‹♥›**EPA**

Systematic Review Protocol for the Polychlorinated Biphenyls (PCBs) Noncancer IRIS Assessment (Preliminary Assessment Materials)

[CASRN 1336-36-3]

December 2019

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

DISCLAIMER

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

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

ADME	absorption, distribution, metabolism, and excretion
BMR	benchmark response
CAS	Chemical Abstracts Service
CASRN	Chemical Abstracts Service registry number
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CPAD	Chemical and Pollutant Assessment Division
CPHEA	Center for Public Health and Environmental Assessment
EPA	Environmental Protection Agency
GLP	good laboratory practices
HAWC	Health Assessment Workspace Collaborative
HEEAD	Health and Environmental Effects Assessment Division
HERO	Health and Environmental Research Online
HPASB	Hazardous Pollutant Assessment and Systems Branch
IRIS	Integrated Risk Information System
LOAEC	lowest-observed-adverse-effect concentration
LOAEL	lowest-observed-adverse-effect level
MOA	mode of action
NAM	new approach method
NHANES	National Health and Nutrition Examination Survey
NOAEC	no-observed-adverse-effect concentration
NOAEL	no-observed-adverse-effect level
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
OLEM	Office of Land and Emergency Management
ORD	Office of Research and Development
PCB	polychlorinated biphenyl
PBPK	physiologically based pharmacokinetic
PECO	Populations, Exposures, Comparators, and Outcomes
PK	pharmacokinetic
POD	point of departure
RfC	inhalation reference concentration
RfD	oral reference dose
ROBINS-I	Risk of Bias in Nonrandomized Studies of Interventions
TEAB-R	Toxic Effects Assessment Branch-RTP
UF	uncertainty factor

AUTHORS | CONTRIBUTORS | REVIEWERS

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

Geniece Lehmann (Co-Assessment Manager)

Laura Carlson (Co-Assessment Manager)

Xabier Arzuaga

Evan Coffman

Allen Davis

Jeff Gift

Dustin Kapraun

Danelle Lobdell

Anuradha Mudipalli

Paul Schlosser

John Stanek

Michele Taylor

Michael Wright

Erin Yost

Technical Experts/Contributors

Michael Bloom	University at Albany School of Public Health
Pam Factor-Litvak	Columbia University Medical Center
Todd Jusko	University of Rochester Medical Center
Aileen Keating	Iowa State University College of Agriculture and Life Sciences
Carolyn Klocke	UC Davis College of Biological Sciences
Pam Lein	UC Davis College of Biological Sciences
John Meeker	University of Michigan School of Public Health
Pradeep Rajan	Pradeep Rajan LLC
Larry Robertson	University of Iowa College of Public Health
Sharon Sagiv	UC Berkeley School of Public Health
Alexander Sergeev	Ohio University College of Health Sciences and Professions
Michal Toborek	University of Miami Miller School of Medicine
Martina Velichkovska	University of Miami Miller School of Medicine

Executive Direction

Wayne Cascio	U.S. EPA/CPHEA/Director
Samantha Jones	U.S. EPA/CPHEA/Associate Director
Kristina Thayer	U.S. EPA/CPHEA/Chemical and Pollutant Assessment Division (CPAD)/Director
John Vandenberg	U.S. EPA/CPHEA/Health and Environmental Effects Assessment Division (HEEAD)/Director
James Avery	U.S. EPA/CPHEA/CPAD/Associate Director
Belinda Hawkins	U.S. EPA/CPHEA/CPAD/Senior Science Advisor
Andrew Kraft	U.S. EPA/CPHEA/CPAD/Senior Science Advisor
Paul White	U.S. EPA/CPHEA/CPAD/Senior Science Advisor
Janice Lee	U.S. EPA/CPHEA/CPAD/Toxic Effects Assessment Branch-RTP (TEAB-R)/Chief
Andrew Hotchkiss	U.S. EPA/CPHEA/HEEAD/Hazardous Pollutant Assessment and Systems Branch (HPASB)/Chief

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

Anna Champlin

Ryan Jones

Dahnish Shams

Vicki Soto

Contractor Support (ICF)

Robyn Blain

Dave Burch

Michelle Cawley

Anna Engstrom

Ali Goldstone

Brandall Ingle

Revathi Muralidharan

Sean Robins

MaryJane Selgrade

Kelly Shipkowski

Raquel Silva

Joanne Trgovcich

Arun Varghese

Nicole Vetter

1. INTRODUCTION

1   In April 2015, the U.S. Environmental Protection Agency (EPA) released scoping and
2   problem formulation materials for a new Integrated Risk Information System (IRIS) assessment to
3   address noncancer human health hazards associated with exposure to complex mixtures of
4   polychlorinated biphenyls (PCBs). An update of the existing evaluation of cancer risk from PCB
5   exposure (https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=294) was not
6   identified as a priority need, and a new assessment of PCB cancer risk is not planned at this time.
7   The scoping and problem formulation materials were presented at a public science meeting on June
8   17–18, 2015 (https://www.epa.gov/iris/iris-public-meeting-jun-2015) to seek input from the
9   scientific community and interested parties on the IRIS Program's scoping and problem
10 formulation conclusions and identification of key areas of scientific complexity (U.S. EPA, 2015b).
11 This protocol document presents the objectives and specific aims of the assessment, the draft PECO
12 (Populations, Exposures, Comparators, and Outcomes) criteria, and methods for conducting the
13 systematic review and dose-response analysis. While the scoping and problem formulation
14 materials described *what* the assessment will cover, this protocol describes *how* the assessment
15 will be conducted (see Figure 1). The IRIS Program posts assessment protocols on its website and
16 in the Zenodo repository (https://zenodo.org/). Public input received is considered during
17 preparation of the draft assessment and any adjustments to the protocol will be reflected in an
18 updated version released in conjunction with the draft assessment. Literature search results are
19 made available in EPA's Health and Environmental Research Online database (HERO). The PCB
20 project literature page will be updated annually with literature updates and can be found online
21 (https://hero.epa.gov/hero/index.cfm/project/page/project_id/384).
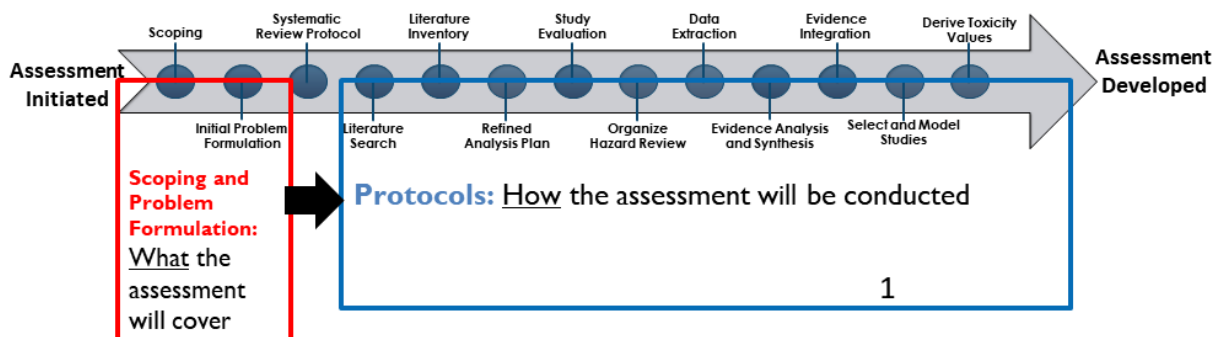
Figure 1. IRIS systematic review problem formulation and method documents.

2. SCOPING AND INITIAL PROBLEM FORMULATION SUMMARY

2.1. BACKGROUND

1   PCBs are a class of synthetic compounds characterized by a biphenyl structure with
2   chlorine substitutions at up to 10 positions, as shown in Figure 2. There are 209 possible PCB
3   congeners based on the various combinations of the numbers and positions of the chlorine
4   substitutions on the biphenyl molecule; PCB congeners vary in structure, stability, and toxicity
5   (Section 2.5.1). PCBs were manufactured and marketed in the United States between about 1930
6   and 1977 under the trade name Aroclor (e.g., Aroclors 1016, 1242, 1248, 1254, 1260). They were
7   used in many industrial applications because of their electrical insulating properties, chemical
8   stability, and relative inflammability. They were widely used in capacitors, transformers, and other
9   electrical equipment, and as coolants and lubricants. Other applications included use in
10   plasticizers, surface coatings, inks, adhesives, flame retardants, pesticide extenders, paints,
11   carbonless duplicating paper, and sealants and caulking compounds ([ATSDR, 2000](#)). EPA issued
12   final regulations banning the manufacture of PCBs and phasing out most PCB uses in 1979 under
13   the Toxic Substances Control Act (TSCA) (40 CFR 761) due to evidence that they persist and
14   accumulate in the environment and can cause toxic effects ([http://www2.epa.gov/aboutepa/epa-](http://www2.epa.gov/aboutepa/epa-bans-pcb-manufacture-phases-out-uses)
15   [bans-pcb-manufacture-phases-out-uses](http://www2.epa.gov/aboutepa/epa-bans-pcb-manufacture-phases-out-uses)). Despite the ban on manufacturing, PCBs continue to be
16   present in environmental media (e.g., air, soil, sediment, food) and are redistributed from one
17   environmental compartment to another ([ATSDR, 2000](#)). They also can be released through the
18   continued use and disposal of PCB-containing products and as a result of inadvertent production
19   during certain manufacturing processes ([Vorkamp, 2015](#)). PCB-containing building materials such
20   as window glazes, fluorescent light ballasts, ceiling tile coatings, caulk, paints, and floor finishes are
21   potential sources of PCBs in the indoor environment ([Lehmann et al., 2015](#)). Over time, the
22   congener profile of PCB mixtures (e.g., Aroclors) can be transformed in the environment, leading to
23   diverse human exposures (Section 2.5.1).

X = 1 to 5, Y = 1 to 5

1
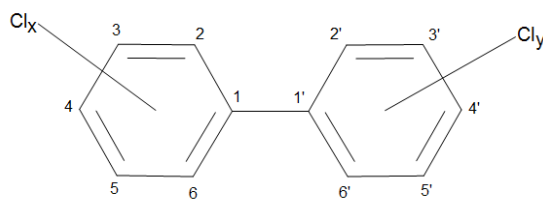
Figure 2. Chemical structure of PCBs (ATSDR, 2000).

2   Occupational exposure to PCBs can occur through inhalation and dermal contact at

3   workplaces where PCBs are present (e.g., handling PCB-containing electrical equipment, spills, or

4   waste-site materials without use of personal protective equipment to limit exposure) (ToxNet

5   Hazardous Substances Data, 2011). In the general population, PCB exposure occurs primarily via

6   dietary intake of contaminated food and inhalation of PCB-contaminated air (Lehmann et al., 2015;

7   ATSDR, 2000). The major contributors to dietary exposure to PCBs include fatty foods such as fish,

8   meat, and dairy products.

9   Inhalation also has been shown to be a contributor to total PCB exposure, especially in

10   indoor settings where PCB sources exist (Lehmann et al., 2015; Harrad et al., 2009). For example,

11   elevated indoor air PCB concentrations have been observed in some public school buildings. The

12   schools at highest risk of having elevated indoor air PCB concentrations are those that were built in

13   the 1950s-1970s and schools that were extensively remodeled during this period (Marek et al.,

14   2017; Thomas et al., 2012). Since September 2009, EPA has released several reports[1] for school

15   administrators and building managers with important information about identifying, and if present,

16   managing airborne PCBs, and tools to help minimize possible exposure.

17   General population exposure also can occur via dermal contact with PCBs in soil or other

18   media or through incidental ingestion of PCB-contaminated soil or dust (ATSDR, 2000). The

19   presence of PCBs in blood, adipose tissue, and breast milk of non-occupationally exposed members

20   of the general population of the United States provides evidence of widespread exposure (Xue et al.,

21   2014; CDC, 2009; ATSDR, 2000).

22   Populations with potentially greater than average exposures include those who consume

23   PCB-contaminated fish or wild game or who eat a higher proportion of food grown in PCB-

---

[1] *Polychlorinated Biphenyls (PCBs) in School Buildings: Sources, Environmental Levels, and Exposures*, EPA-600-R-12-051 (https://www.epa.gov/sites/production/files/2015-08/documents/pcb_epa600r12051_final.pdf);
Fact Sheet for Schools: Caulk Containing PCBs May Be Present in Older Schools and Buildings, EPA-747-F-09-003 (https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P1005C1D.TXT);
Proper Maintenance, Removal, and Disposal of PCB-Containing Fluorescent Light Ballasts (FLBs) in School Buildings: A Guide for School Administrators and Maintenance Personnel
(https://www.epa.gov/sites/production/files/documents/PCBsInBallasts.pdf); and
How to Test for PCBs and Characterize Suspect Material (https://www.epa.gov/pcbs/how-test-pcbs-and-characterize-suspect-materials).

1 contaminated areas ([IARC, 2015](#); [ATSDR, 2000](#)). Because many PCBs tend to accumulate in body

2 lipids and can be transferred to infants via breast milk, nursing infants are another potentially

3 highly exposed population. Certain occupational groups also can have elevated exposures through

4 inhalation, dermal contact, or incidental ingestion of PCB residues from contact with contaminated

5 materials in the workplace, during repair and maintenance of electrical equipment containing PCBs,

6 or from accidents or fires involving PCBs.

7   The IRIS database currently provides assessments for specific Aroclor mixtures:

8 quantitative assessments for Aroclor 1016 and Aroclor 1254 and a qualitative discussion for

9 Aroclor 1248. Although oral reference doses (RfDs) were derived for Aroclor 1016 and Aroclor

10 1254, no inhalation reference concentrations (RfCs) are available for PCBs.

11   • Aroclor 1016 (posted in 1993;
12   [https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=462](#)): The
13   Aroclor 1016 assessment derived an oral RfD based on data reported by ([Schantz et al.,](#)
14   [1989](#)), in which perinatal toxicity and long-term neurobehavioral effects of Aroclor 1016
15   were evaluated in infant rhesus monkeys born to dams exposed at 0.007 or
16   0.028 mg/kg-day for 7 months prior to breeding until offspring were weaned at age
17   4 months. Based on reduced birth weights and neurobehavioral deficits of prenatally
18   exposed monkeys, the 0.028 mg/kg-day dose was identified as the lowest-observed-
19   adverse-effect level (LOAEL). The study no-observed-adverse-effect level (NOAEL) of
20   0.007 mg/kg-day was chosen for the point of departure (POD), yielding an RfD
21   of 7×10^{-5} mg/kg-day after application of a total uncertainty factor (UF) of 100, accounting
22   for intra- and interspecies variability, subchronic study duration, and database limitations.

23   • Aroclor 1248 (posted in 1994;
24   [https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=649](#)): The
25   Aroclor 1248 assessment concluded that the health effect data were inadequate for the
26   derivation of an oral RfD. Derivation of an RfD was not recommended because a frank effect
27   (i.e., infant death) was noted at the lowest dose tested in rhesus monkeys. In the same set of
28   studies used for the Aroclor 1016 assessment, [Schantz et al. (1989)](#) evaluated
29   neurobehavioral performance in the offspring of rhesus monkeys exposed to 0.03, 0.1, and
30   0.2 mg/kg-day of dietary Aroclor 1248 for different durations. Infant death occurred at the
31   lowest dose and appeared to be a dose-responsive effect, leading to the identification of
32   0.03 mg/kg-day as a frank effect level.

33   • Aroclor 1254 (posted in 1994;
34   [https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=389](#)): [Arnold et](#)
35   [al. (1993b)](#), [Arnold et al. (1993a)](#), [Tryphonas et al. (1989)](#), [Tryphonas et al. (1991a)](#), and
36   [Tryphonas et al. (1991b)](#) were used in the derivation of the oral RfD for Aroclor 1254. In
37   these studies, mature female rhesus monkeys were exposed to 0.005, 0.02, 0.04, or
38   0.08 mg/kg body weight Aroclor 1254 each day over 6.5 years. The low dose of
39   0.005 mg/kg-day was identified as the LOAEL based on immunotoxicity and observations of
40   eye exudate, inflammation or prominence of the eyelid tarsal glands, and nail lesions. The
41   RfD of 2×10^{-5} mg/kg-day was calculated by applying a total UF of 300, which accounted for
42   intra- and interspecies variability, subchronic study duration, and the use of a LOAEL as the
43   POD.

2.2.   SCOPING SUMMARY

1   Since the current IRIS assessments for noncancer health effects of Aroclor mixtures were
2   completed in 1993–1994, studies on the noncancer health effects of exposure to environmentally
3   relevant PCB mixtures (e.g., similar to those found in contaminated fish or human milk) have been
4   conducted, and new data are available.
5   The commercial manufacture of PCBs was banned in the United States in 1979. Since that
6   time, their use, manufacture, cleanup, and disposal have been regulated under TSCA (40 CFR 761).
7   However, as discussed above, because of the past widespread use and persistence of PCBs in the
8   environment, humans continue to be exposed to them by inhalation of volatilized PCBs, inhalation
9   of contaminated dust, contact with contaminated dust, contact with primary or secondary sources
10   of PCBs, and ingestion of foods contaminated with PCBs, including breast milk. In addition to
11   regulation under TSCA, PCBs are regulated under the Clean Water Act, the Safe Drinking Water Act,
12   and the Resource Conservation and Recovery Act. Accordingly, PCBs are of interest to several EPA
13   program offices and regional offices due to widespread human exposure to PCBs from many
14   sources and through multiple environmental media.
15   During scoping, the IRIS program met with EPA program and regional offices interested in
16   an IRIS assessment for PCBs to discuss specific assessment needs. Table 1 provides a summary of
17   input from this outreach.

Table 1. EPA program and regional offices interest in a new assessment of PCBs (September 2018)

EPA program or regional office	Oral	Inhalation	Statutes/regulations and anticipated uses/interest
Office of Land and Emergency Management (OLEM) EPA Regions 1–10	✓	✓	Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) and Resource Conservation and Recovery Act (RCRA) PCBs have been identified at numerous contaminated waste sites, including 492 CERCLA National Priority List (NPL) sites.[a] 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. PCB toxicological information may be used to make risk determinations for response actions (e.g., CERCLA short-term removals, CERCLA long-term remedial response actions, or RCRA Corrective Action).

[a] The Superfund Enterprise Management System (SEMS) database identified 492 NPL sites where PCBs were documented as a contaminant in one or more media. SEMS is the official repository for site- and non-site-specific Superfund data in support of CERCLA. It contains information on hazardous waste site assessment and remediation from 1982 to the present. These site numbers are based on contaminant data from the remedy selection administrative records. SEMS only includes remedy data from fiscal years 1982 to 2014 for final and deleted NPL sites, and for sites with a Superfund Alternative Approach (SAA) agreement in place. NPL and SAA status is current as of September 17, 2018. The types of the 492 NPL and SAA sites include Superfund Federal Facility sites and non-Federal Facility sites, Fund-lead sites, and Enforcement sites where CERCLA remedial actions have been proposed. PCBs, identified as PCB, polychlorinated biphenyl, and Aroclor, were documented in SEMS as a contaminant in the evaluation of human health and ecological risks. PCBs were documented as a contaminant in a wide range of media, including air, soil, sediment, surface and ground water, sludge, fish tissue, and debris. Access to site documents or additional information about individual sites can be found at https://cumulis.epa.gov/supercpad/cursites/srchsites.cfm. This

website provides links to site profile pages, which typically include digital copies of site-related decision documents. Regarding the SEMS database, EPA is providing this PCB site data as a public service and does not vouch for the accuracy, completeness, or currency of data. Data provided by external parties is not independently verified by EPA. These data are made available to the public strictly for informational purposes. Data do not represent EPA's official position, viewpoint, or opinion, express or implied. This information is not intended for use in establishing liability or calculating Cost Recovery Statutes of Limitations and cannot be relied upon to create any rights, substantive or procedural, enforceable by any party in litigation with the United States or third parties. EPA reserves the right to change these data at any time without public notice.

1   A new IRIS assessment will identify noncancer human health hazards associated with
2   exposure to complex PCB mixtures (such as those found in the environment) through oral,
3   inhalation, and dermal routes, provided adequate data are available. Discussion of dose-response
4   information for identified hazards also will be included when feasible because this information can
5   be useful for characterizing risks at varying exposure levels and analyzing benefits associated with
6   reducing exposures. Derivation of an RfD for the dermal route of exposure is not planned at this
7   time because oral and inhalation exposure are generally considered the major exposure routes, and
8   relatively few studies of toxicological effects following dermal PCB exposure exist. However,
9   potential risks from dermal exposures can be evaluated using route-to-route extrapolation, and
10   toxicokinetic and other data relevant to dermal exposure will be included in the assessment to
11   support those evaluations. Furthermore, no new assessment for PCB cancer risk is planned. The
12   carcinogenicity of environmentally relevant PCB mixtures is addressed in the IRIS Carcinogenicity
13   Assessment for PCBs posted in 1996
14   (https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=294), and an update of
15   the evaluation of cancer risk from PCB exposure has not been identified as a priority need.

2.3.   PROBLEM FORMULATION

16   Problem formulation information pertaining to the new assessment of PCBs was included in
17   the scoping and problem formulation materials released to the public in April 2015 (U.S. EPA,
18   2015b); a public science meeting was held June 17–18, 2015 to obtain public input on these
19   materials (https://www.epa.gov/iris/iris-public-meeting-jun-2015).
20   As discussed in U.S. EPA (2015b), a preliminary literature survey was performed to identify
21   noncancer health outcomes evaluated for possible associations with PCB exposure. This survey
22   consisted of a search for health assessment information produced by other federal, state, and
23   international health agencies (summarized in Table 2), and an additional broad search of PubMed
24   to locate more recent studies. The review of health assessment information was used to identify
25   health effect categories for consideration in the IRIS assessment and was supplemented by the
26   PubMed search covering dates after the publication of the health assessment. In addition, the
27   preliminary literature survey was used to identify key scientific issues important for assessing
28   human health risk associated with PCB exposure. The PubMed search was not intended to be a
29   comprehensive search of all available literature but was intended to identify noncancer health
30   outcomes that had not been evaluated in prior health assessments.

1   The following health assessments, in addition to EPA's IRIS assessments for Aroclor 1016

2   (https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=462), Aroclor 1248

3   (https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=649), and Aroclor 1254

4   (https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=389), are available from

5   several federal and international health agencies:

6   1) Agency for Toxic Substances and Disease Registry. (ATSDR) (2011). Addendum to the
7   Toxicological Profile for Polychlorinated Biphenyls.
8   http://www.atsdr.cdc.gov/ToxProfiles/pcbs_addendum.pdf

9   2) Agency for Toxic Substances and Disease Registry. ATSDR (2000). Toxicological Profile for
10   Polychlorinated Biphenyls (PCBs). http://www.atsdr.cdc.gov/ToxProfiles/tp17.pdf

11   3) National Institute for Occupational Safety and Health. (NIOSH) (2019b). NIOSH Pocket
12   Guide to Chemical Hazards. RTECS. Chlorodiphenyl (54% chlorine).
13   http://www.cdc.gov/niosh/npg/npgd0126.html

14   4) National Institute for Occupational Safety and Health. (NIOSH) (2019a). NIOSH Pocket
15   Guide to Chemical Hazards. RTECS. Chlorodiphenyl (42% chlorine).
16   http://www.cdc.gov/niosh/npg/npgd0125.html

17   5) Occupational Safety and Health Administration. (OSHA) (2019). Chemical Sampling
18   Information, Chlorodiphenyl (42% Cl).
19   https://www.osha.gov/chemicaldata/chemResult.html?recNo=392

20   6) Occupational Safety and Health Administration. (OSHA) (2018). Chemical Sampling
21   Information, Chlorodiphenyl (54% Cl).
22   https://www.osha.gov/chemicaldata/chemResult.html?recNo=121

23   7) World Health Organization. (WHO) (2003). Concise International Chemical Assessment
24   Document 55. Polychlorinated Biphenyls: Human Health Aspects.
25   http://www.who.int/ipcs/publications/cicad/en/cicad55.pdf

Table 2. Noncancer PCB toxicity values from U.S. federal agencies and international bodies for exposures in the general population

Reference	Risk value or limit	Mixture	Rationale
ATSDR (2011) ATSDR (2000)	Chronic MRL: 0.02 µg/kg-d Intermediate MRL: 0.03 µg/kg-d	Aroclor 1254	Immunological (Tryphonas et al., 1991a; Tryphonas et al., 1989) Neurological (Rice, 1999; Rice and Hayward, 1999; Rice, 1998, 1997; Rice and Hayward, 1997)
WHO (2003)	Chronic MRL: 0.02 µg/kg-d	Aroclor 1254	Based on assessment by ATSDR (2000) and ATSDR (2011)
U.S. EPA (1994a)	RfD: 0.02 µg/kg-d	Aroclor 1254	Immunological, dermal, ocular (Arnold et al., 1993b; Arnold et al., 1993a; Tryphonas et al., 1991a; Tryphonas et al., 1991b; Tryphonas et al., 1989)
U.S. EPA (1993)	RfD: 0.07 µg/kg-d	Aroclor 1016	Reduced birth weight (Schantz et al., 1991; Schantz et al., 1989; Levin et al., 1988; Barsotti and van Miller, 1984)

1   Overall, the *Toxicological Profile for Polychlorinated Biphenyls* (PCBs) (ATSDR, 2011, 2000)
2   was found to be the most comprehensive and current resource, including detailed information on
3   the widest array of health effects and synthesizing evidence from the largest number of primary
4   research articles. Information from other assessments listed above was included in the preliminary
5   literature survey to the extent that it added to the information already presented in (ATSDR, 2011,
6   2000).
7   The preliminary literature survey identified human, animal, and in vitro studies related to
8   multiple noncancer health outcomes, mechanisms of action, mode-of-action (MOA) hypotheses,
9   toxicokinetics, and susceptible lifestages or populations. Each row in Table 3 summarizes whether
10   data are available on a particular broad health effect category or other toxicologically relevant
11   topic. Although the checkmarks in Table 3 indicate the existence of studies that investigated certain
12   health effect categories in the context of PCB exposure, they do not indicate whether the data from
13   those studies support associations between PCB exposure and health effects in those categories.
14   Each column in Table 3 indicates the types of studies that are available with respect to test system
15   (i.e., human, animal, in vitro) and exposure route (i.e., oral or inhalation) for animal studies or
16   exposure setting (i.e., occupational, high fish or seafood consumption,[2] general population) for

---

[2] Studies of populations with "high fish or seafood consumption" were those in which the study authors identified fish or seafood consumption, or both, as the PCB exposure source presumed to be dominant in the study population.

1   human studies. In addition, the table indicates whether animal studies of subchronic, chronic, or

2   developmental design[3] are available.

Table 3. Preliminary literature survey: PCB studies by test system, route of exposure, and health effect category[a]

	Human studies			Animal studies		In vitro studies
	Occupational	High fish/ seafood consumption[b]	General population	Oral	Inhalation	
Health effect categories						
Cardiovascular	✓	✓	✓	✓ (Subchronic, Chronic)		✓
Dermal	✓			✓ (Subchronic, Chronic, Developmental)		
Developmental	✓	✓	✓	✓ (Subchronic, Chronic, Developmental)	✓ (Subchronic)	✓
Endocrine	✓	✓	✓	✓ (Subchronic, Developmental)	✓ (Subchronic)	✓
Gastrointestinal	✓		✓	✓ (Subchronic, Chronic)		
Hematopoietic	✓			✓ (Subchronic, Chronic)	✓ (Subchronic)	
Hepatobiliary	✓	✓	✓	✓ (Subchronic, Chronic, Developmental)	✓ (Subchronic)	✓
Immune System	✓	✓	✓	✓ (Subchronic, Chronic, Developmental)	✓ (Subchronic)	✓

---

[3] In developmental studies, animals are exposed to a chemical during a critical window of development (i.e., the developmental period of vulnerability during which adverse effects can be triggered by exposures to environmental agents or other stressors). The critical windows of development for most biological systems occur during the prenatal and early postnatal periods, but certain systems (e.g., nervous and reproductive systems) do continue to develop throughout early life and adolescence. Studies conducted outside a critical window of development can be characterized by exposure duration: acute (<24 hours), short-term (>24 hours up to 30 days), subchronic (>30 days up to 10% of lifetime), and chronic (up to a lifetime).

	Human studies			Animal studies		In vitro studies
	Occupational	**High fish/ seafood consumption[b]**	**General population**	**Oral**	**Inhalation**	
Metabolic	✓	✓	✓	✓ (Subchronic, Chronic)		✓
Musculoskeletal	✓	✓		✓ (Subchronic, Chronic)		
Nervous System	✓	✓	✓	✓ (Subchronic, Developmental)	✓ (Subchronic)	✓
Ocular	✓			✓ (Subchronic, Chronic, Developmental)		
Reproductive	✓	✓	✓	✓ (Subchronic, Chronic, Developmental)		✓
Respiratory	✓		✓	✓ (Subchronic, Chronic)	✓ (Subchronic)	
Urinary System	✓		✓	✓ (Subchronic, Chronic)	✓ (Subchronic)	✓
Other data and analyses						
ADME	✓	✓	✓	✓	✓	✓
Pharmacokinetic models[c]	✓	✓	✓	✓		✓
MOA hypotheses		✓	✓	✓ (Subchronic, Chronic, Developmental)		✓
Susceptibility data[d]	✓	✓	✓	✓ (Developmental)		
Genotoxicity[e]	✓		✓	✓ (Subchronic)		✓

ADME = absorption, distribution, metabolism, and excretion; MOA = mode of action.

