially relevant to the assessments' objectives and specific aims were tracked during the screening process as "potentially relevant supplemental material" (Table 2-2). These studies were not excluded but, sub-tagged into different categories for consideration when deemed relevant for addressing the specific aims, key science issues, or key scientific uncertainties identified at later stages of assessment development. Important to emphasize is that being tagged as supplemental material does not mean the study necessarily would be excluded from consideration in an assessment. The initial screening-level distinctions between a study meeting the PECO criteria and a supplemental study are often made for practical reasons, and the tagging structure in Table 2-2 is designed to ensure the supplemental studies are categorized for easy retrieval while conducting the assessment. Studies that meet the PECO criteria are those most likely to be used to derive toxicity values and thus will undergo subsequent individual study evaluation and data extraction. In contrast, 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 could emerge as being critically important to the assessment and need to be evaluated and summarized at the individual study level (e.g., cancer MOA or ADME studies). These studies might be helpful to provide context (e.g., summarize current levels of exposure, provide hazard evidence from routes or durations of exposure not pertinent to the PECO) or 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. When tagged during title and abstract screening, whether the chemical of interest is reported in the study (i.e., abstracts might not describe all chemicals investigated) might not be completely clear. In such cases, studies are still tagged with the expectation that additional screening would clarify if the study is pertinent.

[9]HAWC is a free and open-source software application that provides a modular, web-based interface to help develop human health assessments of chemicals: https://hawcproject.org/portal/. Standard operating procedures provided to the reviewers to facilitate consistent and relevant documentation of their judgments using the HAWC software can be found as attachments embedded within the online tool (https://hawcprd.epa.gov/assessment/100000039/).

Title and abstract screening. Following a pilot phase to calibrate screening guidance, two screeners independently performed a title and abstract screen using a structured form in DistillerSR (Evidence Partners; https://distillercer.com/products/distillersr-systematic-review-software/). Studies meeting the PECO criteria with clear titles and abstracts were considered for inclusion and advanced to full-text level. For citations with no abstract, the articles were initially screened on the basis of one or both of the following: title relevance (title should indicate clear relevance) and page numbers (articles two pages or less in length were assumed to be conference reports, editorials, or letters). Screening conflicts were resolved by discussion among the primary screeners with consultation by a third reviewer or technical advisor (if needed) to resolve any remaining disagreements. Eligibility status of non-English studies was assessed using the same approach with online translation tools used as needed to evaluate parts of the study text and assess eligibility at the title and abstract levels.

Studies not meeting title/abstract criteria but identified as "potentially relevant supplemental material" were categorized (i.e., tagged) during the title and abstract screening process. Conflict resolution for supplemental material was similar to title/abstract screening. Conflicts between screeners in applying the supplemental tags are resolved by discussion and consultation with a third reviewer (as needed), erring on the side of over-tagging at the title and abstract levels.

Full-text screening. Records not excluded based on the title and abstract were advanced to full-text review. Full-text copies of these potentially relevant records were retrieved, stored in the HERO database, and independently assessed by two screeners using a structured form in DistillerSR to confirm eligibility. Screening conflicts were resolved by discussion among the primary screeners with consultation by a third reviewer or technical advisor (as needed to resolve any remaining disagreements). As with the title and abstract screening, some studies also were identified as "potentially relevant supplemental material" based on full-text screening. Approaches for language translation included engagement of a native speaker from within EPA or use of fee-based translation services.

Literature inventory. Records considered PECO relevant after full-text review were briefly summarized in DistillerSR to create literature inventories for displaying the extent and nature of the available evidence. Further details of literature inventory are described in Section 4.1.

The results of this screening process are documented in the HERO database (https://hero.epa.gov) and literature flow diagrams (see Figure 3-1), with individual studies "tagged" in HERO according to their appropriate category descriptors (e.g., reference source; screening decisions regarding inclusion, exclusion, or identification as supplemental; type of study).

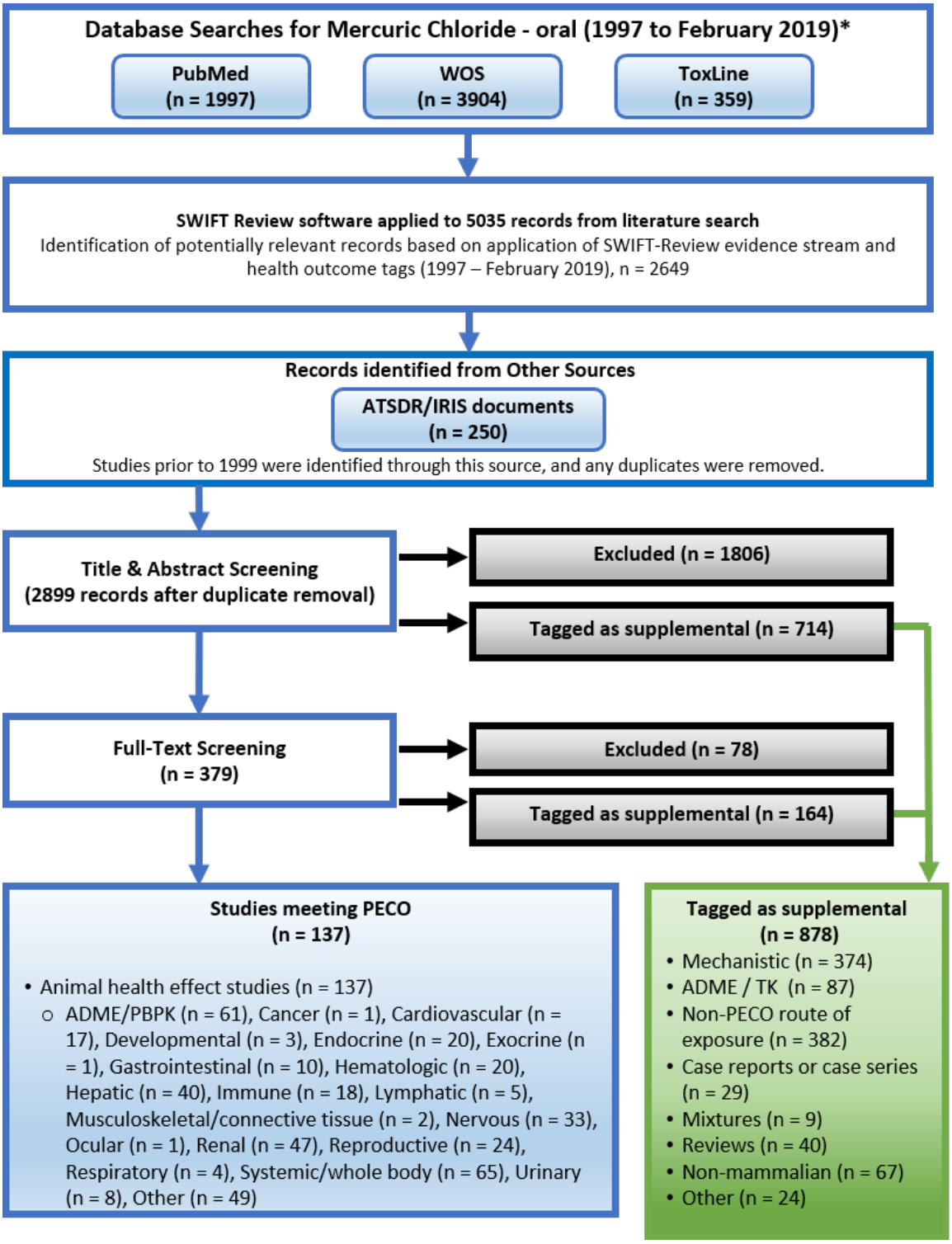
3.2.2. Literature Flow Diagrams

Figure 3-1 presents the literature flow diagrams of mercuric chloride (a), mercuric sulfide (b), and mercurous chloride (c) screened for oral exposures. These figures reflect literature searches from various databases through February 2019 and the screening process, including

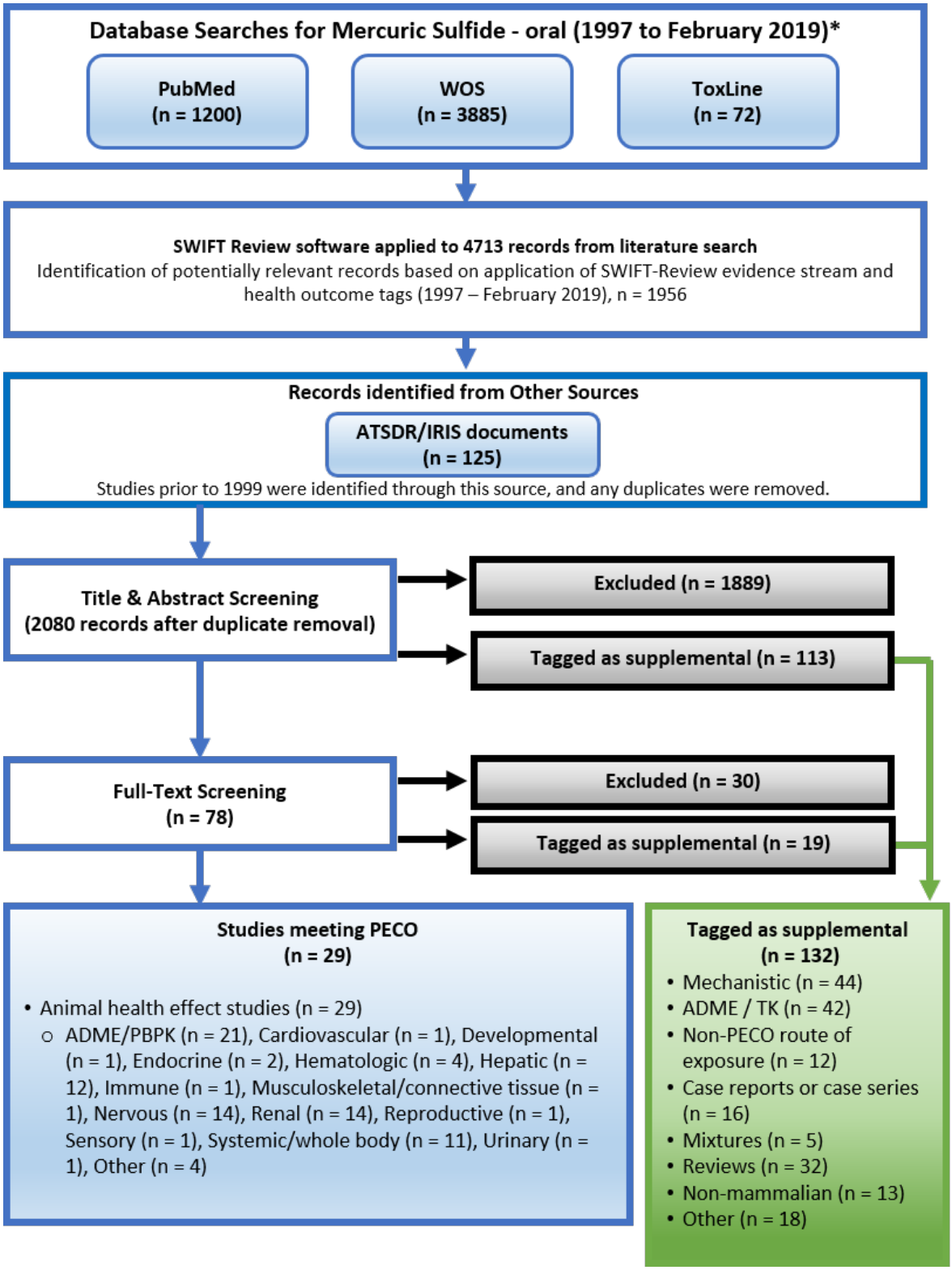
title/abstract screening, full text screening, and literature inventory. In addition, HAWC trees provided in Figure 3-2 include categories such as "inclusion" (meet PECO criteria), "exclusion" (not relevant to PECO criteria or specific aims), and "supplemental materials," and supplemental material subcategories displaying greater detail. The links for the interactive trees is also provided (mercuric chloride: https://hawcprd.epa.gov/lit/assessment/100500164/references/visualization/; mercuric sulfide: https://hawcprd.epa.gov/lit/assessment/100500172/references/visualization/; mercurous chloride: https://hawcprd.epa.gov/lit/assessment/100500173/references/visualization/).

Similar process, title/abstract screening and full text screening was undertaken for the same set of studies for inhalation route of exposure. At the end of the screening process, no relevant studies were identified that met the PECO criteria. The literature flow diagram for inhalation route of exposure for mercuric chloride, mercuric sulfide and mercurous chloride are provided in Appendix C. As noted earlier, due to the absence of information, the mercury salts assessment will not evaluate inhalation exposure.

(a)


Database Searches for Mercuric Chloride - oral (1997 to February 2019)*
PubMed (n = 1997)
WOS (n = 3904)
ToxLine (n = 359)
SWIFT Review software applied to 5035 records from literature search
Identification of potentially relevant records based on application of SWIFT-Review evidence stream and health outcome tags (1997 – February 2019), n = 2649
Records identified from Other Sources
ATSDR/IRIS documents (n = 250)
Studies prior to 1999 were identified through this source, and any duplicates were removed.
Title & Abstract Screening (2899 records after duplicate removal)
Excluded (n = 1806)
Tagged as supplemental (n = 714)
Full-Text Screening (n = 379)
Excluded (n = 78)
Tagged as supplemental (n = 164)
Studies meeting PECO (n = 137)
• Animal health effect studies (n = 137)
○ ADME/PBPK (n = 61), Cancer (n = 1), Cardiovascular (n = 17), Developmental (n = 3), Endocrine (n = 20), Exocrine (n = 1), Gastrointestinal (n = 10), Hematologic (n = 20), Hepatic (n = 40), Immune (n = 18), Lymphatic (n = 5), Musculoskeletal/connective tissue (n = 2), Nervous (n = 33), Ocular (n = 1), Renal (n = 47), Reproductive (n = 24), Respiratory (n = 4), Systemic/whole body (n = 65), Urinary (n = 8), Other (n = 49)
Tagged as supplemental (n = 878)
• Mechanistic (n = 374)
• ADME / TK (n = 87)
• Non-PECO route of exposure (n = 382)
• Case reports or case series (n = 29)
• Mixtures (n = 9)
• Reviews (n = 40)
• Non-mammalian (n = 67)
• Other (n = 24)

(b)


Database Searches for Mercuric Sulfide - oral (1997 to February 2019)*
PubMed (n = 1200)
WOS (n = 3885)
ToxLine (n = 72)
SWIFT Review software applied to 4713 records from literature search
Identification of potentially relevant records based on application of SWIFT-Review evidence stream and health outcome tags (1997 – February 2019), n = 1956
Records identified from Other Sources
ATSDR/IRIS documents (n = 125)
Studies prior to 1999 were identified through this source, and any duplicates were removed.
Title & Abstract Screening (2080 records after duplicate removal)
Excluded (n = 1889)
Tagged as supplemental (n = 113)
Full-Text Screening (n = 78)
Excluded (n = 30)
Tagged as supplemental (n = 19)
Studies meeting PECO (n = 29)
• Animal health effect studies (n = 29)
○ ADME/PBPK (n = 21), Cardiovascular (n = 1), Developmental (n = 1), Endocrine (n = 2), Hematologic (n = 4), Hepatic (n = 12), Immune (n = 1), Musculoskeletal/connective tissue (n = 1), Nervous (n = 14), Renal (n = 14), Reproductive (n = 1), Sensory (n = 1), Systemic/whole body (n = 11), Urinary (n = 1), Other (n = 4)
Tagged as supplemental (n = 132)
• Mechanistic (n = 44)
• ADME / TK (n = 42)
• Non-PECO route of exposure (n = 12)
• Case reports or case series (n = 16)
• Mixtures (n = 5)
• Reviews (n = 32)
• Non-mammalian (n = 13)
• Other (n = 18)

(c)

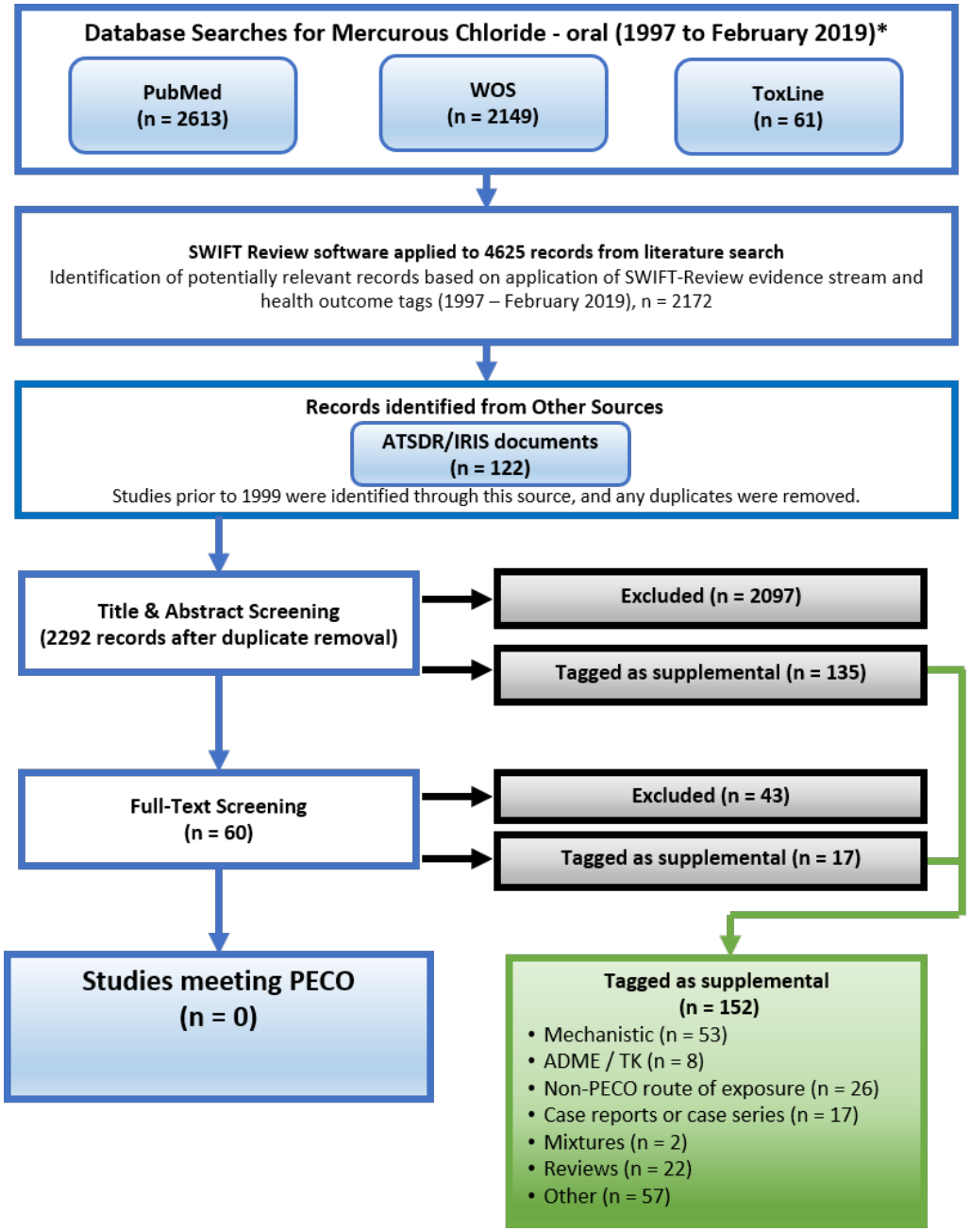
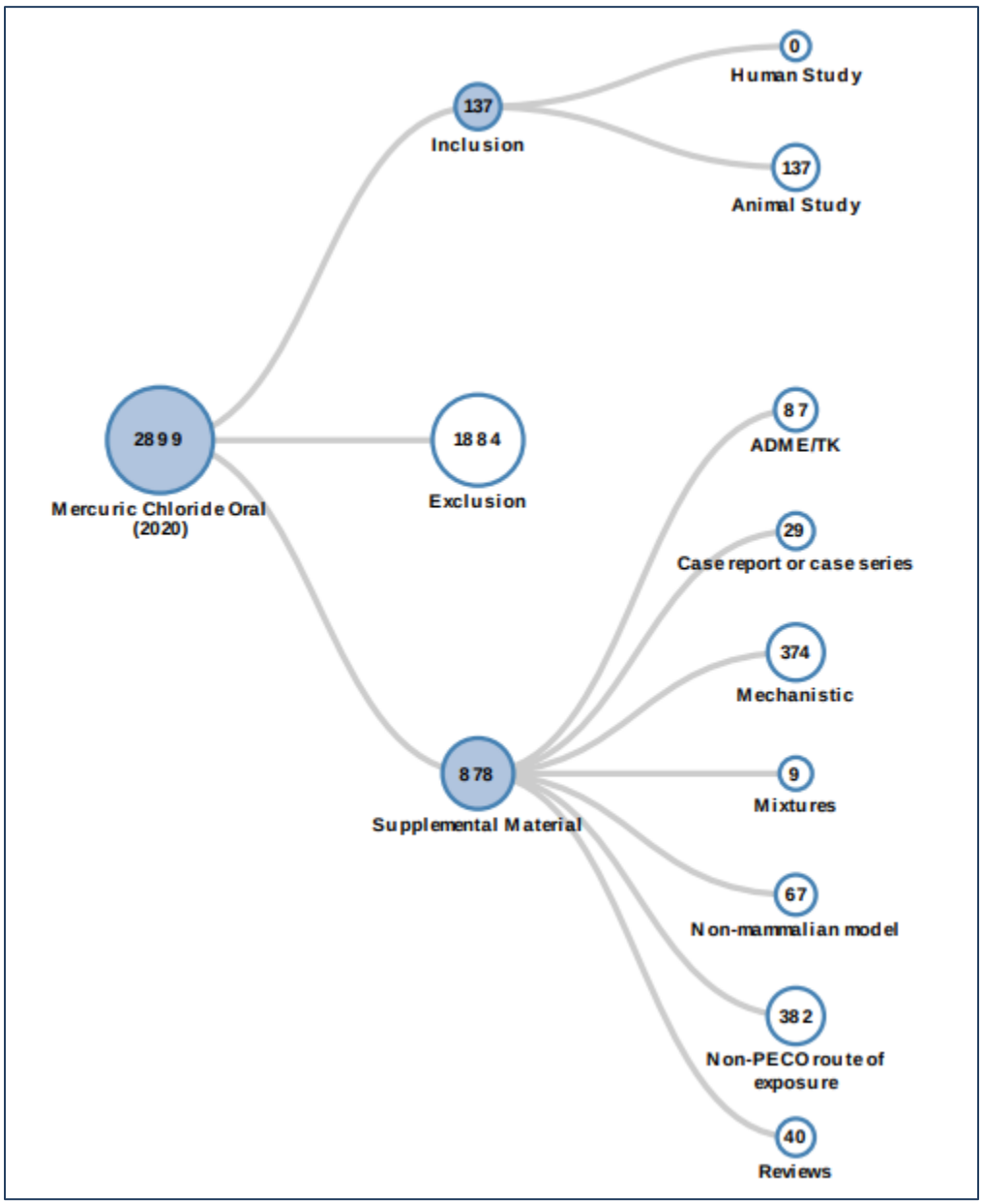
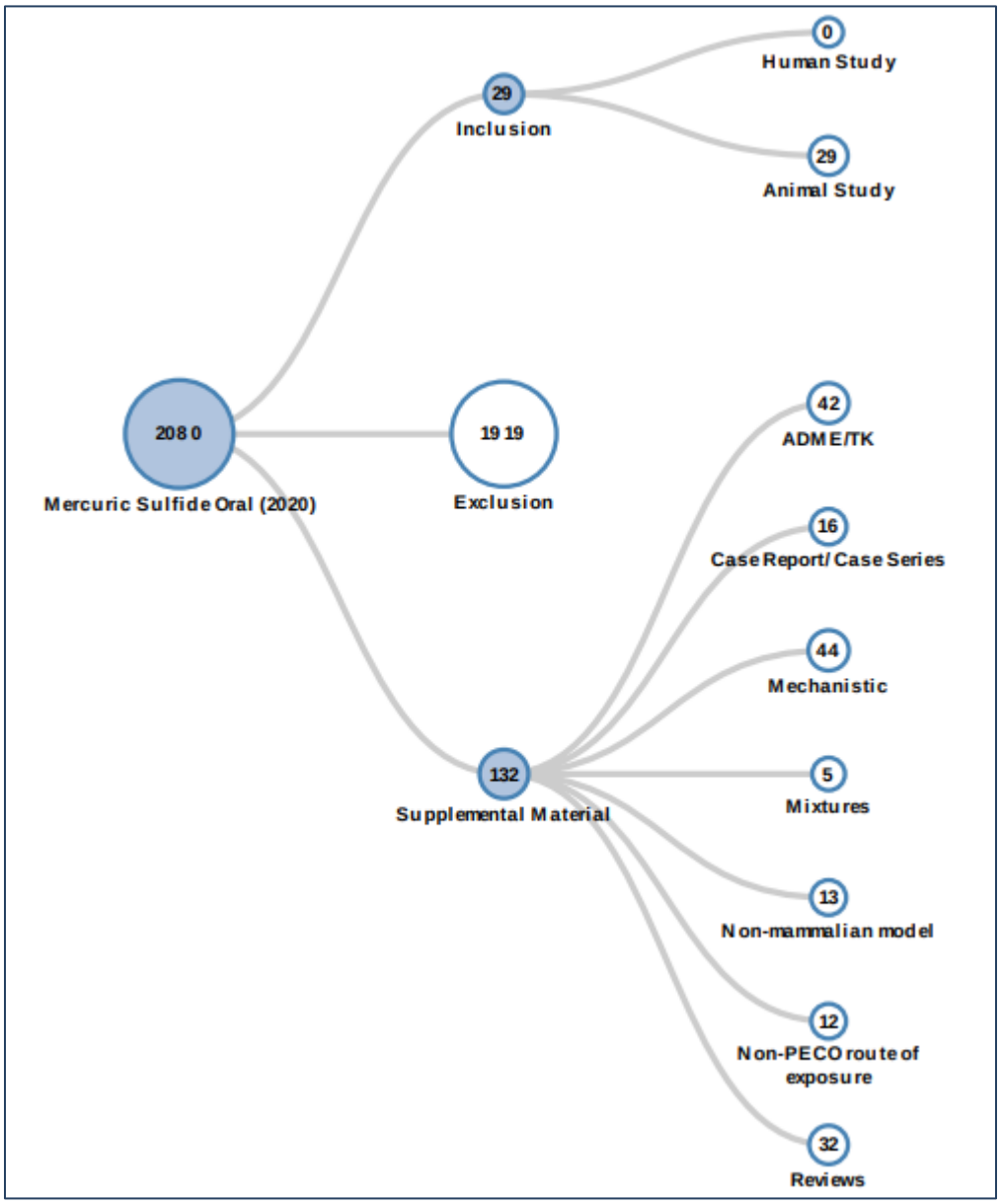


Figure 3-1. Literature flow diagrams for oral exposures of inorganic mercury salts (a), mercuric chloride (b), mercuric sulfide (c), mercurous chloride.