[a] Checkmarks indicate that one or more studies have been identified but do not indicate confidence in the methods used in those studies, or if those studies support associations between PCB exposure and one or more health effect(s) in that category; the absence of a checkmark indicates that no studies were identified for a given health effect category and study design.

[b] Studies of populations with "high fish/seafood consumption" were those in which the study authors identified fish or seafood consumption, or both, as the PCB exposure source presumed to be dominant in the study population.

[c] Earliest physiologically based pharmacokinetic (PBPK) models for PCBs were based on intravenous exposure. Models also exist for dermal exposure.

[d] Individuals who might be more susceptible to toxic effects include young children, especially those who are breastfed.

[e] Includes studies investigating potential epigenetic impacts of PCB exposure.

2.4. ASSESSMENT APPROACH

1   The overall objective of this assessment is to identify adverse human health effects and
2   characterize exposure-response relationships for the effects of PCB mixtures to support the
3   development of oral and inhalation noncancer toxicity values. This assessment will use systematic
4   review methods to evaluate the epidemiological and toxicological literature for PCBs; mechanistic
5   evidence will also be considered, focusing on data informative to analyses of the key science issues
6   identified in Section 2.5 (see Section 9.2). The evaluations conducted in this assessment will be
7   consistent with relevant EPA guidance.[4]
8   The specific approach taken to the assessment of the health effects of PCBs will be based on
9   input received during scoping, a survey of the literature describing the health effects of PCBs, and
10 consideration of the physicochemical properties of PCBs. The literature noted and screened as
11 described in Section 2.3 was used to identify broad categories of potential health effects considered
12 to be most relevant for assessment. U.S. EPA (2015b) proposed that the following list of broad
13 health effect categories be considered for further evaluation to identify specific health endpoints for
14 systematic review: cardiovascular, dermal, developmental, endocrine, gastrointestinal,
15 hematopoietic, hepatobiliary, immune, metabolic, nervous system, ocular and reproductive effects.
16 After consideration of stakeholder input collected at the public science meeting
17 (https://www.epa.gov/iris/iris-public-meeting-jun-2015) on June 17–18, 2015, the decision was
18 made also to include musculoskeletal, respiratory, and urinary effects in this preliminary analysis,
19 the results of which are described in Section 5.

2.5. KEY SCIENCE ISSUES

20 As described in U.S. EPA (2015b) and discussed at the public science meeting on June 17–
21 18, 2015 (https://www.epa.gov/iris/iris-public-meeting-jun-2015), the following key scientific
22 issues were identified that warrant further consideration in this assessment.

2.5.1. Impact of Congener Profile on the Toxicity of PCB Mixtures

23 Humans are environmentally exposed to PCBs as complex mixtures of congeners. PCB
24 congeners differ not only structurally but also qualitatively and quantitatively with respect to
25 biological responses. Prior to human exposure, PCB mixtures in the environment undergo

---

[4] EPA guidance documents: http://www.epa.gov/iris/basic-information-about-integrated-risk-information-system#guidance/. Note: the Agency has initiated a review of, and possible updates to, their guidance documents.

1   processes such as volatilization and preferential bioaccumulation, which can result in dramatic
2   differences in the congener profiles of PCB mixtures found in various exposure sources (e.g., human
3   milk, contaminated fish, indoor air) (IARC, 2015; ATSDR, 2000). Although environmental PCB
4   mixtures can be characterized analytically as if they were Aroclors, this can be imprecise given
5   weathering and degradation.[5] Furthermore, of all possible congener combinations that might exist,
6   a relatively small subset of complex PCB mixtures has been tested in animal studies; important
7   differences might exist between these tested mixtures and the mixtures to which humans are
8   exposed in the environment. Thus, methods for translating toxicological data from tested to
9   untested mixtures would be useful, including methods for addressing PCB mixtures with varying
10 proportions of congeners with diverse modes of action (e.g., "dioxin-like," "estrogenic," "anti-
11 estrogenic," "neurotoxic" (Wolff et al., 1997; Wolff and Toniolo, 1995)).
12   For these reasons, approaches for assessing chemical mixtures will be evaluated for use in
13 this assessment. The *Supplementary Guidance for Conducting Health Risk Assessment of Chemical*
14 *Mixtures* (U.S. EPA, 2000) recommends several approaches to quantitative health risk assessment of
15 a chemical mixture, depending on the type of available data. The preferred approach is to use
16 toxicity data on the mixture of concern. Alternatively, when toxicity data are not available for the
17 mixture of concern, use of toxicity data on a "sufficiently similar" mixture is recommended.
18 Sufficient similarity, as discussed in U.S. EPA (2000), implies that the toxicological consequences of
19 exposure to the mixture of concern are expected to be identical or indistinguishable from those of
20 the mixture for which data are available. Sufficiently similar mixtures are of similar chemical
21 composition, or there is some understanding of chemical differences between the mixtures.
22 Methods for defining sufficient similarity have been developed and are an area of active
23 investigation (Catlin et al., 2018; Rice et al., 2018; Marshall et al., 2013; Feder et al., 2009). The
24 feasibility of using these new methods to support the derivation and application of toxicity values
25 for PCB mixtures is a current research area; if specific methods are identified for use in this
26 assessment, they will be described in a protocol update.
27   As described in U.S. EPA (2015b), the assessment will review the available data and, as
28 feasible, will develop oral and inhalation toxicity values for environmental PCB mixtures by
29 evaluating (1) toxic potencies of complex PCB mixtures (e.g., environmental, commercial) tested for
30 various noncancer health effects in animal bioassays; (2) methods for using toxicological data from
31 a limited set of tested PCB mixtures for human health risk assessment in a wide variety of exposure
32 contexts (e.g., sufficient similarity testing); and (3) approaches for assessing health risk based on

---

[5] There are diverse analytical methodologies for measuring PCBs in the environment and in biological samples. Some of these analytical methods treat all environmental mixtures as Aroclor mixtures (with rigid congener profiles), whereas others measure individual congeners (all 209 congeners or a subset of selected congeners). Aroclor analyses are cost-effective but have limitations since chromatographic patterns and peak ratios can change during environmental weathering. This can be especially challenging when multiple Aroclors are present at a given site (Erickson, 2018; U.S. EPA, 1996b; Alford-Stevens, 1986; Alford-Stevens et al., 1985).

1   measurements of PCB levels in the environment collected using various analytical techniques

2   (e.g., Aroclor analyses vs. congener-specific analyses).

2.5.2. Potential for Hazard Identification and Dose-Response Assessment for PCB Exposure via Inhalation

3   As described in U.S. EPA (2015b), evidence suggests that PCB inhalation can pose a hazard

4   to human health. However, the database of studies investigating health effects resulting from PCB

5   exposure consists primarily of oral exposure studies. Whether the existing database of inhalation

6   studies will be adequate to support human health risk assessment for inhalation exposure to PCBs

7   is not clear (Lehmann et al., 2015). Based on the available data, feasible options for conducting a

8   dose-response assessment for PCB inhalation exposure will be evaluated, considering differences in

9   toxicity of congeners that are inhaled versus ingested and differences between the inhalation and

10   oral exposure routes. Potential options include the use of data from available PCB inhalation

11   studies or the use of kinetic models or default approaches for route-to-route extrapolation from

12   oral PCB exposure data.

2.5.3. Suitability of Available Pharmacokinetic Models for Reliable Route-to-Route, Interspecies, or Intraspecies Extrapolation

13   Because the assessment will address noncancer hazards associated with exposure to

14   complex PCB mixtures, available pharmacokinetic models will be evaluated for their ability to

15   predict the dose metrics of such mixtures. Further information regarding pharmacokinetic model

16   evaluation can be found in Section 6.4. Lipophilicity, binding to liver proteins (e.g., cytochromes,

17   AhR), and rate of elimination (due to metabolism or fecal excretion) are the main determinants of

18   PCB toxicokinetics. Variation of these toxicokinetic determinants among individual PCBs limits the

19   application of congener-specific models in the assessment of a complex PCB mixture. A single set of

20   parameters to describe these determinants for the complex mixture might not be justifiable

21   because significant individual toxicokinetic variation has been observed for different PCB

22   congeners. Additionally, possibilities of toxicokinetic interaction, such as competition at binding

23   sites or synergy in the case of induction of enzymes, could exist between PCB congeners in a

24   complex mixture.

25   As described in U.S. EPA (2015b), the assessment will evaluate (1) existing pharmacokinetic

26   models for their potential ability to support reliable route-to-route, interspecies, or intraspecies

27   extrapolations, including the ability to quantitatively predict transfer of PCBs across the placenta or

28   via breast milk; (2) available information on toxicokinetic differences among PCB congeners and

29   mixtures; and (3) available information on inter- or intraspecies differences in the toxicokinetics of

30   PCBs, including differences across lifestages. All of this information will be carefully considered

31   during evidence synthesis (Section 9) and dose-response assessment (Section 11).

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

3.1. SPECIFIC AIMS

1   The aims of the assessment are to:

2   • Identify epidemiological (i.e., human) and toxicological (i.e., experimental animal) literature
3   reporting effects of exposure to PCBs as outlined in the PECO. The assessment will include
4   evaluations of the evidence relevant to the following noncancer health effect categories:
5   cardiovascular, dermal, developmental, endocrine, gastrointestinal, hematopoietic,
6   hepatobiliary, immune, metabolic, musculoskeletal, nervous system, urinary, reproductive,
7   and respiratory. The systematic review will focus on the highest priority health effect
8   categories and outcomes (see Section 5).

9   • Evaluate mechanistic information (including toxicokinetic understanding) associated with
10   exposure to PCBs to inform the interpretation of findings related to potential health effects
11   in studies of humans and animals. The scope of these analyses of mechanistic information
12   will be determined by the complexity and confidence in the phenotypic evidence in humans
13   and animals, the likelihood of the analyses to impact evidence synthesis conclusions for
14   human health, and the directness or relevance of the available model systems for
15   understanding potential human health hazards (Section 9.2). The mechanistic evaluations
16   will focus primarily on the key science issues identified in Section 2.5.

17   • Conduct study evaluations for individual epidemiology and toxicology studies (evaluating
18   reporting quality, risk of bias, and sensitivity) and PBPK models (scientific and technical
19   review).

20   • Extract data on relevant health outcomes from selected epidemiology and toxicology
21   studies based on the study evaluations. Full data extraction of low confidence studies might
22   not be performed for poorly studied health effects or for health effects for which extensive
23   *medium* and *high* confidence studies are available.

24   • Synthesize the evidence across studies, assessing similar health outcomes using a narrative
25   approach.

26   • For each health outcome (or grouping of outcomes), evaluate the strength of evidence
27   across studies (or subsets of studies) separately for studies of exposed humans and for
28   animal studies. If studies informing mechanisms are synthesized, mechanistic evidence will
29   be used to inform evaluations of the available health effect evidence (or lack thereof).

30   • For each health outcome (or grouping of outcomes), develop an integrated expert judgment
31   across evidence streams as to whether the evidence is sufficient (or insufficient) to indicate

1   that exposure to PCBs has the potential to be hazardous to humans (in rare instances, the
2   evidence may be judged as sufficient to indicate that a hazard is unlikely). The judgment
3   will be directly informed by the evidence syntheses and based on structured review of an
4   adapted set of considerations for causality first introduced by Austin Bradford Hill ([Hill,](#)
5   [1965](#)) (see Sections 9 and 10), including consideration (e.g., based on available mechanistic
6   information) and discussion of biological understanding. As part of the evidence
7   integration narrative, characterize the strength of evidence for the available database of
8   studies and its uncertainties, and identify and discuss issues concerning potentially
9   susceptible populations and lifestages.

10   •   Derive toxicity values (e.g., RfDs, RfCs) as supported by the available data (see Section 10.2),
11   considering the similarities and differences in toxicity among PCB mixtures (see Section
12   2.5.1). Evaluating the applicability and uncertainties of methods to address potential
13   differences will be a key consideration. The feasibility of using these methods to support
14   the derivation and application of PCB mixture-specific toxicity values is a current research
15   area; if specific methods are identified for use in this assessment, they will be described in a
16   protocol update.

17   •   Evaluate the feasibility and applicability of pharmacokinetic and dosimetric modeling to
18   account for interspecies differences and route-to-route extrapolation. In the absence of
19   appropriate models or data, apply default dosimetric adjustments and explore alternative
20   approaches to developing estimates across exposure routes. Given the differences in
21   toxicokinetic properties among PCB congeners (see Sections 2.5.2 and 2.5.3), evaluating the
22   applicability and uncertainties of methods to address these potential differences will be a
23   key consideration.

24   •   Characterize uncertainties and identify key data gaps and research needs such as
25   limitations of the available evidence, limitations of the systematic review, and consideration
26   of dose relevance and toxicokinetic differences when extrapolating findings from higher
27   dose animal studies to lower levels of human exposure.

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

28   The PECO is used to identify the evidence that addresses the specific aims of the assessment

29 and to focus the literature screening, including the inclusion/exclusion criteria, in a systematic

30 review. The PECO criteria for PCBs (see Table 4) are based on (1) nomination of the chemical for

31 assessment; (2) discussions with scientists in EPA program and regional offices to determine the

32 scope of the assessment that will best meet Agency needs; (3) preliminary review of the health

33 effect literature for PCBs (primarily reviews and authoritative health assessment documents as

34 described in Section 2.3) to identify the major health hazards associated with exposure to PCBs and

35 key areas of scientific complexity; and (4) input received during public discussion of preliminary

36 materials released to the public in 2015.

37   In addition to those studies meeting the PECO criteria, studies containing supplemental

38 material that are potentially relevant to the specific aims of the assessment were tracked during the

39 literature screening process. Although these studies did not meet PECO criteria, they were not

1   excluded from further consideration. The categories used to track studies as "potentially relevant

2   supplemental material" during screening and to prioritize these studies for consideration in the

3   assessment based on likelihood to impact evidence synthesis conclusions for human health are

4   described in Section 4.3.

Table 4. Populations, exposures, comparators, outcomes (PECO) criteria

PECO element	Evidence
<u>Populations</u>	**Human:** Any population and lifestage (occupational or general population, including children and other sensitive populations). **Animal:** Nonhuman mammalian animal species (whole organism) of any lifestage (including preconception, in utero, lactation, peripubertal, and adult stages).
<u>Exposures</u>	**Human:** Any exposure to PCBs (in vivo) as determined by: • Controlled exposure • Measured concentration in contact medium (e.g., food, air, dust) • Biomarkers of exposure (e.g., serum PCB levels) • Occupation in a job involving exposure to PCBs (e.g., electric capacitor manufacturing) • Self-reported history of using commercial products containing PCBs (e.g., mixing Aroclors into caulk). **Animal:** One or more oral (gavage, diet, drinking water, intragastric), inhalation (aerosol, vapor, or particle; whole-body or nose-only), dermal (occlusive, semi-occlusive, non-occlusive), or injected (intravenous, subcutaneous, intraperitoneal) treatment(s) with any clearly quantified dosage of PCBs alone administered to a whole animal (in vivo).
<u>Comparators</u>	**Human:** A comparison or referent population exposed to lower levels (or no exposure/exposure below detection limits) of PCBs, or exposure to PCBs for shorter time periods. Case reports and case series will be tracked as "potentially relevant supplemental information." **Animal:** A concurrent control group exposed to vehicle-only treatment or untreated control.
<u>Outcomes</u>	**Human:** Any examination of survival, body weight, or development, or of the structure or function of dermatologic, cardiovascular, endocrine, gastrointestinal, hematologic, hepatic, immune, nervous, ocular, musculoskeletal, renal, respiratory or reproductive cells, tissues or systems. **Animal:** Any examination of survival, body weight, or development, or of the structure or function of dermatologic, cardiovascular, endocrine, gastrointestinal, hematologic, hepatic, immune, nervous, ocular, musculoskeletal, renal, respiratory or reproductive cells, tissues or systems.

4. LITERATURE SEARCH AND SCREENING STRATEGIES

4.1. LITERATURE SEARCH STRATEGIES

1   The literature search strategy relied on terms to gather information on exposure to PCB
2   mixtures and individual PCB congeners ("E" component of PECO). Additional exposure terms were
3   used to identify studies that were not indexed by the chemical name (e.g., poisoning events
4   [Yusho/Yu-Cheng], capacitor manufacturing workers). These exposure terms were intentionally
5   broad and did not prioritize studies in which exposure was quantified; this was considered during
6   screening of the literature (Section 4.3). The search queries did not contain terms for the
7   population ("P"), comparison ("C"), or outcome ("O") components of the PECO statement; these
8   were also considered during screening of the literature.
9   The following databases were searched:

10   • PubMed (National Library of Medicine)

11   • Web of Science (Thomson Reuters)

12   • Toxline (National Library of Medicine)

13   Searches were not restricted by publication date, and no language restrictions were applied.
14   The detailed search strategies are presented in Appendix A. Literature searches were conducted
15   using EPA's HERO database.[6] The HERO page for the PCB assessment contains the literature search
16   results (https://hero.epa.gov/hero/index.cfm/project/page/project_id/384). The literature
17   search will be periodically updated throughout development of the draft assessment to identify
18   literature published during the course of review.[7] The last full literature search update is
19   anticipated to be conducted less than 1 year before the planned release of the draft assessment
20   document for public comment. The results returned (i.e., the number of "hits" from each electronic
21   database or other literature source), including the results of any literature search updates, are

---

[6] Health and Environmental Research Online: https://hero.epa.gov/hero/.
[7] The initial literature search was completed in July 2015 and is updated annually. References retrieved
through August 2016 are accounted for in this protocol. The literature is currently being updated and will be
updated regularly until several months prior to public release of the draft assessment. As such, the methods
for literature search and screening (and some of the approaches to refining the evaluation plan based on the
identified literature; see Section 5) are described in the protocol using the past tense, while the approaches
for other assessment methods are outlined in future tense.

1   documented in the literature flow diagrams (see Figure 6, which also reflect the literature screening

2   decisions (see Section 4.3).

3   The IRIS Program takes extra steps to ensure identification of pertinent studies: by

4   encouraging the scientific community and the public to identify additional studies and ongoing

5   research; by searching for publicly available data submitted under the Toxic Substances Control Act

6   (TSCA) and the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA); and by considering

7   late-breaking studies that would impact the credibility of the conclusions, even during the review

8   process.[8] Release of the PECO-screened literature in parallel with release of the protocol for public

9   comment provides an opportunity for stakeholders to identify any missing studies, which, if

10   identified, will be screened as outlined above for adherence to the PECO criteria. Studies identified

11   after peer review begins will be considered for inclusion only if they are directly relevant to the

12   PECO criteria and could fundamentally alter the assessment's conclusions.

4.2.   NONPEER-REVIEWED DATA

13   IRIS assessments rely mainly on publicly accessible, peer-reviewed studies. However, it is

14   possible that nonpeer-reviewed data directly relevant to the PECO could be identified during

15   assessment development. EPA might obtain external peer review if the owners of the data are

16   willing to have the study details and results made publicly accessible. Consistent with policies and

17   procedures outlined in U.S. EPA's *Science Policy Council Peer Review Handbook* (U.S. EPA, 2015a),

18   this independent, contractor-driven, peer review would include an evaluation of the study similar

19   to that for peer review of a journal publication. The contractor would identify and select two or

20   three scientists knowledgeable in scientific disciplines relevant to the topic as potential peer

21   reviewers. Persons invited to serve as peer reviewers would be screened for conflict of interest. In

22   most instances, the peer review would be conducted by letter review. The study authors would be

23   informed of the outcome of the peer review and given an opportunity to clarify issues or provide

24   missing details. The study and its related information, if used in the IRIS assessment, would

25   become publicly available. In the assessment, EPA would acknowledge that the document

26   underwent peer review, and the names of the peer reviewers would be identified. In certain cases,

27   IRIS will conduct an assessment for utility and data analysis based on having access to a description

28   of study methods and to the raw data that have undergone rigorous quality assurance/quality

29   control review (e.g., ToxCast/Tox21 data, results of National Toxicology Program [NTP] studies)

30   but that have not yet undergone external peer review.

31   Unpublished data from personal communication by the author can supplement a

32   peer-reviewed study provided the information is made publicly available (typically through

33   documentation in HERO).

---

[8] IRIS "stopping rules": https://www.epa.gov/sites/production/files/2014-06/documents/iris_stoppingrules.pdf.

4.3. LITERATURE SCREENING STRATEGY

1   The PECO criteria were used to determine inclusion or exclusion of a reference as a primary
2 source of health effect data. In addition to the inclusion of studies that meet the PECO criteria,
3 studies containing supplemental material that is potentially relevant to the specific aims were
4 tracked during the screening process. Although not considered to directly meet PECO criteria,
5 these studies are not strictly excluded unless otherwise specified. Unlike studies that meet PECO
6 criteria, supplemental studies might not be subject to additional consideration unless they help
7 address specific assessment aims (see Section 3.1). Studies that were categorized as "potentially
8 relevant supplemental material" include the following:

9   • Study materials that have not been peer reviewed;

10   • Study materials published in a language other than English;

11   • Records that do not contain original data, such as other agency assessments, informative
12   scientific literature reviews, editorials, or commentaries;

13   • Studies appearing only as abstracts (e.g., conference abstracts);

14   • Mechanistic studies: Studies reporting measurements related to a health outcome that
15   informs the biological or chemical events associated with phenotypic effects, in both
16   mammalian and nonmammalian model systems, including in vitro, in vivo (by various
17   routes of exposure), ex vivo, and in silico studies;

18   • ADME studies: Studies designed to capture information regarding absorption, distribution,
19   metabolism, and excretion, including toxicokinetic studies and studies describing PBPK
20   models for PCB congeners and mixtures;

21   • Exposure characteristics: Exposure studies that include data unrelated to toxicological
22   endpoints, but which provide information on exposure sources or measurement properties
23   of the environmental agent (e.g., demonstrating a biomarker of exposure);

24   • Susceptible populations: Studies that identify potentially susceptible populations and
25   lifestages, such as studies that focus on a specific demographic, lifestage, or genotype;

26   • Human case reports or case series; and

27   • Studies of PCB exposures and health effects in wildlife populations.

28   Because of the large size of the database and large number of health effects associated with
29 exposure to PCBs, the PECO criteria are expected to be narrowed to focus on the highest priority
30 health outcomes (see Section 5). As described below, the initial literature screen, conducted
31 according to the PECO criteria listed in Section 3.2, was used to identify the studies included in
32 summary-level literature inventories (see Section 4.4).

4.3.1. Electronic Screening

1   The initial literature search described in Section 4.1 identified over 50,000 references.
2   Manual review of every record would have been time and resource intensive, so natural language
3   processing (NLP) and machine learning (ML) techniques were employed to identify the most
4   relevant literature for screening. Studies were prioritized using DoCTER, a publicly available
5   Document Classification and Topic Extraction Resource (https://www.icfdocter.com/index).
6   Details of the NLP and ML methods are described elsewhere (Varghese et al., 2019; Varghese et al.,
7   2017). Briefly, 484 studies
8   (https://hero.epa.gov/hero/index.cfm/project/page/usage_id/18723/format/list/sort/year%
9   20desc/usage_searchType/any/page/1/project_id/384/rows/10) selected as meeting the PECO
10   criteria for inclusion in the assessment were designated as seed[9] references and included in the
11   corpus of 53,801 references identified by the literature searches. In phase one of a two-phase
12   approach, schematically illustrated in Figure 3, titles and abstracts were represented in a
13   mathematical matrix and organized into clusters based on semantic similarity using NLP tools
14   (Figure 3).
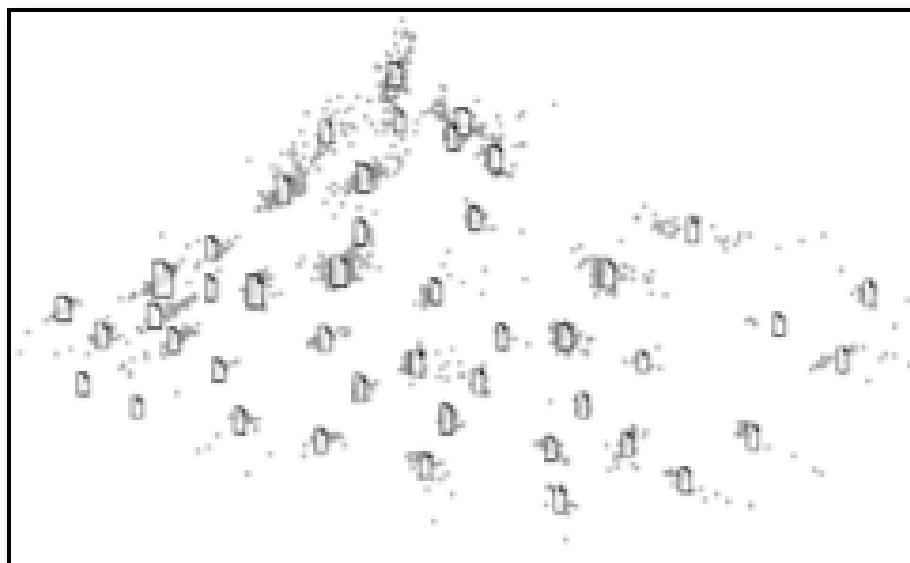
Figure 3. Schematic illustration of electronic prioritization of literature depicting references clustered by similarity using natural language processing.

---

[9] Seed references are a subset of the larger collection of unclassified references that are known to be topic relevant. In a clustering analysis based on NLP, the distribution of seed references can be used to identify clusters most likely to contain relevant references, as the references in the clusters containing seeds are expected to be similar to the seeds (Varghese et al., 2017).

1   Two clustering algorithms (k-means, nonnegative matrix factorization) were applied using

2   cluster sizes of 10, 20, or 30 references for a total of 6 different clustering approaches. Clusters

3   harboring seed references were identified (Figure 4).
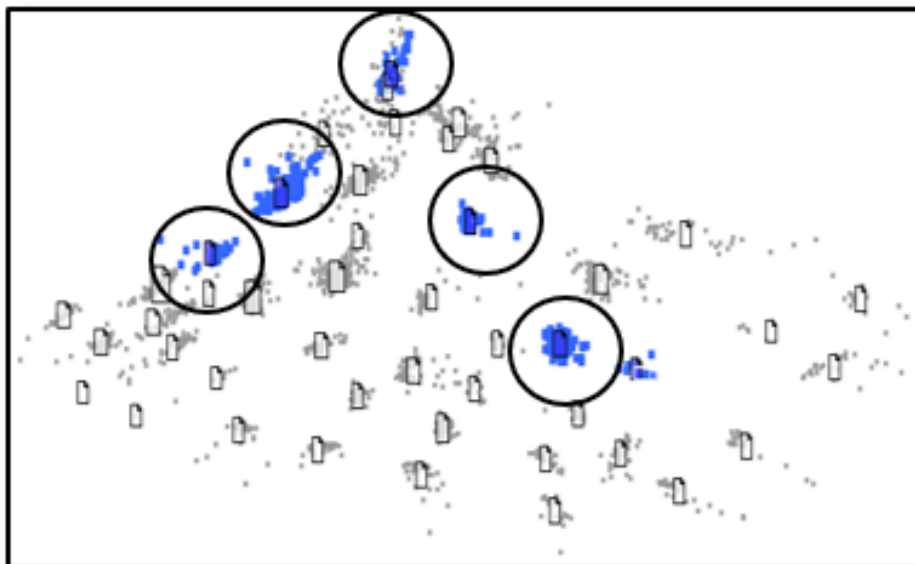
**Figure 4. Illustration depicting clusters containing relevant seed references
(circled blue clusters).** Clusters were ranked by the number of seed studies
included.

4   In each approach, clusters were ranked in decreasing order of the number of seed studies in

5   each cluster, and clusters were accepted in order until 90% of the total set of seed studies was

6   captured. This was repeated for all six approaches; thus, a given study could have appeared in one

7   of the accepted clusters (and thus appear with the greatest fraction of the seed studies) in 0 to 6 of

8   the approaches. Clusters containing seed references were grouped by the number of approaches in

9   which they were identified (groups A–F, Figure 5).

10   Studies that appeared in ranked clusters using 6, 5, 4, or 3 approaches were subjected to

11   title and abstract-level screening, as described below (groups A–D, Figure 5).