*Additional sources described in Appendix B will be searched prior to the draft development. a) mercuric chloride
https://hawcprd.epa.gov/lit/assessment/100500164/references/visualization/



(b) mercuric sulfide:
https://hawcprd.epa.gov/lit/assessment/100500172/references/visualization/



(c) mercurous chloride:
https://hawcprd.epa.gov/lit/assessment/100500173/references/visualization/

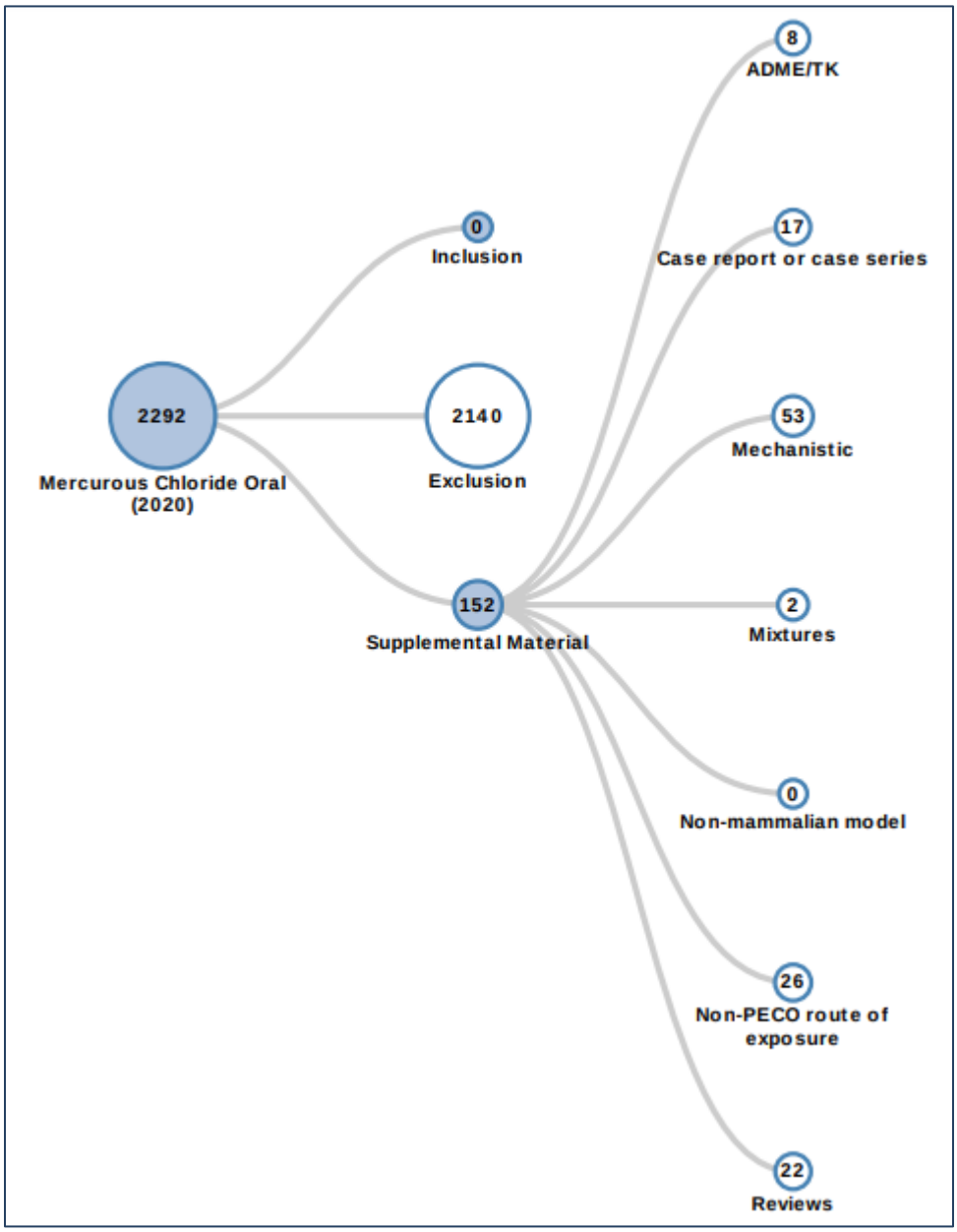


Figure 3-2. HAWC tree diagrams for inorganic mercury salts.

4. LITERATURE INVENTORY AND REFINED EVALUATION PLAN

4.1. LITERATURE INVENTORY PROCESS

During title/abstract or full-text level screening, studies tagged as included (meeting PECO criteria) were further categorized based on features such as evidence type (human, animal, *in vitro*, whether they also contain supplemental content) and the specific inorganic mercury salt addressed (see Appendix A, Table A-4).

Studies that met PECO criteria after full-text review were briefly summarized using DistillerSR to create literature inventories to display the extent and nature of the available evidence. The literature inventory is intended to provide a brief summary of study methods and results to help identify studies with the design features that make them most suitable for characterizing chronic hazard and identifying a POD. The results of the literature inventory are used to inform the refined evaluation plan (Section 5).

For animal studies, the following information was captured: chemical form; exposure timing and duration (acute, short term, subchronic, chronic, multigenerational, peripubertal, developmental); route; species; strain; sex; dose or concentration levels tested; dose units; health system and specific endpoints assessed; and a summary of findings at the health system level (null or no-observed-effect level/low-observed-effect level (NOAEL/LOAEL) based on author-reported statistical significance with an indication of which specific endpoints were affected). For epidemiological studies, the following information was summarized: chemical form; population type (e.g., general population-adult, occupational, pregnant women, infants and children); study type (e.g., cross-sectional, cohort, case-control); sex; major route of exposure (if known); description of how exposure was assessed; health system; and a summary of findings at the health system level (null or an indication of any associations found). Studies were inventoried in DistillerSR by one team member and checked by at least one other team member. These literature inventories are presented in Tableau visualization software (https://www.tableau.com/) and available for download in Excel file format.

These inputs were analyzed with respect to the considerations below to identify studies that plausibly could be considered suitable for identifying a POD. Studies prioritized from this process would undergo study evaluation and full data extraction. Studies not prioritized from this process are typically summarized at the literature inventory level only and will be used primarily to identify data gaps, identify emerging health concerns, or to provide context when interpreting findings from prioritized studies.

- When available, studies with chronic or subchronic exposure durations, or including exposure during reproduction or development, are prioritized over studies with short-term (more than 24 hours but less than 30 days) or acute exposure durations.

- Animal studies with multiple dose groups covering a broad range of dose levels are prioritized over single-dose studies. Single dose studies can be considered for toxicity value derivation, however, if they test phenotypic health outcomes unexamined in multidose studies testing similar levels.

- For human studies (controlled exposure and epidemiological studies), priority is given to studies in which exposure levels are expressed quantitatively and exposure-response quantitative results are presented in sufficient detail (e.g., odds ratios or relative risks with associated confidence intervals, numbers of cases/controls).

- For epidemiological studies, studies that used biomarker measurements in tissues or bodily fluids as the metric for exposure are considered potentially suitable for dose-response analysis if data or PBPK models are available to extrapolate between the reported biomarker measurement and the level of exposure.

- For both animal and human studies, whether the nature of the outcomes/endpoints assessed were interpretable with respect to potential adversity was considered. Typically, apical or clinical measures, when available, are preferred over biochemical or other mechanistic endpoints for dose-response.

4.2. PRELIMINARY LITERATURE INVENTORY RESULTS

Figures 4-1 and 4-2 summarize included studies for mercuric chloride and mercuric sulfide, respectively, with results broken down by study design, species, and evaluated health systems. No mercurous chloride studies were identified for literature inventory that met PECO criteria, nor were any PBPK models for inorganic mercury salts.

Mercuric chloride. The literature search and screening process identified 137 animal studies and no human studies that met PECO criteria for mercuric chloride. Figures 4-3 through 4-5 provide further details on dosing and sex of the animals for the subchronic, chronic, and developmental studies that evaluated hepatic, immune, and renal health outcomes as examples. Detailed information on other health outcomes can be found at https://public.tableau.com/profile/literature.inventory#!/vizhome/MercuricChlorideHgCl2EvidenceMapVisualizations/ReadMe.

Mercuric sulfide. The literature search and screening process identified 29 animal studies and no human studies that met PECO for mercuric sulfide. Of these 29, no chronic studies and one reproductive study were identified. Subchronic studies assessed hematologic, hepatic, nervous, renal, musculoskeletal, sensory, and systemic outcomes; further details on the reproductive and subchronic studies are presented in Figures 4-6 and 4-7.

Health System	acute		short-term			subchronic			chronic		repro..	developmental		Grand Total
	mouse	rat	mouse	rabbit	rat	goat	mouse	rat	mouse	rat	rat	mouse	rat	
Cancer					1			1						1
Cardiovascular		1	1		5		1	4	1	6			1	17
Developmental								1				2		3
Endocrine			2		3		2	10	2	4				20
Exocrine										1				1
Gastrointestinal		1		1	2		1	3	1	1			1	10
Hematologic		3	2	1	9		1	5		1				20
Hepatic	2	1	7		18		3	5	1	2			4	40
Immune			4		5		3	3	1	2		3	1	18
Lymphatic		1	1		1		1						1	5
Musculoskeletal/Connective ..							1		1	1				2
Nervous		1	6		7			10	2	1		3	6	33
Ocular									1					1
Other	2	2	4	2	16		5	11	1	5	2		1	49
Renal	3	2	7		17		4	9	1	5		1	3	47
Reproductive			3		4		4	8	2	2	1	1	1	24
Respiratory			1		2			2	1	1				4
Systemic/Whole Body		3	7		16	1	8	18	2	4	2	3	5	65
Urinary		1			3			1	1	3			1	8
Grand Total	4	6	18	2	45	1	13	30	4	10	2	5	9	137

1 18

Figure 4-1. Inventory of mercuric chloride animal studies that met PECO criteria by study design and health systems assessed.

Click here to view the interactive version of Figure 4-1, which includes a more detailed description of study design and results. The numbers indicate the number of studies that investigated a particular health system, not the number of studies that observed an association with inorganic mercury salts exposure. If a study evaluated multiple species, study designs, or health outcomes, it is shown here multiple times.

Health System	acute	short-term			subchronic		reproductive	Grand Total
	mouse	mouse	rat	guinea pig	mouse	rat	mouse	
Cardiovascular			1					1
Developmental							1	1
Endocrine		2						2
Hematologic		1			1	1	1	4
Hepatic	1	6	3		2			12
Immune		1						1
Musculoskeletal/Connective ..						1		1
Nervous		7	4	2	2		1	14
Other	1	1	2					4
Renal	1	4	4		2	3		14
Reproductive			1					1
Sensory		1			1			1
Systemic/Whole Body		4	3	2		1	1	11
Urinary			1					1
Grand Total	1	12	8	2	4	3	1	29

1 7

Figure 4-2. Inventory of mercuric sulfide animal studies that met PECO criteria by study design and health systems assessed.

Click here to view interactive version of Figure 4-2, which includes a more detailed description of study design and results. The numbers indicate the number of studies that investigated a particular health system, not the number of studies that observed an association with inorganic mercury salts exposure. If a study evaluated multiple species, study designs, or health outcomes, it is shown here multiple times.

Health System	Species	Sex	Dosing Duration	All Dose Levels	Dose Units	Reference	
Hepatic	mouse	both	44 d	0, 32	mg/kg-d	Lu, Wu, Liang, et al., 2011	
		male	7 wk	0, 3, 15, 75	ppm	Dieter et al., 1983	
			12 wk	0, 4	ppm	Afonne et al., 2000	
	rat	female	79 d (7 d/wk)	0, 0.75, 1.5	mg/kg-d	Atkinson et al., 2001	
				0, 0.75, 1.5, 2.5	mg/kg-d	Atkinson et al., 2001	
		male	8 wk	0, 80	mg/L	Zhang, H et al., 2017	
			12 wk	0, 0.4, 1.6	mg/kg	Lengyel et al., 2006	
						Lukacs et al., 2007	
			13 wk	0, 5	ppm	Prakash et al., 2000	
			81 d (7 d/wk)	0, 0.5, 1	mg/kg-d	Atkinson et al., 2001	
				0, 0.5, 1, 1.5	mg/kg-d	Atkinson et al., 2001	
Immune	mouse	female	10 wk	0, 0.5, 1, 2	mg/L	Nielsen and Hultman, 2002	
				0, 0.5, 1, 2, 4, 8	mg/L	Nielsen and Hultman, 2002	
				0, 0.625, 1.25, 2.5, 5	mg/L	Hultman and Eneström, 1992	
		male	7 wk	0, 3, 15, 75	ppm	Dieter et al., 1983	
			10 wk	0, 0.5, 1, 2, 4	mg/L	Nielsen and Hultman, 2002	
				0, 0.5, 1, 2, 4, 8	mg/L	Nielsen and Hultman, 2002	
	rat	male	12 wk	0, 0.4, 1.6	mg/kg	Lengyel et al., 2006	
						Lukacs et al., 2007	
			60 d	0, 4	mg/kg-d	Mohamed, 2018	

Route of Exposure
oral (gavage) oral (water)

Figure 4-3. Summary of subchronic mercuric chloride animal studies evaluating hepatic, immune, and renal health outcomes.

Health System	Species	Sex	Dosing Duration	All Dose Levels	Dose Units	Reference	
Renal	mouse	both	6 wk	0, 0.118	mmol/kg	Lu, Wu, Yan, et al., 2011	
		female	10 wk	0, 0.625, 1.25, 2.5, 5	mg/L	Hultman and Enestrom, 1992	
		male	7 wk	0, 3, 15, 75	ppm	Dieter et al., 1983	
		not reported	12 wk	0, 4	ppm	Afonne et al., 2002	
	rat	both	60 d	0, 0.02	g/kg	Shi et al., 2011	
		female	79 d (7 d/wk)	0, 0.75, 1.5	mg/kg-d	Atkinson et al., 2001	
				0, 0.75, 1.5, 2.5	mg/kg-d	Atkinson et al., 2001	
		male	6 wk	0, 0.15	mg/kg-d	Shin et al., 2017	
			9 wk	0, 8	mg/kg-d	Wang et al., 2010	
			12 wk	0, 0.4, 1.6	mg/kg	Lengyel et al., 2006	
						Lukacs et al., 2007	
				0, 1, 3, 10, 30	mg/kg-d	Takahashi et al., 2000a	
			13 wk	0, 5	ppm	Prakash et al., 2000	
			56 d	0, 80	mg/kg	Tan et al., 2018	
			81 d (7 d/wk)	0, 0.5, 1, 1.5	mg/kg-d	Atkinson et al., 2001	

Route of Exposure
oral (diet) oral (gavage) oral (water)

Figure 4-3 (cont'd). Summary of subchronic mercuric chloride animal studies evaluating hepatic, immune, and renal health outcomes.

Click here to view the interactive version, which includes the complete list of health outcomes and additional study details.

Health System	Species	Sex	Dosing Duration	All Dose Levels	Dose Units	Reference	
Renal	mouse	female	2 y (5 d/wk)	0, 5, 10	mg/kg	NTP, 1993	
			15 mon (5 d/wk)	0, 5, 10	mg/kg	NTP, 1993	
			27 wk (5 d/wk)	0, 1.25, 2.5, 5, 10, 20	mg/kg	NTP, 1993	
		male	2 y (5 d/wk)	0, 5, 10	mg/kg	NTP, 1993	
			15 mon (5 d/wk)	0, 5, 10	mg/kg	NTP, 1993	
			26 wk (5 d/wk)	0, 1.25, 2.5, 5, 10, 20	mg/kg	NTP, 1993	
	rat	both	21 wk	0, 1, 3, 10, 30	mg/k..	Takahashi et al., 2000b	
		female	2 y (5 d/wk)	0, 2.5, 5	mg/kg	Dieter et al., 1992	
						NTP, 1993	
			6 mon (5 d/wk)	0, 0.312, 0.625, 1.25, 2.5, 5	mg/kg	Dieter et al., 1992	
			15 mon (5 d/wk)	0, 2.5, 5	mg/kg	Dieter et al., 1992	
						NTP, 1993	
			26 wk (5 d/wk)	0, 0.312, 0.625, 1.25, 2.5, 5	mg/kg	NTP, 1993	
		male	2 y (5 d/wk)	0, 2.5, 5	mg/kg	Dieter et al., 1992	
						NTP, 1993	
			6 mon	0, 0.05	mg/kg	Argrawal et al., 2014	
			6 mon (5 d/wk)	0, 0.312, 0.625, 1.25, 2.5, 5	mg/kg	Dieter et al., 1992	
			15 mon (5 d/wk)	0, 2.5, 5	mg/kg	Dieter et al., 1992	
						NTP, 1993	
			26 wk (5 d/wk)	0, 0.312, 0.625, 1.25, 2.5, 5	mg/kg	NTP, 1993	
			180 d	0, 200	µg/mL	Carmignani et al., 1992	

Route of Exposure
oral (diet) oral (gavage) oral (water)

Figure 4-4. Summary of chronic mercuric chloride animal studies evaluating hepatic, immune, and renal health outcomes.

Health System	Species	Sex	Dosing Duration	All Dose Levels	Dose Units	Reference	
Hepatic	mouse	female	2 y (5 d/wk)	0, 5, 10	mg/kg	NTP, 1993	
			15 mon (5 d/wk)	0, 5, 10	mg/kg	NTP, 1993	
			27 wk (5 d/wk)	0, 1.25, 2.5, 5, 10, 20	mg/kg	NTP, 1993	
		male	2 y (5 d/wk)	0, 5, 10	mg/kg	NTP, 1993	
			15 mon (5 d/wk)	0, 5, 10	mg/kg	NTP, 1993	
			26 wk (5 d/wk)	0, 1.25, 2.5, 5, 10, 20	mg/kg	NTP, 1993	
	rat	female	15 mon (5 d/wk)	0, 2.5, 5	mg/kg	NTP, 1993	
			26 wk (5 d/wk)	0, 0.312, 0.625, 1.25, 2.5, 5	mg/kg	NTP, 1993	
		male	6 mon	0, 0.05	mg/kg	Argrawal et al., 2014	
			15 mon (5 d/wk)	0, 2.5, 5	mg/kg	NTP, 1993	
			26 wk (5 d/wk)	0, 0.312, 0.625, 1.25, 2.5, 5	mg/kg	NTP, 1993	
Immune	mouse	female	27 wk (5 d/wk)	0, 1.25, 2.5, 5, 10, 20	mg/kg	NTP, 1993	
		male	26 wk (5 d/wk)	0, 1.25, 2.5, 5, 10, 20	mg/kg	NTP, 1993	
	rat	female	26 wk (5 d/wk)	0, 0.312, 0.625, 1.25, 2.5, 5	mg/kg	NTP, 1993	
		male	26 wk (5 d/wk)	0, 0.312, 0.625, 1.25, 2.5, 5	mg/kg	NTP, 1993	
			180 d	0, 200	µg/mL	Carmignani et al., 1992	

Route of Exposure
oral (gavage) oral (water)

Figure 4-4 (cont'd). Summary of chronic mercuric chloride animal studies evaluating hepatic, immune, and renal health outcomes.

Click here to view the interactive version of Figure 4-4, which includes the complete list of health outcomes and additional study details.

Health System	Species	Sex	Dosing Duration	All Dose Levels	Dose Units	Reference	
Hepatic	rat	both	GD0-GD20	0, 0.2, 0.5, 10, 50	µg/mL	Oliveira et al., 2012	
			GD0-PND20	0, 0.2	µg/mL	Feng et al., 2004	
			gestation + PND2-50	0, 1.2	mg/kg-d	Su et al., 2008	
		female (dam)	21 d (gestation)	0, 10, 50	µg/mL	Oliveira et al., 2016	
Immune	mouse	both	35 d (GD8-PND21	0, 50	µM	Zhang et al., 2011	
			35 d (GD8-PND21)	0, 50	µM	Zhang et al., 2011	
			GD0-16	0, 10	ppm	Pilones et al., 2008	
		female	35 d (GD8-PND21)	0, 50	µM	Zhang et al., 2013	
		female (dam)	35 d (GD8-PND21	0, 50	µM	Zhang et al., 2011	
			35 d (GD8-PND21)	0, 50	µM	Zhang et al., 2011	
						Zhang et al., 2013	
		male	35 d (GD8-PND21)	0, 50	µM	Zhang et al., 2013	
	rat	both	GD0-PND20	0, 0.2	µg/mL	Feng et al., 2004	
Renal	mouse	female (dam)	79 d (7 d/wk; 16 d pre-mating + 21 d mating..	0, 0.25, 0.5, 1.0	mg/kg-d	Khan et al., 2004	
		male	61 d (7 d/wk; 40 d pre-mating + 21 d mating)	0, 0.25, 0.5, 1.0	mg/kg-d	Khan et al., 2004	
	rat	both	GD0-PND20	0, 0.2	µg/mL	Feng et al., 2004	
			gestation + PND2-50	0, 1.2	mg/kg-d	Su et al., 2008	
		female (dam)	GD0-GD20	0, 0.2, 0.5, 10, 50	µg/mL	Oliveira et al., 2012	

Route of Exposure
oral (gavage) oral (water)

Figure 4-5. Summary of developmental mercuric chloride animal studies evaluating hepatic, immune, and renal health outcomes.

Click here to view the interactive version of Figure 4-5, which includes the complete list of health outcomes and additional study details.

Health System	Species	Sex	Dosing Duration	All Dose Levels	Dose Units	Reference	
Hematologic	mouse	female	11 wk	0, 10	mg/kg-d	Huang et al., 2007	
		male	11 wk	0, 10	mg/kg-d	Huang et al., 2007	
	rat	both	8 wk	0, 1	g/kg-d	Wang, Wang, Wu, Wang, Wang, et al., 2015	
			12 wk	0, 1	g/kg-d	Wang, Wang, Wu, Wang, Wang, et al., 2015	
Hepatic	mouse	both	44 d	0, 300	mg/kg-d	Lu, Wu, et al., 2011	
		female	11 wk	0, 10	mg/kg-d	Huang et al., 2007	
		male	11 wk	0, 10	mg/kg-d	Huang et al., 2007	
Musculoskeletal/ Connective Tissue	rat	both	8 wk	0, 1	g/kg-d	Wang, Wang, Wu, Wang, Wang, et al., 2015	
			12 wk	0, 1	g/kg-d	Wang, Wang, Wu, Wang, Wang, et al., 2015	
Nervous	mouse	both	10 wk	0, 10	mg/kg-d	Huang et al., 2008	
		female	6 wk	0, 10	mg/kg-d	Huang et al., 2007	
			11 wk	0, 10	mg/kg-d	Huang et al., 2007	
		male	6 wk	0, 10	mg/kg-d	Huang et al., 2007	
			11 wk	0, 10	mg/kg-d	Huang et al., 2007	

Route of Exposure
oral (gavage)

Figure 4-6. Summary of subchronic mercuric sulfide animal studies.

Health System	Species	Sex	Dosing Duration	All Dose Levels	Dose Units	Reference	
Renal	mouse	both	6 wk	0, 0.3	g/kg	Lu et al., 2008	
		female	11 wk	0, 10	mg/kg-d	Huang et al., 2007	
		male	11 wk	0, 10	mg/kg-d	Huang et al., 2007	
	rat	both	8 wk	0, 1	g/kg-d	Wang, Wang, Wu, Wang, Gao, et al., 2015	
						Wang, Wang, Wu, Wang, Wang, et al., 2015	
			12 wk	0, 1	g/kg-d	Wang, Wang, Wu, Wang, Gao, et al., 2015	
						Wang, Wang, Wu, Wang, Wang, et al., 2015	
		male	60 d	0, 0.2	g/kg	Shi et al., 2011	
Sensory	mouse	both	6 wk	0, 10	mg/kg-d	Huang et al., 2008	
			10 wk	0, 10	mg/kg-d	Huang et al., 2008	
Systemic/ Whole Body	rat	male	60 d	0, 0.2	g/kg	Shi et al., 2011	

Route of Exposure
oral (gavage)

Figure 4-6 (cont'd). Summary of subchronic mercuric sulfide animal studies.

Click here to view the interactive version of Figure 4-6, which includes a more detailed description of study design and results.

Health System	Species	Sex	Dosing Duration	All Dose Levels	Dose Units	Reference	
Developmental	mouse	female (dam)	4 wk pre-mating + 7 wk postnatal	0, 10	mg/kg-d	Huang et al., 2012	
Hematologic	mouse	both	4 wk pre-mating + 7 wk postnatal	0, 10	mg/kg-d	Huang et al., 2012	
Nervous	mouse	both	4 wk pre-mating + 7 wk postnatal	0, 10	mg/kg-d	Huang et al., 2012	
Systemic/ Whole Body	mouse	both	4 wk pre-mating + 7 wk postnatal	0, 10	mg/kg-d	Huang et al., 2012	

Route of Exposure
oral (gavage)

Figure 4-7. Summary of reproductive mercuric sulfide animal studies.

Click here to view the interactive version of Figure 4-7, which includes a more detailed description of study design and results.

5. REFINED EVALUATION PLAN

The primary purpose of this step is to outline any potential or expected refinements, which would narrow the scope of studies considered for use in evidence synthesis and beyond. This optional step is typically applied to focus an assessment with a very large number of PECO-relevant studies on review of the most informative evidence (e.g., when many studies examine the same health outcome, focusing on toxicity studies including exposures below a specified range, those studies examining more specific or objective measures of toxicity, or those that address lifestage- or exposure duration-specific knowledge on how the health outcome develops). Given the relatively small databases of animal toxicology studies for these three inorganic mercury salts, this narrowing is not considered applicable to these data.

An additional purpose of this step is to clarify the grouping of different health endpoints for analysis. Based on the health outcomes in ATSDR *Toxicological Profile for Mercury* and the literature inventory, EPA anticipates hazard identification for mercuric chloride will focus on potential renal, immune, nervous system, hepatic, reproductive, hematologic and cancer effects. For mercuric sulfide, hepatic, renal, and nervous system effects likely will be the focus. For mercurous chloride, because no relevant data are available for animal and humans, analogue-based, read-across methodology will be explored to derive toxicity values. Based on the preliminary literature inventory, the specific animal endpoint grouping categories that will be used to frame the evidence syntheses in this assessment are provided in Table 5-1.

Finally, another function of the refined analysis plan is to consider whether any adjustments in the literature compilation are needed to address the key science issues outlined in Section 6. For mercury salts, no additional searches were specifically identified as the studies considered pertinent to address the key science issues would be identified through the screening and tagging approach used to identify PECO and supplemental material studies.

Table 5-1. Animal endpoint grouping categories

Relevant human health effect category[a]	Examples of animal endpoints included
General toxicity	• Body weight (not maternal or pup weights, or weights after developmental-only exposure) • Mortality, survival, or growth curve • Clinical chemistry endpoints and observations (nonbehavioral)
Hepatic (toxicity)	• Liver weight and histopathology • Serum or tissue liver enzymes (e.g., ALT and AST from clinical chemistry), other liver tissue enzyme activity (e.g., catalase) or protein/DNA content, other liver tissue biochemical markers (e.g., albumin; glycogen; glucose), liver-specific serum biochemistry (e.g., albumin; albumin/globulin)
Hematologic (effects)	• Red blood cells, blood hematocrit or hemoglobin, corpuscular volume • Blood platelets or reticulocytes • Blood biochemical measures (e.g., sodium, calcium, phosphorus)
Immune (effects)	Host resistance; allergic, autoimmune or infectious disease; hypersensitivity • General immune assays (e.g., white blood cell counts, immunological factors or cytokines in blood, lymphocyte phenotyping, or proliferation); immune cell counts or immune-specific cytokines in nonlymphoid tissues • Other immune functional assays (e.g., antibody production, natural killer cell function, DTH, MLR, CTL, phagocytosis, or bacterial killing by monocytes) • Any measure in lymphoid tissues (weight, histopathology, cell counts, etc.); immune responses in the respiratory system; stress-related factors in blood (e.g., glucocorticoids or other adrenal markers)
Renal/Urinary (toxicity)	• Kidney weight and histopathology • Urinary measures (e.g., protein, volume, pH, specific gravity) • Clinical chemistry (BUN)
Nervous system (effects)	• Brain weight • Behavioral measures (including FOB and cage-side observations) • Nervous system histopathology
Reproductive (toxicity) Note: Evidence synthesis and evidence integration conclusions in assessment are developed separately for male and female reproductive effects (toxicity)	• Reproductive organ weight and histopathology • Markers of sexual differentiation or maturation (e.g., preputial separation in males, vaginal opening or estrous cycling in females) • Mating parameters (e.g., success, mount latency) • Reproductive hormones

Relevant human health effect category[a]	Examples of animal endpoints included
Developmental (effects) Note: Evidence synthesis of these endpoints in the assessment is termed "offspring growth and early development," but evidence integration conclusions will be drawn on the broader category of "developmental effects" (which also considers organ/system-specific effects after exposure during development)	• Dam health (e.g., weight gain, food consumption) • Pup viability/survival or other birth parameters (e.g., number of pups per litter) • Pup weight or growth (includes measures into adulthood after developmental-only exposure) • Developmental landmarks (eye opening, etc., but not including markers for other organ-/system-specific toxicities)
Carcinogenicity	• Tumors • Precancerous lesions (e.g., dysplasia)

[a]Other health outcomes, unless necessary based on updated literature searches, will not be formally evaluated in this assessment, although short summaries of the evidence might be included for context.

ALT = alanine aminotransferase; AST = aspartate transaminase; BUN = blood urea nitrogen; CTL = cytotoxic T lymphocyte; DNA = deoxyribonucleic acid; DTH = delayed type hypersensitivity; FOB = functional operational battery; LD50 = median lethal dose; MLR = mixed leukocyte reaction.

6. KEY SCIENCE ISSUES

Based on the problem formulation and the preliminary literature inventory (Figures 4-1 and 4-2), the following key scientific issues were identified that warrant focused evaluation in this assessment.

- **Key science issue #1: Consideration of the use of analogue-based, read-across method to inform the assessment of mercurous chloride:** The literature inventory (Figures 4-1 and 4-2) did not identify any animal toxicology or human health studies for mercurous chloride. In the absence of relevant data for hazard and dose-response assessment, an analogue-based, read-across methodology (Wang et al., 2012) will be explored to identify and evaluate analogues for the development of toxicity values for the target chemical, mercurous chloride. If read-across is not feasible, it is likely no toxicity value will be derived. Any read-across analysis presented in an IRIS assessment would undergo targeted discussion during Agency and interagency review. Specific charge questions would be developed as part of external peer-review.

- **Key science issue #2: Characterizing toxicokinetic and toxicodynamic differences across compounds:** Both mercuric chloride and mercuric sulfide are divalent, with mercury in a +2-oxidation state; however, the solubilities of the two salts differ by about four orders of magnitude. Thus, the bioavailability for mercuric sulfide is expected to be low compared with that of mercuric chloride. Mercurous chloride has a +1-oxidation state, and mercurous forms are less soluble and presumably less bioavailable than mercuric forms. A better understanding of the bioavailability of different salts and how it affects the ADME/toxicokinetic properties and toxicodynamic profiles of the mercury salts will aid in determining the potential human health hazards of these salts.

- **Key science issue #3: Characterizing toxicokinetic differences across sexes, species and lifestages:** Some evidence suggests sex, species and age-related differences in the toxicokinetics within individual inorganic mercury salts such as mercuric chloride. For example, female mice tend to retain considerably less mercuric chloride in the kidneys than do male mice. Therefore, evaluating potential influence of sex on mercury salts toxicokinetics is necessary. In addition to sex differences, mechanistic data revealed the toxicokinetics of inorganic mercury salts were age and species dependent. For instance, absorption of inorganic mercury in the sucking pups is prolonged, as compared to adult mice, indicating the uptake of mercury salts could be higher in younger animals. Furthermore, variations in toxicokinetics of inorganic mercury were also observed among animal species and strains, and the strain difference in the toxicokinetics of mercury salts (e.g., mercuric chloride) were primarily attributed to differences in elimination kinetics. Thus, understanding how toxicokinetics of mercury salts could vary by sexes, species/strains, and lifestages is important.

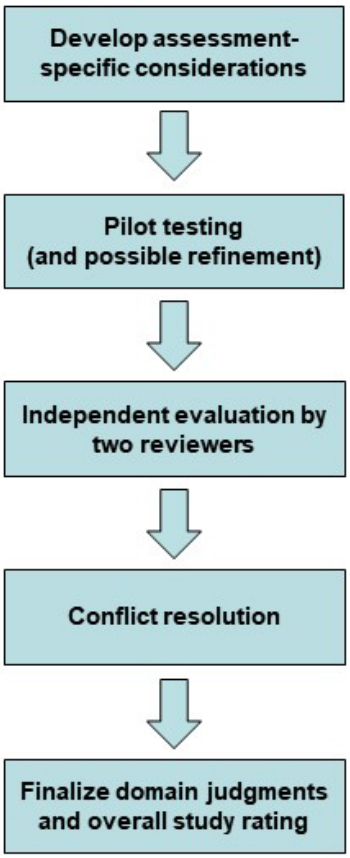
7. STUDY EVALUATION (REPORTING, RISK OF BIAS, AND SENSITIVITY) STRATEGY

The general approach for evaluating PECO-relevant primary health effect studies is described in Section 7.1, and the instructional and informational materials for data extraction and study quality evaluations are available at https://hawcprd.epa.gov/assessment/100000039/. The approach is the same for epidemiology studies and animal toxicology experiments, but the application specifics differ; thus, they are described separately for epidemiology and animal toxicology studies in Sections 7.2 and 7.3, respectively. No controlled human exposure studies for these inorganic mercury salts were identified (see Section 3). No physiologically based pharmacokinetic (PBPK) studies have been identified for inorganic mercury salts; if one becomes available, however, it will be evaluated using methods described in the Quality Assurance Project Plan for PBPK models (U.S. EPA, 2018). Approaches used to evaluate mechanistic studies are described in section 7.4.

7.1. STUDY EVALUATION OVERVIEW FOR HEALTH EFFECT STUDIES

Key concerns for the review of epidemiology and animal toxicology studies are potential bias (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). The study evaluations are aimed at discerning the expected magnitude of any identified limitations (focusing on limitations that could substantively change a result), considering the expected direction of the bias. The study evaluation approach is designed 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 within a study) regarding confidence in the reliability of the results is illustrated in Figure 7-1.

(a) Study evaluation process



(b) Individual evaluation domains

Animal	Epidemiology
Selection and performance • Allocation • Observational bias/blinding	Participant selection
Confounding/variable control	Confounding
Selective reporting and attrition	Selective reporting
Exposure methods sensitivity • Chemical administration and characterization • Exposure timing, frequency, and duration	Exposure measurement
Outcome measures and results display • Endpoint sensitivity and specificity • Results presentation	Outcome ascertainment Analysis
Reporting quality	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 7-1. Overview of Integrated Risk Information System (IRIS) study evaluation process.

(a) An overview of the general evaluation process (b) The evaluation domains and definitions for ratings (i.e., domain and overall judgments, performed on an outcome-specific basis).

At least two reviewers will independently evaluate health effect studies to identify characteristics that bear on the informativeness (i.e., validity and sensitivity) of the results. The independent reviewers will use the structured platform for study evaluation housed within the EPA's version of HAWC (https://hawcprd.epa.gov/assessment/100000039/) to record separate judgments for each domain and the overall study for each outcome, to reach consensus between reviewers, and when necessary, resolve differences by discussion between the reviewers or consultation with additional independent reviewers.

In general, considerations for reviewing a study with regard to its conduct for specific health outcomes are based on using existing guidance documents when available, including EPA guidance for carcinogenicity, neurotoxicity, reproductive toxicity, and developmental toxicity (U.S. EPA, 2005a, 1998c, 1996a, 1991). For some aspects of the study evaluations (e.g., review of exposure assessment in epidemiology studies), additional considerations are developed in consultation with topic-specific technical experts. To calibrate the assessment-specific considerations, the study evaluations will include a pilot phase to assess and refine the evaluation process. As reviewers examine a group of studies, additional chemical-specific knowledge or methodological concerns could emerge, and a second pass of all pertinent studies might become necessary. The study evaluation process will be refined during the pilot phase and subsequent implementation across all relevant studies will be acknowledged as updates to the protocol.

Authors might be queried to obtain critical information, particularly that involving missing reporting quality information or other data (e.g., information on variability) or additional analyses that could address potential study limitations. The decision on whether to seek missing information includes considering what additional information would be useful, specifically regarding any information that could result in a reevaluation of the overall study confidence for an outcome. Outreach to study authors will be documented and considered unsuccessful if researchers do not respond to an email or phone request within one month of the attempt to contact. Only information or data that can be made publicly available (e.g., within HAWC or HERO) will be considered.

When evaluating studies[10] 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,[11] evaluations might be performed at a more granular level if appropriate, but these measures could still be grouped for evidence synthesis.

[10]Note: "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. All evaluations discussed herein, however, are explicitly conducted at the level of an individual outcome within a population or cohort of humans or animals similarly exposed (e.g., unexposed or exposed at comparable lifestages and for the same exposure duration), or to a sample of the study population within a study.
[11]Note: "outcome" will be used throughout these methods; the same methods apply to an endpoint assessed separately within a larger outcome. "Endpoint" refers to a more granular measurement (e.g., for the outcome of liver histopathology, different endpoints might include necrosis and cellular hypertrophy).

During review, for each evaluation, domain reviewers will reach a consensus judgment of *good*, *adequate*, *deficient*, *not reported*, or *critically deficient*. 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 identifying individual hazards is important to stress. Although limitations specific to the usability of the study for dose-response analysis are useful to note to inform those later decisions, they do not contribute to the study confidence classifications.

These four 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 minor deficiencies noted would not be expected to influence the study results.

- *Adequat*e indicates a judgment that methodological limitations related to the evaluation domain might occur, but those limitations are unlikely to be severe or to notably impact the results.

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

- *Not reported* indicates the information necessary to evaluate the domain 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 of unreported items and severity of other limitations identified in the study, reaching out to the study authors for this information might or might not be worthwhile (see discussion above).

- *Critically deficient* reflects a judgment that the study conduct relating to the evaluation domain question introduced a serious flaw interpreted to be the primary driver of any observed effect(s) or makes the study uninterpretable. Studies with a determination of *critically deficient* in an evaluation domain will not be used for hazard identification or dose-response analysis, but they could be used to highlight possible research gaps.

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

- *High* confidence: No notable deficiencies or concerns were 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: Possible deficiencies or concerns were noted, but the limitations are unlikely to be notable. 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: 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 might have a *deficient* rating in domain(s) considered to have less influence on the magnitude or direction of the outcome-specific results). *Low* confidence results are given less weight compared to *high* or m*edium* confidence results during evidence synthesis and integration (see Section 11, Tables 11-2 and 11-3) and are generally not used alone for hazard identification or dose-response analyses unless they are the only studies available or they inform data gaps unexamined in the *high* or *medium* confidence studies. Studies rated as *medium* or *low* confidence only because of sensitivity concerns about bias towards the null will be asterisked or otherwise noted because they might require additional consideration during evidence synthesis. Effects observed in studies biased toward the null might actually increase confidence in the results, assuming the study is otherwise well conducted (see Section 10).

- *Uninformative*: Serious flaw(s) make the study results unusable for hazard identification. 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 also might be considered *uninformative*. *Uninformative* studies will not be considered further in the synthesis and integration of evidence, except perhaps to highlight possible research gaps.

As previously noted, study evaluation determinations reached by each reviewer and the consensus judgment between reviewers will be recorded in HAWC. Final study evaluations housed in HAWC 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 evidence synthesis (see Section 10) to help interpret the results across studies.

7.2. EXPERIMENTAL ANIMAL STUDY EVALUATION

Using the principles described in Section 7.1, the animal studies of health effects to assess risk of bias and sensitivity will be evaluated for the following domains: reporting quality, risk of bias (selection or performance bias, confounding/variable control, and reporting or attrition bias), study sensitivity (exposure methods sensitivity, and outcome measures and results display) (Table 7-2).

The rationale for judgments will be documented clearly and consistently at the outcome level. In addition, similar to the evaluation of epidemiology studies, for domains other than reporting quality, the evaluation documentation in HAWC will include the identified limitations and consider their impact on the overall confidence level. This will, to the extent possible, reflect an interpretation of the potential influence on the outcome-specific results (including the direction or magnitude of influence, or both).

Table 7-1. Considerations to evaluate domains from animal toxicology studies

Evaluation concern	Domain—core question	Prompting questions	General considerations
Reporting quality	**Reporting quality** Does the study report information for evaluating the design and conduct of the study for the endpoints/outcomes of interest? *Note:* *This domain is limited to reporting. Other aspects of the exposure methods, experimental design, and endpoint evaluation methods are evaluated using the domains related to risk of bias and study sensitivity.*	Does the study report the following? **Critical information** necessary to perform study evaluation: • Species, test article name, levels and duration of exposure, route (e.g., oral; inhalation), qualitative or quantitative results for at least one endpoint of interest **Important information** for evaluating the study methods: • Test animal: strain, sex, source, and general husbandry procedures • Exposure methods: source, purity, method of administration • Experimental design: frequency of exposure, animal age and lifestage during exposure and at endpoint/outcome evaluation • Endpoint evaluation methods: assays or procedures used to measure the endpoints/outcomes of interest	A judgment and rationale for this domain generally will be given for the study. In the rationale, reviewers will also indicate when a study adhered to GLP, or to OECD (or similar) testing guidelines. • **Good:** All **critical and important information** is reported or inferable for the endpoints/outcomes of interest. • **Adequate:** All **critical information** is reported, but some **important information** is missing. The missing information, however, is not expected to significantly impact the study evaluation. • **Deficient:** All **critical information** is reported, but **important information** expected to significantly reduce the ability to evaluate the study is missing. • **Critically deficient:** Study report is missing any pieces of **critical information**. Studies that are *Critically Deficient* for reporting are *Uninformative* for the overall rating and not considered further.

Evaluation concern		Domain—core question	Prompting questions	General considerations
Risk of bias	**Selection and performance bias**	**Allocation** Were animals assigned to experimental groups using a method that minimizes selection bias?	For each study: • Did each animal or litter have an equal chance of being assigned to any experimental group (i.e., random allocation[a])? • Is the allocation method described? • Aside from randomization, were any steps taken to balance variables across experimental groups during allocation?	A judgment and rationale for this domain will be given for each cohort or experiment in the study. • **Good**: Experimental groups were randomized, and any specific randomization procedure was described or inferable (e.g., computer-generated scheme). (Note that normalization is not the same as randomization [see response for '*Adequate*'].) • **Adequate**: Authors report that groups were randomized but do not describe the specific procedure used (e.g., "animals were randomized"). Alternatively, authors used a nonrandom method to control for important modifying factors (i.e., with respect to the outcome of interest) across experimental groups (e.g., body-weight normalization). • **Not reported** (interpreted as *Deficient*): No indication of randomization of groups or other methods (e.g., normalization) to control for important modifying factors across experimental groups. • **Critically deficient:** Bias in the animal allocations was reported or inferable.

Evaluation concern		Domain—core question	Prompting questions	General considerations
Risk of bias (continued)	**Selection and performance bias (continued)**	**Observational bias/blinding** Did the study implement measures to reduce observational bias?	For each endpoint/outcome or grouping of outcomes in a study: • Does the study report blinding or other methods/procedures for reducing observational bias, as appropriate for the assays of interest? • If not, did the study use a design or approach for which such procedures can be inferred? • What is the expected impact of failure to implement (or report implementation) of these methods/procedures on results?	A judgment and rationale for this domain will be given for each endpoint/outcome or group of outcomes investigated in the study. • **Good**: Measures to reduce observational bias were described (e.g., blinding to conceal treatment groups during endpoint evaluation; consensus-based evaluations of histopathology lesions[a]). • **Adequate**: Methods for reducing observational bias (e.g., blinding) can be inferred or were reported but described incompletely. • **Not reported:** Measures to reduce observational bias were not described. o (Interpreted as *Adequate*): The potential concern for bias was mitigated based on use of automated/computer-driven systems, standard laboratory kits, relatively simple, objective measures (e.g., body or tissue weight), or screening-level evaluations of histopathology. o (Interpreted as *Deficient*): The potential impact on the results is large (e.g., outcome measures are highly subjective). • **Critically deficient:** Strong evidence for observational bias that impacted the results.

Evaluation concern		Domain—core question	Prompting questions	General considerations
Risk of bias (continued)	**Confounding/variable control**	**Confounding** Are variables with the potential to confound or modify results controlled for and consistent across all experimental groups?	For each study: • Are there differences across the treatment groups (e.g., co-exposures, vehicle, diet, palatability, husbandry, health status, surgery) that could bias the results? • If differences are identified, to what extent are they expected to impact the results?	A judgment and rationale for this domain will be given for each cohort or experiment in the study, noting when the potential for confounding is restricted to specific endpoints/outcomes. • **Good**: Outside of the exposure of interest, variables likely to confound or modify results appear to be controlled for and consistent across experimental groups. • **Adequate**: Some concern that variables likely to confound or modify the results were uncontrolled or inconsistent across groups, but these are expected to have a minimal impact on the results. • **Deficient**: Notable concern that potentially confounding variables were uncontrolled or inconsistent across groups and that they are expected to substantially impact the results. • **Critically deficient**: Confounding variables were presumed to be uncontrolled or inconsistent across groups, and they are expected to be a primary driver of the results.

Evaluation concern		Domain—core question	Prompting questions	General considerations
Risk of bias (continued)	**Selective reporting and attrition bias**	**Selective reporting and attrition** Did the study report results for all prespecified outcomes and tested animals? *Note:* *This domain does not consider the appropriateness of the comparisons/results presentation. This aspect of study quality is evaluated in another domain.*	For each study: Selective reporting bias: • Are all results presented for endpoints/outcomes described in the methods (see note)? Attrition bias: • Do the results account for all animals? • If discrepancies occur, do the authors provide an explanation (e.g., death or unscheduled sacrifice during the study)? • If unexplained results omissions or attrition are identified, what is the expected impact on the interpretation of the results?	A judgment and rationale for this domain will be given for each cohort or experiment in the study. • **Good:** Quantitative or qualitative results were reported for all prespecified outcomes (explicitly stated or inferred), exposure groups and evaluation time points. Data not reported in the primary article are available from supplemental material. If results omissions or animal attrition are identified, the authors provide an appropriate explanation, and the omissions or attrition are not expected to impact the interpretation of the results. • **Adequate:** Quantitative or qualitative results are reported for most prespecified outcomes (explicitly stated or inferred), exposure groups, and evaluation time points. Omissions or attrition not explained, but not expected to significantly impact the interpretation of the results. • **Deficient:** Quantitative or qualitative results are missing for many prespecified outcomes (explicitly stated or inferred), exposure groups and evaluation time points, or high animal attrition occurred; omissions or attrition are not explained and are expected to significantly impact the interpretation of the results. • **Critically deficient:** Extensive results omission or animal attrition are identified and prevent comparisons of results across treatment groups.

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity	**Exposure methods sensitivity**	**Chemical administration and characterization** Did the study adequately characterize exposure to the chemical of interest and the exposure administration methods?	For each study: • Does the study report the source and purity or composition (e.g., identity and percent distribution of different isomers) of the chemical? If not, can the purity or composition be obtained from the supplier (e.g., as reported on the website)? • Was independent analytical verification of the test article purity and composition performed? • Are there concerns about the methods used to administer the chemical (e.g., gavage volume)? • If necessary, on the basis of considering chemical-specific knowledge (e.g., instability in solution; volatility) or exposure design (e.g., the frequency and duration of exposure), were the chemical concentrations in the dosing solutions or diet analytically confirmed?	A judgment and rationale for this domain will be given for each cohort or experiment in the study. • **Good:** Chemical administration and characterization is complete (i.e., source, purity, and analytical verification of the test article are provided). No concerns about the composition, stability, or purity of the administered chemical, or the specific methods of administration, exist. • **Adequate:** Some uncertainties in the chemical administration and characterization are identified, but are expected to have minimal impact on interpreting the results (e.g., source and vendor-reported purity are presented, but not independently verified; purity of the test article is suboptimal but not concerning). • **Deficient:** Uncertainties in the exposure characterization are identified and expected to substantially impact the results (e.g., source of the test article is not reported; levels of impurities are substantial or concerning; deficient administration methods, such as use of a gavage volume considered too large for the species or lifestage at exposure). • **Critically deficient:** Uncertainties in the exposure characterization are identified, and that the results are largely attributable to factors other than exposure to the chemical of interest is reasonably certain (e.g., identified impurities are expected to be a primary driver of the results).

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity (continued)	**Exposure methods sensitivity (continued)**	**Exposure timing, frequency, and duration** Was the timing, frequency, and duration of exposure sensitive for the endpoint(s)/outcome(s) of interest?	For each endpoint/outcome or grouping of outcomes in a study: • Does the exposure period include the full critical window of sensitivity, based on current biological understanding? • Were the duration and frequency of exposure sensitive for detecting the endpoint of interest?	A judgment and rationale for this domain will be given for each endpoint/outcome or group of outcomes investigated in the study. • **Good:** The duration and frequency of the exposure was sensitive, and the exposure included the critical window of sensitivity (if known). • **Adequate:** The duration and frequency of the exposure was sensitive, and the exposure covered most of the critical window of sensitivity (if known). • **Deficient:** The duration or frequency of the exposure is not sensitive and did not include most of the critical window of sensitivity (if known). These limitations are expected to bias the results toward the null. • **Critically deficient:** The exposure design was not sensitive and is expected to strongly bias the results toward the null. The rationale should indicate the specific concern(s).
Sensitivity (continued)	**Outcome measures and results display**	**Endpoint sensitivity and specificity** Are the procedures sensitive and specific for evaluating the endpoint(s)/outcome(s) of interest? *Note:* • Sample size alone is not a reason to conclude an individual study is critically deficient. • Considerations related to adjustments/corrections to endpoint measurements (e.g., organ weight corrected for body weight) are addressed under results presentation.	For each endpoint/outcome or grouping of outcomes in a study: • Are there concerns regarding the sensitivity, specificity, and/or validity of the outcome measurement protocols? • Are there serious concerns regarding the sample size? • Are there concerns regarding the timing of the endpoint assessment?	A judgment and rationale for this domain will be given for each endpoint/outcome or group of outcomes investigated in the study. Examples of potential concerns include: • Selection of protocols insensitive or nonspecific for the endpoint of interest. • Evaluations did not include all treatment groups (e.g., only control and high dose). • Use of unreliable or invalid methods to assess the outcome. • Assessment of endpoints at inappropriate or insensitive ages, or without addressing known endpoint variation (e.g., due to circadian rhythms, estrous cyclicity). • Decreased specificity or sensitivity of the response due to the timing of endpoint evaluation, as compared with exposure (e.g., immediate endpoint assessment after exposure to chemicals with short-acting depressant or irritant effects; insensitivity due to prolonged period of nonexposure before testing).

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity (continued)	**Outcome measures and results display (continued)**	**Results presentation** Are the results presented in a way that makes the data usable and transparent? *Note:* *Potential issues associated with statistical analyses will be flagged for review by EPA statisticians and possible reanalysis (if information is available to do so, any reanalysis will be transparently presented). Any remaining limitations will be discussed during evidence synthesis or dose-response analyses (depending on the identified issue).*	For each endpoint/outcome or grouping of outcomes in a study: • Does the level of detail enable an informed interpretation of the results? • Are the data analyzed, compared, or presented in an inappropriate or misleading way?	A judgment and rationale for this domain will be given for each endpoint/outcome or group of outcomes investigated in the study. Examples of potential concerns include: • Nonpreferred presentation (e.g., developmental toxicity data averaged across pups in a treatment group, when litter responses are more appropriate; presentation of absolute organ-weight data when relative weights are more appropriate). • Failing to present quantitative results either in tables or figures. • Pooling data when responses are known or expected to differ substantially (e.g., across sexes or lifestages). • Failing to report on or address overt toxicity when exposure levels are known or expected to be highly toxic. • Lack of full presentation of the data (e.g., presentation of mean without variance data; concurrent control data are not presented).
Overall confidence		**Overall confidence** Considering the identified strengths and limitations, what is the overall confidence rating for the endpoint(s)/outcome(s) of interest? *Note:* *Reviewers will mark studies that are rated lower than* high *confidence due only to low sensitivity (i.e., bias towards the null) for additional consideration during evidence synthesis. If the study is otherwise well conducted and an effect is observed, the confidence may be increased.*	For each endpoint/outcome or grouping of outcomes in a study: • Were concerns (i.e., limitations or uncertainties) related to the reporting quality, risk of bias, or sensitivity identified? • If yes, what is their expected impact on the overall interpretation of the reliability and validity of the study results, including (when possible) interpretations of impacts on the magnitude or direction of the reported effects?	The overall confidence rating considers the likely impact of the noted concerns (i.e., limitations or uncertainties) in reporting, bias, and sensitivity on the results. A confidence rating and rationale will be given for each endpoint/outcome or group of outcomes investigated in the study.

[a]Several studies have characterized the relevance of randomization, allocation concealment, and blind outcome assessment in experimental studies (Hirst et al., 2014; Krauth et al., 2013; Macleod, 2013; Higgins and Green, 2011).

GLP = good laboratory practice; OECD = Organisation for Economic Co-operation and Development.

7.3. MECHANISTIC STUDY EVALUATION

Sections 10 and 11 outline an approach for focused consideration of information from mechanistic studies (including in vitro, in vivo, ex vivo, and in silico studies) where the specific analytical approach is targeted to the assessment needs, depending in part on the extent and nature of the phenotypic human and animal evidence. In this way, the mechanistic synthesis for a given health effect might range from a high-level summary (or detailed analysis) of potential mode of action to specific, focused questions needed to address important and impactful assessment uncertainties unaddressed by the available phenotypic studies (e.g., expected shape of the dose-response curve in the low-dose region, applicability of the animal evidence to humans, addressing susceptible populations). Individual study-level evaluation of mechanistic studies will not typically be pursued. Identifying assay-specific considerations for study endpoint evaluations on a case-by-case basis, however, might be necessary to provide a more detailed summary and evaluation for the most relevant individual mechanistic studies addressing a key assessment uncertainty. This may be done, for example, when the scientific understanding of a critical mechanistic event or mode of action lacks scientific consensus, when the reported findings on a critical mechanistic endpoint are conflicting, or when in vitro or in silico data or in vivo studies with mechanistic endpoints comprise the bulk of the evidence base and little or no evidence from epidemiological studies or animal bioassays is available. Particularly, for mercurous chloride, where no PECO-relevant epidemiological or animal data are available, ADME and mechanistic information will be carefully evaluated to identify similarities among the three inorganic mercury salts. This may provide insights into potential shared biological mechanisms and inferences that can be drawn from studies of mercuric chloride and mercuric sulfide.

8. ORGANIZING THE HAZARD REVIEW

The organization and scope of the hazard evaluation is determined by the available evidence for each inorganic mercury salt regarding metabolism and distribution, outcomes evaluated, and number of studies pertaining to each outcome, as well as the results of the evaluation of sources of bias and sensitivity. The hazard evaluations will be organized around organ systems (e.g., nervous system) informed by one or multiple related outcomes, and a decision will be made as to what level (e.g., organ system or subsets of outcomes within an organ system) to organize the synthesis.

Table 8-1 lists questions that could be asked of the evidence to assist with this decision. These questions derive from considerations and decisions made during development of the refined evaluation plan. Resolution of these questions will inform critical decisions about the organization of the hazard evaluation and help determine what studies might be useful in dose-response analyses.

Table 8-1. Querying the evidence to organize syntheses for human and animal evidence

Evidence	Questions	Follow-up questions
ADME	Given the known ADME issues for these inorganic mercury salts, do the data appear to differ by sex, species or lifestage? In addition, do toxicokinetic and toxicodynamic differences across compounds such as solubility affect the health outcomes?	Will separate analyses by factors such as sex, route of exposure, or by methods of dosing within a route of exposure (e.g., are large differences expected between gavage and dietary exposures) be needed? Are data available to inform which lifestages and what dosing regimens are more relevant to human exposure scenarios?
Outcomes	What outcomes are reported in studies? Are the data reported in a comparable manner across studies (similar output metrics at similar levels of specificity, such as adenomas and carcinomas quantified separately)?	At what level (hazard, grouped outcomes, or individual outcomes) will the synthesis be conducted? What commonalities will the outcomes be grouped by: • health effect, • exposure levels, • functional or population-level consequences (e.g., endpoints all ultimately leading to decreased fertility or impaired cognitive function), involvement of related biological pathways? How well do the assessed human and animal outcomes relate within a level of grouping?