12   Then, for phase two, ML was used to predict relevance for those studies in the remaining

13   groups of clusters that appeared in one or two approaches (groups E and F, Figure 5). Also

14   included in this approach was one group of studies excluded from initial clustering until abstracts

15   were recovered and a second group of studies with titles only. The training dataset for this

16   secondary analysis included PECO-relevant and nonrelevant studies identified during screening in

17   phase one. Studies predicted to be relevant then were subjected to manual title and abstract-level

18   screening.
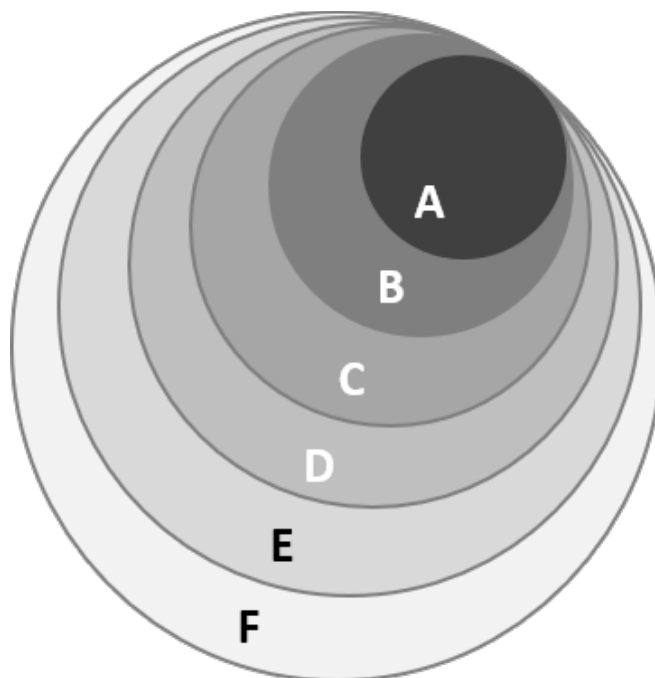
Figure 5. Visualization of identified clusters. Clusters were organized into groups (A–F) based on the number of approaches that identified the cluster such that group A contains clusters harboring seed references identified by six approaches and group F contains clusters harboring seed references identified by a single approach. All references in the top four groups (A–D) were manually screened for inclusion based on PECO criteria. Low-scoring groups (E, F) were subjected to additional machine learning approaches to capture relevant references for manual screening.

1   The number of studies identified using electronic prioritization methods is summarized in
2 Table 5. Studies not reviewed included those not identified by any of the clustering approaches or
3 those identified by one or two approaches but predicted to be nonrelevant during the ML phase. A
4 subset of nonprioritized studies was randomly selected for manual title and abstract-level review;
5 this additional review demonstrated that less than 10% of nonprioritized studies were relevant
6 based on PECO criteria.

Table 5. Electronic prioritization of literature for hazard identification

Group of studies	Prioritization approach	Number of prioritized studies (of 53,798 retrieved in original search)
Groups A–D (Figure 5)	Supervised clustering	4,652
Groups E–F (Figure 5)	Supervised clustering and ML	3,428
Studies with titles only	ML	3,302
Total number electronically prioritized		**11,382**

Table 6. Sources of studies subjected to manual review for relevance to hazard identification

Source	Number of references
DoCTER electronic prioritization (after duplicate removal) (Table 5)	11,382
Seed references used for priority ranking	484
Other sources	8
Stakeholder identified	29
2015–2016 literature update	1,818
Total manually screened	**13,721**

4.3.2. Title and abstract-level screening

Following a pilot phase to calibrate screening guidance, two screeners independently conducted a manual title and abstract screen to identify records that appeared to meet the PECO criteria for studies electronically prescreened as described above. Literature updates, references identified as seed studies, and stakeholder-identified references yielded fewer references and were not subjected to electronic prioritization prior to manual review (summarized in Table 6). References retrieved through August 2016 were screened using structured forms developed for DRAGON (a modular database with integrated literature evaluation and screening tools developed for systematic review) ([ICF, 2018](#)). References identified in search updates after August 2016 will be reviewed using SWIFT-Active Screener (Sciome; https://www.sciome.com/swift-activescreener/), in which manual review is integrated with electronic prioritization using ML and statistical approaches.

For citations with no abstract, articles were screened based on title relevance. 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.

Studies not meeting the PECO criteria but identified as "potentially relevant supplemental material" were categorized (i.e., tagged) during the title and abstract screening process (further described in Section 4.3). Conflict resolution is not required during the screening process to identify supplemental information (i.e., tagging by a single screener is sufficient to identify the study as potentially relevant supplemental material that could be considered during draft development).

4.3.3. Full-text level screening

Records that were not excluded based on the title and abstract 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 to confirm eligibility according to the PECO criteria.

1   Screening conflicts were resolved by discussion between the primary screeners with consultation

2   by a third reviewer or technical advisor (as needed to resolve any remaining disagreements).

3   The results of this screening process were posted on the project page for this assessment in

4   the HERO database (https://hero.epa.gov/hero/index.cfm/project/page/project_id/384), and

5   studies were "tagged" with appropriate category descriptors (e.g., studies eligible for study

6   evaluation, potentially relevant supplemental material, excluded). Results were also annotated

7   and reported in a literature flow diagram (see Figure 6). Figure 6 reflects literature searches

8   through August 2016. Literature search updates will be conducted, and the results will be

9   reflected in the draft assessment; the most current results can be viewed at any time in the HERO

10   project page provided above.
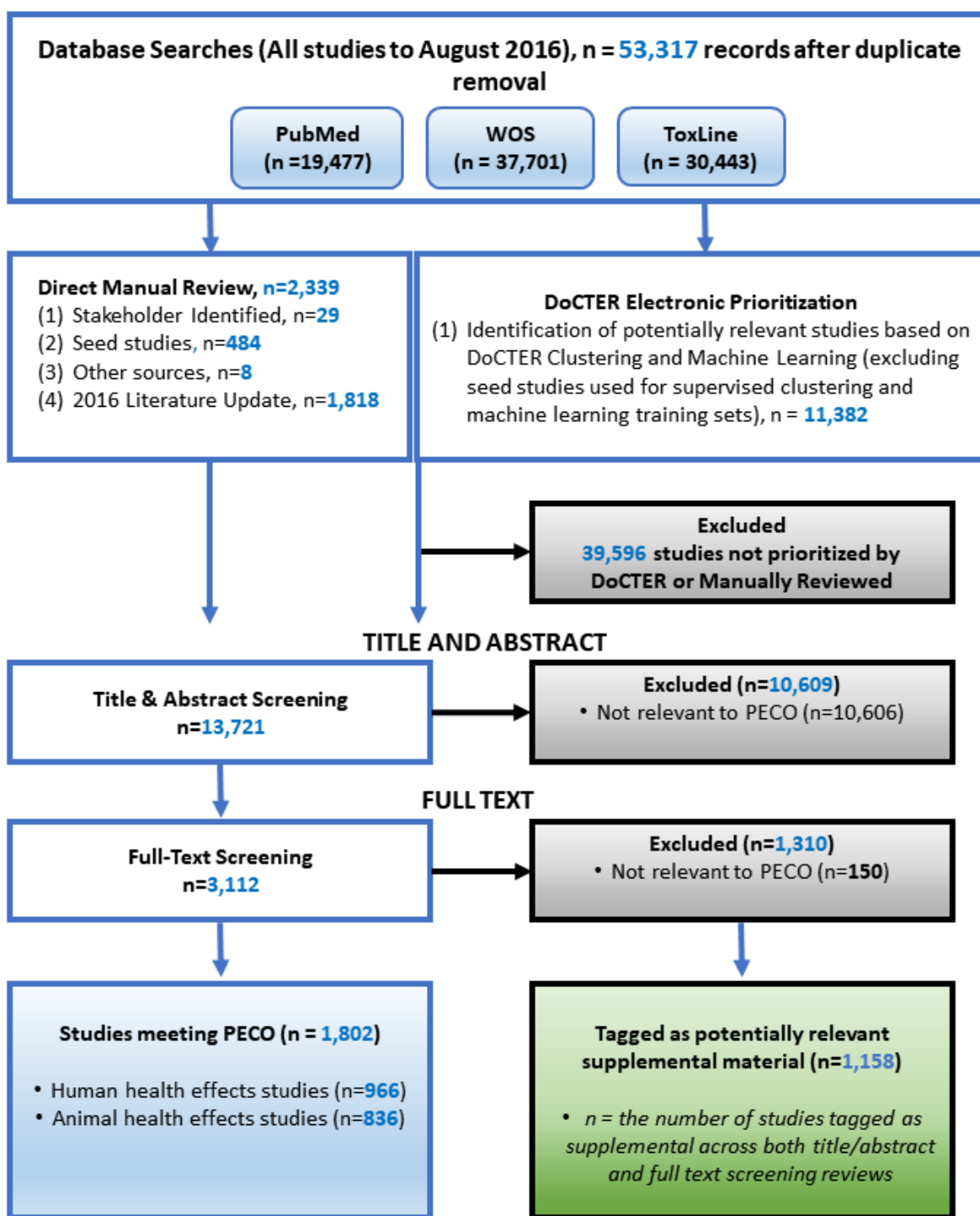
Figure 6. Literature search flow diagram for PCBs.

4.3.4. Multiple Publications of the Same Data

1   For multiple publications using the same or overlapping data, all publications on the
2 research will be included, with one selected for use as the primary study; the others will be
3 considered as secondary publications with annotation indicating their relationship to the primary
4 record during data extraction. For epidemiology studies, the primary publication generally will be
5 the one with the longest follow-up, the largest number of cases, or the most recent publication date.
6 For animal studies, the primary publication typically will be the one with the longest duration of
7 exposure, or the one that assessed the outcome(s) most informative to the PECO. For both
8 epidemiology and animal studies, the assessment will include relevant data from all publications of
9 the study; however, if the same outcome is reported in more than one report, the data will be
10 extracted only once.

4.4. SUMMARY-LEVEL LITERATURE INVENTORIES

11   During full text-level screening, studies tagged based on PECO eligibility were further
12 categorized based on features such as evidence type (i.e., human or animal), health outcome(s), or
13 endpoint measure(s) included in the study. Based on the results of discussions with external
14 stakeholders at the public science meeting held to discuss scoping and problem formulation
15 materials for PCBs on June 17–18, 2015 (https://www.epa.gov/iris/iris-public-meeting-jun-2015),
16 studies were tagged to the following health outcome categories: cardiovascular, dermal,
17 developmental, endocrine, gastrointestinal, hematopoietic, hepatobiliary, immune, metabolic,
18 musculoskeletal, nervous system, ocular, reproductive, respiratory, and urinary. Literature
19 inventories for PECO-relevant studies were created to develop summary-level, sortable lists that
20 include some basic study design information (e.g., study population, exposure information such as
21 doses administered or biomarkers analyzed, age/lifestage[10] of exposure, endpoints examined).
22 These literature inventories facilitate subsequent review of individual studies or sets of studies by
23 topic specific experts.
24   Inventories also will be created for studies that were tagged as "potentially relevant
25 supplemental material" during screening, including mechanistic studies (e.g., in vitro or in silico
26 models), ADME studies, and studies on endpoints or routes of exposure that do not meet the
27 specific PECO criteria but that might still be relevant to the research question(s). Here, the
28 objective is to create an inventory of studies that can be tracked and further summarized as
29 needed—for example, by model system, key characteristic [e.g., of carcinogens (Smith et al., 2016)],
30 mechanistic endpoint, or key event—to support analyses of critical mechanistic questions that arise
31 at various stages of the systematic review (see Section 9.2 for a description of the process for
32 determining the specific questions and pertinent mechanistic studies to be analyzed). ADME data

[10]Age/lifestage of chemical exposure will be considered according to EPA's *Guidance on Selecting Age Groups for Monitoring and Assessing Childhood Exposures to Environmental Contaminants* and EPA's *A Framework for Assessing Health Risk of Environmental Exposures to Children*.

1   and related information can be critical to the next steps of prioritizing or evaluating individual

2   PECO-specific studies and will be reviewed by subject matter experts early in the assessment

3   process. For example, the comprehensive identification of studies relevant to interpreting the

4   ADME or toxicokinetic characteristics of PCB congeners and mixtures will be prioritized.

5. REFINED EVALUATION PLAN

1 The purpose of the refined evaluation plan is to describe refinements to the set of studies
2 meeting PECO criteria to be carried forward to study evaluation and to identify and group the
3 endpoints that will be the primary focus of the outcome-specific evaluations. The process also
4 helps determine which studies tagged as "potentially relevant supplemental material" might need
5 to be considered in the assessment. To focus on the studies most informative to this human health
6 assessment, refinements to the initial PECO criteria were developed based on the literature
7 inventories (shown in Tables 7–10) and explanations provided. The numbers of studies reporting
8 human and animal evidence associated with PCB exposure and specific health endpoints grouped
9 by hazard category are depicted in Figures 7–21.

10 Health outcome categories evaluated using the same P, E, and C criteria are combined into a
11 single table (Table 7). Of note, unique considerations are included in the P, E, or C criteria for
12 developmental, hepatobiliary, and reproductive effects; therefore, the refined PECO criteria for
13 these outcomes are presented as separate tables.

14 - For developmental effects (Table 8), the human and animal populations considered are
15   restricted to those exposed during preconception, in utero, or as neonates, juveniles, or
16   adolescents because these are sensitive windows of exposure for developmental effects
17   (NTP, 2011; U.S. EPA, 1991).

18 - The database of animal studies available to support evaluations of hepatobiliary effects is
19   particularly large (Figure 20). Although these effects have been evaluated in many
20   mammalian species exposed to PCBs, information from studies of nonhuman primates and
21   of well-characterized laboratory species (i.e., rats and mice) likely will be sufficient to
22   support hazard conclusions. Therefore, the animal populations considered for this health
23   effect category are restricted to these species (Table 9).

24 - For animal studies of certain reproductive effects related to fertility and fecundity
25   (e.g., mating, conception, pregnancy rate), the exposure criteria require PCB exposure to
26   have been present prior to mating (Table 10); otherwise, observed effects on the number of
27   litters or offspring are more likely to result from effects on offspring development than on
28   fertility of the parental animals (Foster and Gray, 2013; NTP, 2011; U.S. EPA, 1996a).

Table 7. Refined PECO criteria (cardiovascular, dermal, endocrine, gastrointestinal, hematopoietic, immune, metabolic, musculoskeletal, nervous system, ocular, respiratory, and urinary effects)

PECO element	Evidence
Populations	**Human:** Adults and children with exposure to PCBs at any lifestage. The following study designs will be included: controlled-exposure and randomized intervention, cohort, case-control, and cross-sectional. These studies include those conducted in the general population (e.g., NHANES), in cohorts specifically assembled to assess PCB-related health effects, and in occupational settings, where the incidence of disease is compared to a standard or reference population. Case reports, case series, and ecological studies will be tracked as potentially relevant supplemental material. **Animal:** Nonhuman mammalian animal species (whole organism) exposed during any lifestage will be considered (during any period from in utero through adulthood).
Exposures	Exposure to PCB mixtures containing 4[a] or more congeners, including at least one non-dioxin-like PCB. Such "complex" mixtures include commercial PCB mixtures (e.g., Aroclors), mixtures found in the environment (e.g., in contaminated fish or indoor air), and a range of mixtures administered to animals in the laboratory setting. Because all humans are exposed to PCBs as complex mixtures in the environment, every study of PCB exposure in humans is expected to meet this criterion. **Human:** The following exposure assessment methods/exposure contexts will be considered informative: controlled exposure; measured PCB concentration in contact medium (e.g., food, air, dust); biomarkers of exposure (e.g., serum PCB levels); or occupation in a job involving exposure to PCBs (e.g., electrical capacitor manufacturing). The following exposure assessment methods/exposure contexts will not be considered in the absence of biomarker measurements or estimates derived using scientifically sound methods: Yusho/Yu-Cheng patient status; consumption of fish (or marine mammals or other wildlife); or residential proximity to a PCB-contaminated site. **Animal:** Exposure routes to be considered are any oral, inhalation, dermal, or injection exposures; oral and inhalation exposures will be judged the most informative. Studies employing exposures longer than 28 days or short term, developmental exposures will be considered the most informative. Exposure to a single PCB congener or to a mixture containing fewer than 4 congeners will be considered as mechanistic evidence in support of hazard identification and development of toxicity value(s).
Comparators	**Human:** A comparison population exposed to lower levels (or no exposure/exposure below detection levels). For a cohort study, comparisons are made: (1) between levels within a cohort, or (2) between the cohort and an external cohort, presumed to be unexposed or exposed to a lesser degree. Comparisons are made based on "Exposure" definitions above. **Animal:** A concurrent control group exposed to vehicle-only treatment or untreated control.
Outcomes	***Cardiovascular effects*** (Figure 7) **Human:** Assessments of ischemic heart disease (IHD) and IHD mortality, myocardial infarction, hypertension, atherosclerosis, heart failure, and cerebrovascular disease and cerebrovascular disease mortality. Note: Studies that assessed "diseases of the heart" NOS (not otherwise specified) mortality (14 studies) and subjective complaints (4 studies) will be tracked as potentially relevant supplemental material. Consideration of these outcomes was judged lower priority because they are ill-defined and capture a broad range of conditions with potentially unrelated etiologies. **Animal:** Any examination of changes in size, structure, or function of cardiovascular organs or tissues, including the heart and blood vessels. Measures of cardiac enzyme induction and levels of metals in the heart will be considered as supporting evidence for hazard identification or MOA analysis.

PECO element	Evidence
Outcomes (continued)	***Dermal effects*** (Figure 8) Findings of dermatologic changes, including abnormal pigmentation, irritation, erythema, edema, acne/chloracne, fingernail/toenail abnormalities, or alopecia.
	Endocrine effects (Figure 9) Assessments of hypothalamic-pituitary-thyroid and hypothalamic-pituitary-adrenal axis function (including thyroid hormone levels, ultrasound thyroid nodules or thyroid volume, diagnosis of thyroid disease, thyroid weight and histopathology, glucocorticoid and adrenal sex steroid hormone levels, and adrenal weight and histopathology). Note: Measures of thyroid metabolizing enzymes/gene expression will be considered as supporting evidence for hazard identification or MOA analysis. Studies that assessed hormones outside the hypothalamic-pituitary-thyroid and hypothalamic-pituitary-adrenal axes (e.g., insulin-like growth factor, vitamin D, parathyroid, growth hormone) will be tracked as potentially relevant supplemental material because, compared to the outcomes selected for initial consideration, the number of studies evaluating these outcomes is relatively small. However, if sensitivity evaluations (Section 5) identify these outcomes as particularly sensitive to the effects of PCB exposure, they will be prioritized for further analysis. Hypothalamic-pituitary-gonadal axis function was considered as a reproductive effect. Effects on insulin levels and insulin resistance were considered metabolic effects grouped with diabetes.
	Gastrointestinal effects (Figure 10) Evaluations of gastrointestinal histopathology and abdominal ultrasonography. The database also contains 30 studies with evaluations of digestive system complaints and diseases, such as abdominal pain, nausea/vomiting, changes in bowel habits, bloating, gastric ulcer, indigestion, and loss of appetite. However, these studies will be tracked as potentially relevant supplemental material. Consideration of these outcomes was judged lower priority because they are ill-defined and capture a broad range of conditions with potentially unrelated etiologies.
	Hematopoietic effects (Figure 11) Assessments of red blood cells and associated endpoints (e.g., hemoglobin, mean corpuscular volume, hematocrit), bone marrow histopathology, and platelets/clotting function. Note: Studies that assessed blood disease mortality (5 studies) will be tracked as potentially relevant supplemental material; blood disease mortality is ill-defined and captures a broad range of conditions with potentially unrelated etiologies.
	Immune effects (Figure 12) Assessments of impaired immune function, as shown by changes in infectious morbidity, antigen-specific antibody responses, or assays of white blood cell function/proliferation in human studies or by host resistance or functional immune assays in animal studies, allergy and asthma, autoimmunity, and thymus weight and histopathology (i.e., thymic atrophy). Measures of white blood cell phenotype counts and percentages, non-specific Ig levels, cytokine production, serum complement, and thymosin levels will be considered as supporting evidence for hazard identification or MOA analysis. Note: Studies that assessed endotoxin sensitivity, spleen weight and histopathology, lymph node weight and histopathology, and sepsis will be tracked as potentially relevant supplemental material. The number of studies evaluating endotoxin sensitivity and sepsis is relatively small, and effects on spleen and lymph node weight and histopathology were judged to have lower biological significance than the outcomes selected for consideration. However, if sensitivity evaluations (Section 5) identify these outcomes as particularly sensitive to the effects of PCB exposure, they will be prioritized for further analysis.
	Metabolic effects (Figure 13) Assessments of metabolic syndrome and related health outcomes in humans (e.g., blood triglycerides, body mass index, obesity, waist circumference, and diabetes) and animals (e.g., blood triglycerides, adiposity, pancreatic histopathology, and diabetes). Both human and animal studies of diabetes include those evaluating effects on insulin levels, insulin resistance, and blood and urinary glucose levels. Note: Studies that assessed resting metabolic rate, oxygen consumption, and body temperature will be tracked as potentially relevant supplemental material; due to the relatively small number of studies on these outcomes, the available evidence is unlikely sufficient to identify them as hazards of PCB exposure. Further study might be needed to support associations between PCB exposure and these health effects.

PECO element	Evidence
Outcomes (continued)	***Musculoskeletal effects*** (Figure 14) Evaluations of osteoporosis, bone strength and density, bone histopathology, bone development, effects on dentition, skeletal muscle histopathology, and arthritis. Note: Studies that assessed "musculoskeletal complaints and diseases" will be tracked as potentially relevant supplemental material. Consideration of these outcomes was judged lower priority because they are ill-defined and capture a broad range of conditions with potentially unrelated etiologies. Studies that assessed muscle mass and tone will be tracked as potentially relevant supplemental material; due to the relatively small number of studies on these outcomes, the available evidence is unlikely sufficient to identify them as hazards of PCB exposure. Further study might be needed to support associations between PCB exposure and these health effects.
	Nervous system effects (Figure 15) **Human:** Assessments of attention deficit hyperactivity disorder, autism spectrum disorders, and related behaviors (primarily attention, impulse control, and hyperactivity; also executive function, social cognition); cognitive function (includes general intelligence [IQ]; language/verbal skills, learning and memory, school/academic performance, visual-spatial skills, executive function/attention); neonatal neurological and behavioral function; brain aging disorders (e.g., Parkinson's disease, dementia [including Alzheimer's disease], amyotrophic lateral sclerosis); sensory function (auditory function, olfactory function, visual function); motor/cerebellar function, and emotional state (e.g., depression, anxiety symptoms). Deficits in nerve activity (e.g., nerve conduction, electroencephalography) and structural abnormalities will be considered as supporting evidence for hazard identification or MOA analysis. Note: Studies that assessed mortality (caused by neurological disease), "neurological symptoms," and "play behavior" will be tracked as potentially relevant supplemental material. Consideration of these outcomes was judged lower priority because they are ill-defined or capture a broad range of conditions with potentially unrelated etiologies. **Animal:** Assessments of changes in behavior (including motor, cognitive, sensory, attention and motivation, impulse control and hyperactivity) and significant changes in brain structure. Measures of neurochemistry, electrophysiology, neuropathology, and neurodevelopmental processes, including but not limited to apoptosis, dendritic arborization, and neurogenesis in the brain will be considered as supporting evidence for hazard identification or MOA analysis.
	Ocular effects (Figure 16) Findings of ocular changes, including abnormal pigmentation, irritation, erythema, periorbital edema, Meibomian (tarsal) gland enlargement, or ocular discharge.
	Respiratory effects (Figure 17) Evaluations of pulmonary/lung weight and histopathology, pulmonary function, and chest radiography. Note: Studies that assessed "respiratory disease mortality" or "respiratory complaints/illness history" will be tracked as potentially relevant supplemental material. Consideration of these outcomes was judged lower priority because they are ill-defined and capture a broad range of conditions with potentially unrelated etiologies. Measures of respiratory sounds, sputum analysis, blood gas tension, and respiratory rate will be considered as supporting evidence for hazard identification or MOA analysis.
	Urinary effects (Figure 18) Assessments of kidney weight, serum biomarkers of renal function, urinary system histopathology, and kidney diseases or nephropathies (e.g., nephritis, diabetic nephropathy, nephrotic syndrome, gout, renal failure). Note: Measures of urinalysis and urine output will be considered as supporting evidence for hazard identification or MOA analysis.

NHANES = National Health and Nutrition Examination Survey; MOA = mode of action.

[a] As described in Section 2.4, a major goal of this assessment is to develop noncancer toxicity values for PCB mixtures, especially those most relevant for human exposure (Section 2.5.1). Humans tend to be exposed to complex PCB mixtures that contain many congeners of varied toxic potency and MOA. Studies of single congeners and simple (i.e., binary or ternary) mixtures will be tracked as potentially relevant supplemental material, but this assessment will focus primarily on studies of mixtures that better reflect a typical human exposure scenario.

Figure 7. Number of human and animal studies of PCB exposure and cardiovascular effects. Labels indicate the total number of studies for each health effect. Database contains 57 human studies and 45 animal studies; some studies evaluated more than one type of cardiovascular effect.
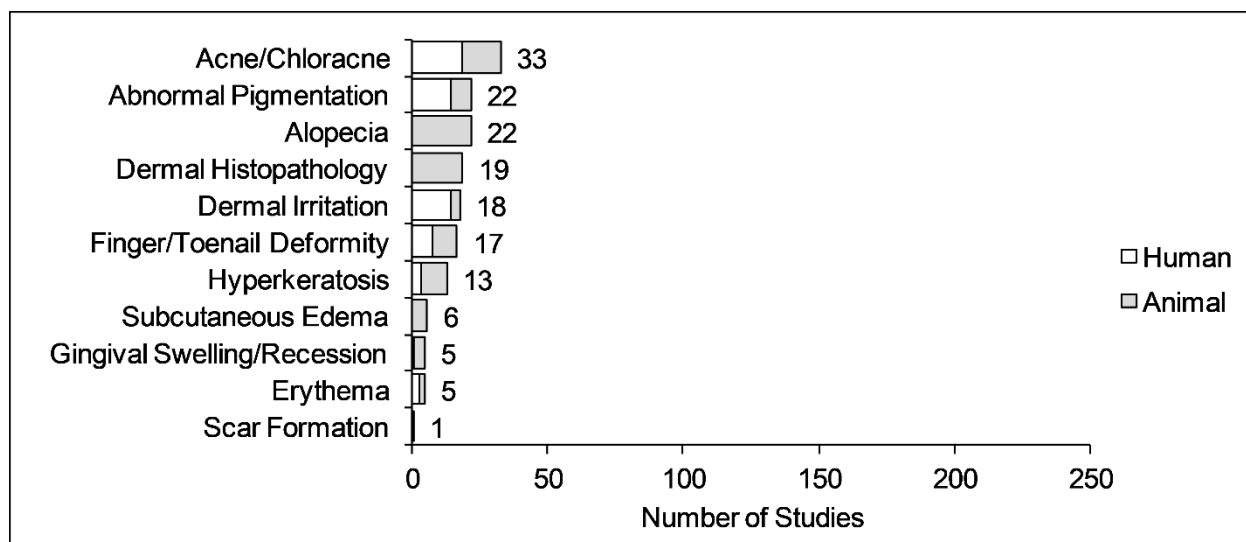
Figure 8. Number of human and animal studies of PCB exposure and dermal effects. Labels indicate the total number of studies for each health effect. Database contains 28 human studies and 40 animal studies; some studies evaluated more than one type of dermal effect.

Figure 9. Number of human and animal studies of PCB exposure and endocrine effects. Labels indicate the total number of studies for each health effect. Database contains 92 human studies and 155 animal studies; some studies evaluated more than one type of endocrine effect.

Figure 10. Number of human and animal studies of PCB exposure and gastrointestinal effects. Labels indicate the total number of studies for each health effect. Database contains 19 human studies and 37 animal studies; some studies evaluated more than one type of gastrointestinal effect.

Figure 11. Number of human and animal studies of PCB exposure and hematopoietic effects. Labels indicate the total number of studies for each health effect. Database contains 18 human studies and 41 animal studies; some studies evaluated more than one type of hematopoietic effect.

Figure 12. Number of human and animal studies of PCB exposure and immune effects. Labels indicate the total number of studies for each health effect. Database contains 101 human studies and 131 animal studies; some studies evaluated more than one type of immune effect.

Figure 13. Number of human and animal studies of PCB exposure and metabolic effects. Labels indicate the total number of studies for each health effect. Database contains 148 human studies and 81 animal studies; some studies evaluated more than one type of metabolic effect.

Figure 14. Number of human and animal studies of PCB exposure and musculoskeletal effects. Labels indicate the total number of studies for each health effect. Database contains 30 human studies and 24 animal studies; some studies evaluated more than one type of musculoskeletal effect.

Figure 15. Number of human and animal studies of PCB exposure and nervous system effects. Labels indicate the total number of studies for each health effect. Database contains 156 human studies and 143 animal studies; some studies evaluated more than one type of nervous system effect.

Figure 16. Number of human and animal studies of PCB exposure and ocular effects. Labels indicate the total number of studies for each health effect. Database contains 16 human studies and 27 animal studies; some studies evaluated more than one type of ocular effect.