Evidence	Questions	Follow-up questions
	Are there interrelated outcomes? If so, consider whether some outcomes are more useful of greater concern than others.	
	Does the evidence indicate greater sensitivity to effects (at lower exposure levels or severity) in certain subgroups (by age, sex, ethnicity, lifestage)? Should the hazard evaluation include a subgroup analysis?	
	Does incidence or severity of an outcome increase with duration of exposure or a particular window of exposure? What exposure time frames are relevant to development or progression of the outcome?	
	Is there mechanistic evidence that informs how outcomes might be grouped?	
	How robust is the evidence for specific outcomes? • What outcomes are reported by both human and animal studies and by one or the other? Were different animal species and sexes (or other important population-level differences) tested? • In general, what are the study confidence conclusions of the studies (*high*, *medium*, *low*, *uninformative*) for the different outcomes? Is there enough evidence from *high* and *medium* confidence studies to draw conclusions about causality?	What outcomes should be highlighted? Should the others be synthesized at all? Would comparisons by specific limitations be informative?
Dose-response	Did some outcomes include better coverage of exposure ranges that may be most relevant to human exposure than others?	What outcomes and studies are informative for developing toxicity values?
	For which outcomes are sufficient data available to draw conclusions about dose-response? Are there outcomes with study results of sufficient similarity (e.g., an established linkage in a biological pathway) to allow examination or calculation of common measures of effect across studies? Do the mechanistic data identify surrogate or precursor outcomes adequate for dose-response analysis?	
	Are there subgroups that exhibit responses at lower exposure levels than others?	
	Are there findings from ADME studies that could inform data-derived extrapolation factors, or link effects between humans and experimental animals? Can findings from toxicokinetics and toxicodynamics inform extrapolation factors	What studies might be used to develop non-default UFs? Is there a common internal dose metric that can be used to compare species or routes of exposure?

ADME = absorption, distribution, metabolism, and excretion; UF = uncertainty factor.

8.1. EPIDEMIOLOGY STUDY EVALUATION

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 negatives (i.e., Types I and II errors), while study sensitivity is typically concerned with identifying the latter.

The principles and framework used for evaluating epidemiology studies are based on the Cochrane Risk of Bias in Nonrandomized Studies of Interventions [ROBINS-I; (Sterne et al., 2016)] but modified to address environmental and occupational exposures. Core and prompting questions, shown in Table 7-1, are used to collect information to guide evaluation of each domain. Core questions represent key concepts, while the prompting questions help the reviewer focus on relevant details under each key domain. Table 7-1 also includes criteria that apply to all exposures and outcomes.

Table 8-2. Questions and criteria for evaluating each domain in epidemiology studies

Domain and core question	Prompting questions	Follow-up questions	Criteria 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 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, 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 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 potential for bias is a concern, is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	**Good** • Valid exposure assessment methods used, which represent the etiologically relevant period of interest. • Exposure misclassification is expected to be minimal. **Adequate** • Valid exposure assessment methods used, which represent the etiologically relevant period of interest. • Exposure misclassification could 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 whether it is influencing the effect estimate is uncertain. • 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 other evidence of exposure misclassification would be expected to notably change the effect estimate. **Critically deficient** • Exposure measurement does not characterize the etiologically relevant period of exposure or is not valid. • Evidence exists 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	Criteria 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 affected by knowledge, 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)?	**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*.

Domain and core question	Prompting questions	Follow-up questions	Criteria that apply to most exposures and outcomes
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 on the basis of knowledge of exposure or preclinical disease symptoms? Was entry into, 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 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 potential for bias is a concern, 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 relative to symptoms? Is there a comparison of participants and nonparticipants to address whether differential selection is likely?	**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, appropriate rationale is given 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 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 is likely to have had 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).

Domain and core question	Prompting questions	Follow-up questions	Criteria that apply to most exposures and outcomes
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 potential for bias is a concern, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	**Good** • Conveys strategy for identifying key confounders. This may include a priori biological consideration, published literature, causal diagrams, or statistical analyses, with the 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 likely to be influential colliders or intermediates on the causal pathway. • Key confounders are evaluated appropriately and considered unlikely sources of substantial confounding. This often will include: o Presenting the distribution of potential confounders by levels of the exposure of interest or the outcomes of interest (with amount of missing data noted); o Consideration that potential confounders were rare among the study population, or were expected to be poorly correlated with exposure of interest; o Consideration of the most relevant functional forms of potential confounders; o Examination of the potential impact of measurement error or missing data on confounder adjustment; or o Presenting a progression of model results with adjustments for different potential confounders, if warranted. **Adequate** • Similar to *good* but might not have included all key confounders, or less detail might be available on the evaluation of confounders (e.g., sub bullets in *good*). That residual confounding could explain part of the observed effect is possible, but concern is minimal. **Deficient** • Does not include variables in the models shown to be influential colliders or intermediates on the causal pathway. • And any of the following: o The potential for bias to explain some results is high based on an inability to rule out residual confounding, such as a lack of demonstration that key confounders of the exposure-outcome relationships were considered;

Domain and core question	Prompting questions	Follow-up questions	Criteria that apply to most exposures and outcomes
			o Descriptive information on key confounders (e.g., their relationship relative to the outcomes and exposure levels) are not presented; or o 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 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 results were most likely due to bias.
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 or lifestages 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 potential for bias is a concern, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	**Good** • Use of an optimal characterization of the outcome variable, including presentation of subgroup- or lifestage-specific comparisons (as appropriate for the outcome). • 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 might be absent (e.g., examination of outliers). **Adequate** • Same as 'Good', except:

Domain and core question	Prompting questions	Follow-up questions	Criteria that apply to most exposures and outcomes
			• Descriptive information about exposure provided (where applicable) but might be incomplete; might not have discussed missing data, cut-points, or shape of distribution(s). • 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. • 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 appropriate justification given for restricting the amount and type of results shown? • Are only statistically significant results presented?	If potential for bias is a concern, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	**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.

Domain and core question	Prompting questions	Follow-up questions	Criteria that apply to most exposures and outcomes
			• Only statistically significant results were reported.
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 or lifestage 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? • Do other aspects related to risk of bias or otherwise raise concerns about sensitivity?		**Adequate** • The range of exposure levels provides adequate variability to evaluate the associations of relevance. • 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, lifestage, 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 effect. • 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.

1

9. DATA EXTRACTION OF STUDY METHODS AND RESULTS

Data extraction and content management will be carried out using the HAWC (web links will be shared in the individual assessment). Data extraction elements that could be collected from epidemiological, controlled human exposure, animal toxicological, and in vitro studies are described in HAWC (https://hawcprd.epa.gov/about/). The level of extraction conducted in HAWC is more detailed than that described during the literature inventory process (see Section 4.1). Not all studies that meet the PECO criteria go through HAWC data extraction. Studies evaluated as being *uninformative* are not considered further and therefore would not undergo data extraction. The same might be true for *low* confidence studies if enough *medium* and *high* confidence studies (e.g., on an outcome) are available. In addition, any outcomes deprioritized when refining the evaluation plan (see Section 5) might have only have a literature inventory level of extraction. During HAWC extraction, all findings are considered for extraction, regardless of statistical significance.

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. (Note: The following browsers are supported for accessing HAWC: Google Chrome [preferred], Mozilla Firefox, and Apple Safari. Errors in functionality occur when viewed with Internet Explorer.) For quality control, data extraction will be performed by one member of the evaluation team and independently verified by at least one other member. Discrepancies in data extraction will be resolved by discussion or consultation with a third member of the evaluation team. Digital rulers, such as WebPlotDigitizer (http://arohatgi.info/WebPlotDigitizer/), will be used to extract numerical information from figures, and their use will be documented during extraction.

As previously described, routine attempts will be made to obtain missing information from study authors regarding human and animal health effect studies, if considered influential during study evaluations (see Section 7) or when information important for dose-response analysis or interpretations of significance (e.g., missing group size or variance descriptors such as standard deviation or confidence interval) can be provided. Missing data from individual mechanistic (e.g., in vitro) studies generally will not be sought. Outreach to study authors or designated contact persons will be documented and considered unsuccessful if researchers do not respond to email or phone requests within 1 month of initial attempt(s) to contact.

9.1. STANDARDIZING REPORTING OF EFFECT SIZES

Quantitative outcomes will be provided in their original units for all study groups. In addition, results from outcome measures will be transformed when possible to a common metric to help compare distinct but related outcomes measured with different scales. These standardized effect size estimates facilitate systematic evaluation and evidence integration for hazard identification. The following summary of effect size metrics by data type outlines issues in selecting the most appropriate common metric for a collection of related endpoints (Vesterinen et al., 2014).

Common metrics for continuous outcomes include:

- *Absolute difference in means.* This metric is the difference between the means in the control and treatment groups, expressed in the units in which the outcome is measured. When the outcome measure and its scale are the same across all studies, this approach is the simplest to implement.

- *Percent control response (or normalized mean difference [NMD]).* Percent control group calculations are based on means. Standard deviation (or standard error) values presented in the studies for these normalized effect sizes also can be estimated if sufficient information has been provided. Note that some outcomes reported as percentages, such as mean percentage of affected offspring per litter, can lead to distorted effect sizes when further characterized as percentage change from control. Such measures are better expressed as absolute difference in means, or preferably transformed to incidences using approaches for event or incidence data (see below).

- *Standardized mean difference.* The NMD approach above is relevant to ratio scales, but sometimes inferring what a "normal" animal would score is not possible, such as when data for animals without lesions are not available. In these circumstances, standardized mean differences can be used. The difference in group means is divided by a measure of the pooled variance to convert all outcome measures to a standardized scale with units of standard deviation. This approach also can be applied to data for which different measurement scales are reported for the same outcome measure (e.g., different measures of lesion size such as infarct volume and infarct area).

Common metrics for event or incidence data include:

- *Percent change from control*. This metric is analogous to the NMD approach described for continuous data above.

- *Binary outcomes.* For binary outcomes such as the number of individuals that developed a disease or died, and with only one treatment evaluated, data can be represented in a 2 × 2 table. Note that when the value in any cell is 0, 0.5 is added to each cell to avoid problems with the computation of the standard error. For each comparison, the odds ratio (OR) and its standard error can be calculated. Odds ratios are normally combined on a logarithmic scale.

An additional approach for epidemiology studies is to extract adjusted statistical estimates when possible rather than rely on unadjusted or raw estimates.

Considering the variability associated with effect size estimates is important, with better powered studies generally showing more precise estimates. Effect size estimation can be influenced, however, by such factors as variances that differ substantially across treatment groups or by lack of information to characterize variance, especially for animal studies in biomedical research (Vesterinen et al., 2014). The assessment will consider the nature of any variance issues and ensure the associated uncertainties are clarified and accounted for during the evidence synthesis process (see Section 10).

9.2. STANDARDIZING ADMINISTERED DOSE LEVELS/CONCENTRATIONS

Exposures will be standardized to common units. Exposure levels in oral studies will be expressed in units of mg/kg-day. When study authors provide exposure levels as concentrations in the diet or drinking water, dose conversions will be made using study-specific food or water consumption rates and body weights when available. Otherwise, EPA defaults will be used (U.S. EPA, 1988b), addressing age and study duration as relevant for the species/strain and sex of the animal of interest. Assumptions used in performing dose conversions will be documented in HAWC.

10. SYNTHESIS OF EVIDENCE

For this assessment, evidence synthesis and integration are considered distinct, but related, processes. The syntheses of separate bodies of evidence (i.e., human, animal, and mechanistic evidence) described in this section will directly inform the integration across all evidence to draw an overall judgment for each assessed human health effect (as described in Section 11). The phrase "evidence integration" used here is analogous to the phrase "weight of evidence" used in some other assessment processes (EFSA, 2017; U.S. EPA, 2017; NRC, 2014; U.S. EPA, 2005a).[12]

For each potential human health effect (or smaller subset of related outcomes), EPA will separately synthesize the available phenotypic human and animal evidence pertaining to that potential health effect. Mechanistic evidence also will be considered in targeted analyses conducted before, during, and after developing syntheses of the phenotypic human and animal evidence. The results of the mechanistic evidence analyses will be used to inform key uncertainties, depending on the extent and nature of the human and animal evidence. Thus, apart from the predefined mechanistic analyses, the human and animal evidence syntheses (or the lack of phenotypic data in humans and animals) help determine the approach to be taken in synthesizing the available mechanistic evidence. In this way, the mechanistic synthesis might range from a high-level summary (or detailed analysis) of potential mechanisms of action to specific, focused questions needed to address key uncertainties unaddressed by the phenotypic human and animal evidence (e.g., shape of the dose-response curve at low doses, applicability of the animal evidence to humans, addressing susceptible populations). Each synthesis will provide a summary discussion of the available evidence that addresses considerations regarding causation. These considerations are adapted from considerations for causality introduced by Austin Bradford Hill (Hill, 1965): consistency, dose-response relationship, strength of the association, temporal relationship, biological plausibility, coherence, and "natural experiments" in humans [see additional discussion in U.S. EPA (2005a) and U.S. EPA (1994)]. Importantly, the evidence synthesis process explicitly considers and incorporates the conclusions from the individual study evaluations.

[12]This revision has been adopted primarily on the basis of the 2014 NAS review of IRIS (NRC, 2014): "The present committee found that the phrase *weight of evidence* has become far too vague as used in practice today and thus is of little scientific use. In some accounts, it is characterized as an oversimplified balance scale on which evidence supporting hazard is placed on one side and evidence refuting hazard on the other... The present committee found the phrase *evidence integration* to be more useful and more descriptive of what is done at this point in an IRIS assessment—that is, IRIS assessments must come to a judgment about whether a chemical is hazardous to human health and must do so by integrating a variety of evidence."

Table 10-1. Information most relevant to describing primary considerations for assessing causality during evidence syntheses

Consideration	Description of the consideration and its application in IRIS syntheses
Study confidence	Description: Incorporates decisions about study confidence within each consideration. Application: In evaluating the evidence for each causality consideration described in the following rows, the syntheses will consider study confidence decisions. *High* confidence studies carry the most weight. The syntheses will consider the specific limitations and strengths identified during study evaluation and describe how these informed each consideration.
Consistency	Description: Examines the similarity of results (e.g., direction; magnitude) across studies. Application: Syntheses will evaluate the homogeneity of findings on a given outcome or endpoint across studies. When inconsistencies exist, the syntheses consider whether results were "conflicting" (i.e., unexplained positive and negative results in similarly exposed human populations or in similar animal models) or "differing" (i.e., mixed results explainable by, for example, differences between human populations, animal models, exposure conditions, or study methods) (U.S. EPA, 2005a). These considerations are based on analyses of potentially important explanatory factors such as: • Confidence in studies' results, including study sensitivity (e.g., some study results that appear to be inconsistent might be explained by potential biases or other attributes that affect sensitivity). • Exposure, including route (if applicable) and administration methods, levels, duration, timing with respect to outcome development (e.g., critical windows), and exposure assessment methods (i.e., in epidemiology studies), including analytical units and specific groups being compared. • Specificity and sensitivity of the endpoint for evaluating the health effect in question (e.g., functional measures can be more sensitive than organ weights). • Populations or species, including consideration of potential susceptible groups or differences across lifestage at exposure or endpoint assessment. • Toxicokinetic information explaining observed differences in responses across route of exposure, other aspects of exposure, species, sexes, or lifestages. The interpretation of consistency will emphasize biological significance, to the extent that it is understood, over statistical significance. Statistical significance from suitably applied tests (which might involve consultation with an EPA statistician) adds weight when biological significance is not well understood. Consistency in the direction of results increases confidence in that association, even in the absence of statistical significance. In some cases, considering the potential for publication bias could be helpful to provide context to interpretations of consistency.[a]

Consideration	Description of the consideration and its application in IRIS syntheses
Strength (effect magnitude) and precision	Description: Examines the effect magnitude or relative risk, based on what is known about the assessed endpoint(s), and considers the precision of the reported results based on analyses of variability (e.g., confidence intervals; standard error). This might include consideration of the rarity or severity of the outcomes. Application: Syntheses will analyze results both within and across studies and could consider the utility of combined analyses (e.g., meta-analysis). While larger effect magnitudes and precision (e.g., $p < 0.05$) help reduce concerns about chance, bias, or other factors as explanatory, syntheses should also consider the biological or population-level significance of small effect sizes.
Biological gradient/ dose-response	Description: Examines whether the results (e.g., response magnitude; incidence; severity) change in a manner consistent with changes in exposure (e.g., level; duration), including consideration of changes in response after cessation of exposure. Application: Syntheses will consider relationships both within and across studies, acknowledging that the dose-response relationship (e.g., shape) can vary depending on other aspects of the experiment, including the biology underlying the outcome and the toxicokinetics of the chemical. Thus, when dose-dependence is lacking or unclear, the synthesis will also consider the potential influence of such factors on the response pattern.
Coherence	Description: Examines the extent to which findings are cohesive across different endpoints related to, or dependent on, one another (e.g., based on known biology of the organ system or disease, or mechanistic understanding such as toxicokinetic/dynamic understanding of the chemical or related chemicals). In some instances, additional analyses of mechanistic evidence from research on the chemical under review or related chemicals that evaluate linkages between endpoints or organ-specific effects might be needed to interpret the evidence. These analyses could require additional literature search strategies. Application: Syntheses will consider potentially related findings, both within and across studies, particularly when relationships are observed within a cohort or within a narrowly defined category (e.g., occupation, strain or sex, lifestage of exposure). Syntheses will emphasize evidence indicative of a progression of effects, such as temporal- or dose-dependent increases in the severity of the type of endpoint observed. If an expected coherence between findings is not observed, possible explanations should be explored including those related to the biology of the effects and the sensitivity and specificity of the measures used.

Consideration	Description of the consideration and its application in IRIS syntheses
Mechanistic evidence related to biological plausibility	Description: There are multiple uses for mechanistic information, and this consideration overlaps with "coherence." This consideration examines the biological support (or lack thereof) for findings from the human and animal health effect studies and becomes more influential on the hazard conclusions when notable uncertainties in the strength of those sets of studies exist. These analyses also can improve understanding of dose- or duration-related development of the health effect. Absent human or animal evidence of apical health endpoints, the synthesis of mechanistic information could drive evidence integration conclusions (when such information is available). Application: Syntheses can evaluate evidence on precursors, biomarkers, or other molecular or cellular changes related to the health effect(s) of interest to describe the likelihood that the observed effects result from exposure. This evaluation will entail an analysis of existing evidence, and not simply whether a theoretical pathway can be postulated. This analysis might not be limited to evidence relevant to the PECO but could include evaluations of biological pathways (e.g., for the health effect; established for other, possibly related, chemicals). Any such synthesis of mechanistic evidence will consider the sensitivity of the mechanistic changes and the potential contribution of alternative or previously unidentified mechanisms of toxicity.
Natural experiments	Description: Specific to epidemiology studies and rarely available, this consideration examines effects in populations that have experienced well-described, pronounced changes in chemical exposure (e.g., lead exposures before and after banning lead in gasoline). Application: Compared to other observational designs, natural experiments have the benefit of dividing people into exposed and unexposed groups without their having influenced their own exposure status. During synthesis, associations in *medium* and *high* confidence natural experiments can substantially reduce concerns about residual confounding.

[a]Publication bias involves the influence of the direction, magnitude, or statistical significance of the results on the likelihood of a paper being published; it can result from decisions made, consciously or unconsciously, by study authors, journal reviewers, and journal editors (Dickersin, 1990). When evidence of publication bias is present for a set of studies, less weight might be placed on the consistency of the findings for or against an effect during evidence synthesis and integration.

PECO = populations, exposures, comparators, and outcomes.

Data permitting, the syntheses also will discuss analyses relating to potential susceptible populations.[13] These analyses will be based on knowledge about the health outcome or organ system affected, demographics, genetic variability, lifestage, health status, behaviors or practices,

[13]Various terms have been used to characterize populations that could be at increased risk of developing health effects from exposure to environmental chemicals, including "susceptible," "vulnerable," and "sensitive." Furthermore, these terms have been inconsistently defined across the scientific literature. The term susceptibility is used in this protocol to describe populations or lifestages at increased risk, focusing on intrinsic biological factors that can modify the effect of a specific exposure, but also considering social determinants or behaviors that might increase susceptibility. Factors resulting in higher exposures to specific groups (e.g., proximity, housing, occupation), however, typically will not be analyzed to describe increased risk among specific populations or subgroups.

and social determinants (Table 10-2). This information will be used to draw conclusions regarding potential susceptibility among specific populations or subgroups in a separate section. This summary will describe concerns across the available evidence for all potential human health effects and will inform both hazard identification and dose-response analyses.

Table 10-2. Individual and social factors that might increase susceptibility to exposure-related health effects

Factor	Examples
Demographic	Gender, age, race/ethnicity, education, income, occupation, geography
Genetic variability	Polymorphisms in genes regulating cell cycle, DNA repair, cell division, cell signaling, cell structure, gene expression, apoptosis, and metabolism
Lifestage	In utero, childhood, puberty, pregnancy, women of childbearing age, old age
Health status	Pre-existing conditions or disease such as psychosocial stress, elevated body mass index, frailty, nutritional status, chronic disease
Behaviors or practices	Diet, mouthing, smoking, alcohol consumption, pica, subsistence or recreational hunting, and fishing
Social determinants	Income, socioeconomic status, neighborhood factors, health care access, and social, economic, and political inequality

EPA ExpoBox Exposure Assessment Tools, based on EPA's *Guidelines for Exposure Assessment* (U.S. EPA, 1992).
DNA = deoxyribonucleic acid.

10.1. SYNTHESES OF HUMAN AND ANIMAL HEALTH EFFECTS EVIDENCE

The syntheses of the human and animal health effects evidence will focus on describing aspects of the evidence that best inform causal interpretations, including the exposure context examined in the sets of studies. Each evidence synthesis will be based primarily on studies of *high* and *medium* confidence. *Low* confidence studies could be used if few or no studies with higher confidence are available to help evaluate consistency, or if the study designs of the *low* confidence studies address notable uncertainties in the set of *high* or *medium* confidence studies on a given health effect. If *low* confidence studies are used, a careful examination of risk of bias and sensitivity with potential impacts on the evidence synthesis conclusions will be included in the narrative.

As previously described, these syntheses will articulate the strengths and the weaknesses of the available evidence organized around the considerations described in Table 11-2 and issues that stem from the evaluation of individual studies (e.g., concerns about bias or sensitivity). If possible, results across studies will be compared using graphs and charts or other data visualization strategies. The analysis will typically include examination of results stratified by any or all of the following: study confidence classification (or specific issues within confidence evaluation domains); population or species; exposures (e.g., level, patterns [intermittent or continuous], duration, intensity); sensitivity (e.g., low vs. high), and other factors that might have been identified during

study evaluation or analyses of key science issues (see Section 6). The number of studies and the differences encompassed by the studies will determine the extent to which specific factors can be examined for use in stratifying study results. Additionally, for both the human and animal evidence syntheses, if supported by the available data, additional analyses across studies (such as meta-analysis) might also be conducted.

10.2. MECHANISTIC INFORMATION

The synthesis of mechanistic information informs the integration of health effects evidence for both hazard identification (i.e., biological plausibility or coherence of the available human or animal evidence; inferences regarding human relevance, or the identification of susceptible populations and lifestages across the human and animal evidence) and dose-response evaluation.

As introduced in prior sections, several key science issues essential to consider in the inorganic mercury salt assessment focused on analysis and synthesis of ADME/mechanistic information. Mechanistic evidence includes any experimental measurement related to a health outcome that provides information about the biological or chemical events associated with phenotypic effects. These measurements can improve understanding of the mechanisms involved in the toxic effects following exposure to a chemical but generally are not considered adverse outcomes. Mechanistic data are reported in a diverse array of observational and experimental studies across species, model systems, and exposure paradigms, including in vitro, in vivo (by various routes of exposure), ex vivo, and in silico studies, and across a wide spectrum of diverse endpoints.

Evaluations of mechanistic information typically differ from evaluations of phenotypic evidence (e.g., from routine toxicology studies). This is primarily because mechanistic data evaluations consider the support for and involvement of specific events or sets of events within the context of a broader research question (e.g., support for a hypothesized mechanism; consistency with known biological processes), rather than evaluations of individual apical endpoints considered in relative isolation. Such analyses are complicated because a chemical might operate through multiple mechanistic pathways, even if one hypothesis dominates the literature (U.S. EPA, 2005a). Similarly, multiple mechanistic pathways might interact to cause an adverse effect. Thus, pragmatic and stepwise approaches will be considered in reviewing this evidence for this inorganic mercury salts assessment. The format of these syntheses is expected to vary from a short narrative summary of existing knowledge to an in-depth analysis and weighing of the evidence underlying multiple mechanistic events, depending on data availability and the criticality of the assessment-specific uncertainty(ies).

10.2.1. Toxicokinetic/ADME Information and Pharmacokinetic (PK)/Physiologically Based Pharmacokinetic (PBPK) Models

One key mechanistic issue for inorganic mercury salts pertains to the toxicokinetics and ADME of these chemicals. For example, some reports suggest that animal strains, sex, and dose

influence toxicokinetics and sensitivity related to toxicity of inorganic mercury compounds (Ekstrand et al., 2010; Nielsen, 1992). Furthermore, reports have suggested that the chemical form of mercury is a major determinant of mercury disposition and toxicity and that cinnabar (HgS) was generally considered less toxic than mercuric chloride (Liu et al., 2016; Havarinasab et al., 2007). Despite the underlying mechanisms that remain to be elucidated, it is possible that different toxicity of various inorganic mercurial compounds are attributed to the toxicokinetic and toxicodynamic dissimilarities. Such information will be used to further understand the ADME and toxicokinetic differences, all of which will be characterized in the evidence synthesis sections for the three salts.

Evidence also suggests sex, species, and age-related differences in the toxicokinetics within individual inorganic mercury salts such as mercuric chloride (Ekstrand et al., 2010; Nielsen, 1992) (Nielsen and Andersen, 1990). For instance, female mice (Bom:NMRI and CBA/Bom strains) tended to retain considerably less mercuric chloride in the kidneys than did male mice, presumably because of testosterone regulation of mercury retention (Nielsen and Andersen, 1990). Sex difference in renal mercury deposition further has been attributed to reduced sulfhydryl groups in female as compared to male mice (Muraoka and Itoh, 1980). Sex differences likewise were found in hepatic mercury content (Nielsen et al., 2006). As the kidneys, followed by the liver, are the dominating organ for accumulation and distribution of mercury salts (Berlin et al., 2007; ATSDR, 1999), evaluation of the potential influence of sex on mercury salts toxicokinetics is necessary.

In addition to sex differences, mechanistic data also revealed that the toxicokinetics of inorganic mercury salts was age and species dependent. For instance, absorption of inorganic mercury in sucking pups is prolonged as compared to adult mice, indicating the uptake of mercury salts could be higher in younger animals (Sundberg et al., 1999). The age difference in gastrointestinal absorption in infants has been attributed to the higher gastric pH in younger animals (Sundberg et al., 1999; Bearer, 1995). Another in vitro study using intestinal sacs and brush border membrane vesicles of male Wistar rats also made similar observations: The transport of mercuric chloride through the intestinal wall increases with pH values ranging from pH 5.5 to 7.4 (Endo et al., 1990).

Of note, variations in toxicokinetics of inorganic mercury also were observed among animal species and strains (Ekstrand et al., 2010). For instance, in a study by Ekstrand et al. (2010), inbred, H-2-congenic A.SW and B10.S mice and their F1 and F2 hybrids were given mercuric chloride (2.0 mg Hg/L drinking water) orally with traces of ^{203}Hg. Whole body retention and renal retention increased by 20–30% and 2- to 5-fold, respectively, in A.SW males relative to that in A.SW females and B10.S mice, despite similar mercury intake among mice strains. Thus, strain difference in the toxicokinetics of mercury salts (e.g., mercuric chloride) were attributed primarily to differences in elimination kinetics rather than absorption among these genetically heterogeneous mice (Ekstrand et al., 2010).

10.2.2. Strategies To Identify Analogues To Inform Read-Across for Mercurous Chloride

Based on the preliminary literature search, appropriate data for conducting hazard identification and dose-response analysis are not available for mercurous chloride. Thus, an analogue-based, read-across approach will be attempted for this salt to calculate toxicity values.