Figure 17. Number of human and animal studies of PCB exposure and respiratory effects. Labels indicate the total number of studies for each health effect. Database contains 28 human studies and 42 animal studies; some studies evaluated more than one type of respiratory effect.
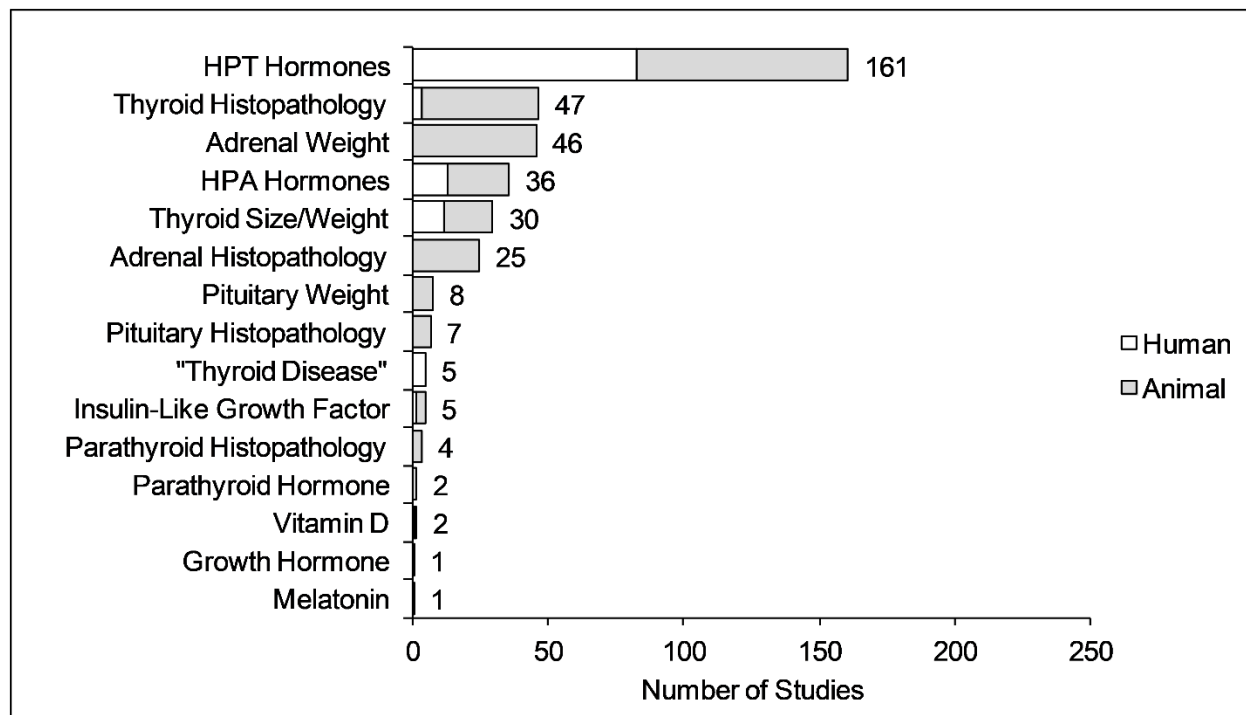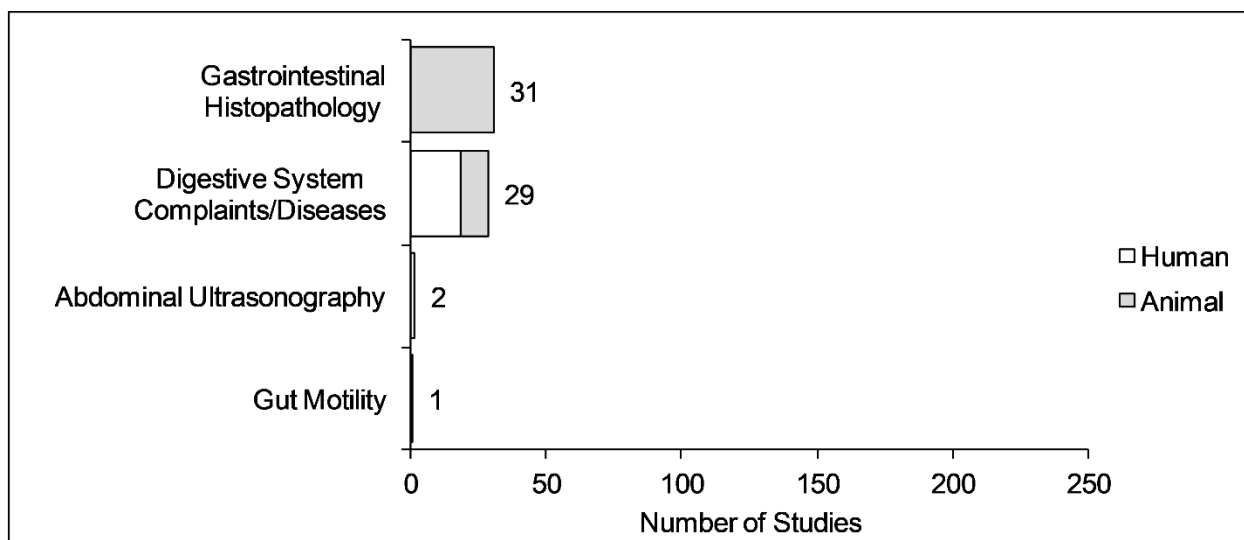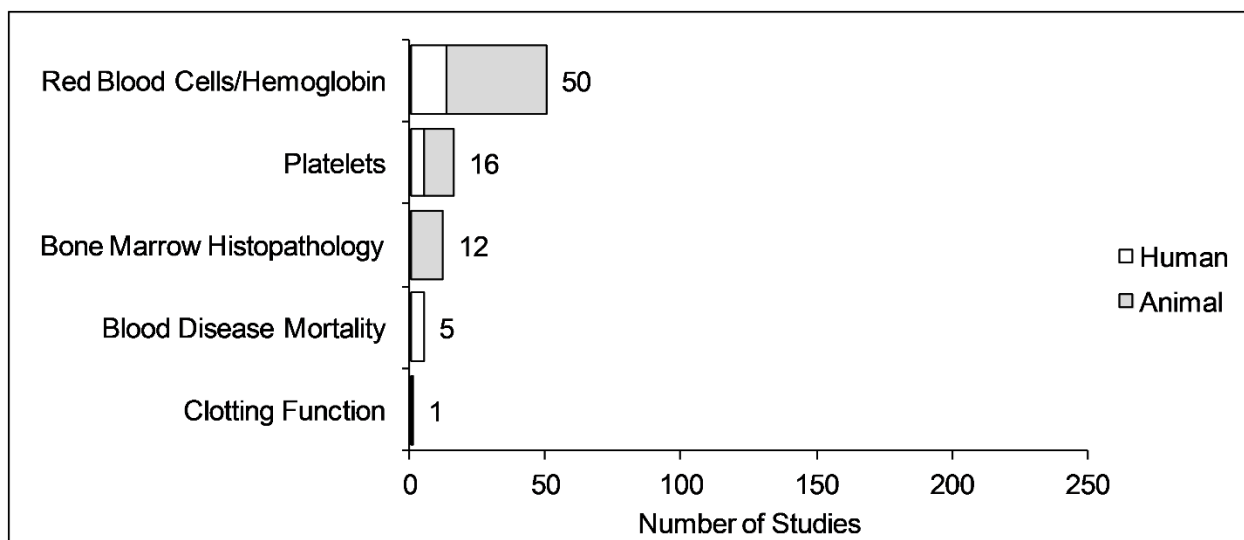
Figure 18. Number of human and animal studies of PCB exposure and urinary system effects. Labels indicate the total number of studies for each health effect. Database contains 24 human studies and 92 animal studies; some studies evaluated more than one type of urinary system effect.

Table 8. Refined PECO criteria (developmental effects)

PECO element	Evidence
Populations	**Human:** Humans with exposure to PCBs during preconception, in utero, infancy, childhood, and adolescence. The following study designs will be included: controlled-exposure and randomized intervention, cohort, case-control, and cross-sectional. These studies include those conducted in the general population (e.g., NHANES), in cohorts specifically assembled to assess PCB-related health effects, and in occupational settings, where the incidence of disease is compared to a standard or reference population. Case reports, case series, and ecological studies will be tracked as potentially relevant supplemental material. **Animal:** Nonhuman mammalian animal species (whole organism) exposed during preconception, in utero, or as neonates, juveniles, or adolescents.

PECO element	Evidence
Exposures	Exposure to PCB mixtures containing 4 or more congeners, including at least one non-dioxin-like PCB. Such "complex" mixtures include commercial PCB mixtures (e.g., Aroclors), mixtures found in the environment (e.g., in contaminated fish or indoor air), and a range of mixtures administered to animals in the laboratory setting. Because all humans are exposed to PCBs as complex mixtures in the environment, every study of PCB exposure in humans is expected to meet this criterion.

Human: The following exposure assessment methods/exposure contexts will be considered informative: controlled exposure; measured PCB concentration in contact medium (e.g., food, air, dust); biomarkers of exposure (e.g., serum PCB levels); or occupation in a job involving exposure to PCBs (e.g., electrical capacitor manufacturing).

The following exposure assessment methods/exposure contexts will not be considered in the absence of biomarker measurements or estimates derived using scientifically sound methods: Yusho/Yu-Cheng patient status; consumption of fish (or marine mammals or other wildlife); or residential proximity to a PCB-contaminated site.

Animal: Exposure routes to be considered are any oral, inhalation, dermal, or injection exposures; oral and inhalation exposures will be considered the most informative. Studies employing exposures longer than 28 days or short-term, developmental exposures will be considered the most informative. Exposure to a single PCB congener or to a mixture containing fewer than 4 congeners will be considered as mechanistic evidence in support of hazard identification and development of toxicity values. |
| **Comparators** | **Human:** A comparison population exposed to lower levels (or no exposure/exposure below detection levels). For a cohort study, comparisons are made: (1) between levels within a cohort, or (2) between the cohort and an external cohort, presumed to be unexposed or exposed to a lesser degree. Comparisons are made based on "Exposure" definitions above.

Animal: A concurrent control group exposed to vehicle-only treatment or untreated control. |
| **Outcomes** | ***Developmental effects*** (Figure 19)
Assessments of birth weight/small for gestational age (related anthropometrics measured in utero (e.g., ultrasound) or at birth (e.g., head circumference) also will be considered); physical growth (height, weight, BMI, overweight, obesity) at various stages of infancy, childhood, or adolescence; birth defects; and pregnancy loss/offspring mortality. Note: Studies that assessed sex ratio, anogenital distance, developmental milestones, and placental weight and histopathology will be tracked as potentially relevant supplemental material. However, if sensitivity evaluations (Section 5) identify these outcomes as particularly sensitive to the effects of PCB exposure, they will be prioritized for further analysis. Due to the relatively small number of studies on Apgar scores, the available evidence is unlikely sufficient to identify an associated hazard of PCB exposure. |

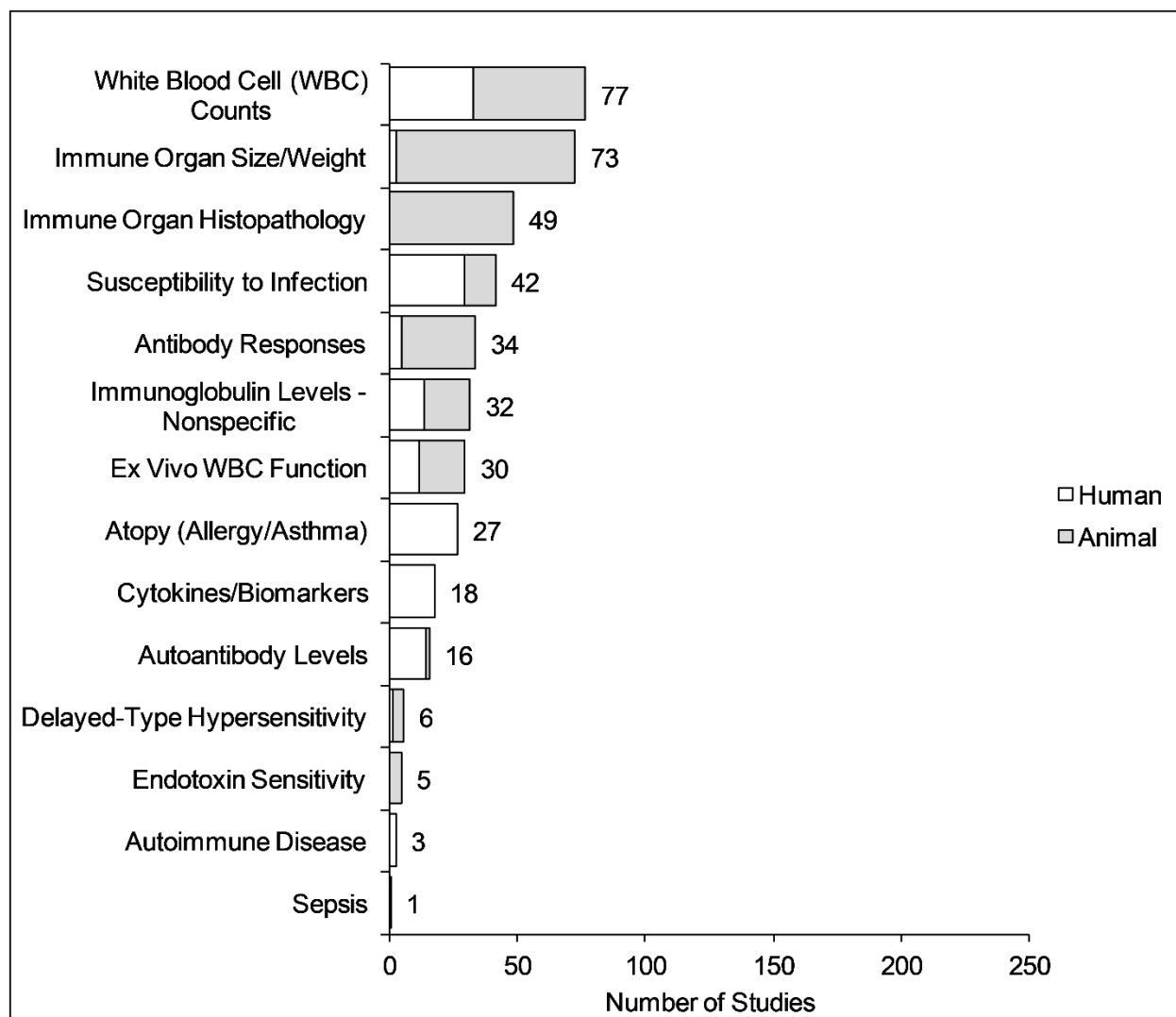NHANES = National Health and Nutrition Examination Survey.
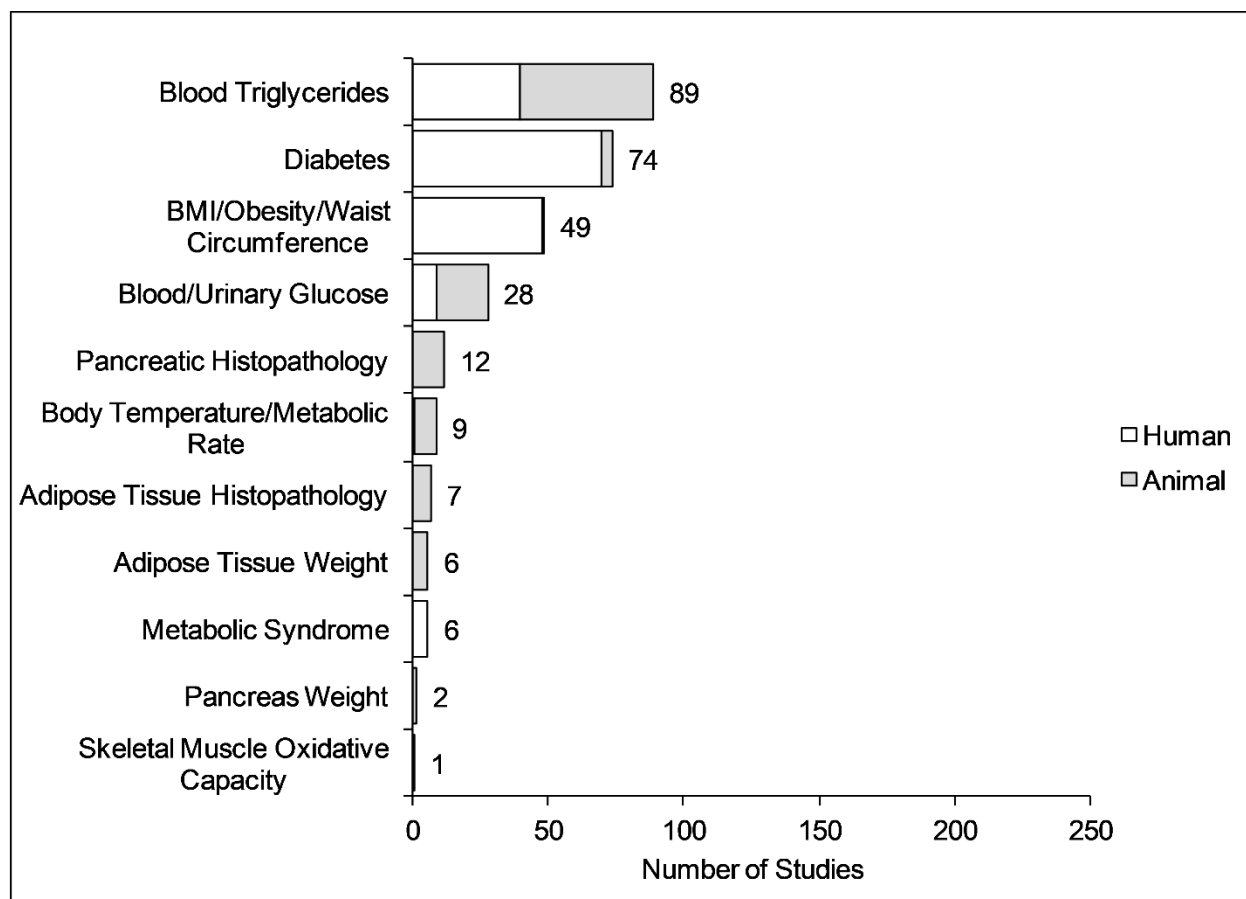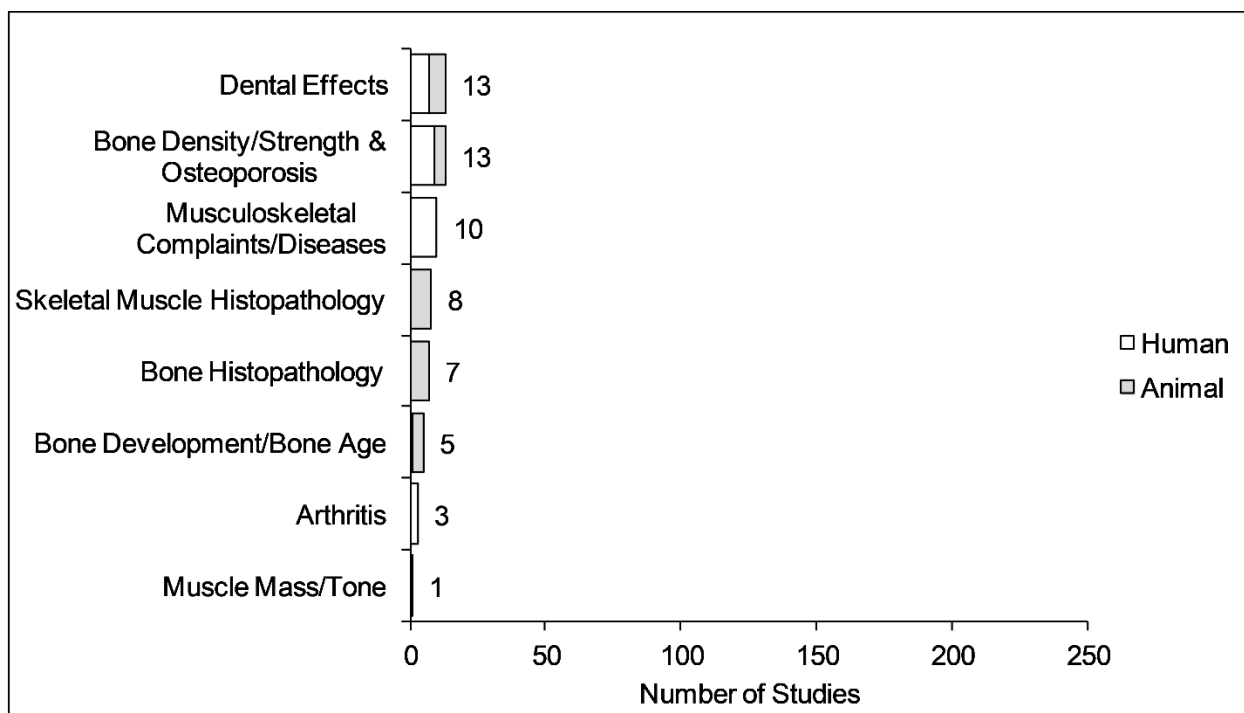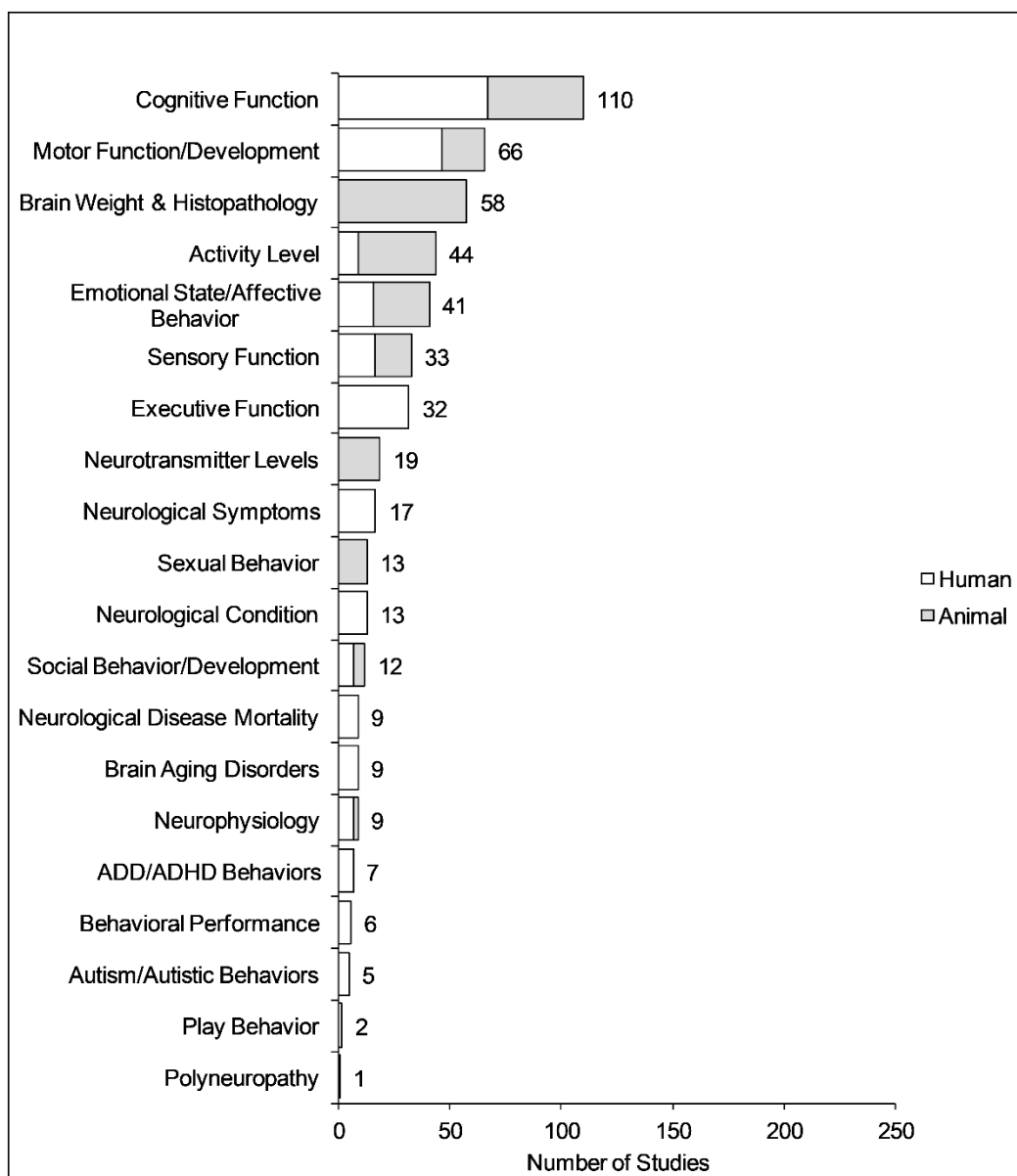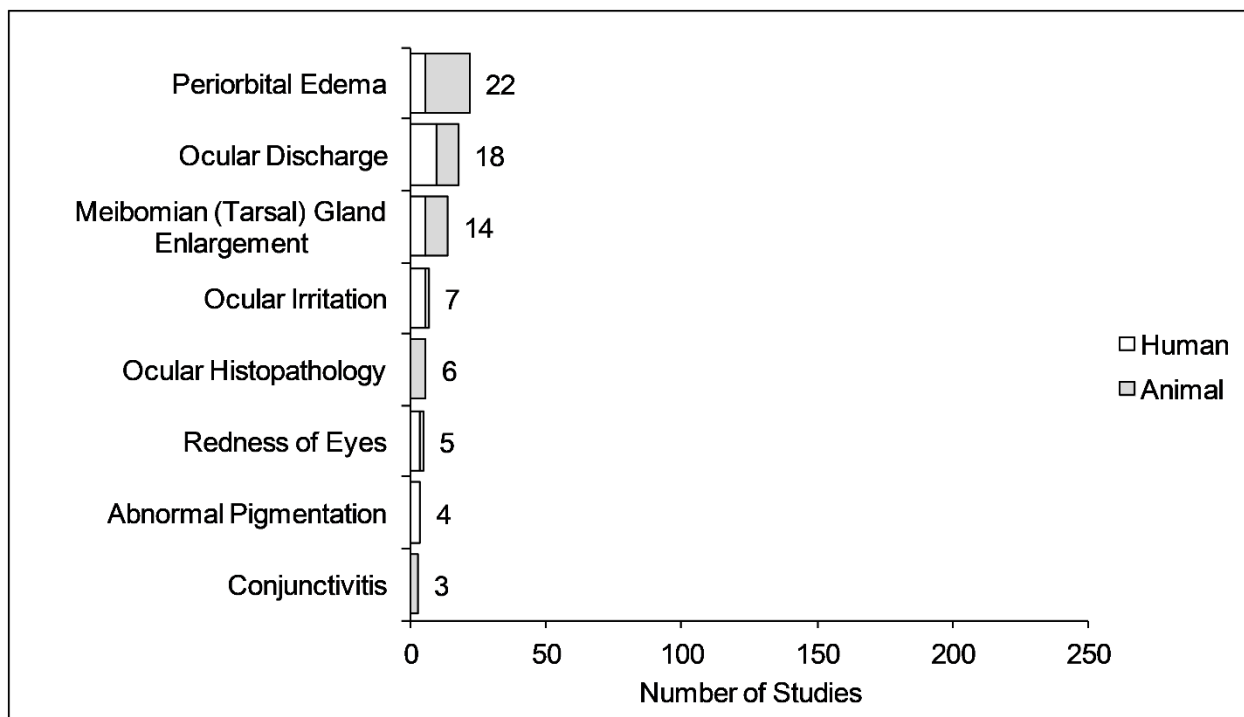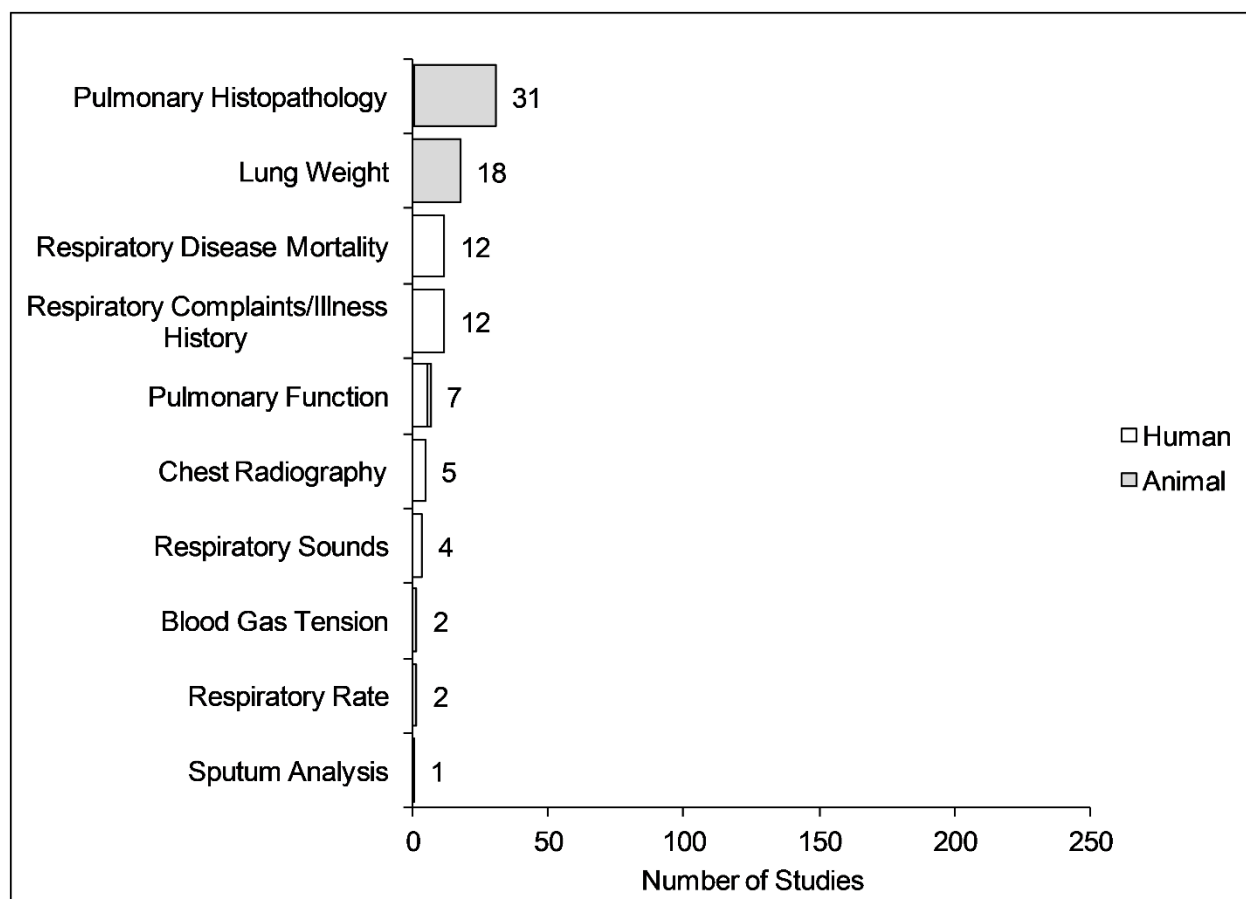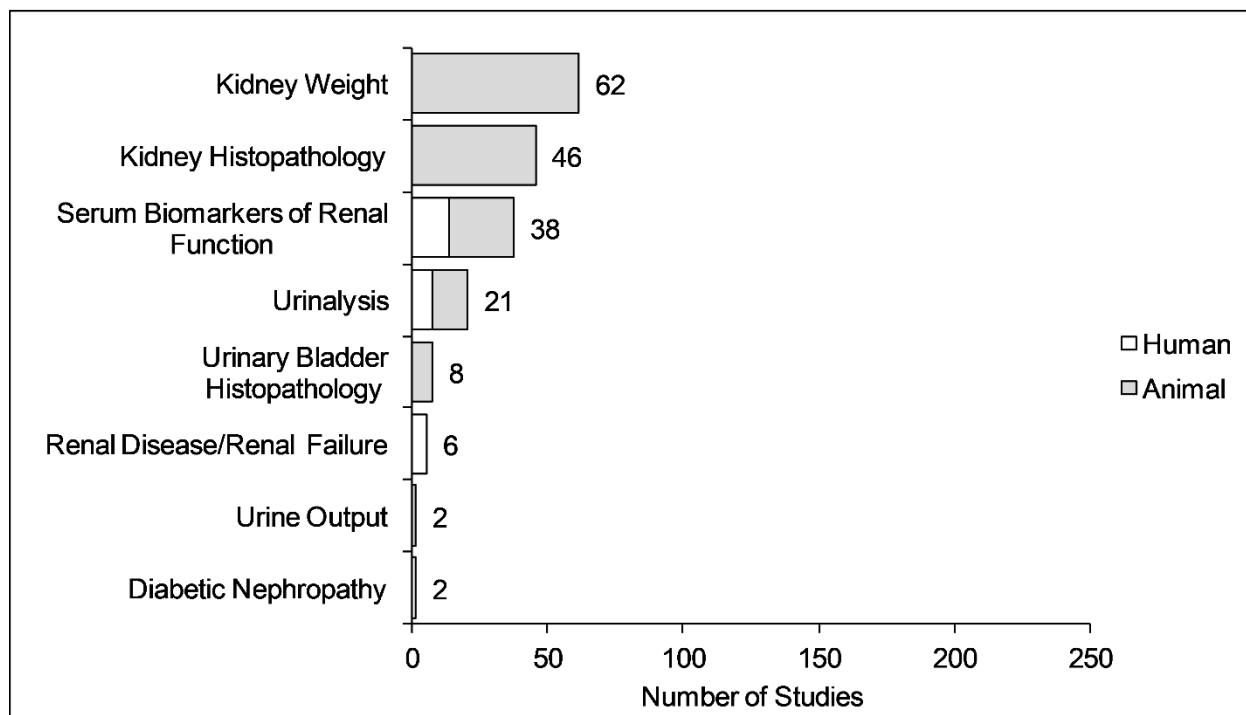
Figure 19. Number of human and animal studies of PCB exposure and developmental effects. Labels indicate the total number of studies for each health effect. Database contains 107 human studies and 161 animal studies; some studies evaluated more than one type of developmental effect.
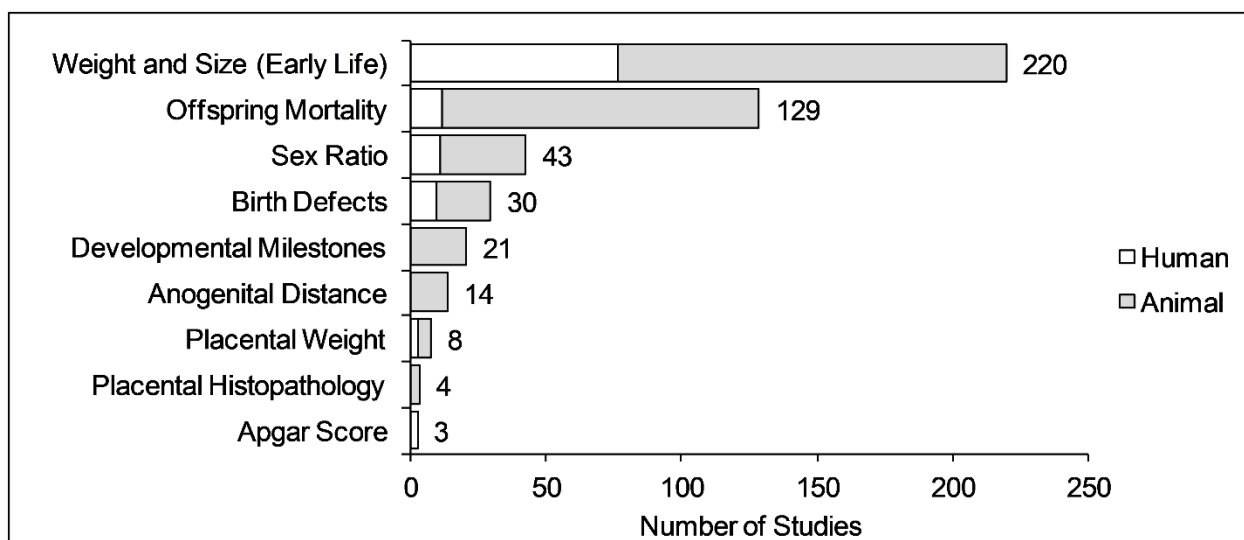
Table 9. Refined PECO criteria (hepatobiliary effects)

PECO element	Evidence
Populations	**Human:** Adults and children with exposure to PCBs at any lifestage. The following study designs will be included: controlled-exposure and randomized intervention, cohort, case-control, and cross-sectional. These studies include those conducted in the general population (e.g., NHANES), in cohorts specifically assembled to assess PCB-related health effects, and in occupational settings, where the incidence of disease is compared to a standard or reference population. Case reports, case series, and ecological studies will be tracked as potentially relevant supplemental material. **Animal:** Rats, mice, and nonhuman primates (whole organism) exposed during any lifestage will be considered (anywhere during the period from in utero through adulthood). Hepatobiliary effects have been evaluated in other nonhuman mammalian species exposed to PCBs but given the large number of studies in this health effect category, studies of nonhuman primates and of well-characterized laboratory species (i.e., rats and mice) have been prioritized.

PECO element	Evidence
Exposures	Exposure to PCB mixtures containing 4 or more congeners, including at least one non-dioxin-like PCB. Such "complex" mixtures include commercial PCB mixtures (e.g., Aroclors), mixtures found in the environment (e.g., in contaminated fish or indoor air), and a range of mixtures administered to animals in the laboratory setting. Because all humans are exposed to PCBs as complex mixtures in the environment, every study of PCB exposure in humans expected to meet this criterion. **Human:** The following exposure assessment methods/exposure contexts will be considered informative: controlled exposure; measured PCB concentration in contact medium (e.g., food, air, dust); biomarkers of exposure (e.g., serum PCB levels); or occupation in a job involving exposure to PCBs (e.g., electrical capacitor manufacturing). The following exposure assessment methods/exposure contexts will not be considered in the absence of biomarker measurements or estimates derived using scientifically sound methods: Yusho/Yu-Cheng patient status; consumption of fish (or marine mammals or other wildlife); or residential proximity to a PCB-contaminated site. **Animal:** Exposure routes to be considered are any oral, inhalation, dermal, or injection exposures; oral and inhalation exposures will be considered the most informative. Studies employing exposures longer than 28 days or short-term, developmental exposures will be considered the most informative. Exposure to a single PCB congener or to a mixture containing fewer than 4 congeners will be considered as mechanistic evidence in support of hazard identification and development of toxicity value(s).
Comparators	**Human:** A comparison population exposed to lower levels (or no exposure/exposure below detection levels). For a cohort study, comparisons are made: (1) between levels within a cohort, or (2) between the cohort and an external cohort, presumed to be unexposed or exposed to a lesser degree. Comparisons are made based on "Exposure" definitions above. **Animal:** A concurrent control group exposed to vehicle-only treatment or untreated control.
Outcomes	*Hepatobiliary effects* (Figure 20) Assessments of blood cholesterol levels, serum biomarkers of liver health and function, liver weight and hepatomegaly, and liver disease, including cirrhosis and steatosis in humans and studies evaluating liver histopathology or liver lipids in animals. Note: Studies that assessed bile acid content/excretion, metabolic enzyme induction, liver cell proliferation, porphyrins/porphyria, gall bladder histopathology, and hepatic levels of vitamins and micronutrients will be tracked as potentially relevant supplemental material. The number of studies evaluating bile acid content/excretion, liver cell proliferation, porphyrins/porphyria, and gall bladder histopathology is relatively small, and effects on metabolic enzyme induction and hepatic levels of vitamins and micronutrients were judged to have lower biological significance than the outcomes selected for consideration. However, if sensitivity evaluations (Section 5) identify these outcomes as particularly sensitive to the effects of PCB exposure, they will be prioritized for further analysis.

NHANES = National Health and Nutrition Examination Survey.

Figure 20. Number of human and animal studies of PCB exposure and hepatobiliary effects. Labels indicate the total number of studies for each health effect. Database contains 64 human studies and 313 animal studies; some studies evaluated more than one type of hepatobiliary effect.
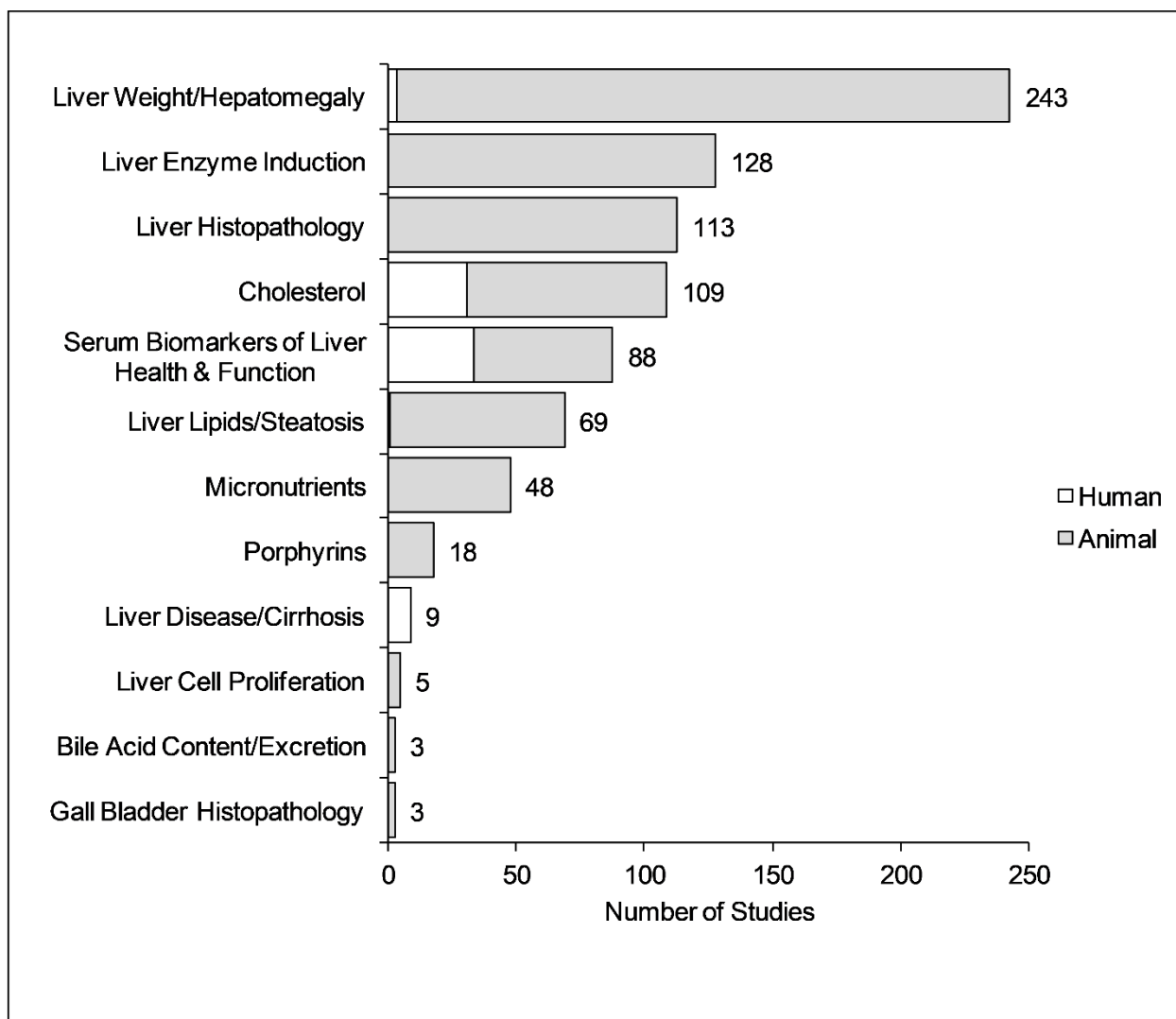
Table 10. Refined PECO criteria (reproductive effects)

PECO element	Evidence
<u>Populations</u>	**Human:** Adults and children with exposure to PCBs at any lifestage. The following study designs will be included: controlled-exposure and randomized intervention, cohort, case-control, and cross-sectional. These studies include those conducted in the general population (e.g., NHANES), in cohorts specifically assembled to assess PCB-related health effects, and in occupational settings, where the incidence of disease is compared to a standard or reference population. Case reports, case series, and ecological studies will be tracked as potentially relevant supplemental material. **Animal:** Nonhuman mammalian animal species (whole organism) exposed during any lifestage will be considered (anywhere during the period from in utero through adulthood).
<u>Exposures</u>	Exposure to PCB mixtures containing 4 or more congeners, including at least one non-dioxin-like PCB. Such "complex" mixtures include commercial PCB mixtures (e.g., Aroclors), mixtures found in the environment (e.g., in contaminated fish or indoor air), and a range of mixtures administered to animals in the laboratory setting. Because all humans are exposed to PCBs as complex mixtures in the environment, every study of PCB exposure in humans is expected to meet this criterion. **Human:** The following exposure assessment methods/exposure contexts will be considered informative: controlled exposure; measured PCB concentration in contact medium (e.g., food, air, dust); biomarkers of exposure (e.g., serum PCB levels); or occupation in a job involving exposure to PCBs (e.g., electrical capacitor manufacturing). The following exposure assessment methods/exposure contexts will not be considered in the absence of biomarker measurements or estimates derived using scientifically sound methods: Yusho/Yu-Cheng patient status; consumption of fish (or marine mammals or other wildlife); or residential proximity to a PCB-contaminated site. **Animal:** Exposure routes to be considered are any oral, inhalation, dermal, or injection exposures; oral and inhalation exposures will be considered the most informative. Studies employing exposures longer than 28 days or short term, developmental exposures will be considered the most informative. Exposure to a single PCB congener or to a mixture containing fewer than 4 congeners will be considered as mechanistic evidence in support of hazard identification and development of toxicity values. For evaluation of fertility (e.g., mating and pregnancy rate) in parental females, studies will be considered if the PCB exposure began prior to mating.
<u>Comparators</u>	**Human:** A comparison population exposed to lower levels (or no exposure/exposure below detection levels). For a cohort study, comparisons are made: (1) between levels within a cohort, or (2) between the cohort and an external cohort, presumed to be unexposed or exposed to a lesser degree. Comparisons are made based on "Exposure" definitions above. **Animal:** A concurrent control group exposed to vehicle-only treatment or untreated control.

PECO element	Evidence
<u>Outcomes</u>	*Reproductive effects* (Figure 21) **<u>Human:</u>** Assessments of time-to-pregnancy, gestation duration/preterm birth, endometriosis, and semen quality (including sperm count, morphology, and motility). Note: Studies that assessed male and female sexual development; menstrual cycle characteristics, including age at menarche/menopause; and maternal body weight gain during pregnancy will be tracked as potentially relevant supplemental material. However, if sensitivity evaluations (Section 5) identify these outcomes as particularly sensitive to the effects of PCB exposure, they will be prioritized for further analysis. Measures of reproductive hormone levels and function will be considered as supporting evidence for hazard identification or MOA analysis. **<u>Animal:</u>** Assessments of endpoints measuring characteristics needed for successful reproduction, as defined by U.S. EPA (1996. *Guidelines for Reproductive Toxicity Risk Assessment* (EPA/630/R-96/009) (U.S. EPA, 1996a). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum.https://www.epa.gov/sites/production/files/2014-11/documents/guidelines_repro_toxicity.pdf). Such endpoints include, but are not limited to, mating and pregnancy rate, interval between offspring/litters, gestation duration, sperm count, sperm morphology, sperm motility, and histopathology of reproductive organs (female and male). Measures of reproductive hormone levels and function and reproductive organ weights will be considered as supporting evidence for hazard identification or MOA analysis. Measures of maternal body weight gain will be tracked as potentially relevant supplemental material and could be considered as supporting evidence to inform interpretation of primary outcomes.

NHANES = National Health and Nutrition Examination Survey; MOA = mode of action.
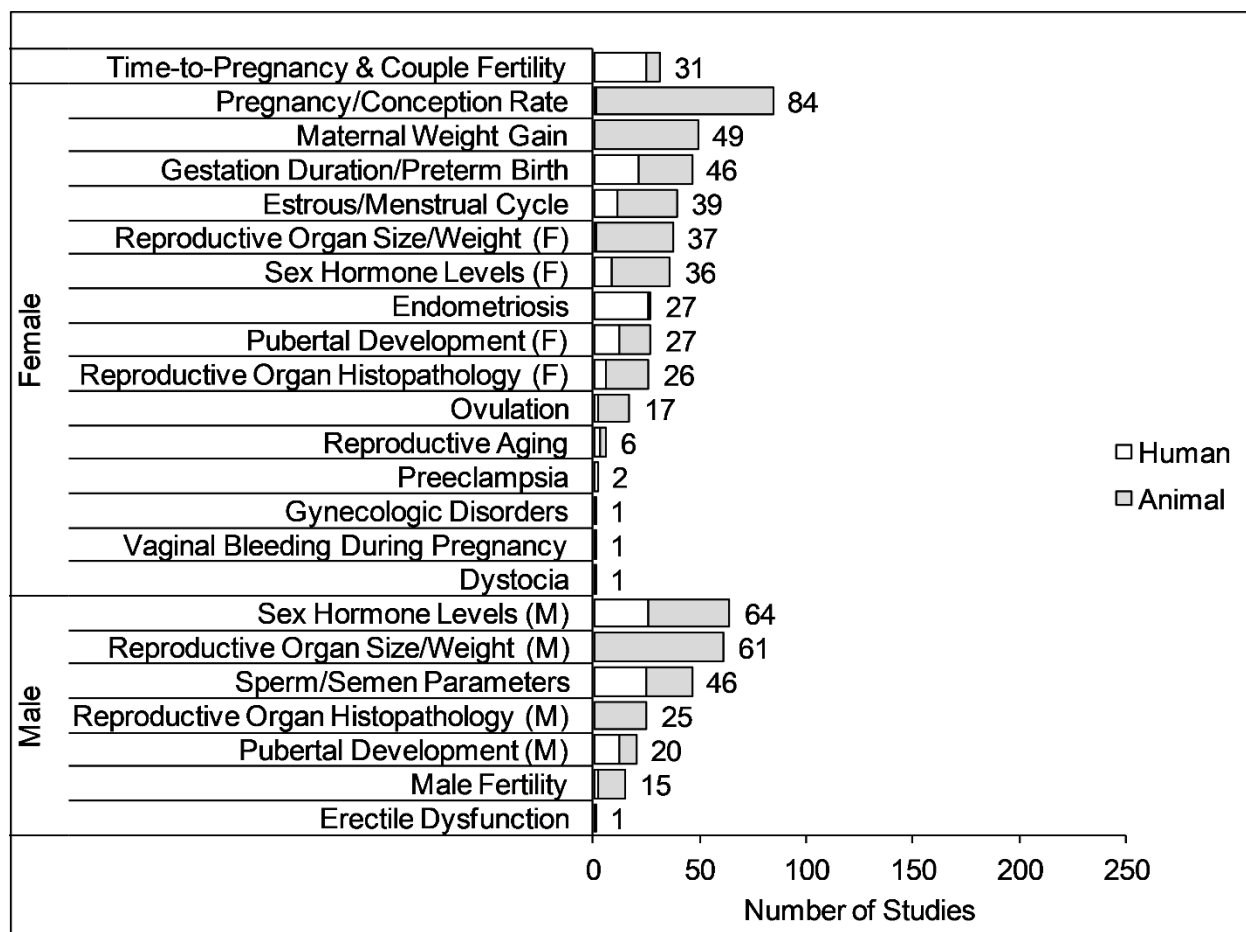
Figure 21. Number of human and animal studies of PCB exposure and reproductive effects. Labels indicate the total number of studies for each health effect. Database contains 153 human studies and 205 animal studies; some studies evaluated more than one type of reproductive effect.

1   The next step in the refinement process will be to evaluate the potential sensitivity to PCB

2   exposure of each health outcome based on dose-response data from animal toxicology studies;

3   these sensitivity evaluations will consider differences in effect levels observed across PCB mixtures.

4   Health outcomes in each category then will be grouped and sorted into priority tiers based on

5   factors including potential sensitivity to PCB exposure, size of the evidence base (as a preliminary

6   measure of the potential strength of evidence), and relative biological significance. Outcomes

7   observed exclusively or primarily in humans will be assigned to tiers based on size of the evidence

8   base, relative biological significance, and expected etiological similarity to sensitive outcomes

9   observed in animal toxicology studies. Details of the sensitivity evaluation and resulting

10   refinements to the assessment approach will be documented in the updated protocol released with

11   the draft assessment.

12   In addition to prioritizing specific health outcomes within each health effect category, entire

13   health effect categories might also be prioritized based on relative sensitivity and potential strength

1   of evidence. As shown in Figure 22, the following health effect categories contain over 100 studies

2   evaluating potential effects of PCB exposure on outcomes: hepatobiliary, reproductive, nervous

3   system, developmental, endocrine, immune, metabolic, urinary system, and cardiovascular. Top-

4   tier outcomes from each of these health effect categories will be subject to full systematic review,

5   including additional, targeted literature searches (as described in Section 4.3) and steps described

6   below, beginning with study evaluation (Section 6). For the remaining health effect categories with

7   relatively small databases of supporting evidence (i.e., respiratory, dermal, hematopoietic,

8   gastrointestinal, musculoskeletal, ocular effects), the results of epidemiology and animal toxicology

9   studies will be tracked and recorded in inventories, but these outcomes likely will not be subject to

10   further analysis unless identified as particularly sensitive to the effects of PCB exposure.

11   If discernable differences in potential sensitivity are observed across entire health effect

12   categories, this assessment might use a modular approach to evaluating health effects. The first

13   module would focus on the most sensitive health effect categories, which would form the basis for

14   reference values to assess overall health risk resulting from PCB exposure. While developing the

15   first module, EPA would also evaluate the potential utility of conducting hazard assessments and

16   deriving reference values for less sensitive health effect categories (e.g., to support cumulative risk

17   assessments of exposures to PCBs in combination with other agents or stressors). If identified as a

18   priority need, a second module would be developed and released separately from the first module.
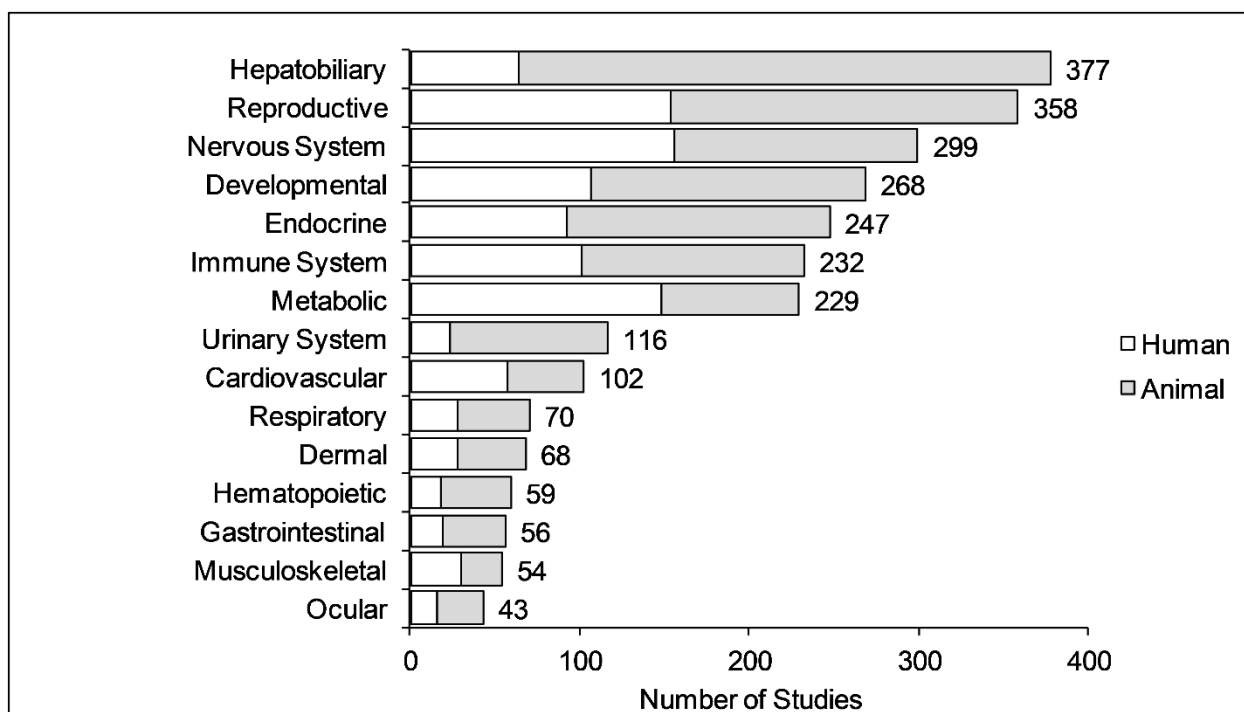
Figure 22. Number of studies evaluating health outcomes in each health effect category based on the results of the literature search illustrated in Figure 6.

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

1   The general approach for evaluating PECO-relevant primary health effect studies of all
2 study types is described in Section 6.1. However, the specifics of applying the approach differ; thus,
3 they are described separately for epidemiology and animal toxicology studies in Sections 6.2 and
4 6.3, respectively. Different approaches will be used for evaluation of PBPK models (see Section 6.4)
5 and mechanistic studies (see Sections 6.5 and 9.2).

6.1. STUDY EVALUATION OVERVIEW FOR HEALTH EFFECT STUDIES

6   Key concerns for the review of epidemiology and animal toxicology studies are risk of bias,
7 which is the assessment of internal validity (factors that affect the magnitude or direction of an
8 effect in either direction) and insensitivity (factors that limit the ability of a study to detect a true
9 effect; low sensitivity is a bias toward the null when an effect exists). Reporting quality is evaluated
10 to determine the extent the available information allows for evaluating these concerns. The study
11 evaluations are aimed at discerning the severity of any identified limitations (focusing on
12 limitations that could substantively change a result), considering also the expected direction of the
13 bias. The study evaluation considerations described below can be refined to address a range of
14 study designs, health effects, and chemicals. The general approach for reaching an overall judgment
15 for the study (or a specific analysis in a study) regarding confidence in the reliability of the results
16 is illustrated in Figure 23.
17   At least two reviewers will independently evaluate the studies to identify characteristics
18 that bear on the validity and sensitivity of the results and provide additional chemical- or
19 outcome-specific knowledge or methodological concerns.
20   Considerations for evaluating studies will be developed in consultation with topic-specific
21 technical experts. Existing guidance documents are used when available, including EPA guidance
22 for carcinogenicity, neurotoxicity, reproductive toxicity, and developmental toxicity ([U.S. EPA,](#)
23 [2005](#), [1998](#), [1996a](#), [1991](#)). The independent evaluations include a pilot phase to assess and refine
24 the evaluation process. During this phase, decisions will be compared and a consensus reached
25 between reviewers, and when necessary, differences will be resolved by discussion among the
26 reviewers, the chemical assessment team, or technical experts. As reviewers examine a group of
27 studies, additional chemical-specific knowledge or methodological concerns could emerge, and a
28 second pass might become necessary. Refinements to the study evaluation process made during
29 the pilot phase and subsequent implementation will be acknowledged as updates to the protocol.