The analogue approach allows for the use of data from related compounds to calculate toxicity values when data for the compound of interest are limited or unavailable. Details regarding searches and methods for surrogate analysis are presented in Wang et al., (2012). Three types of potential surrogates (structural, metabolic, and toxicity-like) will be identified to facilitate the final surrogate chemical selection. The surrogate approach might or might not be route specific or applicable to multiple routes of exposure. All information will be considered together as part of the final weight-of-evidence (WOE) approach to select a potentially suitable surrogate both toxicologically and chemically.

This WOE approach will be used to evaluate information from potential candidate surrogates, as described by Wang et al., (2012). Commonalities in structural/physicochemical properties, toxicokinetics, metabolism, toxicity, or MOA between potential surrogates and chemical(s) of concern will be identified. Emphasis will be given to toxicological or toxicokinetic similarity over structural similarity. Surrogate candidates will be excluded if they do not have commonality or demonstrate significantly different physicochemical properties, and toxicokinetic profiles that set them apart from the pool of potential surrogates and chemical(s) of concern. From the remaining potential surrogates, the most appropriate surrogate will be selected. The selection will be based on consideration of the biologically and toxicologically relevant analogues, structural similarities as well as sensitivity of toxicological values.

11. EVIDENCE INTEGRATION

For the analysis of human health outcomes that might result from chemical exposure, the inorganic mercury salts assessment will draw integrated judgments across available evidence for each assessed health effect. The evidence integration judgments include interpretations drawn regarding the support provided by the individual lines of evidence (i.e., human, animal, and mechanistic evidence) based on the structured application of an adapted set of considerations first introduced by Austin Bradford Hill (Hill, 1965), which are directly informed by the summary discussions of each line of evidence during evidence synthesis (see Section 10). This includes evaluations of mechanistic evidence relevant to the identified key science issues (see Section 6) prior to or in parallel with evaluations of the phenotypic data in human and animal studies. During evidence integration, a structured and documented, two-step process will be used, as follows (and depicted in Figure 11-1):

- Step 1: Judgments regarding the strength of the evidence from the available human and animal studies will be made in parallel, but separately. Building from the separate syntheses of the human and animal evidence, the strength of the evidence from the available human and animal studies will be judged using a structured evaluation of an adapted set of considerations first introduced by Austin Bradford Hill (Hill, 1965). Table 11-2 describes these structured evaluations and the explicit consideration of study confidence within each evaluation domain. Based on the approaches and considerations described in Section 10, these judgments will incorporate the relevant mechanistic evidence (or MOA understanding) that informs the biological plausibility and coherence within the available human or animal health effect studies. Note that at this stage, the animal evidence judgment does not yet consider the human relevance of that evidence.

- Step 2: The animal and human evidence judgments will be combined to draw an overall evidence integration judgment(s). As described in section 10, this step will incorporate inferences drawn based on information on the human relevance of the animal and mechanistic evidence, coherence across the human and animal lines of evidence, and other important information (e.g., judgments regarding susceptibility). Note that without evidence to the contrary, the human relevance of animal findings is assumed.

- The summary judgments as to whether and to what extent the available evidence for each potential human health effect indicates that inorganic mercury salts exposure has the potential to be hazardous to humans will be characterized fully in the evidence integration narrative and abbreviated using the shorthand described in Figure 11-1.

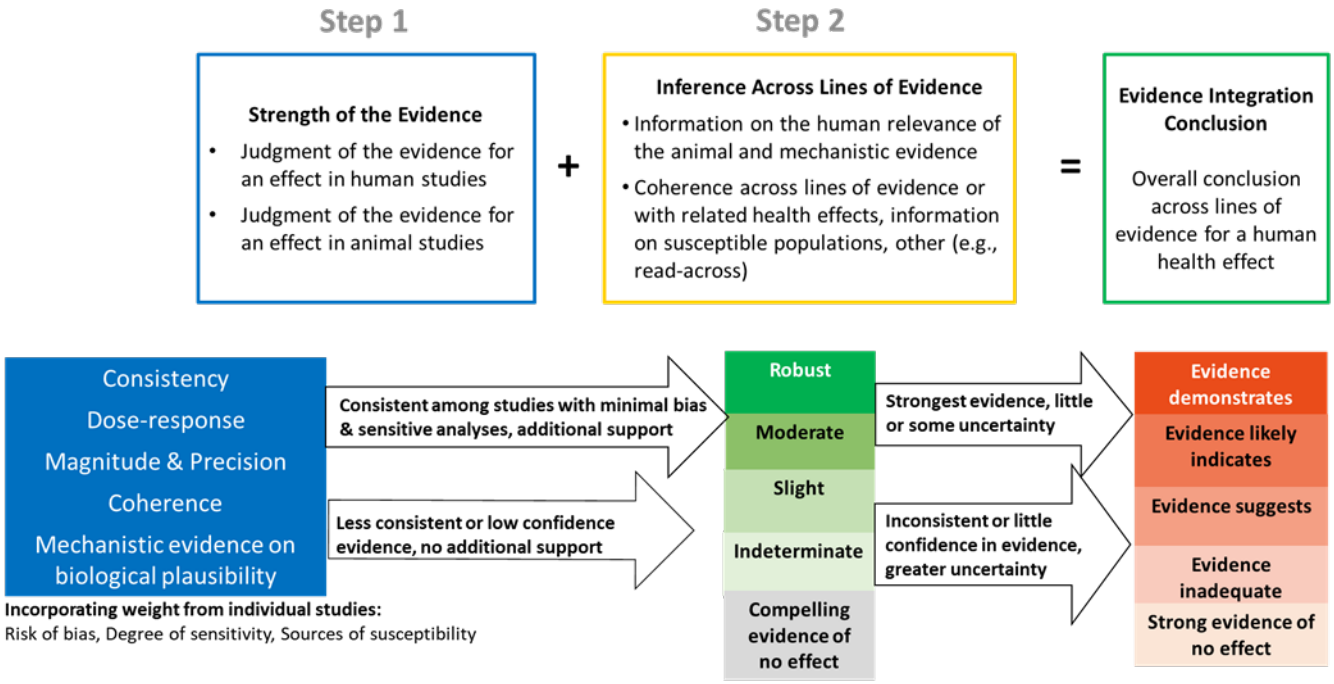


Figure 11-1. Process for evidence integration.

The decision points within the structured two-step evidence integration process will be summarized in an evidence profile table for each health effect or category of effects (see Table 11-1 for a template version) in support of the evidence integration narrative. The specific decision frameworks for the structured evaluation of the human and animal evidence (Step 1) and for drawing the overall evidence integration judgment (Step 2) are described in Section 11.1 and 11.2, respectively. This process is similar to that used by the Grading of Recommendations Assessment, Development, and Evaluation [GRADE; (Morgan et al., 2016; Guyatt et al., 2011; Schünemann et al., 2011)], which arrives at an overall integration conclusion based on consideration of the body of evidence. As described in Section 10, the human, animal, and mechanistic evidence syntheses serve as inputs, providing a foundation for the evidence integration decisions; thus, the major conclusions from these syntheses will be summarized in the evidence profile table (see Table 11-1 for a template version) supporting the evidence integration narrative. The evidence profile tables on each potential human health effect evaluated will summarize the judgments and their evidence basis for each step of the structured evidence integration process. In the evidence profile table, separate sections will be included for the summarizing the human and animal evidence and drawing Step 1 judgments, for the inferences drawn across lines of evidence, and for the overall evidence integration judgment. Overall, the evidence profile table presents a summary of the expert judgments as well as the key information from the different lines of evidence that informs each decision.

Table 11-1. Evidence profile table template

Evidence Summary and Interpretation					Inferences and Summary Judgment
Studies, outcomes, and confidence	**Summary of key findings**	**Factors that increase certainty**	**Factors that decrease certainty**	**Judgments and rationale**	Describe overall evidence integration judgment(s): ⊕⊕⊕ Evidence demonstrates ⊕⊕⊙ Evidence indicates (likely) ⊕⊙⊙ Evidence suggests ⊙⊙⊙ Evidence inadequate – – – Strong evidence supports no effect o Summarize the models and range of dose levels upon which the judgment(s) were primarily reliant Address human relevance of findings in animals Summarize cross-stream coherence Summarize potential susceptibility
Evidence from studies of exposed humans (may be separated by exposure route or other study design characteristic[a])					
<u>May be separate rows by outcome</u> References (or link) Study confidence Study design description (if informative)	Description of the primary results across human epidemiological and controlled exposure studies[c], and any human mechanistic evidence informing biological plausibility (e.g., precursor events linked to adverse outcomes)	Consistency Dose-response gradient Coherence of effects Large or concerning magnitude of effect Mechanistic evidence providing plausibility *Medium* or *high* confidence studies[b]	Unexplained inconsistency Imprecision Lack of expected coherence *Low* confidence studies[b] *Evidence demonstrating implausibility*	Describe the strength of the evidence from human studies: ⊕⊕⊕ *Robust* ⊕⊕⊙ *Moderate* ⊕⊙⊙ *Slight* ⊙⊙⊙ *Indeterminate* – – – *Compelling evidence of no effect* o Summarize any important interpretations, and the primary basis for the judgment(s)	
Evidence from animal studies (may be separated by exposure route or other study design characteristic[a])					
<u>May be separate rows by outcome</u> References (or link) Study confidence Study design description (if informative)	Description of the primary results across animal toxicological studies[c], and any human mechanistic evidence informing biological plausibility (e.g., precursor events linked to adverse outcomes)	Consistency, replication Dose-response gradient Coherence of effects Large or concerning magnitude of effect Mechanistic evidence providing plausibility *Medium* or *high* confidence studies[b]	Unexplained inconsistency Imprecision Lack of expected coherence *Low* confidence studies[b] *Evidence demonstrating implausibility*	Describe the strength of the evidence from animal studies: ⊕⊕⊕ *Robust* ⊕⊕⊙ *Moderate* ⊕⊙⊙ *Slight* ⊙⊙⊙ *Indeterminate* – – – *Compelling evidence of no effect* o Summarize any important interpretations, and the primary basis for the judgment(s)	

Mechanistic evidence and supplemental information—may be separated (e.g., by exposure route or key uncertainty addressed)			Summarize any other critical inferences: o E.g., from MOA analysis o E.g., from read-across comparison
Biological events or pathways (or other)	**Summary of key findings and interpretation**	**Judgment(s) and rationale**	
May be separate rows by biological events or other feature of the approach used for analysis Generally, will cite evidence synthesis (e.g., for references; for detailed analysis) Does not have to be chemical-specific (e.g., read-across)	May include separate summaries, for example by study type (e.g., new approach methods vs. in vivo biomarkers), dose, or design *Interpretation*: Summary of expert interpretation for the body of evidence and supporting rationale *Key findings*: Summary of findings across the body of evidence (may focus on or emphasize highly informative designs or findings), including key sources of uncertainty or identified limitations of the study designs tested (e.g., regarding the biological event or pathway being examined)	Overall summary of expert interpretation across the assessed set of biological events, potential mechanisms of toxicity, or other analysis approach (e.g., AOP). Includes the primary evidence supporting the interpretation(s) Describes and substantiates the extent to which the evidence influences inferences across evidence streams Characterizes the limitations of the evaluation and highlights existing data gaps May have overlap with factors summarized for other streams	

[a]In addition to exposure route, the summaries of each evidence stream may include multiple rows (e.g., by study confidence, population, or species, if they informed the analysis of results heterogeneity or other features of the evidence). When data within an evidence stream are lacking or otherwise not informative to the evidence integration decisions, the summary subrows for that evidence stream may be abbreviated to more easily present this information.

[b]Study confidence, based on evaluation of risk of bias and study sensitivity (see Section 7), and information on susceptibility will be considered when evaluating the other factors that increase or decrease certainty (e.g., consistency). Notably, lack of findings in studies deemed insensitive neither increases nor decreases certainty. Typically, *medium* confidence in only a single study is not a factor that increases certainty, whereas *high* confidence in a single, extensive or rigorous study (e.g., a guideline study) is such a factor.

[c]If sensitivity issues were identified, describe the impact on reliability of the reported findings

11.1. INTEGRATION WITHIN THE HUMAN AND ANIMAL EVIDENCE

Before drawing overall evidence integration judgments about whether exposure to one of the inorganic mercury salts could cause certain health effect(s) in humans, separate judgments will be drawn regarding the strength of the available human and animal evidence. For each assessed health effect, the relevant mechanistic evidence in exposed humans and animals (or in their cells or other relevant new approach methods [NAMS] including in silico models), which will be synthesized based on the approaches and considerations in Sections 10.1 and 10.2, will be integrated with the evidence from the available studies of phenotypic effects in humans and animals. The evaluation of the strength of the human or animal health effects evidence (i.e., based on the considerations in Table 11-2) will preferably occur at the most specific health outcome level possible. If studies on a target system are sparse or varied, or if the evidence strength relies largely on the interpretation of coherence across related outcomes, then the analyses may need to be conducted at a broader health effect (or category of health effects) level. The factors judged to increase or decrease the interpreted certainly in the findings (i.e., strength of the evidence) will be summarized in tabular format using the evidence profile table template in Table 11-1 to transparently convey expert judgments made throughout the evidence synthesis and integration processes. The evidence profile table allows for consistent documentation of the supporting rationale for each decision.

Table 11-2. Considerations that inform evaluations of the strength of the human and animal evidence

Consideration	Increased evidence strength (of the human or animal evidence)	Decreased evidence strength (of the human or animal evidence)
The structured categories and criteria in Tables 11-3 and 11-4 will guide the application of strength-of-evidence judgments for an outcome or health effect. Evidence synthesis scenarios that do not warrant an increase or decrease in evidence strength for a given consideration will be considered "neutral" and are not described in this table (and, in general, will not be captured in the assessment-specific evidence profile tables).		
Risk of bias; sensitivity (across studies)	• An evidence base of *high* or *medium* confidence studies increases strength. Typically, *medium* confidence in only a single study is not a factor that increases certainty, whereas high confidence in a single, extensive or rigorous study (e.g., a guideline study) is such a factor.	• An evidence base of mostly *low* confidence studies decreases strength. An exception to this is an evidence base of studies in which the primary issues resulting in *low* confidence are related to insensitivity. This may increase evidence strength in cases where an association is identified because the expected impact of study insensitivity is towards the null. • Decisions to increase strength for other considerations in this table should generally not be made if there are serious concerns for risk of bias.
Consistency	• Similarity of findings for a given outcome (e.g., of a similar magnitude, direction) across independent studies or experiments increases strength, particularly when consistency is observed across populations (e.g., geographical location) or exposure scenarios in human studies, and across laboratories, populations (e.g., species), or exposure scenarios (e.g., duration; route; timing) in animal studies.	• Unexplained inconsistency [i.e., conflicting evidence; see (U.S. EPA, 2005a)] decreases strength. Generally, strength should not be decreased if discrepant findings reasonably can be explained by study confidence conclusions; variation in population or species, sex, or lifestage; exposure patterns (e.g., intermittent or continuous); exposure levels (low or high); or exposure duration.

Consideration	Increased evidence strength (of the human or animal evidence)	Decreased evidence strength (of the human or animal evidence)
Strength (effect magnitude) and precision	• Evidence of a large magnitude effect (considered either within or across studies) can increase strength. Effects of a concerning rarity or severity also can increase strength, even if they are of a small magnitude. • Precise results from individual studies or across the set of studies increases strength, noting that biological significance is prioritized over statistical significance.	• Strength may be decreased if effect sizes that are small in magnitude are concluded not to be biologically significant, or if there are only a few studies with imprecise results.
Biological gradient/dose-response	• Evidence of dose-response increases strength. Dose-response may be demonstrated across studies or within studies and it can be dose- or duration-dependent. It also may not be a monotonic dose-response (monotonicity should not necessarily be expected, e.g., different outcomes may be expected at low vs. high doses because of activation of different mechanistic pathways or induction of systemic toxicity at very high doses). • Decreases in a response after cessation of exposure (e.g., symptoms of current asthma) also may increase strength by increasing certainty in a relationship between exposure and outcome (this is most applicable to epidemiology studies because of their observational nature).	• A lack of dose-response when expected based on biological understanding and having a wide range of doses/exposures evaluated in the evidence base can decrease strength. • In experimental studies, strength may be decreased when effects resolve under certain experimental conditions (e.g., rapid reversibility after removal of exposure). However, many reversible effects, are of high concern. Deciding between these situations is informed by factors such as the toxicokinetics of the chemical and the conditions of exposure [see (U.S. EPA, 1998b)] endpoint severity, judgments regarding the potential for delayed or secondary effects, as well as the exposure context focus of the assessment (e.g., addressing intermittent or short-term exposures). • In rare cases, and typically only in toxicological studies, the magnitude of effects at a given exposure level might decrease with longer exposures (e.g., due to tolerance or acclimation). Like the discussion of reversibility above, a decision about whether this decreases evidence strength depends on the exposure context focus of the assessment and other factors. • If the data are not adequate to evaluate a dose-response pattern, then strength is neither increased nor decreased.

Consideration	Increased evidence strength (of the human or animal evidence)	Decreased evidence strength (of the human or animal evidence)
Coherence	• Biologically related findings within an organ system, or across populations (e.g., sex) increase strength, particularly when a temporal- or dose-dependent progression of related effects is observed within or across studies, or when related findings of increasing severity are observed with increasing exposure.	• An observed lack of expected coherent changes (e.g., well-established biological relationships) will typically decrease evidence strength. However, the biological relationships between the endpoints being compared and the sensitivity and specificity of the measures used need to be carefully examined. The decision to decrease evidence strength depends on the availability of evidence across multiple related endpoints for which changes would be anticipated, and it considers factors (e.g., dose and duration of exposure, strength of expected relationship) across the studies of related changes.
Mechanistic evidence related to biological plausibility	• Mechanistic evidence of precursors or health effect biomarkers in well-conducted studies of exposed humans or animals, in appropriately exposed human or animal cells, or other relevant human, animal, or in silico models (including new approach methods, NAMs) increases strength, particularly when this evidence is observed in the same cohort/population exhibiting the phenotypic health outcome. • Evidence of changes in biological pathways or support for a proposed MOA in appropriate models also increases strength, particularly when support is provided for rate-limiting or key events or conserved across multiple components of the pathway or MOA.	• Mechanistic understanding is not a prerequisite for drawing a conclusion that a chemical causes a given health effect (NTP, 2015; NRC, 2014); thus, an absence of knowledge should not be used as basis for decreasing strength. When mechanistic evidence does not exist or is inconclusive and the findings in humans or animals are judged not to conflict with current biological understanding, those findings are presumed to be real unless proven otherwise. • Mechanistic evidence in well-conducted studies (see examples of evidence types at left) that demonstrates the health effect(s) are unlikely to occur, or only likely to occur under certain scenarios (e.g., above certain exposure levels), can decrease evidence strength. A decision to decrease strength depends on an evaluation of the strength of the mechanistic evidence for and against biological plausibility, as well as the strength of the health effect-specific findings (e.g., stronger health effect data require more certainty in mechanistic evidence opposing plausibility).

For human and animal evidence, the analyses of each consideration in Table 11-2 will be used to develop a strength-of-evidence judgment. Tables 11-3 and 11-4 provide the example-based criteria that will guide how to draw the judgments for each health effect, and the terms that will be used to summarize those judgments. These terms are applied to human and animal evidence separately. Briefly, the terms describe judgments of the evidence strength as follows:

Robust and *Moderate* are standardized characterizations for judgments that the relevant effect(s) observed in humans or animals result from exposure to inorganic mercury salts; these two terms are differentiated by the quantity and quality of information available to rule out alternative explanations for the results. For example, repeated observations of effects by independent studies examining various aspects of exposure or response (e.g., different exposure settings, dose levels or patterns, populations or species, and related endpoints) will result in a stronger strength-of-evidence judgment.

Slight indicates situations in which there is some evidence indicating an association, but substantial uncertainties in the data exist to prevent judgments that that the relevant effect(s) observed in humans or animals can be reliably attributed to exposure to inorganic mercury salt of interest.

Indeterminate reflects evidence stream judgments when no studies are available, or situations when the evidence is inconsistent and/or primarily of *low* confidence.

Compelling evidence of no effect represents a situation in which extensive evidence across a range of populations and exposures has identified no effects/associations. This last scenario is seldom used because it requires a high degree of confidence in the conduct of individual studies, including consideration of study sensitivity, and comprehensive assessments of health outcomes and lifestages of exposure.

Publication bias can potentially result in strength-of-evidence judgments that are stronger than would be merited if the entire body of research were available. However, the existence of publication bias can be difficult to determine and is not a component of the strength-of-evidence framework for human or animal studies presented in this protocol. If potential publication bias is evaluated for an outcome, it may inform the level of certainty regarding the completeness of the assessment database for that outcome.

Table 11-3. Framework for strength-of-evidence judgments from studies in humans

Strength-of-evidence judgment	Description
Robust (⊕⊕⊕) …evidence in human studies (strong signal of effect with little residual uncertainty)	A set of *high* or *medium* confidence independent studies reporting an association between the exposure and the health outcome, with reasonable confidence that alternative explanations, including chance, bias, and confounding, can be ruled out across studies. The set of studies is primarily consistent, with reasonable explanations when results differ; and an exposure-response gradient is demonstrated. Supporting evidence, such as associations with biologically related endpoints within or across human studies (coherence) or large estimates of risk or severity of the response, may help to rule out alternative explanations. Mechanistic evidence from exposed humans, if available, may add support by informing considerations such as exposure response, temporality, coherence, and biological plausibility (i.e., evidence consistent/inconsistent with mechanistic understanding of how chemical exposure could cause the health effect based on current biological knowledge), thus raising the level of certainty to *robust* for a set of studies that otherwise would be described as *moderate*.

Strength-of-evidence judgment	Description
Moderate (⊕⊕⊙) …evidence in human studies (signal of effect with some uncertainty)	A smaller number of studies (at least one *high* or *medium* confidence study with supporting evidence) that do not reach the certainty required for *robust*. For multiple studies, there is primarily consistent evidence of an association, but with some residual uncertainty due to potential chance, bias, or confounding (e.g., effect estimates of low magnitude or small effect sizes given what is known about the endpoint; uninterpretable patterns with respect to exposure levels). For a single *high* or *medium* confidence study, there is supporting evidence increasing certainty in the findings such as a large magnitude or severity of the effect, a dose-response gradient, or other factors that increase the evidence strength, without serious residual uncertainties. In both scenarios, associations with related endpoints, including mechanistic evidence from exposed humans, can address uncertainties relating to exposure response, temporality, coherence, and biological plausibility, and any conflicting evidence is not from a comparable body of higher confidence, sensitive studies.[a]
Slight (⊕⊙⊙) …evidence in human studies (signal of effect with large amount of uncertainty)	One or more studies reporting an association between exposure and the health outcome, where considerable uncertainty exists. In general, the evidence is limited to a set of consistent *low*-confidence studies, a single *high* or *medium* confidence study without supporting evidence, or higher confidence studies with unexplained heterogeneity [e.g., comparable studies of similar confidence and sensitivity provide conflicting evidence, or the differences cannot be reasonably explained by, for example, the populations or exposure levels studied. This includes scenarios in which there are serious residual uncertainties across studies (these uncertainties typically relate to exposure characterization or outcome ascertainment, including temporality) in a set of largely consistent *medium* or *high* confidence studies.[a] Strong mechanistic evidence in well-conducted studies of exposed humans (*medium* or *high* confidence) or human cells (including NAMs), in the absence of other substantive data, where an informed evaluation has determined that the data are reliable for assessing toxicity relevant to humans and the mechanistic events have been causally linked to the development of the health effect of interest may be independently interpreted as *slight*.[b] On the other hand, strong human mechanistic evidence demonstrating that the effect is unlikely to occur may reduce to *slight* evidence that would otherwise be characterized as *moderate* (see Table 11-2). This category serves primarily to encourage additional study where evidence exists that might provide some support for an association, but for which the evidence does not reach the degree of confidence required for *moderate*.
Indeterminate (⊙⊙⊙) …evidence in human studies (signal cannot be determined for or against an effect)	No studies of exposed humans or well-conducted studies of human cells, or situations when the evidence is highly inconsistent and primarily of *low* confidence. In addition, this may include situations where higher confidence studies exist, but unexplained heterogeneity exists, and there are additional outstanding concerns such as effect estimates of low magnitude, uninterpretable patterns with respect to exposure levels, or uncertainties or methodological limitations that result in an inability to discern effects from exposure. A set of largely null studies could be concluded to be *indeterminate* if the evidence does not reach the level required for *compelling evidence of no effect*.

Strength-of-evidence judgment	Description
Compelling evidence of no effect (- - -) …in human studies (strong signal for lack of an effect with little uncertainty)	Several *high* confidence studies showing null results (for example, an odds ratio of 1.0), ruling out alternative explanations including chance, bias, and confounding with reasonable confidence. Each of the studies should have used an optimal outcome and exposure assessment and adequate sample size (specifically for higher exposure groups and for susceptible populations). The overall set should include the full range of levels of exposures that human beings are known to encounter, and an evaluation of an exposure-response gradient.

[a]Scenarios with unexplained heterogeneity across sets of studies with similar confidence and sensitivity can be considered either *slight* or *moderate*, depending on the expert judgment of the strength of the available evidence. Specifically, this judgment considers the level of support (or lack thereof) provided by evaluations of the magnitude or severity of the effects, coherence of related findings (including mechanistic evidence), dose-response, and biological plausibility, as well as the comparability of the supporting and conflicting evidence (e.g., the specific endpoints tested, or the methods used to test them; the specific sources of bias or insensitivity in the respective sets of studies). The evidence-specific factors supporting either judgment will be clearly articulated in the evidence integration narrative.

[b]Scientific understanding of toxicity mechanisms and of the human implications of new toxicity testing methods (e.g., from high-throughput screening, from short-term in vivo testing of alternative species, or from new in vitro and in silico testing and other NAMs) will continue to increase. Thus, the sufficiency of mechanistic evidence alone for identifying potential human health hazards is expected to increase as the science evolves. The evidence integration decisions based on these data represent expert judgments dependent on the state-of-the-science at the time of review.

1

2

3

Table 11-4. Framework for strength-of-evidence judgments from studies in animals

Strength-of-evidence judgment	Description
Robust (⊕⊕⊕) …evidence in animals (strong signal of effect with little residual uncertainty)	A set of *high* or *medium* confidence experiments with consistent findings of adverse or toxicologically significant effects across multiple laboratories, exposure routes, experimental designs (e.g., a subchronic study and a two-generation study), or species; and the experiments reasonably rule out the potential for nonspecific effects to have caused the effects of interest. Any inconsistent evidence (evidence that cannot be reasonably explained based on study design or differences in animal model) is from a set of experiments of lower confidence or sensitivity. To reasonably rule out alternative explanations, multiple additional factors in the set of experiments exist, such as: coherent effects across biologically related endpoints; an unusual magnitude of effect, rarity, age at onset, or severity; a strong dose-response relationship; or consistent observations across animal lifestages, sexes, or strains. Similarly, mechanistic evidence (e.g., precursor events linked to adverse outcomes) in animal models may exist to address uncertainties in the evidence base. Experimental support for an MOA that defines a causal relationship with reasonable confidence may raise the level of certainty to *robust* for evidence that otherwise would be described as *moderate* or, exceptionally, *slight*.
Moderate (⊕⊕⊙) …evidence in animals (signal of effect with some uncertainty)	A set of evidence that does not reach the degree of certainty required for *robust*, but which includes at least one *high* or *medium* confidence study with supporting information increasing the strength of the evidence. Although the results are largely consistent, notable uncertainties remain. However, in scenarios when inconsistent evidence or evidence indicating nonspecific effects exist, it is not judged to reduce or discount the level of concern regarding the positive findings, or it is not from a comparable body of higher confidence, sensitive studies.[a] The additional support provided includes either consistent effects across laboratories or species; coherent effects across multiple related endpoints; an unusual magnitude of effect, rarity, age at onset, or severity; a strong dose-response relationship; or consistent observations across exposure scenarios (e.g., route, timing, duration), sexes, or animal strains. Mechanistic evidence in animals may serve to provide this support or otherwise address residual uncertainties such that it raises the level of certainty to *moderate* for evidence that otherwise would be described as *slight*.