(a) Study evaluation process

Refined evaluation plan

↓

Criteria development

↓

Pilot testing/refine criteria

↓

Evaluation by two reviewers

↓

Conflict resolution

↓

Final domain judgments and overall study rating

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

1   For studies that examine more than one outcome, the evaluation process will be performed

2   separately for each outcome because the utility of a study can vary for different outcomes. If a

3   study examines multiple endpoints for the same outcome,[11] evaluations might be performed at a

4   more granular level if appropriate, but these measures could still be grouped for evidence

5   synthesis.

---

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

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

9   For each outcome in a study,[12] reviewers will reach a consensus judgment of *good*,
10 *adequate*, *deficient*, *not reported*, or *critically deficient* for each evaluation domain. If a consensus is
11 not reached, a third reviewer will perform conflict resolution. It is important to stress that these
12 evaluations are performed in the context of the study's utility for identification of individual
13 hazards. Although limitations specific to the usability of the study for dose-response analysis are
14 useful for later decisions, they do not contribute to the study confidence classifications. These
15 categories are applied to each evaluation domain for each study as follows:

16 • *Good* represents a judgment that the study was conducted appropriately relative to the
17   evaluation domain, and any deficiencies, if present, are minor and would not be expected to
18   influence the study results.

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

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

25 • *Not reported* indicates the information necessary to evaluate the domain question was not
26   available in the study. Generally, this term carries the same functional interpretation as
27   *deficient* for the purposes of the study confidence classification (described below).
28   Depending on the number and severity of other limitations identified in the study,
29   contacting the study authors to obtain this information might or might not be useful (see
30   discussion above).

• *Critically deficient* reflects a judgment that the study conduct introduced a serious flaw that
  makes the study uninterpretable. Studies with a determination of critically deficient in an
  evaluation domain almost always will be considered overall *uninformative*.

---

[12] "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, sensitivity, etc. However, all
evaluations discussed herein are explicitly conducted at the level of an individual outcome within an
(un)exposed group of animals or humans, or on a sample of the population within a study.

1 Once the evaluation domains have been rated, the identified strengths and limitations will

2 be considered to reach a study confidence classification of *high*, *medium*, or *low* confidence, or

3 *uninformative* for each specific health outcome. This classification is based on the reviewer

4 judgments across the evaluation domains and includes consideration of the likely impact the noted

5 deficiencies in bias and sensitivity or inadequate reporting have on the results. The classifications,

6 which reflect a consensus judgment between reviewers, are defined as follows:

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

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

15   • *Low* confidence: A study where deficiencies or concerns are noted, and the potential for bias
16   or inadequate sensitivity could have a significant impact on the study results or their
17   interpretation. Typically, *low* confidence studies have a *deficient* evaluation for one or more
18   domains, although some *medium* confidence studies might have a *deficient* rating in
19   domain(s) considered to have less influence on the magnitude or direction of effect
20   estimates. Generally, *low* confidence results are given less weight compared to *high* or
21   *medium* confidence results during evidence synthesis and integration (see Section 10.1,
22   Tables 14 and 15) and are generally not used as the primary sources of information for
23   hazard identification or derivation of toxicity values unless they are the only studies
24   available. Studies rated as *low* confidence only because of sensitivity concerns about bias
25   toward the null would require additional consideration during evidence synthesis.
26   Observing an effect in these studies could increase confidence, assuming the study is
27   otherwise well conducted (see Section 9).

28   • *Uninformative*: A study where serious flaw(s) make the study results unusable for informing
29   hazard identification. Studies with *critically deficient* judgments in any evaluation domain
30   are almost always classified as *uninformative* (see explanation above). Studies with
31   multiple *deficient* judgments across domains also might be considered *uninformative*.
32   *Uninformative* studies will not be considered further in the synthesis and integration of
33   evidence for hazard identification or dose-response but might be used to highlight possible
34   research gaps.

35 Study evaluation determinations reached by each reviewer and the consensus judgment

36 between reviewers will be documented in EPA's version of Health Assessment Workspace

37 Collaborative (HAWC), a free and open-source web-based software application.[13] Final study

38 evaluations housed in HAWC will be made available when the draft is publicly released. The study

---

[13]HAWC: A modular web-based interface to facilitate development of human health assessments of chemicals (https://hawcproject.org/portal/).

1   confidence classifications and their rationales will be carried forward and considered as part of
2   evidence synthesis (see Section 9), to aid in interpreting results across studies.

6.2.   EPIDEMIOLOGY STUDY EVALUATION

3   Evaluation of epidemiology studies of health effects to assess risk of bias and study
4   sensitivity will be conducted for the following domains: exposure measurement, outcome
5   ascertainment, participant selection, potential confounding, analysis, study sensitivity, and selective
6   reporting. Bias can result in false positives and false negatives, while study sensitivity is typically
7   concerned with identifying the latter.
8   The principles and framework used for evaluating epidemiology studies are adapted from
9   the principles in the *Cochrane Risk of Bias in Nonrandomized Studies of Interventions* [ROBINS-I;
10 (Sterne et al., 2016)], modified to address environmental and occupational exposures. The
11 underlying philosophy of ROBINS-I is to describe attributes of an "ideal" study relative to each
12 evaluation domain (e.g., exposure measurement, outcome classification). Core and prompting
13 questions are used to collect information to guide evaluation of each domain.
14 Core and prompting questions are presented in Table 11 with additional considerations
15 that apply to most outcomes for each domain. Core questions represent key concepts while the
16 prompting questions help the reviewer focus on relevant details under each key domain. Exposure-
17 and outcome-specific criteria to use during evaluation of studies will be developed using the core
18 and prompting questions and refined during a pilot phase with engagement from topic-specific
19 experts. The types of information that could be the focus of those criteria are listed in Table 12.
20 Several considerations will be applied when evaluating the exposure domain to assess the
21 exposure measurement and assessment methods for PCBs used in epidemiology studies. In
22 epidemiology studies of PCBs, exposure is commonly characterized using current measures of PCB
23 congeners or their metabolites in biological matrices, including blood serum/plasma, breast
24 milk/colostrum, and adipose tissue. These could be expressed on a whole-tissue basis
25 (e.g., ng of PCB/mL of serum) or might be lipid standardized (i.e., ng of PCB/g of lipid). These
26 studies often rely on a limited number of measured congeners or metabolites. For biomarkers of
27 exposure, exposure assessments that quantify multiple PCB congeners with a wide range of
28 chlorination levels characterizing current or previous exposure by various routes (e.g., oral,
29 inhalation, dermal) will receive higher ratings.
30 Interpretation of PCB exposure measurements will need to account for the following issues:

31 • The half-life and elimination characteristics of PCB congeners vary significantly. Current
32   measures of PCBs in serum/plasma, adipose tissue, or breast milk/colostrum reflect
33   cumulative exposure to persistent PCB congeners, but only recent exposure to labile
34   congeners. In recent years, the full scope of human exposure to PCBs in the general
35   environment has become more appreciated, including the potential for significant
36   inhalation and dermal exposure to lower-chlorinated, less-persistent congeners. Single
37   time-point estimates of tissue PCB concentrations might therefore capture only a portion of

1 past exposure levels and could impact the sensitivity of the study to detect associations that
2 might be present. An exception would be studies that measure exposure during a discrete
3 period with relatively short duration (e.g., prenatal exposure where specific critical
4 exposure windows are known and correspond to sample collection).

5 • Most studies do not measure all PCB congeners present in biological tissues; analyzed
6 congeners are generally selected because of their relative occurrence in biological samples
7 or the ability to detect them using a given analytical method—not because of their
8 biological activity or their potential to induce a particular health effect. Again, use of this
9 approach results in an incomplete exposure assessment that easily could miss important
10 relationships between exposure and effect. Nevertheless, tissue levels of PCB congeners
11 generally correlate with overall total PCB exposure ([Devoto et al., 1997](#)); therefore, these
12 studies inform the potential for health hazard to result from exposure to PCBs, especially
13 when the analysis is based on a relatively large number of congeners.

14 • Even for persistent congeners that are routinely measured in epidemiology studies
15 (e.g., PCBs 138, 153, 180), a current, single time-point estimate of tissue PCB levels might
16 not represent the composition of PCB congeners with biological activity during the relevant
17 period for the development of toxicity. Depending on the endpoint of concern, the timing of
18 exposure could be just as important as the magnitude. Therefore, assessing for exposure
19 during the relevant developmental window or within a relatively short time before or after
20 that period is important. If PCB exposure is assessed years after the critical window has
21 passed, it is possible to envision several different exposure scenarios that could lead to the
22 observed PCB levels. And, for each scenario, although the resulting PCB levels are the same,
23 the toxicological implications could be very different.

24 • Limits of detection for PCB analytical methods vary. Regardless of the analytical method
25 used, confidence in exposure measures is increased by reporting on limits of detection and
26 methods used to account for values below those limits. PCB analytical methods also differ
27 in accuracy and in the type of information they provide.

28 • PCBs are lipophilic compounds and, depending on the matrix, adjusting for lipids in these
29 studies could be important. General population studies differ in how they account for lipids:
30 studies lipid-standardize PCB exposure measures, include a measure of lipids as a covariate
31 in multivariable exposure-outcome models, or simply do not adjust for lipids. The most
32 appropriate method for addressing lipid levels in PCB analyses depends on the causal
33 structure of the exposure-outcome association ([O'Brien et al., 2016](#); [Schisterman et al.,](#)
34 [2005](#)).

35 To address the issues listed above, criteria will be developed to evaluate the type of

36 analytical methods used by a study, the accuracy of the analytical method for the PCB congeners

37 assessed (e.g., total PCBs, individual congeners, Aroclor mixtures), and how the exposure

38 measurement will be interpreted in terms of the exposure period represented by the mixture, route

39 of exposure, and relevance to the window of susceptibility for each health effect. Criteria will be

40 developed to evaluate methods for lipid adjustment on an outcome-specific basis. These criteria

41 and others, both PCB specific and outcome specific, developed for use in this assessment, will be

42 documented in the updated protocol released with the draft assessment.

Table 11. Questions to guide the development of criteria for each domain in epidemiology studies

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
Exposure measurement Does the exposure measure reliably distinguish between levels of exposure in a time window considered most relevant for a causal effect with respect to the development of the outcome?	For all: • Does the exposure measure capture the variability in exposure among the participants, considering intensity, frequency, and duration of exposure? • Does the exposure measure reflect a relevant time window? If not, can the relationship between measures in this time and the relevant time window be estimated reliably? • Was the exposure measurement likely to be affected by a knowledge of the outcome? • Was the exposure measurement likely to be affected by the presence of the outcome (i.e., reverse causality)? For case-control studies of occupational exposures: • Is exposure based on a comprehensive job history describing tasks, setting, time	Is the degree of exposure misclassification likely to vary by exposure level? If the correlation between exposure measurements is moderate, is there an adequate statistical approach to ameliorate variability in measurements? If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations require customization to the exposure and outcome (relevant timing of exposure) **Good** • Valid exposure assessment methods used, which represent the etiologically relevant 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 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 exists that 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 indicates that reverse causality is very likely to account for the observed association. • Exposure measurement was not independent of outcome status.

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
	period, and use of specific materials? For biomarkers of exposure, general population: • Is a standard assay used? What are the intra- and interassay coefficients of variation? Is the assay likely to be affected by contamination? Are values less than the limit of detection dealt with adequately? • What exposure time period is reflected by the biomarker? If the half-life is short, what is the correlation between serial measurements of exposure?		
Outcome ascertainment Does the outcome measure reliably distinguish the presence or absence (or degree of severity) of the outcome?	For all: • Is outcome ascertainment likely to be affected by knowledge of, or presence of, exposure (e.g., consider access to health care, if based on self-reported history of diagnosis)? For case-control studies: • Is the comparison group without the outcome (e.g., controls in a case-control study) based on objective criteria with little or no likelihood of inclusion of people with the disease?	Is there a concern that any outcome misclassification is nondifferential, differential, or both? What is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations require customization to the outcome **Good** • High certainty in the outcome definition (i.e., specificity and sensitivity), minimal concerns with respect to misclassification. • Assessment instrument was validated in a population comparable to the one from which the study group was selected. **Adequate** • Moderate confidence that outcome definition was specific and sensitive, some uncertainty with respect to misclassification but not expected to greatly change the effect estimate. • Assessment instrument was validated but not necessarily in a population comparable to the study group.

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
	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 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 interassay variability? Is the sensitivity of the assay appropriate for the outcome measure in this study population?		**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 will not be automatically construed to be critically deficient.
Participant selection Does evidence indicate that selection into or out of the study (or analysis sample) was jointly related to exposure and to outcome?	For longitudinal cohort: • Did participants volunteer for the cohort based on knowledge of exposure or preclinical disease symptoms? Was entry into the cohort or continuation in the cohort related to exposure and outcome? For occupational cohort:	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?	These considerations could require customization to the outcome. This might include determining what study designs effectively allow analyses of associations appropriate to the outcome measures (e.g., design to capture incident vs. prevalent cases, design to capture early pregnancy loss). **Good** • Minimal concern for selection bias based on description of recruitment process (e.g., selection of comparison population, population-based random sample selection, recruitment from sampling frame including current and previous employees). • Exclusion and inclusion criteria specified and would not induce bias.

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
	• 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 populations and time periods from which cases were drawn? • Are hospital controls selected from a group whose reason for admission is independent of exposure? • Could recruitment strategies, eligibility criteria, or participation rates result in differential participation relating to both disease and exposure? For population-based survey: • Was recruitment based on advertisement to people with	Is there a comparison of participants and nonparticipants to address whether differential selection is likely?	• Participation rate is reported at all steps of study (e.g., initial enrollment, follow-up, selection into analysis sample). If rate is not high, there is appropriate rationale for why it is unlikely to be related to exposure (e.g., comparison between participants and nonparticipants or other available information indicates differential selection is not likely). **Adequate** • Sufficient description of the recruitment process to be comfortable 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 or participation; or aspects of these processes raise the potential for bias (e.g., healthy worker effect, survivor bias). **Critically deficient** • Aspects of the processes for recruitment, selection strategy, sampling framework, or participation result in concern that selection bias resulted in a large impact on effect estimates (e.g., convenience sample with no information about recruitment and selection, cases and controls are recruited from different sources with different likelihood of exposure, recruitment materials stated outcome of interest and potential participants are aware of or are concerned about specific exposures)

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
	knowledge of exposure, outcome, and hypothesis?		

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), 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)?	These considerations require customization to the exposure and outcome, but this could be limited to identifying key covariates.
			
			 Good
			
			 • Conveys strategy for identifying key confounders. This might include: a priori biological considerations, published literature, causal diagrams, or statistical analyses; with recognition that not all "risk factors" are confounders.
			
			 • Inclusion of potential confounders in statistical models not based solely on statistical significance criteria (e.g., $p < 0.05$ from stepwise regression).
			
			 • Does not include variables in the models 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.
			
			 Adequate
			
			 • Similar to *good* but might not have included all key confounders or less detail might be available on the evaluation of confounders

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
			(e.g., sub-bullets in *good*). Residual confounding could explain part of the observed effect, but concern is minimal.
			Deficient
			• Does not include variables in the models likely to be influential colliders or intermediates on the causal pathway.
			And any of the following:
			• 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.
			• Descriptive information on key confounders (e.g., their relationship relative to the outcomes and exposure levels) is not presented.
			• Strategy of evaluating confounding is unclear or is not recommended (e.g., based only 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.
			• Confounding is likely present and not accounted for, indicating all of the results were most likely due to bias.
			• Presenting a progression of model results with adjustments for different potential confounders, if warranted.
Analysis Does the analysis strategy and presentation convey the necessary familiarity with the data and assumptions?	• Are missing outcome, exposure, and covariate data recognized, and if necessary, accounted for in the analysis? • Does the analysis appropriately consider variable distributions and modeling assumptions?	If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations might require customization to the outcome. This could include the optimal characterization of the outcome variable and ideal statistical test (e.g., Cox regression). **Good** • Use of an optimal characterization of the outcome variable.

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
	- Does the analysis appropriately consider subgroups of interest (e.g., based on variability in exposure level or duration or susceptibility)? - Is an appropriate analysis used for the study design? - Is effect modification considered, based on considerations developed a priori? - Does the study include additional analyses addressing potential biases or limitations (i.e., sensitivity analyses)?		- 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 limits of detection (and percentage below the limits of detection), 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: - Descriptive information about exposure provided (where applicable) but might be incomplete; might not have discussed missing data, cutpoints, or shape of distribution. - Includes analyses that address robustness of findings (examples in *good*), but some important analyses are not performed. **Deficient** - Does not conduct analysis using optimal characterization of the outcome variable. - Descriptive information about exposure levels not provided (where applicable).

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

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

Table 12. Information relevant to evaluation domains for epidemiology studies

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

6.3. EXPERIMENTAL ANIMAL STUDY EVALUATION

1   The evaluation of experimental animal studies applies similar principles as those described
2 above for the evaluation of epidemiology studies. The evaluation process focuses on assessing
3 aspects of the study design and conduct through three broad types of evaluations: reporting quality,
4 risk of bias, and study sensitivity. A set of domains with accompanying core questions falls under
5 each evaluation type and directs individual reviewers to evaluate specific study characteristics. For
6 each domain and core question pairing, basic considerations provide additional guidance on how a
7 reviewer might evaluate and judge a study for that domain.

1   Table 13 provides the standard domains and core questions, along with some basic
2   considerations for guiding the evaluation. Some domain considerations will need to be tailored to
3   the chemical and endpoint/outcome, while others are generalizable across assessments
4   (e.g., considerations for reporting quality). Assessment teams work with subject matter experts to
5   develop the assessment-specific considerations. These specific considerations are determined
6   prior to performing study evaluation, although they might be refined as the study evaluation
7   proceeds (e.g., during pilot testing). Assessment-specific considerations are documented and made
8   publicly available with the assessment.
9   Each domain receives a consensus judgment of *good, adequate, deficient, not reported,* or
10   *critically deficient* (as described in Section 6.1), accompanied by a rationale for the judgment. Once
11   all domains are rated, an overall confidence classification of *high, medium,* or *low* confidence or
12   *uninformative* is assigned (as described in Section 6.1). The rationale for the classification,
13   including a brief description of any identified strengths or limitations from the domains and their
14   potential impact on the overall confidence determination, will be documented clearly and
15   consistently. This rationale will, to the extent possible, reflect an interpretation of the potential
16   influence on the results (including the direction or magnitude of influence).
17   One of the key uncertainties in this assessment relates to the impact of congener profile on
18   the toxicity of PCB mixtures. As a result, chemical-specific considerations for reporting quality will
19   be applied in evaluations of studies of PCB exposure. For example, studies that administer PCB
20   mixtures should provide the name, source, purity, and lot number of the mixture to receive the
21   highest evaluation rating (*good*) for chemical administration and characterization. The congener
22   profiles of different lots of Aroclor 1254 can vary, resulting in differences in biological activity
23   ([Kodavanti et al., 2001](#)); therefore, reporting the lot number of the administered PCB mixture is
24   important for fully characterizing its chemical composition. However, if the identity of the PCB
25   mixture (e.g., Aroclor 1254) is known, a lack of information on the lot number usually will not be
26   considered a significant limitation and, by itself, is unlikely to have a significant impact on overall
27   study confidence ratings.
28   A wide variety of outcomes have been assessed in animal studies of PCBs. Considerations
29   specific to each hazard domain and outcome are not included in this protocol; these will be
30   documented in the updated protocol released with the draft assessment. However, examples of
31   specific considerations that could be applied include better domain ratings for studies that address
32   potential differences in timing (e.g., time of day) for evaluations of specific behaviors or for
33   evaluations of hormone levels (due to fluctuations with circadian rhythms) and for studies that
34   address fasting status for measurements related to metabolism.

Table 13. Questions to guide the development of criteria for each domain in experimental 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 endpoint(s)/outcome(s) of interest? *Notes:* *Reviewers will attempt to contact authors to obtain missing information when studies are considered key for hazard evaluation or dose-response.* • 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	These considerations typically do not need to be refined, although in some instances the **important information** could be refined depending on the endpoints/outcomes of interest or the chemical under investigation. A judgment and rationale for this domain will be given for the study. Typically, these will not change, regardless of the endpoints/outcomes investigated by the study. In the rationale, reviewers will indicate whether the study adhered to good laboratory practices, OECD, or other 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. However, the missing information is not expected to significantly impact the study evaluation. • **Deficient**: All **critical information** is reported but **important information** is missing that is expected to significantly reduce the ability to evaluate the study. • **Critically deficient**: Study report is missing pieces of **critical information**. Studies critically deficient for reporting are uninformative for the overall rating and not considered further for evidence synthesis and integration.

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?	These considerations typically need not be refined. 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 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 (cont.)	**Selection and performance bias (cont.)**	**Observational bias/blinding**
		
		Did the study implement measures to reduce observational bias?
			For each endpoint/outcome or grouping of endpoints/outcomes in a study:
			
			• Does the study report blinding or other methods/procedures for reducing observational bias?
			
			• 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?
				These considerations typically do not need to be refined. (Note that it can be useful for teams to identify highly subjective measures of endpoints/outcomes where observational bias might strongly influence results prior to performing evaluations.)
				A judgment and rationale for this domain will be given for each endpoint/outcome or group of endpoints/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[b]).
				
				• **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.
				
				    ◦ (Interpreted as *adequate*) The potential concern for bias was mitigated based on the 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.
				
				    ◦ (Interpreted as *deficient*) The potential impact on the results is major (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 (cont.)	**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, etc.) that could bias the results? • If differences are identified, to what extent are they expected to impact the results?	These considerations might need to be refined, as the specific variables of concern can vary by experiment or chemical. 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**: Beyond 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 results were uncontrolled or inconsistent across groups but are expected to have a minimal impact on the results. • **Deficient**: Notable concern that potentially confounding variables were uncontrolled or inconsistent across groups and are expected to substantially impact the results. • **Critically deficient**: Confounding variables were presumed to be uncontrolled or inconsistent across groups and are expected to be a primary driver of the results.

Evaluation concern		Domain – core question	Prompting questions	General considerations
Risk of bias (cont.)	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 analysis/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*:* • Are all animals accounted for in the results? • If discrepancies exist, do authors provide an explanation (e.g., death or unscheduled sacrifice during the study)? • If results omissions or attrition are identified, what is the expected impact on the interpretation of the results?	These considerations typically do not need to be refined. 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 explanation, and these 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 are not explained but are 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; omissions or attrition could significantly impact the interpretation of the results. • **Critically deficient**: Extensive results omission or animal attrition are identified, preventing comparisons of results across treatment groups.

Evaluation concern		Domain – core question	Prompting questions	General considerations
Sensi-tivity	Exposure methods sensitivity	**Chemical administration and characterization** Did the study adequately characterize exposure to the chemical of interest and the exposure administration methods? *Note:* *Consideration of the appropriateness of the route of exposure is not evaluated at the individual study level. Relevance and utility of the routes of exposure are considered in the PECO criteria for study inclusion and during evidence synthesis.*	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? • Did the authors take steps to ensure the reported exposure levels were accurate? • Are there concerns about the methods used to administer the chemical (e.g., inhalation chamber type, gavage volume)? For inhalation studies: • Were target concentrations confirmed using reliable analytical measurements in chamber air?	It is essential that these criteria are considered, and potentially refined, as the specific variables of concern can vary by chemical (e.g., stability could be an issue for one chemical but not another). 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). There are no concerns about the composition, stability, or purity of the administered chemical or the specific methods of administration. For inhalation studies, chemical concentrations in the exposure chambers are verified using reliable analytical methods. • **Adequate**: Some uncertainties in the chemical administration and characterization are identified but these are expected to have minimal impact on interpretation of 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; for inhalation studies, actual exposure concentrations are missing or verified with less reliable methods).

Evaluation concern		Domain – core question	Prompting questions	General considerations
Sensi-tivity (cont.)	Exposure methods sensitivity (cont.)	**Chemical administration and characterization (cont.)**	For oral studies: • If necessary based on consideration of chemical-specific knowledge (e.g., instability in solution; volatility) or exposure design (e.g., the frequency and duration of exposure), were chemical concentrations in the dosing solutions or diet analytically confirmed?	• **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 the use of static inhalation chambers or a gavage volume considered too large for the species or lifestage at exposure). • **Critically deficient**: Uncertainties in the exposure characterization are identified and there is reasonable certainty that the results are largely attributable to factors other than exposure to the chemical of interest (e.g., identified impurities are expected to be a primary driver of the results).
		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 endpoints/outcomes in a study: • Does the exposure period include the critical window of sensitivity? • Was the duration and frequency of exposure sensitive for detecting the endpoint of interest?	Considerations for this domain are highly variable depending on the endpoint(s)/outcome(s) of interest and must be refined. A judgment and rationale for this domain will be given for each endpoint/outcome or group of endpoints/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 the majority of the critical window of sensitivity (if known). These limitations are expected to bias the results towards the null. • **Critically deficient**: The exposure design was not sensitive and is expected to strongly bias the results toward the null. The rationale will indicate the specific concern(s).

Evaluation concern		Domain – core question	Prompting questions	General considerations
Sensi-tivity (cont.)	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 endpoints/outcomes in a study: • Are there concerns regarding the specificity and validity of the protocols? • Are there serious concerns regarding the sample size? • Are there concerns regarding the timing of the endpoint assessment?	Considerations for this domain are highly variable, depending on the endpoint(s)/outcome(s) of interest, and must be refined. A judgment and rationale for this domain will be given for each endpoint/outcome or group of endpoints/outcomes investigated in the study. Examples of potential concerns include: • Selection of protocols that are insensitive or nonspecific for the endpoint of interest. • Use of unreliable 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 to exposure (e.g., short-acting depressant or irritant effects of chemicals; insensitivity due to prolonged period of nonexposure prior to testing).

Evaluation concern		Domain – core question	Prompting questions	General considerations
Sensi-tivity (cont.)	Outcome measures and results display (cont.)	**Results presentation** Are the results presented in a way that makes the data usable and transparent?	For each endpoint/outcome or grouping of endpoints/outcomes in a study: • Does the level of detail allow for an informed interpretation of the results? • Are the data analyzed, compared, or presented in an inappropriate or misleading way?	Considerations for this domain are highly variable, depending on the outcomes of interest, and must be refined. A judgment and rationale for this domain will be given for each endpoint/outcome or group of endpoints/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. • Pooling data when responses are known or expected to differ substantially (e.g., across sexes or ages). • 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).

Evaluation concern	Domain – core question	Prompting questions	General considerations
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 toward the null) for additional consideration during evidence synthesis. If the study is otherwise well conducted and an effect is observed, the confidence could be increased.*	For each endpoint/outcome or grouping of endpoints/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 endpoints/outcomes investigated in the study. Confidence ratings are described above (see Section 6.1.).

OECD = Organisation for Economic Cooperation and Development.

[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).

[b] For nontargeted or screening-level histopathology outcomes often used in guideline studies, blinding during the initial evaluation of tissues is generally not recommended as masked evaluation can make "the task of separating treatment-related changes from normal variation more difficult" and "there is concern that masked review during the initial evaluation might result in missing subtle lesions." Generally, blinded evaluations are recommended for targeted secondary review of specific tissues or in instances when a predefined set of outcomes is known or predicted to occur (Crissman et al., 2004).

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

1 PBPK (or classical pharmacokinetic [PK]) models might be used in an assessment when an
2 applicable one exists and no equal or better alternative for dosimetric extrapolation is available. As
3 described in Sections 2.5.2 and 2.5.3, pharmacokinetic models will be considered for use in this
4 assessment to support route-to-route (i.e., oral-to-inhalation) and interspecies extrapolations and
5 to quantitatively predict transfer of PCBs across the placenta or via breast milk. Any models used
6 will represent current scientific knowledge and accurately translate the science into computational
7 code reproducibly and transparently. For a specific target organ/tissue, employing or adapting an
8 existing PBPK model or developing a new PBPK model or an alternative quantitative approach
9 might be possible. Data for PBPK models could come from studies across various species and might
10 be in vitro or in vivo in design.

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

11 Key information from identified models will be summarized in tabular format (see example
12 Table 14 below).

Table 14. Example descriptive summary for a physiologically based pharmacokinetic (PBPK) model study

Study detail	Description/notes				
Author	LastName et al. (2003)				
Contact email	xxxxx@email.com				
Contact phone	xxx-xxx-xxxx				
Sponsor	N/A				
Model summary					
Species	Rat				
Strain	F433				
Sex	Male and female				
Lifestage	Adult				
Exposure routes	Inhalation	Oral	I.V.	Skin	
Tissue dosimetry	Blood	Liver	Kidney	Urine	Lung
Model evaluation					
Language	ACSL 11.8				
Code available	YES	Effort to recreate model		COMPLETE	
Code received	YES	Effort to migrate to open software		SIGNIFICANT	

Study detail	Description/notes
Structure evaluated	YES
Math evaluated	YES
Code evaluated	YES. Issue (minor): Incorrect units listed in comments for liver metabolism (line 233). Issue (major): Mass balance error in stomach compartment.
Available PK data	Urine (cumulative amount excreted) and blood (concentration) time course data for oral (gavage) and inhalation (6 hr/day for 4 days) exposure. In vitro skin permeation.

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

1   Once summarized, available PBPK models will be evaluated in accordance with criteria
2   outlined by U.S. EPA (2018). Judgments on the suitability of a model are separated into two
3   categories: scientific and technical (see Table 16). The scientific criteria focus on whether the
4   biology, chemistry, and other information available for chemical MOAs are justified (i.e., preferably
5   with citations to support use) and represented by the model structure and equations. The scientific
6   criteria are judged based on information presented in the publication or report that describes the
7   model and do not require evaluation of the computer code. Preliminary technical criteria include
8   availability of the computer code and completeness of parameter listing and documentation.
9   Studies that meet the preliminary scientific and technical criteria are then subjected to an in-depth
10   technical evaluation, which includes a thorough review and testing of the computational code. The
11   in-depth technical and scientific analyses focus on the accurate implementation of the conceptual
12   model in the computational code, use of scientifically supported and biologically consistent
13   parameters in the model, and reproducibility of model results reported in journal publications and
14   other documents. This approach stresses (1) clarity in the documentation of model purpose,
15   structure, and biological characterization; (2) validation of mathematical descriptions, parameter
16   values, and computer implementation; and (3) evaluation of each plausible dose metric. The
17   in-depth analysis is used to evaluate the potential value and cost of developing a new model or
18   substantially revising an existing one. PBPK models developed by EPA during the course of the
19   assessment will be peer reviewed, either as a component of the draft assessment or by publication
20   in a journal article.

Table 15. Criteria for evaluating physiologically based pharmacokinetic (PBPK) models

Category	Specific criteria
Scientific	Biological basis for the model is accurate. • Consistent with mechanisms that significantly impact dosimetry. • Predicts dose metric(s) expected to be relevant. • Applicable for relevant route(s) of exposure.

	Consideration of model fidelity to the biological system strengthens the scientific basis of the assessment relative to standard exposure-based extrapolation (default) approaches.
	
	• Ability of model to describe critical behavior, such as nonlinear kinetics in a relevant dose range, better than the default (i.e., $BW^{3/4}$ [body weight scaling to the ¾ power] scaling).
	
	• Model parameterization for critical lifestages or windows of susceptibility. Evaluation of these criteria also will consider the model's fidelity vs. default approaches and possible use of an intraspecies UF in conjunction with the model to account for variations in sensitivity between lifestages.
	
	• Predictive power of model-based dose metric vs. default approach, based on exposure.
	
	    ○ Specifically, model-based metrics might correlate better than the applied doses with animal/human dose-response data.
	
	    ○ The degree of certainty in model predictions vs. default is also a factor. For example, although target tissue metrics are generally considered better than blood concentration metrics, lack of data to validate tissue predictions when blood data are available might lead to choosing the latter.

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

6.5. MECHANISTIC STUDY EVALUATION

1   Sections 9 and 10 outline an approach for considering information from mechanistic studies
2 (including in vitro, in vivo, ex vivo, and in silico studies) where the specific analytical approach is
3 targeted to the assessment needs, depending in part on the extent and nature of the phenotypic
4 human and animal evidence. In this way, the mechanistic synthesis for a given health effect might
5 range from a high-level summary (or detailed analysis) of potential mechanisms of action to
6 specific, focused questions needed to address important and impactful assessment uncertainties
7 unaddressed by the available phenotypic studies (e.g., expected shape of the dose-response curve in
8 the low-dose region, applicability of the animal evidence to humans, addressing susceptible
9 populations). Individual study-level evaluation of mechanistic endpoints typically will not be
10 pursued. However, for some chemical assessments, it may be necessary to identify assay-specific
11 considerations for study endpoint evaluations on a case-by-case basis to provide a more detailed
12 summary and evaluation for the most relevant individual studies. This might be done, for example,
13 when the scientific understanding of a critical mechanistic event or MOA is less established or lacks
14 scientific consensus, the reported findings on a mechanistic endpoint are conflicting, the available
15 mechanistic evidence addresses a complex and influential aspect of the assessment, or in vitro or in
16 silico data make up the bulk of the evidence base and little or no evidence from epidemiology
17 studies or animal bioassays is available. Any considerations used to evaluate mechanistic studies
18 will be documented in the assessment.

7. ORGANIZING THE HAZARD REVIEW

1   The organization and scope of the hazard evaluation is determined by the available
2 evidence for the chemical regarding routes of exposure, metabolism and distribution, outcomes
3 evaluated, and number of studies pertaining to each outcome and by the results of the evaluation of
4 sources of bias and sensitivity. The hazard evaluations will be organized around organ systems
5 (e.g., respiratory, nervous system) informed by one or multiple related outcomes, and a decision
6 will be made as to what level (e.g., organ system or subsets of outcomes within an organ system) to
7 organize the synthesis.
8   Table 16 lists some questions that might be asked of the evidence to aid this decision. These
9 questions extend from considerations and decisions made during development of the refined
10 evaluation plan to include review of the concerns raised during individual study evaluations and
11 the direction and magnitude of the study-specific results. Resolution of these questions then will
12 inform critical decisions about the organization of the hazard evaluation and what studies might be
13 useful in dose-response analyses.

Table 16. Querying the evidence to organize syntheses for human and animal evidence

Evidence	Questions	Follow-up questions
ADME	Are absorption, distribution, metabolism, or excretion different by route of exposures studied, lifestage when exposure occurred, or dosing regimens used?	Will separate analyses be needed by route of exposure, or by methods of dosing within a route of exposure (e.g., are large differences expected between gavage and dietary exposures)? Which lifestages and what dosing regimens are more relevant to human exposure scenarios?
	Is there toxicity information for metabolites that also should be evaluated for hazard?	What exposures will be included in the evaluation?
	Is the parent chemical or metabolite also produced endogenously?	
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?

Evidence	Questions	Follow-up questions
	Are there interrelated outcomes? If so, consider whether some outcomes are more useful or of greater concern than others.	By what commonalities will the outcomes be grouped: • 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?
	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 any of the outcomes and how might they 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, not informative*) for the different outcomes? Is there enough evidence from *high* and *medium* confidence studies for particular outcomes 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 could be most relevant to human exposure than others?	What outcomes and study characteristics are informative for development of toxicity values?
	Which outcomes have sufficient data available to draw conclusions about dose-response? Are any outcomes with study results sufficiently similar (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 sufficient for dose-response analysis?	
	Do some subgroups exhibit responses at lower exposure levels than others?	
	Could findings from ADME studies inform data-derived extrapolation factors, or link toxicity observed via different routes of exposure, or link effects between humans and experimental animals?	Can a common internal dose metric be used to compare species or routes of exposure?

8. DATA EXTRACTION OF STUDY METHODS AND RESULTS

1 Data extraction and content management will be carried out using HAWC or DRAGON. Data
2 extraction elements that might be collected from epidemiology, controlled human exposure, animal
3 toxicology, and in vitro studies are listed in Appendix B. Data extraction elements that might be
4 collected from PBPK studies are listed in Table 14. The content of the data extraction might be
5 revised following the identification of the studies included in the review as part of a pilot phase to
6 assess the data extraction workflow. Not all studies that meet the PECO criteria will be subject to
7 data extraction. Studies evaluated as being *uninformative* are not considered further and would,
8 therefore, not undergo data extraction. In addition, outcomes determined to be less relevant during
9 PECO refinement might not go through data extraction or could have only minimal data extraction.
10 The same could be true for *low* confidence studies if sufficient *medium* and *high* confidence studies
11 are available. All findings are considered for extraction, regardless of statistical significance,
12 although the level of extraction for specific outcomes within a study might differ (i.e., ranging from
13 a narrative to full extraction of dose-response effect size information). Similarly, decisions about
14 data extraction for *low* confidence studies are typically made during implementation of the protocol
15 based on consideration of the quality and extent of the available evidence. The version of the
16 protocol released with the draft assessment will outline how *low* confidence studies were treated
17 for extraction and evidence synthesis.
18 The data extraction results for included studies will be presented in the assessment and
19 made available for download from EPA HAWC in Excel format when the draft is publicly released.
20 Note that the following browsers are fully supported for accessing HAWC: Google Chrome
21 (preferred), Mozilla Firefox, and Apple Safari; errors in functionality occur when viewed with
22 Internet Explorer. Data extraction will be performed by one member of the evaluation team and
23 checked by one or two other members. Discrepancies in data extraction will be resolved by
24 consultation with a third member of the evaluation team. Once the data have been verified, they
25 will be "locked" to prevent accidental changes. Digital rulers, such as WebPlotDigitizer
26 (https://automeris.io/WebPlotDigitizer/), are used to extract numerical information from figures.
27 Use of digital rulers is documented during extraction.
28 As previously described, routine attempts will be made to obtain information missing from
29 human and animal health effect studies if missing information is considered influential during study
30 evaluations (see Section 6) or when required to conduct a meta-analysis (e.g., missing group size or
31 variance descriptors such as standard deviation or confidence interval). Missing data from
32 individual mechanistic (e.g., in vitro) studies generally will not be sought. Outreach to study

1 authors will be documented and considered unsuccessful if researchers do not respond to email or
2 phone requests after one or two attempts.

8.1. STANDARDIZING REPORTING OF EFFECT SIZES

3 In addition to providing quantitative outcomes in their original units for all study groups,
4 results from outcome measures will be transformed to a common metric, when possible, to help
5 compare distinct but related outcomes measured with different scales. These standardized effect
6 size estimates facilitate systematic evaluation and evidence integration for hazard identification
7 and meta-analysis when feasible for an assessment (see Section 9.1). Based on metrics across the
8 available studies, a common metric might be used and the calculation presented in the assessment.
9 For epidemiology studies, the typical approach is to extract adjusted statistical estimates
10 when possible, rather than unadjusted or raw estimates.
11 It is important to consider the variability associated with effect size estimates, with stronger
12 studies generally showing more precise estimates. However, effect size estimation can be
13 influenced by such factors as variances that differ substantially across treatment groups or by lack
14 of information to characterize variance, especially for animal studies in biomedical research
15 ([Vesterinen et al., 2014](#)).

8.2. STANDARDIZING ADMINISTERED DOSE LEVELS/CONCENTRATIONS

16 Exposures will be standardized to common units. Exposure levels in oral studies will be
17 expressed in units of mg PCB/kg-day. When study authors provide exposure levels as
18 concentrations in the diet, dose conversions will be made using study-specific food consumption
19 rates and body weights when available. Otherwise, EPA defaults will be used ([U.S. EPA, 1988](#)),
20 addressing age and study duration as relevant for the species/strain and sex of the animal of
21 interest. Exposure levels in inhalation studies will be expressed in units of mg/m^3. Assumptions
22 used in performing dose conversions will be documented.
23 Unless otherwise reported by study authors, the background level in experimental animal
24 studies is assumed 0 ppm (0 mg/kg-day).

9. SYNTHESIS OF EVIDENCE

1   For the purposes of this assessment, evidence synthesis and integration are considered
2   distinct, but related, processes. The syntheses of separate bodies of evidence (i.e., human, animal,
3   and mechanistic evidence) described in this section will directly inform the integration across all
4   evidence to draw an overall judgment for each assessed human health effect (described in
5   Section 10). The phrase "evidence integration" used here is analogous to the phrase "weight of
6   evidence" used in some other assessment processes ([EFSA, 2017](#); [U.S. EPA, 2017](#); [NRC, 2014](#); [U.S.
7   EPA, 2005](#)).[14]
8   For each potential health hazard or smaller subset of related outcomes, the available
9   phenotypic human and animal health effect evidence will be synthesized separately. Mechanistic
10 evidence also will be considered, although the specific analytical approach is targeted to the
11 assessment needs, depending on the extent and nature of the phenotypic human and animal
12 evidence (see Sections 9.2 and 10). The results of the analyses of mechanistic evidence will be used
13 to inform key uncertainties; as a result, the scope of the mechanistic analyses will generally depend
14 on the extent and nature of the human and animal evidence (see Sections 9.2 and 10). Thus, apart
15 from the pre-defined mechanistic analyses (see Sections 9.2.1-9.2.3), the human and animal
16 evidence syntheses (or the lack of phenotypic data in humans and animals) help determine the
17 approach to be taken in synthesizing the available mechanistic evidence (see Section 9.2.4). In this
18 way, a mechanistic evidence synthesis might range from a high level summary of potential toxicity
19 mechanisms discussed in the published literature to a detailed analysis of multiple potential
20 modes-of-action, or it might evaluate specific, focused questions that inform key uncertainties
21 unaddressed by the phenotypic human and animal evidence (e.g., shape of the dose response curve
22 at low doses, applicability of the animal evidence to humans, addressing susceptible populations).
23 Each synthesis will provide a summary discussion of the available evidence that addresses
24 considerations adapted from considerations for causality introduced by Austin Bradford Hill ([Hill,
25 1965](#)): consistency, exposure-response relationship, strength of the association, temporal
26 relationship, biological plausibility, coherence, and "natural experiments" in humans [([U.S. EPA,
27 2005](#), [1994b](#)); see Table 17]. Importantly, the evidence synthesis process explicitly considers and
28 incorporates the conclusions from the individual study evaluations (see Section 6).

---

[14] This revision has been adopted primarily based on 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 17. Information most relevant to describing primary considerations informing 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, syntheses will consider study confidence decisions. *High* confidence studies carry the most weight. Syntheses will consider specific limitations and strengths of studies and how they inform 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 explained by differences between human populations, animal models, exposure conditions, or study methods) ([U.S. EPA, 2005](#)) based on analyses of potentially important explanatory factors such as: • Confidence in the studies' results, including study sensitivity (e.g., some study results that appear to be inconsistent could 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, and exposure assessment methods (i.e., in epidemiology studies). • 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, or lifestages. The interpretation of consistency will emphasize biological significance, to the extent it is understood, over statistical significance. Statistical significance from suitably applied tests 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 and providing context to interpretations of consistency could be helpful.[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). Although larger effect magnitudes and precision (e.g., $p < 0.05$) help reduce concerns about chance, bias, or other factors as explanatory, syntheses also will 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 (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-response is lacking or unclear, the synthesis also will consider the potential influence of such factors on the response pattern.
Coherence	**Description:** Examines the extent to which findings are cohesive across different endpoints that are 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 will be explored including 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 (see Section 9.2) and this consideration overlaps with "coherence." This examines the biological support (or lack thereof) for findings from the human and animal health effect studies and becomes more impactful on the hazard conclusions when notable uncertainties in the strength of those sets of studies exist. These analyses can also improve understanding of dose- or duration-related development of the health effect. In the absence of human or animal evidence of apical health endpoints, the synthesis of mechanistic information could drive evidence integration judgments (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 the observed effects result from exposure. This will be 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 also could include evaluations of biological pathways (e.g., for the health effect, established for other, possibly related, chemicals). The synthesis 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, one benefit of natural experiments is that people are divided into exposed and unexposed groups without influencing their own exposure status. During synthesis, associations in *medium* and *high* confidence natural experiments can substantially reduce concerns about residual confounding.

PECO = populations, exposures, comparators, and outcomes.

[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.

1

2   Data permitting, the syntheses will also discuss analyses relating to potential susceptible

3   populations.[15] These analyses will be based on knowledge about the health outcome or organ

4   system affected, demographics, genetic variability, lifestage, health status, behaviors or practices,

---

[15] 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." Further, these terms have been inconsistently defined across the scientific literature. The term susceptibility is used in this protocol to describe populations at increased risk, focusing on biological (intrinsic) factors and social and behavioral determinants that can modify the effect of a specific exposure. However, certain factors resulting in higher exposures to specific groups (e.g., proximity, occupation, housing) might not be analyzed to describe potential susceptibility among specific populations or subgroups.

1 social determinants, and exposure to other pollutants (see Table 18). This information will be used
2 to describe potential susceptibility among specific populations or lifestages in a separate section
3 (see Section 10.3). This summary will describe concerns across the available evidence for all
4 potential human health effects and will inform hazard identification and dose-response analyses.

Table 18. Individual and social factors that could 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, elderly
Health status	Preexisting 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

9.1. SYNTHESES OF HUMAN AND ANIMAL HEALTH EFFECT EVIDENCE

5 The syntheses of the human and animal health effect evidence will focus on describing
6 aspects of the evidence that best inform causal interpretations, including the exposure context
7 examined in the sets of studies. These syntheses (or the lack of data within these bodies of
8 evidence) help determine the approach to be taken in synthesizing the available mechanistic
9 evidence (see Section 9.2).
10 Evidence synthesis will be based primarily on studies of *high* and *medium* confidence. *Low*
11 confidence studies might be used, if few or no studies with higher confidence are available, to help
12 evaluate consistency, or if the study designs of the *low* confidence studies address notable
13 uncertainties in the set of *high* or *medium* confidence studies on a given health effect. If *low*
14 confidence studies are used, a careful examination of risk of bias and sensitivity with potential
15 impacts on the evidence synthesis conclusions will be included in the narrative.
16 As previously described, these syntheses will articulate the strengths and the weaknesses of
17 the available evidence organized around the considerations described in Table 17 and issues that
18 stem from the evaluation of individual studies (e.g., concerns about bias or sensitivity). If possible,
19 results across studies will be compared using graphs and charts or other data visualization
20 strategies. The analysis typically will include examination of results stratified by any or all of the
21 following: study confidence classification (or specific issues within confidence evaluation domains);
22 population or species; exposures (e.g., level, patterns [intermittent or continuous]; duration;

1   intensity); sensitivity (e.g., low vs. high); and other factors that might have been identified in the
2   refined evaluation plan (e.g., sex, lifestage, other demographic). The number of studies and the
3   differences encompassed by the studies will determine the extent to which specific types of factors
4   can be examined to stratify study results. Additionally, for both the human and animal evidence
5   syntheses, if supported by the available data, additional analyses across studies (such as
6   meta-analysis) also might be conducted.

9.2.   MECHANISTIC INFORMATION

7   Mechanistic information includes any experimental measurement related to a health
8   outcome that informs the biological or chemical events associated with phenotypic effects; these
9   measurements can improve understanding of the mechanisms involved in the biological effects
10   following exposure to a chemical but generally are not, by themselves, considered adverse
11   outcomes. Mechanistic data are reported in a diverse array of observational and experimental
12   studies across species, model systems, and exposure paradigms, including in vitro, in vivo (by
13   various routes of exposure), ex vivo, and in silico studies. The evidence available to describe
14   mechanistic events or MOAs (U.S. EPA, 2005) is typically aggregated from numerous studies, often
15   involving a diverse range of exposure paradigms and models and a wide spectrum of diverse
16   endpoints. In addition, a chemical could operate through multiple mechanistic pathways (U.S. EPA,
17   2005). Similarly, multiple mechanistic pathways might interact to cause an adverse effect. In
18   contrast to the defined scope of the evaluation and syntheses of PECO-specific human or animal
19   health effect studies, the potential utility and interpretation of mechanistic information can be quite
20   broad and hard to define. Thus, to be pragmatic and provide clear and transparent syntheses of the
21   most useful information, the mechanistic syntheses for most health outcomes will focus on a subset
22   of the most relevant mechanistic studies. It should be stressed that the process of evaluating
23   mechanistic information differs fundamentally from evaluations of the other evidence streams.
24   More specifically, the mechanistic analysis for any specific substance will depend on evaluating the
25   confidence that the relevant data are consistent with a plausible biological understanding of how a
26   chemical exposure might generate an adverse outcome, rather than focusing on evaluations of
27   individual studies.
28   The synthesis of mechanistic information informs the integration of health effect evidence
29   for both hazard identification (i.e., biological plausibility or coherence of the available human or
30   animal evidence, inferences regarding human relevance, or the identification of susceptible
31   populations and lifestages across the human and animal evidence) and dose-response evaluation.
32   As introduced in Section 2.5, several key science issues essential to consider in this
33   assessment will involve a focused analysis and synthesis of mechanistic information. One such
34   issue is the identification of toxicokinetic parameters for use in pharmacokinetic models of PCBs,
35   particularly PCB congener half-life values needed to support route-to-route, interspecies, and
36   intraspecies extrapolations. Half-lives are known to vary significantly among the congeners for

1   which they have been determined and are critical determinants of PCB body burden given long-
2   term exposure. Another key issue is the evaluation of the relative contributions of individual PCB
3   congeners to the toxicity of complex PCB mixtures. Because toxic potency can vary independently
4   from half-life, the mechanistic analyses will need to identify both half-life data and toxicity data
5   available for specific congeners and to estimate (e.g., using QSAR methods) half-lives and relative
6   potencies of congeners for which no data are available.

7   Other analyses within the syntheses of mechanistic information will focus on the evidence
8   most useful for informing key uncertainties in the human or animal health effect evidence. This
9   means that, for example, if extensive and consistent *high* confidence human or animal evidence is
10 available, the need to synthesize all available mechanistic evidence will likely be diminished. In
11 such cases, the synthesis will focus on the analysis and interpretation of smaller sets of mechanistic
12 studies that specifically address controversial issues that are anticipated to have a substantial
13 impact on the assessment conclusions. For example, data related to applicability of animal evidence
14 to humans when the human evidence is weak, or the shape of the dose-response curve at low
15 exposure levels when this understanding is highly uncertain and data informing this uncertainty
16 are available. Thus, consideration of biological understanding represents an important component
17 of the evidence analysis. However, mechanistic understanding is not a prerequisite for drawing a
18 conclusion that a chemical causes a given health effect (NTP, 2015; NRC, 2014).

19   To identify the focused set(s) of studies for use in analyses of critical mechanistic questions,
20 the synthesis will apply a phased approach that progressively focuses the scope of the mechanistic
21 information to be considered. This stepwise focusing, which begins during the literature search
22 and screening steps based on problem formulation decisions, depends primarily on the potential
23 hazard signals that arise from the human or animal health effect studies, or from mechanistic
24 studies that signal potential hazards that have not been examined in health effect studies (Table
25 19). Table 20 lists examples of the focused questions or scenarios triggering these mechanistic
26 evaluations and when, during the systematic review, they are likely to apply. Although the specific
27 methods for evaluating the sets of studies relevant to each question will vary, some general
28 considerations are provided below.

Table 19. Preparation for the analysis of mechanistic evidence

Assessment stages of identifying mechanistically relevant information	Examples of evidence to review and key considerations

Scoping and problem formulation materials	• For the chemical under review, identify existing chemical-specific MOAs from other agency assessments or review articles. If summary information is lacking, are there structurally similar chemicals that are better studied mechanistically?
	
	• Are there indications that a specific mechanistic analysis will be warranted? For example, are there recognized areas of scientific controversy or predefined assessment questions that are already known to require a mechanistic evaluation (e.g., chemicals with a potential mutagenic MOA)?
	
	   o If so, consider whether additional, targeted literature searches would be informative.
	
	   o If mechanistic information relevant to a key scientific controversy or to address a mutagenic MOA is lacking, consider whether inferences can be drawn from structure-activity relationships or other "data-poor" approaches.
	
	• What is the active moiety of the agent? Are there metabolites that should be considered? Are there indications that the purity is critically important? Is the chemical endogenously produced?

Literature inventory of toxicokinetic, ADME, and physicochemical information	• Based on ADME differences across species, does information exist that suggests a lack of relevance of the animal exposure scenarios to human situations? Is there evidence that the active moiety would not be expected to reach the target tissue(s) in some species? • If exposure and risk need to be evaluated for routes of exposure not included in existing PBPK models, how should this disconnect be addressed?
Literature inventories of human, animal, and mechanistic information (including in vitro and in silico studies)	• Which human health hazards (both cancer and noncancer) appear to be well studied in the mechanistic inventory? For cancer, which key characteristics of carcinogens are indicated by the database? o Are there mechanistic studies on an organ system, hazard, or key characteristic that were not examined by human or animal studies meeting the PECO criteria? *If so, consider evidence mapping or similar approaches to highlight these knowledge gaps.* • Are there mechanistic endpoints identified from human and animal studies meeting PECO criteria that could be added to the mechanistic inventory?

Assessment stages of identifying mechanistically relevant information	Examples of evidence to review and key considerations

Human and animal evidence syntheses	• For the health effects of primary concern, is an in-depth mechanistic evaluation(s) warranted to inform the available evidence in humans or animals? Typically, this consideration would focus on health effects that show some indication of an association in epidemiology studies or causality in experimental studies. Based on the literature inventory, consider whether mechanistic data are available to inform the specific, key uncertainties that remain. Examples of specific scenarios for evaluation could include:
	
	   o If cancer has been observed and tumor types appear to differ across populations (e.g., species or sex), can mechanistic evaluations inform potential explanations (noting that site concordance is not a requirement for determining the relevance of animal data for humans)?
	
	   o When notable uncertainties in the human or animal findings occur for a health effect (e.g., outstanding methodological limitations), is evidence of biological precursors in humans or animals linked to the observed outcome? Precursors in the same studies or populations provide stronger evidence.
	
	   o Were questions of relevance raised that could be addressed by an evaluation of the mechanistic evidence to establish the human relevance of effects observed in animal studies?
	
	   o Were pronounced, unexplained differences in susceptibility observed that might be explained by an analysis of toxicokinetic or toxicodynamic differences across lifestages or populations (e.g., animal strain, human demographic)?

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

1   The information collected (e.g., in sortable inventories) will be used to identify studies
2 available for consideration in addressing the specific gaps in understanding identified as critical to
3 address through the application of the questions in Table 20, including postulated mechanistic
4 pathways or MOAs that might be involved in the toxicity of the chemical. Subsequently, from the
5 studies available to potentially address the identified gaps in understanding, the synthesis will
6 focus on those considered most impactful to the specified evaluation based on study design
7 characteristics (which might or might not encompass all studies relevant for a particular question),
8 with transparent documentation of the rationale for the focusing. As the potential influence of the
9 information provided by these studies can vary depending on the hazard question(s) or the
10 associated mechanistic events or pathways, the level of rigor also will depend on the potential
11 impact of increased understanding to hazard identification or dose-response decisions, and could
12 range from overviews of potential mechanisms or cursory insights drawn from sets of unanalyzed
13 results to detailed evaluations of a subset of the most relevant mechanistic studies.

1　　　　Although the application of this approach cannot be predefined, for the small subsets of
2　　studies that best address the key mechanistic questions, the synthesis will first prioritize the
3　　studies based on their toxicological relevance to answering the specific question (e.g., model
4　　system, specificity of the assay for the effect of interest). The path for focusing the mechanistic
5　　database will be documented in the updated protocol released with the draft assessment.
6　　　　More rigorous analyses will be particularly important when the sets of studies available to
7　　inform influential mechanistic conclusions are inconsistent and potentially conflicting, or when the
8　　studies include experiments that directly challenge the necessity of proposed mechanistic
9　　relationships between exposure and an apical effect (e.g., altering a receptor-mediated pathway
10　　through chemical intervention or using knock-out animals). More detailed analyses also might be
11　　useful when the study design aspects in the available studies are likely to have significant flaws or
12　　introduce important uncertainties (e.g., potential shortcomings identified during the evaluation of
13　　exposure methods might be clarified using mechanistic studies). In some instances, additional
14　　literature searches could be warranted, targeting mechanistic events or biological pathways that
15　　are not specific to one chemical.
16　　　　For the more rigorous mechanistic analyses, the review will be facilitated using pathway-
17　　based organizational methods and established evidence evaluation frameworks. These approaches
18　　provide transparency and objectivity to the integration and interpretation of mechanistic events
19　　and pathways anchored to the specific questions that have been identified (e.g., anchored to a
20　　specific health effect) across diverse sets of relevant data (e.g., human, animal, in vitro studies).
21　　　　The mechanistic analyses will inform the evidence integration and dose-response analyses,
22　　described in Sections 11 and 12. Examples of how mechanistic information can inform these steps
23　　are summarized in Table 20.

Table 20. Examples of iterative questions and considerations that focus the synthesis and application of mechanistic information for evidence integration and dose-response analysis

Systematic review step	Mechanistic synthesis triggers and example actions
Human and animal evidence syntheses (see Section 9.1).	• Did the sets of studies report findings that appear to be biologically related to the health effects of interest? *Consider whether these findings might serve as precursors informing an association between exposure and effect; if the set of studies has notable uncertainties (e.g., they are all low confidence), consider a focused analysis of precursors to inform strength of evidence; if the data amenable to dose-response analysis are weak[a] or if responses are observed only at high exposure levels, consider evaluating the precursor data for quantitative analysis.* • Do the results appear to differ by categories that indicate the apparent presence of susceptible populations (e.g., across demographics, species, strains, sexes, or lifestages)? *Consider analyses to better characterize the sources and impact of potential susceptibilities that might be explained by mechanistic information (e.g., due to genetic polymorphisms or metabolic differences).* • Were other key uncertainties or data gaps identified during the analyses of the sets of available human or animal health effect studies? If so, does the literature inventory of mechanistic studies indicate the likelihood of a reasonable number of studies on the topic? *If yes, a focused analysis of these studies could be informative. If no, consider whether an additional focused search of mechanistic information might be worthwhile (i.e., to identify other informative studies not captured by the initial PECO).*
Evidence integration (see Section 10.1): Information relating to biological plausibility	• Are there notable uncertainties in the sets of human or animal health effect studies for which related mechanistic information is available? *An understanding of mechanistic pathways (e.g., by identifying mechanistic precursor events linked qualitatively or quantitatively to apical health effect[s]) can increase the strength of the evidence integration judgments.*

Systematic review step	Mechanistic synthesis triggers and example actions

Evidence integration (see Section 10.2): Considering human relevance of animal findings	• When human evidence is lacking or has results that differ from animal studies, is there evidence that the mechanisms underlying the effects in animals operate in humans? *Analyses of the mechanisms underlying the animal response in relation to those presumed to operate in humans, or the suitability of the animal models to a specific human health outcome, can inform the extent to which the animal response is likely to be directly relevant to humans.*
	
	• The analysis will focus on evaluations of the following issues. The extent of the analysis will vary depending on the impact of the animal evidence to the conclusions.
	
	   o Evidence for a plausible mechanistic pathway or MOA, within which the key events and relationships are evaluated regarding the likelihood of similarities (e.g., in presence or function) across species.
	
	   o Coherence of mechanistic changes observed in exposed humans (or a demonstrated lack of changes that would be expected, e.g., that are known to be linked to the health effect) with animal evidence of mechanistic/toxicological changes.
	
	   o ADME information describing similarities across species, primarily relating to distribution (e.g., to the likely target tissue).