Strength-of-evidence judgment	Description
Slight (⊕⊙⊙) …evidence in animals (signal of effect with large amount of uncertainty)	Scenarios in which there is a signal of a possible effect, but the evidence is conflicting or weak. Most commonly, this includes situations in which only *low* confidence experiments are available, but largely consistent. It also applies when there is single *high* or *medium* confidence experiment in the absence of information increasing the strength of the evidence (e.g., corroboration within the same study or from other studies). Lastly, this includes scenarios in which there is evidence that would typically be characterized as *moderate*, but inconsistent evidence (evidence that cannot be reasonably explained by the respective study design or differences in animal model) from a set of experiments of higher confidence exists[a], or strong mechanistic evidence demonstrates that the effect is unlikely to occur (see Table 11-2). Strong mechanistic evidence in well-conducted studies of animals or animal cells (including NAMs), in the absence of other substantive data, where an informed evaluation has determined the assays are reliable for assessing toxicity relevant to humans and the mechanistic events have been causally linked to the development of the health effect may also be independently interpreted as *slight*.[b] This category served primarily to encourage additional research by describing situations for which evidence does exist that might provide some support for an association but is insufficient for a judgment of *moderate*.
Indeterminate (⊙⊙⊙) …evidence in animals (signal cannot be determined for or against an effect)	No animal studies or well-conducted studies of animal cells were available, the available endpoints are not informative to the hazard question under evaluation, or the evidence is highly inconsistent and primarily of *low* confidence. In addition, this may include situations in which higher confidence studies exist, but there is unexplained heterogeneity and additional concerns, such as small effect sizes (given what is known about the endpoint) or a lack of dose dependence. A set of largely null studies could be concluded to be *indeterminate* if the evidence does not reach the level required for *compelling evidence of no effect*.
Compelling evidence of no effect (- - -) …in animals (strong signal for lack of an effect with little uncertainty)	A set of *high* confidence experiments examining a reasonable spectrum of endpoints relevant to a type of toxicity that demonstrate a lack of biologically significant effects across multiple species, both sexes (if applicable), and a broad range of exposure levels. The data are compelling in that the experiments have examined the range of scenarios across which health effects in animals could be observed, and an alternative explanation (e.g., inadequately controlled features of the studies' experimental designs; inadequate sample sizes) for the observed lack of effects is not available. The experiments were designed to specifically test for the effects of interest, including suitable exposure timing and duration, post- exposure latency, and endpoint evaluation procedures. Mechanistic data in animals (*in vivo* or *in vitro*) that address the above considerations or that provide information supporting the lack of an association between exposure and effect with reasonable confidence may provide additional support for this judgment.

1

11.2. OVERALL EVIDENCE INTEGRATION JUDGMENTS

The second and final step of evidence integration combines the judgments regarding the strength of the animal and human evidence (from step 1) with considerations regarding mechanistic information on the human relevance of the animal evidence, relevance of the mechanistic evidence to humans (especially in cases where animal evidence is lacking), coherence across bodies of evidence, and information on susceptible populations and lifestages, all of which can be informed based on the considerations and analyses outlined in Section 11.2. This evidence integration decision process will culminate in an evidence integration narrative that summarizes the judgments regarding the evidence for each potential health effect (i.e., each noncancer health effect and specific type of cancer, or broader grouping of related outcomes). For each health effect, this narrative will include:

- A descriptive summary of the primary judgments about the evidence informing the potential for health effects in exposed humans, based on the following analyses:
    - Judgments regarding the strength of the available human and animal evidence (see Section 11.1);
    - consideration of the coherence of findings (i.e., the extent to which the evidence for health effects and relevant mechanistic changes are similar) across human and animal studies;
    - other information on the human relevance of findings in animals; and
    - conclusions drawn based on mechanistic analyses, as well as those based on analyses identified during stepwise consideration of the health effect-specific evidence during draft development.
- A summary of key evidence supporting these judgments, highlighting the evidence that was the primary driver of these judgments and any notable issues (e.g., data quality, coherence of the results), and a narrative expression of confidence (a summary of strengths and remaining uncertainties) for these judgments.
- Information on the general conditions of expression of these health effects (e.g., exposure routes and levels in the studies that were the primary drivers of these judgments), noting that these conditions will be clarified during dose-response analysis (see Section 12).
- Indications of potentially susceptible populations or lifestages (i.e., an integrated summary of the available evidence on potentially susceptible populations and lifestages drawn across the syntheses of the human, animal, and mechanistic evidence).
- A summary of key assumptions used in the analysis, which are generally based on EPA guidelines and which are largely captured in this protocol.
- Strengths and limitations of the evidence integration judgments, including key uncertainties and data gaps, and the limitations of the systematic reviews.

In short, the evidence integration narrative will present a qualitative summary of the strength of each evidence stream and an overall judgment across all relevant evidence, with exposure context provided. For each health effect or specific cancer type of potential concern, the first sentence of the evidence integration narrative will include the summary judgment [see description below for how these judgments help inform selection of a descriptor for carcinogenicity (U.S. EPA, 2005a)]. The assessment will also include an evidence profile table (see Table 11-1) to support the evidence integration narrative by providing the major decisions and supporting rationale. Table 11-5 describes the categories of evidence integration judgments that will be used in inorganic mercury salts assessment and presents examples of the database scenarios that fit each category of evidence. These summary judgments provide a succinct and clear representation of the decisions from the more detailed analyses of whether the evidence strength indicates that exposure to inorganic mercury salts could cause the human health effect(s) under the necessary conditions of exposure. Consistent with EPA noncancer and cancer guidelines, a judgment that the evidence supports an apparent lack of an effect of inorganic mercury salts exposure on the health effect(s) will be used only when the available data are considered extensive and definitive for deciding no basis for human hazard concern; lesser levels of evidence suggesting a lack of an effect will be characterized as "***evidence inadequate***."

Table 11-5. Evidence integration judgments for characterizing potential human health hazards in the evidence integration narrative

Evidence integration judgment[a] in narrative	Evidence integration judgment level	Explanation and example scenarios[b]
The currently available ***evidence demonstrates*** that [chemical] causes [health effect] in humans[c] under relevant exposure circumstances. This conclusion is based on studies of [humans or animals] that assessed [exposure or dose] levels of [range of concentrations or specific cutoff level concentration[d]].	***Evidence demonstrates***	A strong evidence base demonstrating that [chemical] exposure causes [health effect] in humans. • This judgment level is used if there is *robust* human evidence supporting an effect. • This judgment level could also be used with *moderate* human evidence and *robust* animal evidence if there is strong mechanistic evidence that an MOA(s) or key precursors identified in animals are expected to occur and progress in humans.
The currently available ***evidence indicates*** that [chemical] likely causes [health effect] in humans under relevant exposure circumstances. This conclusion is based on studies of [humans or animals] that assessed [exposure or dose] levels of [range of concentrations or specific cutoff level concentration].	***Evidence indicates (likely)[e]***	An evidence base that indicates that [chemical] exposure likely causes [health effect] in humans, although there may be outstanding questions or limitations that remain. The currently available evidence is insufficient for the highest judgment level. • This judgment level is used if there is *robust* animal evidence supporting an effect and *slight* or *indeterminate* human evidence, or with *moderate* human evidence when strong mechanistic evidence is lacking. • This judgment level could also be used with *moderate* human evidence supporting an effect and *slight* or *indeterminate* animal evidence, or with *moderate* animal evidence supporting an effect and *slight or indeterminate* human evidence. In these scenarios, any uncertainties in the *moderate* evidence are not sufficient to substantially reduce confidence in the reliability of the evidence, or mechanistic evidence in the *slight* or *indeterminate* evidence base (e.g., precursors) exists to increase confidence in the reliability of the *moderate* evidence. A decision between judgment levels of "***evidence indicates***" and "***evidence suggests***" considers the extent to which findings are coherent or biologically consistent across evidence streams (Table 11-2), and may incorporate other supplemental evidence (e.g., structure-activity data; chemical class information).

Evidence integration judgment[a] in narrative	Evidence integration judgment level	Explanation and example scenarios[b]
The currently available ***evidence suggests*** but is not sufficient to infer that [chemical] may cause [health effect] in humans under relevant exposure circumstances. This conclusion is based on studies of [humans or animals] that assessed [exposure or dose] levels of [range of concentrations or specific cutoff level concentration].	***Evidence suggests*** but is not sufficient to infer	An evidence base that suggests that [chemical] exposure may cause [health effect] in humans, but there are very few studies that contributed to the evaluation, the evidence is weak or conflicting, and/or the methodological conduct of the studies is poor. • This judgment level is used if there is *slight* human evidence and *indeterminate* or *slight* animal evidence. • This judgment level is also used with *slight* animal evidence and *indeterminate* or *slight* human evidence. • This judgment level could also be used with *moderate* human evidence and *slight* or *indeterminate* animal evidence, or with *moderate* animal evidence and *slight* or *indeterminate* human evidence. In these scenarios, there are outstanding issues regarding the *moderate* evidence that substantially reduced confidence in the reliability of the evidence, or mechanistic evidence in the *slight* or *indeterminate* evidence base (e.g., null results in well-conducted evaluations of precursors) exists to decrease confidence in the reliability of the *moderate* evidence. • Exceptionally, when there is general scientific understanding of mechanistic events that result in a health effect, this judgment level could also be used if there is strong mechanistic evidence that is sufficient to highlight potential human toxicity[f]—in the absence of informative conventional studies in humans or in animals (i.e., *indeterminate* evidence in both).

Evidence integration judgment[a] in narrative	Evidence integration judgment level	Explanation and example scenarios[b]
The currently available ***evidence is inadequate*** to assess whether [chemical] may cause [health effect] in humans under relevant exposure circumstances.	***Evidence inadequate***	This conveys either a lack of information or an inability to interpret the available evidence for [health effect]. On an assessment-specific basis, a single use of this "***evidence inadequate***" judgment might be used to characterize the evidence for multiple health effect categories.[g] • This judgment level is used if there is *indeterminate* human and animal evidence. • This judgment level is also used with *slight* animal evidence and *compelling evidence of no effect* human evidence. • This judgment level could also be used with *slight* or *robust* animal evidence and *indeterminate* human evidence if strong mechanistic information indicated that the animal evidence is unlikely to be relevant to humans. A judgment of "***evidence inadequate***" is not a determination that the agent does not cause the indicated human health effect(s). It simply indicates that the available evidence is insufficient to reach judgment(s) regarding the potential for the agent to cause the effect(s).
Strong evidence supports no effect in humans under relevant exposure circumstances. This conclusion is based on studies of [humans or animals] that assessed [exposure or dose] levels of [range of concentrations].	***Strong evidence supports no effect***[h]	This represents a situation in which extensive evidence across a range of populations and exposure levels has identified no effects/associations. This scenario requires a high degree of confidence in the conduct of individual studies, including consideration of study sensitivity, and comprehensive assessments of the endpoints and lifestages of exposure potentially relevant to the heath effect of interest. • This judgment level is used if there is *compelling evidence of no effect* in human studies *and compelling evidence of no effect* or *indeterminate* animal evidence. • This judgment level is also used if there is *indeterminate* human evidence and *compelling evidence of no effect* animal evidence in models judged as relevant to humans. • This judgment level could also be used with *compelling evidence of no effect* in human studies and *moderate* or *robust* animal evidence if strong mechanistic information indicated that the animal evidence is unlikely to be relevant to humans.

[a]As described in EPA guidance documents [(U.S. EPA, 2005a, 1996a, 1991, 1988b)], evidence integration depends heavily on expert judgment (note: as applied herein, "evidence integration" is synonymous with "weight of evidence"). The overall evidence integration judgment for each assessed health effect will be included as part of an evidence integration narrative, with the specific documentation of the various expert decisions and evidence-based (or default) rationales summarized in an evidence profile table, and the judgment contextualized based on the primary supporting evidence (experimental model or observed population, and exposure levels tested or estimated). Importantly, as discussed in Section 11, these judgments may be based on analyses of

grouped outcomes at different levels of granularity (e.g., motor activity vs. neurobehavioral effects vs. nervous system effects) depending on the specifics of the health effect evidence base. Evidence integration judgments are typically developed at the level of the health effect when there are sufficient studies on the topic to evaluate the evidence at that level; this should always be the case for "***evidence demonstrates***" and "***strong evidence supports no effect***," and typically for "***evidence indicates (likely)***." However, some databases only allow for evaluations at the category of health effects examined (e.g., nervous system effects); this will more frequently be the case for judgment levels of "***evidence suggests***" and "***evidence inadequate***." For all judgments, but particularly for those based on borderline evidence scenarios, the assessments will characterize the strengths and uncertainties in the evidence base within the evidence integration narrative and convey those interpretations to subsequent steps, including any toxicity values developed based on those effects. Health effects with judgments of "***evidence demonstrates***" and "***evidence indicates (likely)***" will be evaluated for use in dose-response assessment (see Section 12). When the database includes at least one well-conducted study and a hazard characterization judgment of "***evidence suggests"*** is drawn, quantitative analyses may be useful for some purposes (e.g., providing a sense of the magnitude and uncertainty of estimates for health effects of potential concern, ranking potential hazards, or setting research priorities), but not for others [see related discussions in (U.S. EPA, 2005a)]. When quantitative analyses are performed for "***evidence suggests***," it is critical to transparently convey the extreme uncertainty in any such estimates.

[b]Terminology of "is" refers to the default option; terminology of "could also be" refers to situational options (e.g., dependent on mechanistic understanding).

[c]In some assessments, these judgments might be based on data specific to a particular lifestage of exposure, sex, or population (or another specific group). In such cases, this would be specified in the overall summary judgment, with additional detail provided in the narrative text. This applies to all judgment levels.

[d]If concentrations cannot be estimated, an alternative expression of exposure level such as "occupational exposure levels," will be provided. This applies to all judgment levels.

[e]For some applications, such as benefit-cost analysis, to better differentiate the categories of "***evidence demonstrates***" and "***evidence indicates (likely)***," the latter category should be interpreted as evidence that supports an exposure-effect linkage that is likely to be causal.

[f]As discussed in Section 10.2, scientific understanding of toxicity mechanisms and of the human implications of new toxicity testing methods (e.g., from high-throughput screening, from short-term in vivo testing of alternative species, or from new in vitro and in silico testing and other NAMs) will continue to increase. Thus, the sufficiency of mechanistic evidence alone for identifying potential human health hazards is expected to increase as the science evolves. The evidence integration decisions based on these data represent expert judgments dependent on the state of the science at the time of review.

[g]Specific narratives for each of the health effects meeting this judgment level may also be deemed unnecessary.

hThe criteria for this category are intentionally more stringent than those justifying a conclusion of "evidence demonstrates" consistent with the "difficulty of proving a negative" [as discussed in (U.S. EPA, 1996a, 1991, 1988b)].

1

Evaluations of carcinogenicity will be consistent with EPA's Cancer Guidelines (U.S. EPA, 2005a). One of EPA's standardized cancer descriptors will be used as a shorthand characterization of the evidence integration narrative, describing the overall potential for human carcinogenicity across all potential cancer types. These are: (1) *carcinogenic to humans*, (2) *likely to be carcinogenic to humans*, (3) *suggestive evidence of carcinogenic potential*, (4) i*nadequate information to assess carcinogenic potential*, or (5) *not likely to be carcinogenic to humans*. More than one descriptor can be used when a chemical's effects differ by exposure level or route (U.S. EPA, 2005a); if the database supports such an analysis, these decisions will be clarified based on a more thorough review of the mechanistic evidence or more detailed dose-response analysis (see Section 12). In some cases, mutagenicity also will be evaluated (e.g., when evidence of carcinogenicity), because it influences the approach to dose-response assessment and subsequent application of adjustment factors for exposures early in life (U.S. EPA, 2005a, b).

An appropriate cancer descriptor will be selected as described in EPA Cancer Guidelines (U.S. EPA, 2005a). The cancer descriptor will consider the interrelatedness of cancer types potentially due to inorganic mercury salt exposure, consistency across the human and animal evidence for any cancer type [noting that site concordance is not required (U.S. EPA, 2005a)], and the uncertainties associated with assessment-specific conclusion. In general, however, if a systematic review of more than one cancer type was conducted, then the overall judgment and discussion of evidence strength in the evidence integration narrative for the cancer type(s) with the strongest evidence for hazard will be used to inform selection of the cancer descriptor, with each assessment providing a transparent description of the decision rationale. The cancer descriptor and evidence integration narrative for potential carcinogenicity, including application of the MOA framework, will consider the conditions of carcinogenicity, including exposure (e.g., route; level) and susceptibility (e.g., genetics; lifestage), as the data allow (Farland, 2005; U.S. EPA, 2005a, b).

11.3. HAZARD CONSIDERATIONS FOR DOSE-RESPONSE

This section outlines how the assessment will consider and describe the transition from hazard identification to dose-response analysis, highlighting (1) information that will inform the selection of outcomes or broader health effect categories for which toxicity values will be derived, (2) information to be considered and addressed during toxicity value derivation relating to specific populations or lifestages identified as susceptible, (3) how dose-response modeling will be informed by toxicokinetic information, and (4) information aiding the identification of biologically based benchmark response (BMR) levels. The pool of outcomes and study-specific endpoints will be discussed to identify which categories of effects and study designs are considered the strongest and most appropriate for quantitative assessment of a given health effect. Health effects analyzed in human studies in relation to exposure levels within or closer to the range of exposures encountered in the environment will be considered particularly informative, as will animal studies that test a broad range of exposure levels and include levels in the lower dose region. When there are multiple endpoints for an organ/system, considerations for characterizing the overall impact on this organ/system will be discussed, including the severity and longevity of the effects. For example, if there are multiple histopathological alterations relevant to liver function changes, liver necrosis might be selected as the most representative endpoint to consider for dose-response analysis. This section may review or clarify which endpoints or combination of endpoints in each organ/system characterize the overall effect for dose-response analysis. For cancer types, consideration will be given to the overall risk of multiple types of tumors. Multiple tumor types (if applicable) will be discussed and a rationale given for any grouping.

Biological considerations important for dose-response analysis (e.g., that could help with selection of a BMR) will be discussed. The impact of route of exposure on toxicity to different organs/systems will be examined, if appropriate. The existence and validity of physiologically based pharmacokinetic (PBPK) models or toxicokinetic information that could allow the estimation of internal dose for route-to-route extrapolation will be presented when available. In addition, mechanistic evidence analyses that will influence the dose-response analyses will be highlighted, for example, evidence related to susceptibility or potential shape of the dose-response curve.

This section also will describe the evidence regarding populations and lifestages that appear susceptible to the health hazards identified and factors likely to increase the risk of developing (or exacerbating) these health effects, depending on the available evidence. This section will include this discussion even if there are no specific data on the effects of exposure to the inorganic mercury salt of interest in the potentially susceptible population. At a minimum, consideration will be given to discussion of information relevant to infants and children, pregnant women, and women of childbearing age.

The section will consider options for using susceptible population data in the dose-response analysis. In particular, an attempt will be made to highlight where it might be possible to develop

separate risk estimates for a specific population or lifestage or to determine whether evidence is available to select a data-derived uncertainty factor.

12. DOSE-RESPONSE ASSESSMENT: SELECTING STUDIES AND QUANTITATIVE ANALYSIS

The previous sections of this protocol describe how systematic review principles will be applied to evaluate studies (for potential bias and sensitivity) and reach evidence integration conclusions on potential human health effects associated with exposure to inorganic mercury salts. Selection of specific data sets for dose-response assessment and performance of the dose-response assessment will be conducted after hazard identification is complete and involves database- and chemical-specific biological judgments that build from decisions made at earlier stages of assessment development. Several EPA guidance and support documents describe data requirements and other considerations for dose-response modeling, especially EPA's *Benchmark Dose Technical Guidance* (U.S. EPA, 2012), EPA's *Review of the Reference Dose and Reference Concentration Processes* (U.S. EPA, 2002b), *Guidelines for Carcinogen Risk Assessment* (U.S. EPA, 2005a), and *Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens* (U.S. EPA, 2005b). This section of the protocol provides an overview of considerations for conducting the dose-response assessment, particularly statistical considerations specific to dose-response analysis. Importantly, these considerations do not supersede existing EPA guidance.

As discussed in Section 1.2 "Scoping and Problem Formulation Summary" for inorganic mercury salts assessment, EPA's IRIS Program will conduct the assessments with a goal of developing oral toxicity values that the available data reasonably support, based on judgments of the evidence drawn during hazard identification and the suitability of studies for dose-response analysis. As stated in Section 1, this assessment will focus only on the oral route of exposures.

Dose-response assessment will be performed for both noncancer and cancer health hazards, for oral route of exposure,[14] if supported by existing data. For noncancer hazards, an oral reference dose (RfD) will be derived when possible. An RfD is an estimate, with uncertainty spanning perhaps an order of magnitude, of an exposure to the human population (including susceptible subgroups) likely to be without an appreciable risk of deleterious health effects over a lifetime (U.S. EPA, 2002b). When lifetime data are not available, less-than-lifetime (subchronic) data will be considered for deriving RfD values. These health effects could also include cancer effects [e.g., in a case where a nonlinear mode of action (MOA) is concluded that indicates a key precursor event necessary for carcinogenicity does not occur below a specific exposure level (U.S.

[14]For most health outcomes, dose-response assessments will be preferably based on studies of chronic exposure. Analyses also will be conducted for shorter durations, however, particularly when the evidence base for an inorganic mercury salt indicates potential risks associated with shorter exposures (U.S. EPA, 2002b).

EPA, 2005a). Reference values are not predictive risk values; that is, they provide no information about risks at higher or lower exposure levels. These values, generally related to noncancer effects, can be considered alongside environmentally measured exposures to characterize risk. That is, the RfD serves as a reference point from which to gauge the potential effects of the chemical at environmentally measured exposures. Usually, exposures less than the RfD are not likely to be associated with adverse health risks and are therefore less likely to be of regulatory concern. As the frequency and/or magnitude of the exposures exceeding the RfD increase, the probability of adverse effects in a human population increases.

When low-dose linear extrapolation for cancer effects is supported, particularly if data with direct mutagenic activity exist or if the data indicate a linear component below the POD, an oral slope factor (OSF) will be used to estimate human cancer risks. In general, this case also will occur when no data are available to inform the evaluation of linearity. An OSF is a plausible upper bound lifetime cancer risk from chronic ingestion of a chemical per unit of mass consumed per unit body weight per day (mg/kg-day). An OSF can be used in conjunction with exposure information to predict cancer risk at a given dose.

The derivation of reference values and cancer risk estimates will depend on the nature of the health hazard conclusions drawn during evidence integration (see Section 11.2). Specifically, EPA generally conducts dose-response assessment and derives cancer values for chemicals that are classified as *carcinogenic* or *likely to be carcinogenic* to humans. When there is *suggestive evidence* of carcinogenicity to humans, EPA generally would not conduct a dose-response assessment or derive a cancer value except when the evidence includes a well-conducted study and quantitative analyses could be useful for some purposes, for example, providing a sense of the magnitude and uncertainty of potential risks, ranking potential hazards, or setting research priorities (U.S. EPA, 2005a). A parallel approach will be taken for potential noncancer health effects in this assessment. Specifically, for noncancer outcomes this assessment will attempt dose-response assessments when the evidence integration judgments indicate stronger evidence of a hazard (i.e., "***evidence demonstrates***" and "***evidence indicates [likely]***"), and quantitative analyses generally will not be attempted for other evidence integration conclusions (with exceptions described in Section 11.2).

12.1. SELECTING STUDIES FOR DOSE-RESPONSE ASSESSMENT

The dose-response assessment will begin with a review of the important health effects highlighted during hazard identification, particularly among the studies of highest quality and that exemplify the study attributes summarized in Table 12-1. This review also will consider whether opportunities for quantitative evidence integration exist, although the data available will allow for the assessment on the basis of the preliminary literature inventory is considered unlikely.

Some studies used qualitatively for hazard identification might or might not be considered useful quantitatively for dose-response analysis in inorganic mercury salts assessment because of factors like lack of quantitative measures of exposure or variability measures for response data. If

the needed information cannot be located (e.g., by contacting study authors and making any information publicly available), a semiquantitative analysis (e.g., via no-observed-adverse-effect level [NOAEL]/lowest-observed-adverse-effect level [LOAEL]) will be considered. Studies of low sensitivity may be considered less useful if they failed to detect an effect or reported points of departure with wide confidence limits, but such studies will still be considered for inclusion in a meta-analysis.

Table 12-1. Attributes used to evaluate studies for deriving toxicity values

Study attributes		Considerations	
		Human studies	**Animal studies**
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).	Animal studies provide supporting evidence when adequate human studies are available and are considered principal studies when adequate human studies are not available. For some hazards, studies of animal species known to respond similarly to humans would be preferred over studies of other species.
Relevance of exposure paradigm	Exposure route	Studies involving human environmental exposures (oral).	Studies by a route of administration relevant to human environmental exposure are preferred. A validated toxicokinetic model also can be used to extrapolate across exposure routes.
	Exposure duration	When developing a chronic toxicity value, chronic- or subchronic-studies are preferred over studies of acute exposure. Exceptions exist, such as when a susceptible population or lifestage is more sensitive in a certain time window (e.g., developmental exposure).	
	Exposure level	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 they can provide information about the shape of the exposure-response relationship [see the EPA *Benchmark Dose Technical Guidance*, § 2.1.1; (U.S. EPA, 2012)] 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 (e.g., matching procedures, blocking) 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 a causal effect with respect to the development of the outcome are preferred. Exposure assessment methods that reduce measurement error and methods that provide measurement of exposure at the level of the individual are preferred. Measurements of exposure should not be influenced by knowledge of health outcome status.	Studies providing actual measurements of exposure (e.g., analytical inhalation concentrations versus target concentrations) are preferred. Relevant internal dose measures could facilitate extrapolation to humans, as would availability of a suitable animal PBPK model in conjunction with an animal study reported in terms of administered exposure.

Study attributes	Considerations	
	Human studies	**Animal studies**
Measurement of health outcome	Studies that can distinguish the presence or absence (or degree of severity) of the outcome reliably are preferred. Outcome ascertainment methods using generally accepted, standardized approaches are preferred.	
	Studies with individual data generally are preferred. Examples include characterizing experimental variability more realistically and characterizing overall incidence of individuals affected by related outcomes (e.g., phthalate syndrome).	
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 the context 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 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).

Among the studies that support the evidence integration conclusions, those most useful for dose-response analysis generally will 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. Such studies generally also will have 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 by reducing statistical uncertainty in the POD and minimizing the need for low-dose extrapolation. In addition to these more general considerations, specific issues that might be considered for their potential to impact the feasibility of dose-response modeling for individual data sets are described in more detail in *Benchmark Dose Technical Guidance* (U.S. EPA, 2012).

For cases where the data are limited or unavailable, such as mercurous chloride, the surrogate approach will be used, which allows for using data from related compounds to calculate values when data for the compound of interest are limited or unavailable. Details regarding searches and methods for surrogate analysis are presented in Wang et al., (2012). Three types of potential surrogates (structural, metabolic, and toxicity-like) will be identified to facilitate the final surrogate chemical selection (see Section 10.2.2). For example, structural analogs to mercurous chloride will be identified with available oral toxicity values and >50% similarity scores from at least two of the structure activity relationship (SAR) databases. Metabolic surrogates based on ADME of mercurous chloride will be identified, when possible. In addition, identification of toxicity-like potential surrogate candidates will be attempted. Among the candidate surrogates, a sensitive target organ of toxicity will be used.

A WOE approach will be used to evaluate information from potential candidate surrogates, as described by Wang et al., (2012). Commonalities in structural/physicochemical properties, toxicokinetics, metabolism, toxicity, or MOA between potential surrogates and chemical(s) of concern will be identified. Toxicological or toxicokinetic similarity will be emphasized over structural similarity. Surrogate candidates will be excluded if they have no commonality or demonstrate significantly different physicochemical properties and toxicokinetic profiles that distinguish them from the pool of potential surrogates or chemical(s) of concern. From the remaining potential surrogates, the most appropriate surrogate (most biologically or toxicologically relevant analog chemical) with the highest structural similarity or more sensitive toxicity value will be selected.

12.2. CONDUCTING DOSE-RESPONSE ASSESSMENT

Consistent with EPA practice, the inorganic mercury salts assessment will apply 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, 2005a):

1. Within the observed dose range, the preferred approach will be to use dose-response modeling to incorporate as much of the data set as possible into the analysis. This modeling to derive a POD should include an exposure level ideally near the lower end of the range of observation, without significant extrapolation to lower exposure levels (see Section 12.2.1).

2. As derivation of cancer risk estimates and reference values nearly always involves extrapolation to exposures lower than the POD; the approaches to be applied in the assessment are described in more detail in Section 12.2.2 and Section 12.2.3, respectively.

When sufficient and appropriate human and laboratory animal data are available for the same outcome, human data will be generally preferred for the dose response assessment because its use eliminates the need to perform interspecies extrapolations.

For reference values, the assessment will typically derive a candidate value from each suitable data set, whether in humans or animals (see Section 12.1). Evaluation of these candidate values grouped within a given organ/system will yield a single organ-/system-specific value for each organ/system under consideration. Next, evaluation of these organ/system-specific values will result in the selection of a single overall reference value to cover all health outcomes across all organs/systems.

For cancer, if multiple tumor sites can be quantified individually, the final cancer risk estimate(s) will typically address overall cancer risk, to the extent the data allow.

For both cancer and noncancer toxicity values, uncertainties in these estimates will be transparently characterized and discussed.

12.2.1. Dose-Response Analysis in the Range of Observation

As indicated previously, human data are preferred over animal data to eliminate interspecies extrapolation uncertainties (e.g., in toxicodynamics, relevance of specific health outcomes to humans). Human data are extremely limited for the mercury salts under consideration, however, and are not likely to be useful for establishing reference values.

Toxicodynamic ("biologically based") modeling is generally preferred when sufficient, reliable data are available 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. Such data, however, do not appear available for inorganic mercury salts.

Because neither human data nor a toxicodynamic model will be available for dose-response assessment, empirical modeling of animal toxicological bioassay data will be used (on the apical outcome or a key precursor event) in the range of observation. For this purpose, EPA has developed its benchmark dose software (BMDS) (http://www.epa.gov/ncea/bmds). BMDS is designed to help model dose-response datasets in accordance with EPA *Benchmark Dose Technical Guidance* (U.S. EPA, 2012). For noncancer (and nonlinear cancer) datasets, EPA recommends (1) application of a preferred set of models that use maximum likelihood estimation (MLE) methods (default models in BMDS) and (2) selection of a POD from a single model based on criteria designed to limit model selection subjectivity (auto-implemented in BMDS version 3 and higher).

For the linear analysis of cancer datasets, EPA recommends (1) application of the Multistage MLE model or a multitumor model that appropriately estimates combined tumor risk, both of which are available in BMDS, and (2) selection of a single Multistage degree based on criteria outlined in an EPA statistical workgroup technical memo available on the BMDS website (https://cfpub.epa.gov/ncea/bmds/recordisplay.cfm?deid=308382).[15] The standard set of models considered for noncancer and nonlinear cancer analyses are the default models in BMDS 3.2 and are detailed in the "Model Descriptions" section of the BMDS User Manual (http://www.epa.gov/ncea/bmds/documentation/BMDS_Manual.pdf).

BMDS 3.2 also provides an alternative modeling approach that uses Bayesian model averaging for dichotomous modeling average (DMA) and continuous modeling average (CMA) response data.[16] EPA makes DMA and CMA available as alternative approaches but has not yet finalized guidance for their use. In situations where alternative models with significant biological support are available, the decision maker will be informed by the presentation of these alternatives in the assessment(s) along with the models' strengths and uncertainties.

For each modeled response, a POD from the observed data will 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 will be used as the starting point for subsequent extrapolations and analyses. For linear extrapolation of cancer risk, the POD will be used to calculate an oral slope factor (OSF), and for nonlinear extrapolation, the POD will be used in the calculation of an RfD after the application of UFs.

The response level at which the POD is calculated will be guided by the severity of the endpoint and the power of the study to detect the effect. If linear extrapolation is used, standard values near the low end of the observable range generally will be used. For nonlinear analyses of dichotomous data, a response level of 10% extra risk generally will be used for minimally adverse effects and 5% or lower for more severe effects. For continuous data, a response level ideally will be based on an established definition of biological significance. The point of departure will be the 95% lower bound on the dose associated with the selected response level.

EPA has developed standard approaches to determine the relevant dose for use in dose-response modeling in the absence of appropriate toxicokinetic modeling. These standard approaches also can aide comparison across exposure patterns and species in the absence of a validated pharmacokinetic (PK) model (see below). The general approaches and considerations to be used to extrapolate inorganic mercury salts dosimetry from (1) shorter to longer durations within studies, (2) from animals to humans, and (3) across routes of exposure are outlined below:

[15]The Multistage degree selection process outlined in the memo is auto-implemented in the BMDS multitumor model, which can be run on one or more tumor data sets, but only the noncancer model selection process is auto-implemented for individual Multistage model runs in the current version, BMDS 3.2).
[16]DMA has been fully tested and externally peer reviewed, but CMA in BMDS 3.2 is a beta version.

- Intermittent study exposures will be standardized to a daily average over the duration of exposure. For chronic effects, daily exposures will be averaged over the life span. Exposures during a critical period, however, will not be averaged over longer durations (U.S. EPA, 2005a, 1991). Note this typically will be done after modeling because the conversion is linear.

- The preferred approach for dosimetry extrapolation from animals to humans will be through PBPK or PK modeling.

- Based on the selection of half-life as the preferred metric and a POD identified from a health-effects study in animals, the human equivalent dose (HED) will be calculated as:

  $$HED = (T_{0.5,A[s]}/T_{0.5,H[s]}) \times POD$$

  - Here, the [s] in the subscript indicates the value could be sex specific. When values are sex specific (significant differences between males and females) in both animals and humans, the $T_{0.5}$ values for females would be used to extrapolate health effects in female animals to women, and the $T_{0.5}$ values for males would be used to extrapolate health effects in male animals to men. The lower of the two human HEDs then would be used as the value for the more sensitive subpopulation.

- In the absence of PK data/half-lives, oral doses will be scaled allometrically using $mg/kg^{3/4}$-day as the equivalent dose metric across species. Allometric scaling pertains to equivalence across species, not across lifestages, and will not be used to scale doses from adult humans or mature animals to infants or children (U.S. EPA, 2011a, 2005a, 1994). Using this approach, the HED will be calculated as:

  $$HED = (BW_H/BW_A)^{0.25} \times POD \text{ mg/kg-day})$$

- In the absence of study-specific data on, for example, intake rates or body weight, EPA has developed recommended values for use in dose-response analysis (U.S. EPA, 1988b).

12.2.2. Extrapolation: Oral Slope Factor

An OSF will be used to estimate human cancer risks when low dose linear extrapolation for cancer effects is supported by the inorganic mercury salts specific evidence. Low dose linear extrapolation also will be used as a default when the data are insufficient to establish the mode of action If the inorganic mercury salts specific data are sufficient to ascertain that one or more modes of action are consistent with low dose nonlinearity, or to support their biological plausibility, low dose extrapolation will use the reference value approach when suitable data are available (U.S. EPA, 2005a).

Differences in susceptibility will be considered for use in deriving multiple slope factors, with separate estimates for susceptible populations and lifestages (U.S. EPA, 2005a). If appropriate chemical-specific data on susceptibility from early life exposures are available, these data will be used to develop cancer slope factors that specifically address any potential for differential potency in early lifestages (Farland, 2005; U.S. EPA, 2005a). If such data are not available, the evidence integration analyses supports a mutagenic MOA for carcinogenicity, and the extrapolation approach

is linear, the dose-response assessment will indicate to decision makers that in developing risk estimates, the default age-dependent adjustment factors should be used with the cancer slope factor and age-specific estimates of exposure (U.S. EPA, 2005a, b). In this scenario, the final cancer risk value presented in the assessment(s) will reflect this adjustment, with the requisite calculations provided.

12.2.3. Extrapolation: Reference Values

Reference value derivation is EPA's most frequently used nonlinear extrapolation method, and it will be used in these inorganic mercury salts assessment for noncancer effects. This approach also will be used for cancer effects if the available data are sufficient to ascertain the MOA and conclude that it is not linear at low doses (see Section 12.2.2). In this case, reference values for the oral route of exposure will be developed following EPA's established practices (U.S. EPA, 2005a); in general, the reference value will be based not on tumor incidence, but on a key precursor event in the MOA that is necessary for tumor formation. The derivation of an RfD (if feasible) conducted as part of the assessment for mercuric chloride, mercuric sulfide, and mercurous chloride will be performed in a manner consistent with EPA guidance.

For each data set selected, reference values will be estimated by applying relevant adjustments (i.e., UFs) to the PODs to account for the conditions of the reference value definition. These factors account for human variation, extrapolation from animals to humans, extrapolation to chronic exposure duration, extrapolation to a minimal level of risk (if not observed in the data set), and database deficiencies, as outlined below. Increasingly, data-based adjustments (U.S. EPA, 2014), probabilistic approaches (Chiu et al., 2018; Chiu and Slob, 2015), and Bayesian methods for characterizing population variability (NAS, 2014) are becoming feasible and can be distinguished from the UF considerations outlined below, if such data exist for inorganic mercury salts. The assessment will discuss the scientific bases (or lack thereof) for each selected UF, including any data-based adjustments based on the following considerations.

- *Animal-to-human extrapolation*: If animal results are used to make inferences about humans, the reference value derivation will incorporate the potential for cross-species differences, which could arise from differences in toxicokinetics or toxicodynamics. The POD will be standardized to equivalent human terms or be based on toxicokinetic or dosimetry modeling, which could range from detailed chemical-specific to default approaches (U.S. EPA, 2014, 2011a). A factor of $10^{0.5}$ (rounded to 3) will be applied to account for the residual toxicokinetic uncertainty after application of toxicokinetic or dosimetry modeling, as well as uncertainty involving toxicodynamic differences. Data-derived adjustments for toxicodynamic differences across species might include qualitative decisions regarding key science issues.

- *Human variation*: The assessment will account for variation in susceptibility across the human population and the possibility that the available data might not represent individuals who are most susceptible to the effect. If 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 also will be considered (U.S. EPA, 2002b).[17, 18] When sufficient data are available, an intraspecies UF either less than or greater than 10-fold might be justified (U.S. EPA, 2002b). A reduction in this UF will be considered if the POD is derived from or adjusted specifically for susceptible individuals, but not for a general population that includes both susceptible and nonsusceptible individuals (U.S. EPA, 2002b, 1998c, 1996a, 1994, 1991). In general, when the use of such data or modeling is not supported, a UF with a default value of 10 will be used.

- *LOAEL to NOAEL*: When a POD is based on a LOAEL, the assessment will include an adjustment to an exposure level at which 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 generally will be applied to extrapolate to a lower exposure expected to be without appreciable effects. A factor other than 10 could also be considered, depending on the magnitude and nature of the response and the shape of the dose-response curve (U.S. EPA, 2002b, 1998c, 1996a, 1994, 1991).

- *Subchronic-to-chronic exposure*: When using studies of less-than-chronic exposure to make inferences about chronic/lifetime exposure, the assessment will consider whether lifetime exposure reasonably could be interpreted to result in effects at lower levels of exposure, including consideration of the specific health outcome(s) in question. A factor of up to 10 will be considered, depending on the duration of the studies and the nature of the response {(U.S. EPA, 2002b, 1998c, 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 lifestage, the assessment will apply a database UF (U.S. EPA, 2002b, 1998c, 1996a, 1994, 1991). The size of the factor will depend on the nature of the database deficiency. For example, 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 $10^{0.5}$ (i.e., 3) if either one or the other is missing (U.S. EPA, 2002b).

- The POD for a particular reference value will be divided by the product of these factors. Based on the RfD/RfC review (U.S. EPA, 2002b) recommendation that any composite factor exceeding 3,000 represents excessive uncertainty, values with >3,000 UF_C will not be used to derive reference values. An RfD/RfC could be based on the POD for a single endpoint within a study, or on a collection of related PODs within or across studies, if such biological relationships are substantiated by the evidence.

[17]Examples of adjusting the toxicokinetic portion of interhuman variability include the Integrated Risk Information System (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 first percentile of the dose metric distribution for each POD (U.S. EPA, 2011b).

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

13. PROTOCOL HISTORY

This section is a placeholder for tracking information on the original protocol release and any potential protocol updates.

REFERENCES

Andres, P. (1984). IgA-IgG disease in the intestine of Brown-Norway rats ingesting mercuric chloride. Clin Immunol 30: 488-494. http://dx.doi.org/10.1016/0090-1229(84)90034-5

ATSDR (Agency for Toxic Substances and Disease Registry). (1990). Final report: Technical assistance to the Tennessee department of health and environment. Office of Toxic Substance and Disease.

ATSDR (Agency for Toxic Substances and Disease Registry). (1999). Toxicological profile for mercury [ATSDR Tox Profile]. Atlanta, GA. https://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=115&tid=24

Barregard, L; Sallsten, G; Schutz, A; Attewell, R; Skerfving, S; Jarvholm, B. (1992). Kinetics of mercury in blood and urine after brief occupational exposure. Arch Environ Occup Health 47: 176-184. http://dx.doi.org/10.1080/00039896.1992.9938347

Bearer, CF. (1995). How are children different from adults? [Review]. Environ Health Perspect 103: 7-12.

Berlin, M; Zalups, RK; Fowler, BA. (2007). Mercury. In GF Nordberg; BA Fowler; M Nordberg; LT Friberg (Eds.), Handbook on the toxicology of metals (3rd ed., pp. 675-729). Amsterdam, Netherlands: Academic Press. http://dx.doi.org/10.1016/B978-012369413-3/50088-4

Bernaudin, JF; Druet, E; Druet, P; Masse, R. (1981). Inhalation or ingestion of organic or inorganic mercurials produces auto-immune disease in rats. Clin Immunol 20: 129-135. http://dx.doi.org/10.1016/0090-1229(81)90170-7

Budavari, S, O'Neil, M.,.J. Smith, A., Heckelman, P.,.E. (1989). The Merck index: An encyclopedia of chemicals, drugs, and biologicals. In SONMJSAHPE Budavari (Ed.), (11th ed.). Rahway, NJ: Merck Research Laboratories Division of Merck & Co., Inc.

Chiu, WA; Guyton, KZ; Martin, MT; Reif, DM; Rusyn, I. (2018). Use of high-throughput in vitro toxicity screening data in cancer hazard evaluations by IARC Monograph Working Groups. ALTEX 35: 51-64. http://dx.doi.org/10.14573/altex.1703231

Chiu, WA; Slob, W. (2015). A unified probabilistic framework for dose-response assessment of human health effects [Review]. Environ Health Perspect 123: 1241-1254. http://dx.doi.org/10.1289/ehp.1409385

Dantith, J. (2008). Mercurous compounds [Dictionary]. In J Dantith (Ed.), A Dictionary of Chemistry (6th ed.). University of Michigan: Oxford University Press. https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100151937

DEQ. (2015). Chemical update worksheet: Mercury, inorganic (as mercuric chloride). (CAS #: 7487-94-7).

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. JAMA 263: 1385-1389.

Drake, HJ. (1981). Mercury. In Kirk-Othmer encyclopedia of chemical technology: v 15, matches to N-nitrosamines (3rd ed ed.). New York, NY: John Wiley & Sons.

Druet, P; Druet, E; Potdevin, F; Sapin, C. (1978). Immune type glomerulonephritis induced by HgCl2 in the Brown Norway rat. Annales d'Immunologie 129C: 777-792.

EFSA (European Food Safety Authority). (2017). Guidance on the use of the weight of evidence approach in scientific assessments. EFSA J 15: 1-69. http://dx.doi.org/10.2903/j.efsa.2017.4971

Ekstrand, J; Nielsen, JB; Havarinasab, S; Zalups, RK; Söderkvist, P; Hultman, P. (2010). Mercury toxicokinetics--dependency on strain and gender. Toxicol Appl Pharmacol 243: 283-291. http://dx.doi.org/10.1016/j.taap.2009.08.026

Endo, T; Nakaya, S; Kimura, R. (1990). Mechanisms of absorption of inorganic mercury from rat small intestine. III. Comparative absorption studies of inorganic mercuric compounds in vitro. Pharmacol Toxicol 66: 347-353. http://dx.doi.org/10.1111/j.1600-0773.1990.tb00761.x

Farland, WH. (2005). [Memo to Science Policy council regarding implementation of the cancer guidelines and accompanying supplemental guidance - Science Policy Council Cancer Guidelines. Implementation Workgroup communication I: Application of the mode of action framework in mutagenicity determinations for carcinogenicity]. Available online at https://www.epa.gov/sites/production/files/2015-01/documents/cgiwgcommuniation_i.pdf

Gill, GA; Bruland, KW. (1990). MERCURY SPECIATION IN SURFACE FRESH-WATER SYSTEMS IN CALIFORNIA AND OTHER AREAS. Environ Sci Technol 24: 1392-1400.

Guyatt, G; Oxman, AD; Akl, EA; Kunz, R; Vist, G; Brozek, J; Norris, S; Falck-Ytter, Y; Glasziou, P; DeBeer, H; Jaeschke, R; Rind, D; Meerpohl, J; Dahm, P; Schünemann, HJ. (2011). GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 64: 383-394. http://dx.doi.org/10.1016/j.jclinepi.2010.04.026

Havarinasab, S; Bjorn, E; Ekstrand, J; Hultman, P, er. (2007). Dose and Hg species determine the T-helper cell activation in murine autoimmunity. Toxicology 229: 23-32. http://dx.doi.org/10.1016/j.tox.2006.09.006

Health Canada. (2010). Federal contaminated site risk assessment in Canada Part II: Health Canada toxicological reference values (TRVs) and chemical-specific factors, version 2.0. Ottawa, Canada. http://www.hc-sc.gc.ca/ewh-semt/pubs/contamsite/part-partie_ii/index-eng.php

Higgins, JPT; Green, S. (2011). Cochrane handbook for systematic reviews of interventions. Version 5.1.0 (Updated March 2011): The Cochrane Collaboration. Retrieved from http://handbook.cochrane.org/

Hill, AB. (1965). The environment and disease: Association or causation? Proc R Soc Med 58: 295-300.

Hirst, JA; Howick, J; Aronson, JK; Roberts, N; Perera, R; Koshiaris, C; Heneghan, C. (2014). The need for randomization in animal trials: an overview of systematic reviews [Review]. PLoS ONE 9: e98856. http://dx.doi.org/10.1371/journal.pone.0098856

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

Hudson, PJ; Vogt, RL; Brondum, J; Witherell, L; Myers, G; Paschal, DC. (1987). Elemental mercury exposure among children of thermometer plant workers. Pediatrics 79: 935-938.

IARC (International Agency for Research on Cancer). (1993). Mercury and mercury compounds. In IARC monographs on the evaluation of carcinogenic risk to humans Beryllium, cadmium, mercury and exposures in the glass manufacturing industry (pp. 311-325). https://monographs.iarc.fr/wp-content/uploads/2018/06/mono58-8E.pdf

Kingman, A; Albertini, T; Brown, LJ. (1998). Mercury concentrations in urine and whole blood associated with amalgam exposure in a US military population. J Dent Res 77: 461-471. http://dx.doi.org/10.1177/00220345980770030501

Krabbenhoft, DP; Babiarz, CL. (1992). THE ROLE OF GROUNDWATER TRANSPORT IN AQUATIC MERCURY CYCLING. Water Resour Res 28: 3119-3128. http://dx.doi.org/10.1029/92WR01766

Krauth, D; Woodruff, TJ; Bero, L. (2013). Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review [Review]. Environ Health Perspect 121: 985-992. http://dx.doi.org/10.1289/ehp.1206389

Langworth, S; Elinder, CG; Sundquist, KG; Vesterberg, O. (1992). Renal and immunological effects of occupational exposure to inorganic mercury. Br J Ind Med 49: 394-401. http://dx.doi.org/10.1136/oem.49.6.394

Liu, J; Lu, YF; Li, WK; Zhou, ZP; Li, YY; Yang, X; Li, C; Du, YZ; Wei, LX. (2016). Mercury sulfides are much less nephrotoxic than mercury chloride and methylmercury in mice. Toxicol Lett 262: 153-160. http://dx.doi.org/10.1016/j.toxlet.2016.10.003

Macleod, MR. (2013). Systematic reviews of experimental animal studies. Presentation presented at Workshop on weight of evidence; US National Research Council Committee to review the Integrated Risk Information System (IRIS) process, March 27-28, 2013, Washington, DC.

Mckelvey, W; Jeffery, N; Clark, N; Kass, D; Parsons, PJ. (2011). Population-based inorganic mercury biomonitoring and the identification of skin care products as a source of exposure in New York City. Environ Health Perspect 119: 203-209. http://dx.doi.org/10.1289/ehp.1002396

Morgan, RL; Thayer, KA; Bero, L; Bruce, N; Falck-Ytter, Y; Ghersi, D; Guyatt, G; Hooijmans, C; Langendam, M; Mandrioli, D; Mustafa, RA; Rehfuess, EA; Rooney, AA; Shea, B; Silbergeld, EK; Sutton, P; Wolfe, MS; Woodruff, TJ; Verbeek, JH; Holloway, AC; Santesso, N; Schünemann, HJ. (2016). GRADE: Assessing the quality of evidence in environmental and occupational health. Environ Int 92-93: 611-616. http://dx.doi.org/10.1016/j.envint.2016.01.004

Muraoka, Y; Itoh, F. (1980). Sex difference of mercuric chloride-induced renal tubular necrosis in rats--from the aspect of sex differences in renal mercury concentration and sulfhydryl levels. J Toxicol Sci 5: 203-214.

NAS (National Academy of Sciences). (2014). Review of EPA's Integrated Risk Information System (IRIS) process. Washington, DC: National Academies Press (US). http://dx.doi.org/10.17226/18764

NDEP (Nevada Division of Environmental Protection). (2017). Basic Comparison Levels. https://ndep.nv.gov/uploads/documents/july-2017-ndep-bcls.pdf

Nielsen, JB. (1992). Toxicokinetics of mercuric chloride and methylmercuric chloride in mice. J Toxicol Environ Health 37: 85-122. http://dx.doi.org/10.1080/15287399209531659

Nielsen, JB; Andersen, O. (1990). Disposition and retention of mercuric chloride in mice after oral and parenteral administration. J Toxicol Environ Health 30: 167-180. http://dx.doi.org/10.1080/15287399009531420

Nielsen, JB; Ekstrand, J; Zalups, RK; Söderkvist, P; Hultman, P. (2006). Inheritance of susceptibility: Mercury kinetics in two mouse strains (A.SW and B10.S) and their F1 generation. In MC Alpoim; PV Morais; MA Santos (Eds.), Metal ions in biology and medicine: Volume 9 (pp. 351-355). Montrouge, France: John Libbey Eurotext.

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

NTP (National Toxicology Program). (1993). Toxicology and carcinogenesis studies of mercuric chloride (CAS no 7487-94-7) in F344 rats and B6C3F1 mice (gavage studies) (pp. 1-260). (ISSN 0888-8051NTP TR 408). Research Triangle Park, NC. http://ntp.niehs.nih.gov/?objectid=070985B6-9C9D-8C67-4E459578E228B376

NTP (National Toxicology Program). (2015). Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration. U.S. Dept.

of Health and Human Services, National Toxicology Program. https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf

NYSDEC (New York State Department of Environmental Conservation). (2006). New York State Brownfield Cleanup Program: Development of soil cleanup objectives: Technical support document. https://www.dec.ny.gov/docs/remediation_hudson_pdf/techsuppdoc.pdf

OEHHA (California Office of Environmental Health Hazard Assessment). (2008). Mercury reference exposure levels. In TSD for noncancer RELs Appendix D1: Summaries using this version of the hot spots risk assessment guidelines.

Park, JD; Zheng, W. (2012). Human exposure and health effects of inorganic and elemental mercury [Review]. 45: 344-352. http://dx.doi.org/10.3961/jpmph.2012.45.6.344 https://www.jpmph.org/upload/pdf/jpmph-45-344.pdf

PubChem. (2018). PubChem. Available online at https://pubchem.ncbi.nlm.nih.gov/

RIVM (National Institute for Public Health and the Environment (Netherlands)). (2001). Re-evaluation of human-toxicological maximum permissible risk levels (pp. 297). (711701025). Bilthoven, Netherlands: National Institute for Public Health and the Environment (RIVM). https://www.rivm.nl/bibliotheek/rapporten/711701025.pdf

Schünemann, H; Hill, S; Guyatt, G; Akl, EA; Ahmed, F. (2011). The GRADE approach and Bradford Hill's criteria for causation. J Epidemiol Community Health 65: 392-395. http://dx.doi.org/10.1136/jech.2010.119933

Sidle, WC. (1993). Naturally occurring mercury contamination in a pristine environment? Environ Geol 21: 42-50. http://dx.doi.org/10.1007/BF00775049

Sommar, J; Lindqvist, O; Strömberg, D. (2000). Distribution equilibrium of mercury (II) chloride between water and air applied to flue gas scrubbing. J Air Waste Manag Assoc 50: 1663-1666. http://dx.doi.org/10.1080/10473289.2000.10464192

Spectrum (Spectrum Laboratories). (2016). Safety Data Sheet: Mercuric sulfide, red powder, reagent. https://www.spectrumchemical.com/MSDS/M1190_AGHS.pdf

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.

Stokinger, HE. (1981). Mercury, Hg. In GD Clayton; FE Clayton (Eds.), Patty's industrial hygiene and toxicology: v 2A, toxicology 3rd rev ed (3rd rev. ed ed.). New York, NY: John Wiley & Sons.

Sundberg, J; Jonsson, S; Karlsson, MO; Oskarsson, A. (1999). Lactational exposure and neonatal kinetics of methylmercury and inorganic mercury in mice. Toxicol Appl Pharmacol 154: 160-169. http://dx.doi.org/10.1006/taap.1998.8566

U.S. Bureau of Mines (U.S Bureau of Mines). (1993). Mercury. in: Mineral commodity summaries (pp. 110-111). USGS.

U.S. EPA (U.S. Environmental Protection Agency). (1980). Ambient water quality criteria for mercury [EPA Report]. (EPA 440/5-80-058). Washington, DC. http://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=2000LMXR.txt

U.S. EPA (U.S. Environmental Protection Agency). (1988a). Drinking water criteria document for inorganic mercury [EPA Report]. (ECAO-CIN-025).

U.S. EPA (U.S. Environmental Protection Agency). (1988b). Recommendations for and documentation of biological values for use in risk assessment [EPA Report] (pp. 1-395). (EPA/600/6-87/008). Cincinnati, OH: U.S. Environmental Protection Agency, Office of Research and Development, Office of Health and Environmental Assessment. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=34855

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). (1992). Guidelines for Exposure Assessment [EPA Report]. In Risk Assessment Forum. (EPA/600/Z-92/001). Washington, DC: U. S. Environmental Protection Agency. https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=15263

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). (1995). Integrated risk information system (IRIS) chemical assessment summary for Mercuric chloride (HgCl2). Washington, DC. https://cfpub.epa.gov/ncea/iris/iris_documents/documents/subst/0692_summary.pdf

U.S. EPA (U.S. Environmental Protection Agency). (1996a). 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). (1996b). Mercury study report to Congress. Volume IV. Health effects of mercury and mercury compounds. SAB review draft [EPA Report]. (EPA452/R-96-001d). Washington, DC. https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/PB96184650.xhtml

U.S. EPA (U.S. Environmental Protection Agency). (1997). Locating and estimating air emissions from sources of mercury and mercury compounds. (EPA-454/R-97-012). Research Triangle Park, NC.

U.S. EPA (U.S. Environmental Protection Agency). (1998a). 1998 Toxics Release Inventory [EPA Report].

U.S. EPA (U.S. Environmental Protection Agency). (1998b). Assessment of thyroid follicular cell tumors [EPA Report] (pp. 1-51). (EPA/630/R-97/002). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. https://www.epa.gov/sites/production/files/2014-11/documents/thyroid.pdf

U.S. EPA (U.S. Environmental Protection Agency). (1998c). 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). (2002a). Provisional Peer-Reviewed Toxicity Values for mercuric sulfide (CASRN 1344-48-5): Derivation of a chronic oral RfD [EPA Report]. Cincinnati, OH.

U.S. EPA (U.S. Environmental Protection Agency). (2002b). 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). (2005a). Guidelines for carcinogen risk assessment [EPA Report]. (EPA/630/P-03/001B). 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). (2005b). Supplemental guidance for assessing susceptibility from early-life exposure to carcinogens [EPA Report]. (EPA/630/R-03/003F). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. https://www3.epa.gov/airtoxics/childrens_supplement_final.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2011a). Recommended use of body weight 3/4 as the default method in derivation of the oral reference dose (pp. 1-50). (EPA/100/R-11/0001). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum, Office of the Science Advisor. https://www.epa.gov/sites/production/files/2013-09/documents/recommended-use-of-bw34.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2011b). 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). (2017). Guidance to assist interested persons in developing and submitting draft risk evaluations under the Toxic Substances Control Act. (EPA/740/R17/001). Washington, DC: U.S Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention. https://www.epa.gov/sites/production/files/2017-06/documents/tsca_ra_guidance_final.pdf

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.

U.S. EPA (U.S. Environmental Protection Agency). (2019a). ChemView, the database [Website]. https://chemview.epa.gov/chemview

U.S. EPA (U.S. Environmental Protection Agency). (2019b). CompTox chemicals dashboard, the database [Website]. https://comptox.epa.gov/dashboard

USGS (U.S. Geological Survey). (1970). Mercury in the environment. (Professional Paper 713). http://dx.doi.org/10.3133/pp713

USGS (U.S. Geological Survey). (1997). Mercury.

Vesterinen, HM; Sena, ES; Egan, KJ; Hirst, TC; Churolov, L; Currie, GL; Antonic, A; Howells, DW; Macleod, MR. (2014). Meta-analysis of data from animal studies: a practical guide. J Neurosci Methods 221: 92-102. http://dx.doi.org/10.1016/j.jneumeth.2013.09.010

Wang, NC; Zhao, QJ; Wesselkamper, SC; Lambert, JC; Petersen, D; Hess-Wilson, JK. (2012). Application of computational toxicological approaches in human health risk assessment. I. A

tiered surrogate approach. Regul Toxicol Pharmacol 63: 10-19. http://dx.doi.org/10.1016/j.yrtph.2012.02.006

WHO (World Health Organization). (2003). Elemental mercury and inorganic mercury compounds. Human health aspects. In Concise International Chemical Assessment Document. https://www.who.int/ipcs/publications/cicad/en/cicad50.pdf

WHO/UNEP (World Health Organization/United Nations Environment Programme). (2008). Guidance for identifying populations at risk from mercury exposure. Geneva, Switzerland: Inter-Organization Programme for the Sound Management of Chemicals (IOMC).

Wilburn, DR. (2013). Changing patterns in the use, recycling, and material substitution of mercury in the United States. (Scientific Investigations Report 2013–5137). Reston, VA: U.S. Geological Survey. https://pubs.usgs.gov/sir/2013/5137/

Zirschky, J. (1990). Employee Transported Contaminant Releases. Hazardous Waste and Hazardous Materials 7.

APPENDIX A. SEARCH AND SCREENING STRATEGIES

Table A-1. Mercuric chloride database search strategy

Source	Search terms	Dates of search
PubMed	((((("Bichloride of mercury" OR "Calochlor" OR "Corrosive sublimate" OR "Dichloromercury" OR "HgCl2" OR "Mercuric chloride" OR "Mercuric perchloride" OR "Mercury bichloride" OR "Mercury chloromercurate (II)" OR "Mercury dichloride" OR "Mercury perchloride" OR "Mercury (II) chloride"))) AND ("2018/01/01"[Date - Publication] : "2019/02/15"[Date - Publication]))	1997–Feb 2019 Search results: 1,997
Web of Science	TS=("Bichloride of mercury" OR "Calochlor" OR "Corrosive sublimate" OR "Dichloromercury" OR "HgCl2" OR "Mercuric chloride" OR "Mercuric perchloride" OR "Mercury bichloride" OR "Mercury chloromercurate (II)" OR "Mercury dichloride" OR "Mercury perchloride" OR "Mercury (II) chloride" OR "7487-94-7") AND PY=2018-2019	1997–Feb 2019 Search results: 3,904
Toxline	@OR+("Bichloride+of+mercury"+Calochlor+"Corrosive+sublimate"+Dichloromercury+HgCl2+"Mercuric+chloride"+"Mercuric+perchloride"+"Mercury+bichloride"+"Mercury+chloromercurate+(II)"+"Mercury+dichloride"+"Mercury+perchloride"+"Mercury+(II)+chloride"+@TERM+@rn+7487-94-7)+@NOT+@org+pubmed+pubdart+@AND+@RANGE+yr+2018+2019	1997–Feb 2019 Search results: 359

Table A-2. Mercuric sulfide database search strategy

Source	Search terms	Dates of search
PubMed	((alpha-HgS OR Chinese red OR Cinnabar OR Ethiops mineral OR Aethiops mineral OR HgS OR Mercuric sulfide OR Mercury (II) sulfide OR Mercury (II) sulfide black OR Mercury (II) sulfide red OR Mercury sulfide OR Mercury sulphide OR Vermilion)) AND ("2018/01/01"[Date - Publication] : "2019/02/15"[Date - Publication])	1997–Feb 2019 Search results: 1,200
Web of Science	TS=("alpha-HgS" OR "Chinese red" OR "Cinnabar" OR "Ethiops mineral" OR "HgS" OR "Mercuric sulfide" OR "Mercury (II) sulfide" OR "Mercury (II) sulfide black" OR "Mercury (II) sulfide red" OR "Mercury sulfide" OR "Mercury sulphide" OR "Vermilion") AND PY=2018-2019	1997–Feb 2019 Search results: 3,885
Toxline	@OR+("alpha-HgS"+"Chinese+red"+"Cinnabar"+"Ethiops+mineral"+"HgS"+"Mercuric+sulfide"+"Mercury+(II)+sulfide"+"Mercury+(II)+sulfide+black"+"Mercury+(II)+sulfide+red"+"Mercury+sulfide"+"Mercury+sulphide"+"Vermilion"+@TERM+@rn+1344-48- 5)+@NOT+@org+pubmed+pubdart+@AND+@RANGE+yr+2018+2019	1997–Feb 2019 Search results: 72

Table A-3. Mercurous chloride database search strategy

Source	Search terms	Dates of search
PubMed	((calogreen OR calomel OR chloromercuri OR Cl2Hg2 OR mercury dichloride OR Hg2Cl2 OR hydrochloric acid mercury salt OR mercurous chloride OR mercury (I) chloride OR mercury chloride OR mercury monochloride OR mercury protochloride OR mercury subchlorides OR mild mercury chloride)) AND ("2018/01/01"[Date - Publication] : "2019/02/15"[Date - Publication])	1997–Feb 2019 Search results: 2,613
Web of Science	TS=("Calogreen" OR "Calomel" OR "Chloromercuri" OR "Cl2Hg2" OR "Dimercury dichloride" OR "Hg2Cl2" OR "Hydrochloric acid mercury salt OR Mercurous chloride" OR "Mercury (I) Chloride" OR "Mercury chloride" OR "Mercury monochloride" OR "Mercury protochloride" OR "Mercury subchloride" OR "Mild mercury chloride") AND PY=2018-2019	1997–Feb 2019 Search results: 2,149
Toxline	(@OR+("Calogreen"+"Calomel"+"Chloromercuri"+"Cl2Hg2"+"Dimercury+dichloride"+"Hg2Cl2" +"Hydrochloric+acid+mercury+salt"+ "Mercurous+chloride"+"Mercury+(I)+Chloride"+"Mercury+chloride"+"Mercury+ monochloride"+"Mercury+protochloride"+"Mercury+subchloride"+"Mild+mercury+chloride" +@TERM+@rn+10112-91- 1)+@AND+@RANGE+yr+1999+2018)+@NOT+@org+pubmed+pubdart	1997–Feb 2019 Search results: 61

Table A-4. Example DistillerSR form questions used for title/abstract and full text level screening

	Used in title/abstract screening			Used in full-text screening				
Question	**Does the article meet PECO criteria?**	**If meets PECO, what type of evidence?**	**If supplemental, what type of information?**	**Does the article contain PECO evidence?**	**If PECO or supplemental, what type of evidence?**	**If PECO or supplemental, which health outcome(s) apply?**	**If supplemental, what type of information?**	**If no, what is the reason for exclusion?**
Answer options (can select multiple options)	• Yes • No • Unclear (e.g., no abstract) • No, but potentially relevant supplemental information	• Human • Animal • Other	• Non-oral routes of exposure • Case report and case series • Mechanistic data including in vitro, in silico • ADME/PBPK • Mixtures • Reviews • Nonmammalian model systems • Bioavailability • Other	• Yes • No • Supple-mental material	• Human • Animal (mammalian models) • In vitro/modeled/ in silico	• Acute toxicity/Poisoning • ADME/Toxicokinetic/ PBPK • Body weight • Cancer • Cardiovascular • Clinical chemistry/Biochemical/ Cytotoxicity/Cellular function • Endocrine (hormone) • Gastrointestinal • Gene expression/omics • Genotoxicity • Growth (early life) and development • Hematological • Hepatic • Immune/Inflammation • Mortality • Musculoskeletal/Motor function/Bone • Neurological/Behavior • Nutrition and metabolic • Ocular and sensory • PBPK model • Renal/Urinary • Reproductive • Respiratory • Sensory • Skin and connective tissues • Other	• Non-PECO route of administration • Case reports or case series • Mechanistic studies • ADME/Toxico-kinetic • Exposure characteristics • Mixture studies • Records with no original data (reviews, editorials, etc.) • Non-mammalian model • Bioavailability • Other	• Not PECO relevant • Reviews, editorials, commentaries, meta-analyses with no original data • Conference abstract • Unable to translate • Unable to obtain full text • Other

ADME = absorption, distribution, metabolism, and excretion; PBPK = physiologically based pharmacokinetic; PECO = populations, exposures, comparators, and outcomes.

APPENDIX B. PROCESS FOR SEARCHING AND COLLECTING EVIDENCE FROM SELECTED OTHER RESOURCES

B.1. REVIEW OF REFERENCE LISTS FROM EXISTING ASSESSMENTS (FINAL OR PUBLICLY AVAILABLE DRAFT), JOURNAL REVIEW ARTICLES, AND STUDIES CONSIDERED RELEVANT TO PECO BASED ON FULL-TEXT SCREENING

Citation reference lists are typically reviewed manually because they are not available in a file format (e.g., RIS) that permits uploading into screening software applications. Manual review entails scanning the title, study summary, or study details as presented in the resource for those that appear to meet the PECO criteria. Any records identified that were not identified from the other sources are formatted in an RIS file format, imported into DistillerSR, annotated with respect to source, and screened as outlined in Section 3.2. For tracking assessments or reviews, the name of the source citation and the number of records imported into DistillerSR will be noted. The reference list of any study included in the literature inventory will be reviewed manually to identify titles that appeared relevant to the PECO criteria. These citations will be tracked in a spreadsheet, compared against the literature base to determine if they are unique to the project, and then added to DistillerSR to be screened at the title and abstract stage for PECO relevance.

B.2. EPA COMPTOX CHEMICALS DASHBOARD (TOXVAL)

ToxVal will be searched in the EPA CompTox Chemicals Dashboard (U.S. EPA, 2019b), and data available from the Hazard tab will be exported from the CompTox File Transfer Protocol site. Using both the human health POD summary file and the Record Source file, citations will be identified that apply to human-health PODs. A citation for each referenced study will be generated in HERO and verified that it was not already identified from the database search (or searches of "other sources consulted") prior to moving forward to screening in DistillerSR. Full texts will be retrieved where possible; if full texts are not available, data from the ToxVal dashboard will be entered and the citation annotated accordingly for Tableau and HAWC visualizations by adding "(ToxVal)" to the citation.

B.3. ECHA

A search of the ECHA-registered substances database will be conducted using the CASRN number. The registration dossier associated with the CASRN number will be retrieved by

navigating to and clicking the eye-shaped view icon displayed in the chemical summary panel. The general information page and all subpages included under the Toxicological Information tab will be downloaded in PDF format, including all nested reports that have unique URLs. In addition, the data will be extracted from each dossier page and used to populate an Excel tracking sheet with these data. Extracted fields include data from the general information page regarding the registration type and publication dates and, on a typical study summary page, the primary fields reported in the administrative data, data source, and effect levels sections. Each study summary could result in more than one row in the tracking sheet if more than one data source or effect level was reported.

At this stage, each study summary will be reviewed for inclusion on the basis of the PECO criteria. In addition, study summaries identified as without administrative data information will be excluded from review, and study summaries labeled "read across" (if any) will be screened and considered supplemental material. When a study summary considered relevant reported data from a study or lab report, a citation for the full study will be generated in HERO and verified that it is not already identified from the database search (or searches of "other sources consulted") prior to moving forward to screening. When citation information is not available and full text cannot be retrieved, the generated PDF will be used as the full text for screening and extraction and the citation annotated accordingly for Tableau and HAWC visualizations by adding "(ECHA Summary)" to the citation.

B.4. EPA CHEMVIEW

A search of the EPA ChemView database (U.S. EPA, 2019a) using the chemical CASRN number will be conducted. The prepopulated CASRN match and the "Information Submitted to EPA" output option filter will be selected before generating results. If results are available, the square-shaped icon under the "Data Submitted to EPA" column will be selected, and the following records will be included:

- High Production Volume Challenge Database (HPVIS)
- Human Health studies (Substantial Risk Reports)
- Monitoring (Includes environmental, occupational and general entries)
- TSCA Section 4 (Chemical testing results)
- TSCA Section 8(d) (Health and safety studies)
- TSCA Section 8(e) (Substantial Risk)
- FYI (Voluntary documents)

All records for ecotoxicology and physical & chemical property entries will be excluded. When results are available, extractors navigate into each record until a substantial risk report link is identified and saved as a PDF file. If the report cannot be saved, due to file corruption or broken links, the record will be excluded during full-text review as "unable to obtain record." Most substantial risk reports contain multiple document IDs, thus citations will be derived by concatenating the unique report numbers (OTS, 8EHD Num, DCN, TSCATS RefID, CIS) associated with each document along with the typical author organization, year, and title. Once a citation is generated, the study will be moved forward to DistillerSR, where it will be screened according to PECO and supplemental material criteria.

B.5. NTP CEBS

This CEBS database will be searched using the chemical CASRN number (https://manticore.niehs.nih.gov/cebssearch). All non-NTP data will be excluded using the "NTP Data Only" filter. Data tables for reports undergoing peer review also will be searched for studies that have not been finalized (https://ntp.niehs.nih.gov/data/tables/index.html) on the basis of a manual review of chemical names.

B.6. OECD ECHEM PORTAL

The OECD Echem Portal (https://hpvchemicals.oecd.org/UI/Search.aspx) will be searched using the chemical CASRN number. Only database entries from the following sources will be included, and entries from all other databases will be excluded in the search. Final assessment reports and other relevant SIDS reports embedded in the links will be captured and saved as a PDF file.

- OECD HPV
- OECD SIDS IUCLID
- SIDS UNEP

B.7. ECOTOX DATABASE

EPA's ECOTOX Knowledgebase (https://cfpub.epa.gov/ecotox/search.cfm) will be searched using the chemical CASRN number. Results will be refined to terrestrial mammalian studies by selecting the terrestrial tab at the top of the search page and sorting the results by species group. A citation for each referenced study will be generated in HERO and verified that it was not already identified from the database search (or searches of "other sources consulted") before moving it forward to screening in DistillerSR.

B.8. EPA COMPTOX CHEMICAL DASHBOARD VERSION TO RETRIEVE A SUMMARY OF ANY TOXCAST OR TOX21 HIGH THROUGHPUT SCREENING INFORMATION

Version 3.0.9 of the CompTox Chemicals Dashboard (U.S. EPA, 2019b) will be accessed for high throughput screening (HTS) data by searching the Dashboard by chemical CASRN number. Next, the "Bioactivity" section will be selected and the availability of ToxCast/Tox21 HTS data for active and inactive assays will be examined in the "TOXCAST: Summary" tab. If active assays are reported, the figure will be copied for presentation in the systematic evidence map. This figure will present (i) a scatterplot of scaled assay responses vs. AC_{50} values for each active assay endpoint, and (ii) a cytotoxicity limit as a vertical line. More detailed information on the results of ToxCast and Tox21 assays is available in the CompTox Chemicals Dashboard section "ToxCast/Tox21," which includes chemical analysis data, dose-response data and model fits, and "flags" assigned by an automated analysis, which suggest false positivity/negativity or indicate other anomalies in the data. This information will not be summarized further for the purposes of the systematic evidence map, which will be focused on identifying the extent of available evidence.

B.9. COMPARATIVE TOXICOGENOMICS DATABASE (CTDB)

This CTDB database (http://ctdbase.org/) will be searched using the chemical CASRN number in the "keyword search" with the pull-down menu set to "Chemicals." The query results in the "Gene Interactions" tab, which provides the list of genes and proteins reported in the published references as interacting with the query chemical. Human and rodent genes/proteins interacting with a query chemical will be identified and their numbers will be provided. If information is available, a figure presenting the top 10 interacting genes available in the "Basics" tab will be copied to the systematic evidence map. These top interacting genes represent genes for which their interactions with query chemical are supported by most available references. Details on interaction types and interaction degrees are provided at http://ctdbase.org/help/ixnQueryHelp.jsp. The top scoring pathway relevant to the identified interacting genes will be provided in the systematic evidence map. The reference list of studies reporting gene/protein interactions with the query chemical will be compared to existing references in DistillerSR. Unique references will be added to DistillerSR and screened according to PECO and supplemental material criteria.

B.10. GENE EXPRESSION OMNIBUS AND ARRAYEXPRESS

Public repositories of omics data, Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) and ArrayExpress (https://www.ebi.ac.uk/arrayexpress/), will be queried to identify available gene expression datasets relevant to the exposure by the chemical of interest. The GEO will be queried using the custom search string shown below, in which the letters "XX" will be replaced by the chemical name as represented in the PubMed search strategy from Appendix A. Filters will be applied to select human, mouse, and rat datasets. Note,

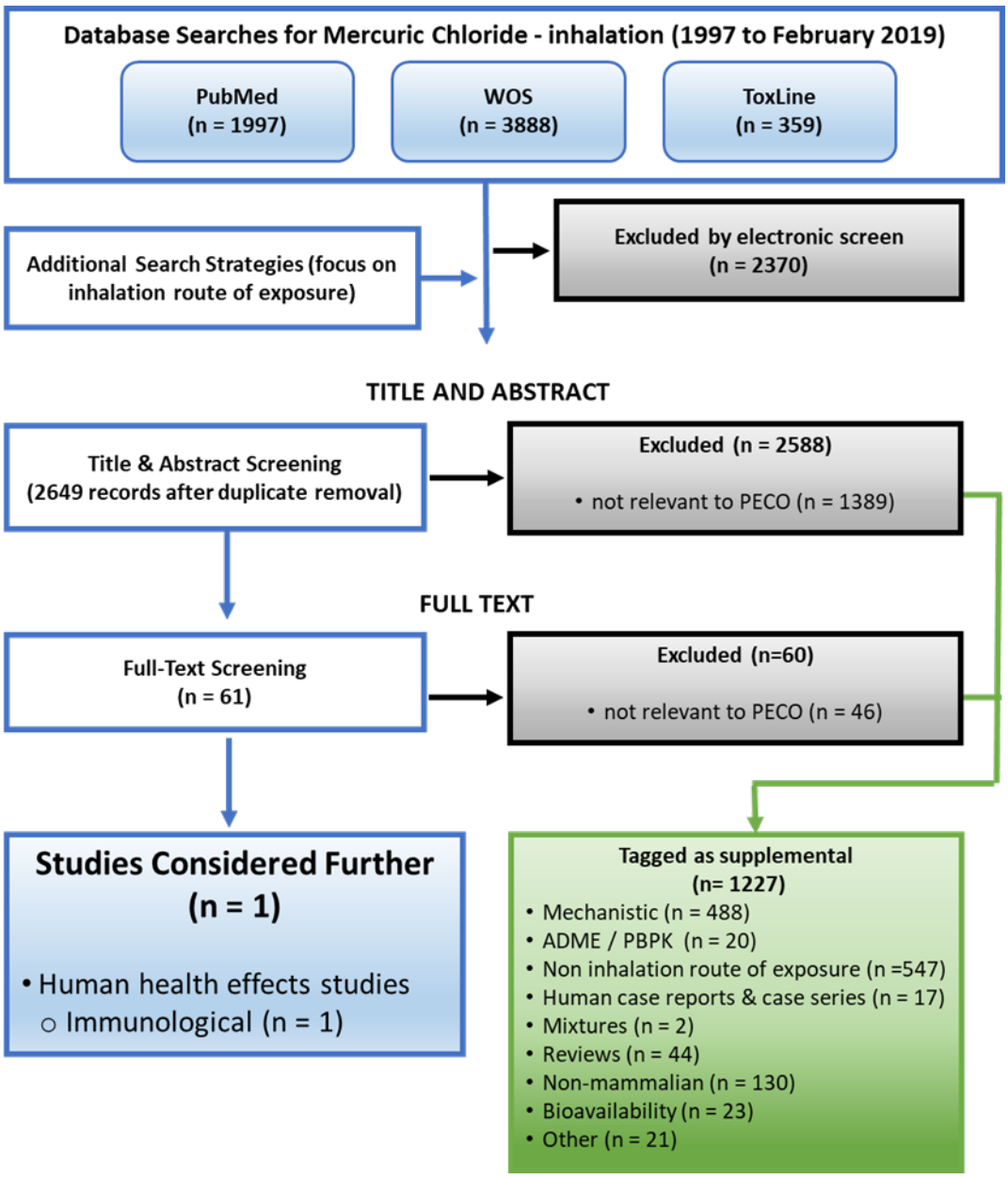
the filter options might not be available in the GEO website if no human, mouse, or datasets exist. Retrieved studies will be reviewed, and the studies that reported gene expression data for chemical exposures of interest will be reported in the systematic evidence map as follows: series accession number (GSExxxxx), title of study, species, route of exposure, platform, and tissue type.

GEO Search string
("XX"[MeSH Terms] OR XX[All Fields]) AND ("Expression profiling by RT-PCR"[Filter] OR "Expression profiling by MPSS"[Filter] OR "Expression profiling by SAGE"[Filter] OR "Expression profiling by SNP array"[Filter] OR "Expression profiling by array"[Filter] OR "Expression profiling by genome tiling array"[Filter] OR "Expression profiling by high throughput sequencing"[Filter] OR "Protein profiling by Mass Spec"[Filter] OR "Protein profiling by protein array"[Filter])

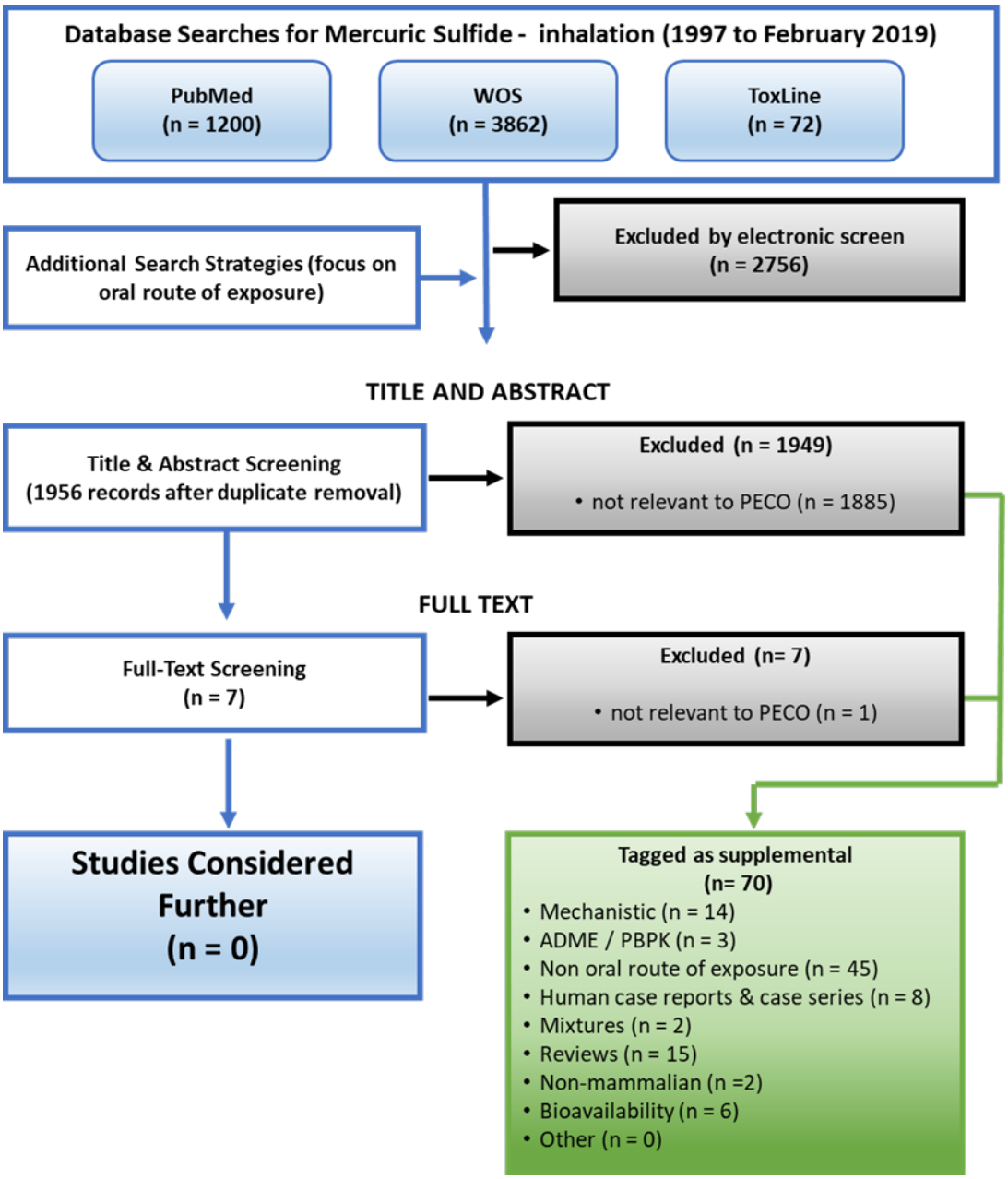
The ArrayExpress repository will be queried using chemical name as a keyword and filtered to limit datasets to *Homo sapiens*, *Rattus norvegicus*, and *Mus musculus* species. All studies reporting RNA-seq, transcription profiling, proteomic profiling or translation profiling data will be reported.

APPENDIX C. LITERATURE FLOW DIAGRAMS FOR INHALATION ROUTE OF EXPOSURE

C.1. LITERATURE FLOW DIAGRAM FOR MERCURIC CHLORIDE THROUGH INHALATION ROUTE OF EXPOSURE



C.2. LITERATURE FLOW DIAGRAM FOR MERCURIC SULFIDE THROUGH INHALATION ROUTE OF EXPOSURE



C.3. LITERATURE FLOW DIAGRAM FOR MERCUROUS CHLORIDE THROUGH INHALATION ROUTE OF EXPOSURE