Evidence integration (see Sections 10.2 and 10.3): Characterizing potential susceptible populations or lifestages	• A mechanistic understanding of how a health outcome develops, even without a full MOA, can clarify characteristics of important events (e.g., their presence or sensitivity across lifestages or across genetic variations) and helps identify susceptible populations. • Identification of lifestages or groups likely to be at greatest risk can clarify hazard descriptions and identify key data gaps including whether the most susceptible populations or lifestages have been adequately tested. If a proposed mechanistic pathway or MOA indicates a sensitive population or lifestage in humans, consider whether the appropriate analogous exposures and populations or lifestages were adequately represented in the human or animal database. • When there is evidence of susceptibilities, but specific studies addressing these susceptibilities are unavailable for quantitative analysis, susceptibility data might support refined human variability UFs or probabilistic uncertainty analyses.

Systematic review step	Mechanistic synthesis triggers and example actions

Dose-response analysis (see Section 11):
Biological understanding, including the identification of precursor events
	• A biological understanding of mechanistic events/MOAs, including the identification of precursor events in humans and the exposure conditions expected to result in these effects, can inform the use of
	
	  o particular dose-response models (e.g., models integrating data across several related outcomes or incorporating toxicokinetic knowledge)
	
	  o proximal measures of exposure (e.g., external vs. internal metrics)
	
	  o surrogate endpoints (e.g., use of well-established precursors in lieu of direct observation of apical endpoints)
	
	  o improved characterization of responses (e.g., combination of related outcomes, such as benign and malignant tumors resulting from the same MOA).

PECO = populations, exposures, comparators, and outcomes; MOA = mode of action; UF = uncertainty factor.
[a] Note that "weak" here refers to the study's usability for dose-response analysis specifically. Such studies might be judged to be of *medium* or *high* confidence for the purposes of identifying potential hazards but possess limitations preventing their use for deriving reliable quantitative estimates.
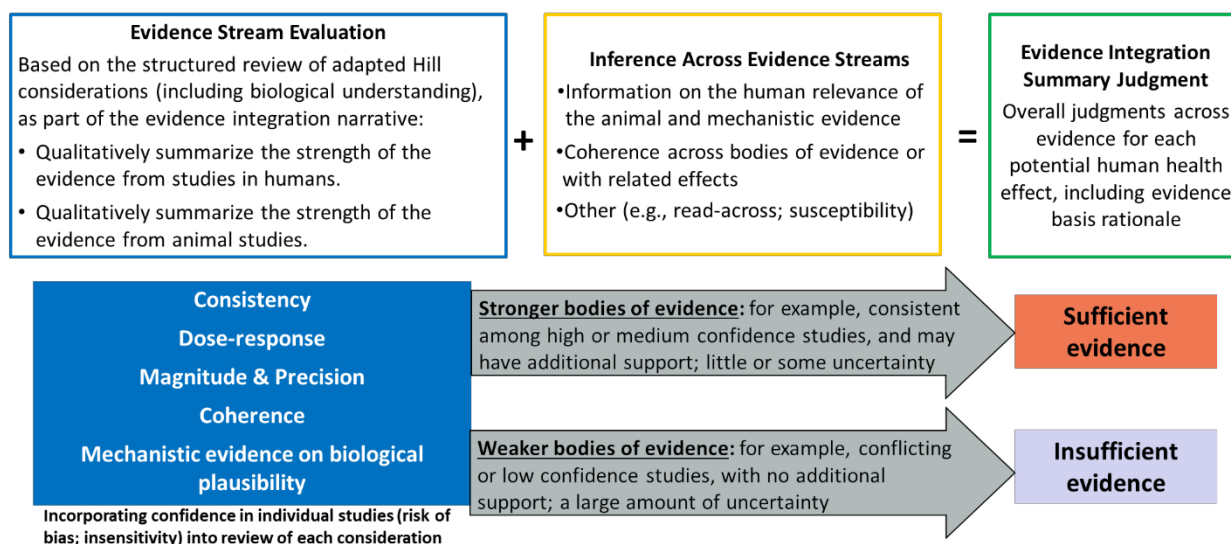
10. EVIDENCE INTEGRATION

1   For the analysis of human health outcomes that might result from chemical exposure, IRIS
2 assessments draw integrated judgments across human, animal, and mechanistic evidence for each
3 assessed health effect (see Section 9). During evidence integration, a structured and documented
4 process will be used, as follows (and depicted in Figure 24):

5   • Building from the separate syntheses of the human and animal evidence (see Section 9.1),
6   the strength of the evidence from the available human and animal health effect studies will
7   be summarized in parallel, but separately, using a structured evaluation of an adapted set of
8   considerations first introduced by Sir Bradford Hill ([Hill, 1965](#)). Table 22 describes these
9   structured evaluations and the explicit consideration of study confidence within each
10   evaluation domain. Based on the approaches and considerations described in Section 9.2,
11   these summaries will incorporate mechanistic evidence (or MOA understanding) that
12   informs the biological plausibility and coherence within the available human or animal
13   health effect studies.

14   • The strength of the animal and human evidence will be considered together in light of
15   inferences across evidence streams. Specifically, the inferences considered during this
16   integration include the human relevance of the animal and mechanistic evidence, coherence
17   across the separate bodies of evidence, and other important information (e.g., judgments
18   regarding susceptibility). Without evidence to the contrary, the human relevance of animal
19   findings is assumed.

20   • A summary judgment is drawn as to whether the available evidence base for each potential
21   human health effect as a whole is sufficient (or insufficient) to indicate that PCB exposure
22   has the potential to be hazardous to humans.[16]

---

[16] Due to the expected rarity of scenarios where there is "sufficient evidence to judge that a hazard is unlikely" (see description in Table 23 and section 10.2) and to improve readability, this judgment is not specified in some instances.

Figure 24. Process for evidence integration. Note that "sufficient evidence" could indicate a judgment of "sufficient evidence for hazard" or "sufficient evidence to judge that a hazard is unlikely", depending on the nature and extent of the available evidence (see Table 23).

1   The decision points within the structured evidence integration process will be summarized
2   in an evidence profile table for each health effect category (see Table 21 for a preliminary template
3   version) in support of the evidence integration narrative. The specific decision frameworks for the
4   structured evaluation of the strength of the human and animal evidence streams and for drawing
5   the overall evidence integration judgment are described in Section 10.1. This process is similar to
6   that used by the *Grading of Recommendations Assessment, Development and Evaluation* (Morgan et
7   al., 2016; Guyatt et al., 2011; Schünemann et al., 2011), which arrives at an overall integration
8   conclusion based on consideration of the body of evidence. As described in Section 9, the human,
9   animal, and mechanistic syntheses serve as inputs providing a foundation for the evidence
10   integration decisions; thus, the major conclusions from these syntheses will be summarized in the
11   evidence profile table (see Table 21 for a preliminary template version) supporting the evidence
12   integration narrative. The evidence profile tables for each potential human health effect evaluated
13   will summarize the judgments and their evidence basis for each step of the structured evidence
14   integration process. Separate sections are included for summarizing the human and animal
15   evidence, for the inference drawn across evidence streams, and for the overall evidence integration
16   judgment. The table presents the key information from the different bodies of evidence that
17   informs each decision.

Table 21. Evidence profile table template

Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of evidence streams	Inferences across evidence streams	Overall evidence integration judgment

Evidence from studies of humans (may be presented by exposure route)[a]				• Human relevance of findings in animals
				• Cross-stream coherence
				• Other inferences:
				  ○ Information on susceptibility
				  ○ MOA analysis inferences
				  ○ Relevant information from other sources (e.g., read across)
					Describe judgment regarding whether there is sufficient (or insufficient) evidence to identify a potential human health hazard, integrating evidence across streams and including a summary of the models and range of dose levels upon which the judgment is primarily reliant.

• References • Study confidence • Study design description	• Consistency • Dose-response gradient • Coherence of observed effects • Effect size • Mechanistic evidence providing plausibility • *Medium* or *high* confidence studies[b]	• Unexplained inconsistency • Imprecision • *Low* confidence studies[b] • Evidence demonstrating implausibility	Qualitative summary of the strength of the evidence from human studies based on the factors at left, including the primary evidence basis and considering: • Results across human epidemiological and controlled exposure studies • Human mechanistic evidence informing biological plausibility (e.g., precursor events linked to adverse outcomes)		
Evidence from animal studies (may be presented by exposure route)[a]					
• References • Study confidence • Study design description	• Consistency or replication • Dose-response gradient • Coherence of observed effects • Effect size • Mechanistic evidence providing plausibility • *Medium* or *high* confidence studies[b]	• Unexplained inconsistency • Imprecision • *Low* confidence studies[b] • Evidence demonstrating implausibility	Qualitative summary of the strength of the evidence for an effect in animals based on the factors at left, including the primary evidence basis and considering: • Results across animal toxicology studies • Animal mechanistic evidence informing biological plausibility (e.g., precursor events linked to adverse outcomes)		

MOA = mode of action.

[a] In addition to exposure route, the summaries of the strength of each evidence stream may include multiple rows (e.g., by study confidence, population, or species) if this informs the analysis of results heterogeneity.

[b] Study confidence, based on evaluation of risk of bias and study sensitivity (see Section 6), and information on susceptibility will be considered when evaluating each of the other factors that increase or decrease strength (e.g., consistency). Notably, lack of findings in studies deemed insensitive neither increases nor decreases strength.

10.1. EVALUATING THE STRENGTH OF THE HUMAN AND ANIMAL EVIDENCE STREAMS

1   As summarized above, prior to drawing overall evidence integration judgments about
2   whether exposure to PCBs has the potential to cause certain health effect(s) in humans given
3   relevant exposure circumstances, the strength of evidence for the available human and animal
4   evidence will be evaluated and summarized. For each assessed health effect or health effect
5   grouping (see Section 5 for examples of the endpoints that will be considered within each health
6   effect category), the relevant mechanistic evidence in exposed humans and animals (or in their
7   cells, relevant new approach methods [NAMs], or in silico models), which will be synthesized based
8   on the approaches and considerations in Section 9.2, will be integrated with the evidence from the
9   available studies of phenotypic effects in humans and animals. The considerations outlined in
10   Table 17 (the different features of the evidence considered and summarized during evidence
11   synthesis; see Section 9) will be evaluated in the context of how they impact judgements of the
12   strength of evidence (see Table 22), which will directly inform the overall evidence integration
13   judgment (see section 10.2). The evaluation of the strength of the human or animal health effects
14   evidence (i.e., based on the considerations in Table 22) will preferably occur at the most specific
15   health outcome level possible (e.g., an analysis at the level of decreased pulmonary function is
16   generally preferable to an analysis of respiratory system effects), if there is an adequate set of
17   studies for analyses at this level and considering the interrelatedness of the available outcomes. If
18   studies on a target system are sparse or varied, or if the interpretation of evidence strength relies
19   largely on the consideration of coherence across related outcomes, then the analyses may need to
20   be conducted at a broader health effect level.   The factors judged to increase or decrease the
21   strength of the evidence will be summarized in tabular format using the evidence profile table
22   template in Table 21 to transparently convey, for each health effect or outcome grouping, expert
23   judgments made throughout the evidence synthesis and integration processes. The evidence
24   profile table allows for consistent documentation of the supporting rationale for each decision.

Table 22. 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 Table 23 (section 10.2) 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.	• An evidence base of mostly *low* confidence studies decreases strength. An exception is an evidence base of studies where the primary issues resulting in *low* confidence are related to insensitivity. This might increase evidence strength in cases where an association is identified because the expected impact of study insensitivity is toward the null. • Decisions to increase strength for other considerations in this table will 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., 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 (conflicting evidence) decreases strength. Generally, strength will not be decreased if discrepant findings can be explained reasonably by factors including study confidence conclusions, variation in population or species, sex, lifestage, exposure patterns (e.g., intermittent or continuous), exposure levels (low or high), exposure duration, or exposure intensity.
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 can also increase strength, even if the magnitude is small. • Precise results from individual studies or across the set of studies increases strength, noting that biological significance is prioritized over statistical significance.	• Strength might 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.

Consideration	Increased evidence strength (of the human or animal evidence)	Decreased evidence strength (of the human or animal evidence)
Biological gradient/ dose-response	• Evidence of dose-response increases strength. Dose-response can be demonstrated across studies or within studies and can be dose or duration dependent. The dose-response need not be monotonic (monotonicity is not necessarily expected, e.g., different outcomes could be expected at low vs. high doses due to 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 might increase strength by increasing certainty in a relationship between exposure and outcome (this is especially useful for interpreting evidence drawn from 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. • If the data are not adequate to evaluate a dose-response pattern, strength is neither increased nor decreased.
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) typically will 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 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 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 is 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; thus, an absence of knowledge will not be used as a basis for decreasing strength (NTP, 2015; NRC, 2014). • Mechanistic evidence in well-conducted studies (see examples of evidence types at left) that demonstrates the health effect(s) are unlikely to occur, or likely to occur only under certain scenarios (e.g., above certain exposure levels), can decrease evidence strength. A decision to decrease depends on an evaluation of the strength of the mechanistic evidence supporting vs. opposing biological plausibility and on the strength of the health effect-specific findings (e.g., stronger health effect data require more certainty in mechanistic evidence opposing plausibility).

MOA = mode of action; NAM = new approach method.

ª Publication bias can 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. If strong evidence of publication bias exists for an outcome, the increase in evidence strength resulting from considering the consistency of the evidence across studies could be reduced.

1   For human and animal evidence, the analyses of each consideration in Table 22 will be used
2   to qualitatively summarize the strength of evidence for the separate evidence streams in the
3   evidence integration narrative. Table 23 provides the criteria that will guide how to develop the
4   judgment for each health effect, and the terms that will be used to summarize those evidence
5   integration judgments.

10.2. OVERALL EVIDENCE INTEGRATION JUDGMENTS

6   Evidence integration combines decisions regarding the strength of the animal and human
7   evidence with considerations regarding mechanistic information on the human relevance of the
8   animal evidence, relevance of the mechanistic evidence to humans (especially in cases where
9   animal evidence is lacking), coherence across bodies of evidence, and information on susceptible
10   populations and lifestages. This evidence integration decision process will culminate in an
11   evidence integration narrative that summarizes the judgments regarding the evidence for each
12   potential health effect evaluated. For each health effect, this narrative will include
13   •   A descriptive summary of the primary judgments about the evidence informing the
14   potential for health effects in exposed humans, based on the following analyses:
15   o   evaluations of the strength of the available human and animal evidence (see Section
16   10.1);
17   o   consideration of the coherence of findings (i.e., the extent to which the evidence for
18   health effects and relevant mechanistic changes are similar) across human and animal
19   studies;
20   o   other information on the human relevance of findings in animals (see Section 9.2); and
21   o   conclusions drawn based on mechanistic analyses (see Section 9.2).
22   •   A summary of key evidence supporting these judgments, highlighting the evidence that was
23   the primary driver of these judgments and any notable issues (e.g., data quality; coherence
24   of the results), and a narrative expression of confidence (a summary of strengths and
25   remaining uncertainties) for these judgments.
26   •   Information on the general conditions of expression of these health effects (e.g., exposure
27   routes and levels in the studies that were the primary drivers of these judgments), noting
28   that these conditions will be clarified during dose response analysis (see Section 11).
29   •   Indications of potentially susceptible populations or lifestages (i.e., an integrated summary
30   of the available evidence on potential susceptible populations and lifestages drawn across
31   the syntheses of the human, animal, and mechanistic evidence).
32   •   A summary of key assumptions used in the analysis, which are generally based on EPA
33   guidelines and which are largely captured in this protocol.
34   •   Strengths and limitations of the evidence integration judgments, including key uncertainties
35   and data gaps, as well as the limitations of the systematic review.

1   In short, the evidence integration narrative will present a qualitative summary of the
2   strength of each evidence stream and an overall judgment across all relevant evidence, with
3   exposure context provided. For each health effect, the first sentence of the evidence integration
4   narrative will include the summary judgment. The assessment will also include evidence profile
5   tables (see Table 21) to support the evidence integration narrative by providing the major
6   decisions and supporting rationale. Table 23 describes the categories of evidence integration
7   judgments that will be used in this assessment and provides examples of database scenarios that fit
8   each category of evidence. These summary judgments provide a succinct and clear representation
9   of the decisions from the more detailed analyses of whether the evidence strength indicates that
10   PCB exposure has the potential to cause the human health effect(s) under specified exposure
11   conditions. Consistent with EPA guidelines, a judgment that the evidence supports an apparent lack
12   of an effect of PCB exposure on the health effect(s) will only be used when the available data are
13   considered robust for deciding that there is no basis for human hazard concern; lesser levels of
14   evidence suggesting a lack of an effect will be characterized as "insufficient."

Table 23. Evidence integration judgments for characterizing potential human health hazards in the evidence integration narrative

Evidence integration judgment[a]	Evidence in studies of humans	Evidence in animal studies	Inferences across evidence streams
Sufficient evidence for hazard	A judgment of *sufficient evidence for hazard* requires that a scenario below is met for either the *evidence in studies of humans* OR *evidence in animal studies*, incorporating the considerations outlined under *inferences across evidence streams*. The scenarios justifying this judgment span a broad range of overall evidence strength, and examples are provided below, starting with the weakest evidence.[b]		
	• 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.[c] • A single *high* or *medium* confidence study without supporting coherent evidence or concern for unexplained inconsistency. Specifically, there are no comparable studies of similar confidence and sensitivity providing conflicting evidence, or the differences can be reasonably explained by, e.g., the populations or exposure levels studied (U.S. EPA, 2005). • Multiple studies showing generally consistent findings, including at least one *high* or *medium* confidence study and supporting evidence, 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). Alternatively, a single *high* or *medium* confidence study with a large magnitude or severity of the effect, a dose-response gradient, or other factors that increase the evidence strength, without serious residual	*Stronger Evidence Stream Scenarios* ↓ • 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.[c] • A single *high* or *medium* confidence experiment in the absence of comparable experiment(s) of similar confidence and sensitivity providing conflicting evidence (i.e., evidence that cannot be reasonably explained, e.g., by respective study designs or differences in animal model (U.S. EPA, 2005)).[d] • 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.[d] 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	• Supplemental evidence (e.g., structure-activity data; chemical class information; other NAMs) is judged to increase the strength of limited or near-equivocal, chemical-specific human or animal evidence to *sufficient evidence for hazard*. • Coherent or biologically consistent findings across evidence streams increases the strength of limited or near-equivocal human or animal evidence (e.g., single or few *high* or *medium* confidence studies with some conflicting evidence) to *sufficient evidence for hazard*. • The strength of the evidence is decreased because mechanistic information (even if it does not provide MOA understanding) raises uncertainties regarding the human and/or animal evidence, but overall the evidence is still considered strong enough to result in a judgment of *sufficient evidence for hazard*. • The strength of the evidence is decreased because findings across evidence streams are conflicting (U.S. EPA, 2005) or biologically inconsistent, but a judgment of *sufficient evidence for hazard* is

	Stronger Evidence Stream Scenarios	
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.[d] • 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 in human studies (coherence) or large estimates of risk or severity of the response, may help to rule out alternative explanations. Similarly, mechanistic evidence from exposed humans may serve to address uncertainties relating to exposure-response, temporality, coherence, and biological plausibility (i.e., providing evidence consistent with an explanation for how exposure could cause the health effect based on current biological knowledge).	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. • 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 (i.e., 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.	supported by review of the adversity and human relevance (prioritizing findings relevant to human toxicity) of the effects. • The strength of the evidence is neither increased nor decreased due to a lack of experimental information on the human relevance of the animal evidence or mechanistic understanding (mechanistic evidence may exist, but it is inconclusive); in these cases, the animal data are judged not to conflict with current biological understanding and thus are assumed to be relevant, while findings in humans and animals are presumed to be real unless proven otherwise. • For the strongest animal evidence, there is mechanistic understanding that the findings are expected to occur and progress in humans. Most notably, an MOA interpreted with reasonable certainty would rule out alternative explanations. • For the strongest evidence, there is adequate testing of potentially susceptible lifestages and populations, based on the effect(s) of interest and chemical knowledge (e.g., toxicokinetics).
Insufficient evidence	A judgment of *insufficient evidence* requires that a scenario below is met for both the *evidence in studies of humans* AND *evidence in animal studies*, incorporating the considerations outlined under *inferences across evidence streams*.	
	• A body of evidence, including scenarios with one or more *high* or *medium* confidence studies reporting an association between exposure and the health outcome, where either (1) conflicting evidence exists in studies of similar confidence and sensitivity[d,e] OR (2) considerable methodological uncertainties remain across the body of evidence (typically related to exposure or outcome	• A body of evidence, including scenarios with one or more *high* or *medium* confidence experiments reporting effects but without supporting coherent evidence that increases the overall evidence strength, where conflicting evidence exists from a set of sensitive experiments of similar or higher confidence (can include mechanistic evidence).[d,e]	• The *evidence in animal studies* meets a scenario for *sufficient evidence for hazard*, but strong experimental evidence (e.g., an MOA interpreted with reasonable certainty) indicates the findings in animals are unlikely to be relevant to humans.

DRAFT—DO NOT CITE OR QUOTE

	ascertainment, including temporality), AND there is no supporting coherent evidence that increases the overall evidence strength. • A set of only *low* confidence studies. • No studies of exposed humans or well-conducted studies of human cells. • A set of largely null studies that does not meet a scenario for *sufficient evidence to judge that a hazard is unlikely*.	• A set of only *low* confidence experiments. • No animal studies or well-conducted studies of animal cells. • The available endpoints are not informative to the hazard question under evaluation. • A set of largely null studies that does not meet the criteria for *sufficient evidence to judge that a hazard is unlikely*.	• The evidence meets a scenario for *sufficient evidence to judge that a hazard is unlikely*, but there is inadequate testing of susceptible populations and lifestages, or data conflict across evidence streams. • The *evidence in animal studies* meets a scenario for *sufficient evidence to judge that a hazard is unlikely*, but the database lacks experimental support that the models are relevant to humans for the effect of interest.
***Sufficient evidence to judge that a hazard is unlikely*[f]**	A judgment of *sufficient evidence to judge that a hazard is unlikely* requires that a scenario below is met for either the *evidence in studies of humans* OR *evidence in animal studies*, incorporating the considerations outlined under *inferences across evidence streams*.		
	• Several *high* confidence studies showing null results (e.g., an odds ratio of 1.0), ruling out alternative explanations including chance, bias, and confounding with reasonable confidence. Each of the studies will have used an optimal outcome and exposure assessment and adequate sample size (specifically for higher exposure groups and for susceptible populations). The overall set will include the full range of levels of exposures that human beings are known to encounter and an evaluation of an exposure-response gradient.	• 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 study 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.	• There is adequate testing of susceptible populations and lifestages. • When the *evidence in animal studies* meets a scenario for this judgment, there is experimental support that the models are relevant to humans for the effect of interest and no conflicting human evidence exists. • When the *evidence in studies of humans* meets a scenario for this judgment and conflicting animal data exist, mechanistic information indicates that the animal data are unlikely to be relevant to humans. • When multiple *high* confidence animal experiments and studies in humans indicate lack of an effect, but the evidence does not meet a scenario for *sufficient evidence to judge that a hazard is unlikely*, strong mechanistic evidence in models relevant to humans supports lack of an effect such that the totality of evidence supports this judgment.

NAM = new approach method.

[a] These categories are based on those indicated for use in hazard characterization from the existing EPA guidelines for noncancer health effects (U.S. EPA, 1998, 1996a, 1991) and, as described in those guidance documents, they depend heavily on expert judgment (note: as applied herein, the process of 'evidence integration' is synonymous with 'weight of evidence'). The evidence integration judgment for each assessed health effect will be included as part of an evidence integration narrative, the specific documentation of the various expert decisions and evidence-based (or default) rationales are summarized in an evidence profile table, and the judgement will be contextualized based on the primary supporting evidence (experimental model or observed population, and exposure levels tested or estimated). Importantly, as discussed in Section 10.1, these judgments may be based on analyses of grouped outcomes at different levels of granularity (e.g., motor activity versus neurobehavioral effects versus nervous system effects) depending on the specifics of the health effect evidence base. Health effects characterized as having *sufficient evidence for hazard* will be evaluated for use in dose-response assessment.

[b] Qualitative descriptions of differences in the strength of the evidence across different health effects judged as having *sufficient evidence for hazard* are useful for other assessment decisions, including prioritizing outcomes in quantitative analyses and characterizing assessment uncertainties. Thus, for all evidence scenarios, but particularly for those in the lower end of this range, it is important to characterize the uncertainties in the evidence base within the evidence integration narrative and to convey the evidence strength to subsequent steps, including toxicity values developed based on those effects. Existing guidance defines the minimum evidence necessary to judge that a health hazard could exist as one adverse endpoint from one well-conducted study (U.S. EPA, 1998); this has been expanded in this table to better incorporate mechanistic evidence, including NAM data.

[c] 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 decision to identify a potential human hazard based on these data is an expert judgment dependent on the state-of-the-science at the time of review.

[d] Scenarios with unexplained heterogeneity across sets of studies with similar confidence and sensitivity can be considered either *sufficient evidence for hazard* or *insufficient evidence*, depending on the expert judgment of the overall weight of 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 evidence integration judgment will be clearly articulated in the evidence integration narrative.

[e] When the database includes at least one well-conducted study and a hazard characterization judgment of *insufficient evidence* is drawn, quantitative analyses may still 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 (2005)). It is critical to transparently convey the extreme uncertainty in any such estimates.

[f] The criteria for this category are intentionally more stringent than those justifying a conclusion of *sufficient evidence for hazard*, consistent with the "difficulty of proving a negative" (as discussed in U.S. EPA (1991), U.S. EPA (1996a), and U.S. EPA (1998)).

10.3. HAZARD CONSIDERATIONS FOR DOSE-RESPONSE

1   This section provides a transition from hazard identification to the dose-response section,
2   highlighting (1) information that will inform the selection of outcomes or broader health effect
3   categories for which toxicity values will be derived, (2) whether toxicity values can be derived to
4   protect specific populations or lifestages, (3) how dose-response modeling will be informed by
5   toxicokinetic information, and (4) information aiding the identification of biologically based
6   benchmark response (BMR) levels. The pool of outcomes and study-specific endpoints will be
7   discussed to identify which categories of effects and study designs are considered the strongest and
8   most appropriate for quantitative assessment of a given health effect. Health effects analyzed
9   relative to exposure levels within or closer to the range of exposures encountered in the
10 environment are particularly informative. When multiple endpoints are available for an
11 organ/system, considerations for characterizing the overall impact on this organ/system will be
12 discussed. For example, if multiple histopathological alterations relevant to changes in liver
13 function are indicated, liver necrosis might be selected as the most representative endpoint to
14 consider for dose-response analysis. This section can review or clarify which endpoints or
15 combination of endpoints in each organ/system characterize the overall effect for dose-response
16 analysis.
17 Biological considerations important for dose-response analysis (e.g., that could help with
18 selection of a BMR) will be discussed. The impact of route of exposure on toxicity to different
19 organs/systems will be examined, if appropriate. The existence and validity of PBPK models or
20 toxicokinetic information that might allow the estimation of internal dose for route-to-route
21 extrapolation will be presented. In addition, mechanistic evidence presented in Section 9 that will
22 influence the dose-response analyses will be highlighted, for example, evidence related to
23 susceptibility or potential shape of the dose-response curve (i.e., linear, nonlinear, threshold
24 model). Mode(s) of action will be summarized including any interactions between them relevant to
25 understanding overall risk.
26 This section will also draw from Sections 9 and 10 to describe the evidence (i.e., human,
27 animal, mechanistic) regarding populations and lifestages susceptible to the hazards identified and
28 factors that increase risk of the hazards. This section will include a discussion of the populations
29 that, in general, could be susceptible to the health effects identified as hazards of exposure to the
30 assessed chemical, even if no specific data on effects of exposure to that chemical in the potentially
31 susceptible population are available. Background information about biological mechanisms or
32 ADME and biochemical and physiological differences among lifestages can be used to guide the
33 selection of populations and lifestages to consider. At a minimum, particular consideration will be
34 given to infants and children, pregnant women, and women of childbearing age. Evidence on
35 factors that contribute to increased responses to chemical exposure in some population groups or
36 factors that contribute to increases in exposure or dose will be summarized and evaluated relative

1   to patterns across studies pertinent to consistency, coherence, and the magnitude and direction of

2   effect measures. Relevant factors could include intrinsic factors (e.g., age, sex, genetics, health

3   status, behaviors); extrinsic factors (e.g., socioeconomic status, access to health care); and

4   differential exposure levels or frequency (e.g., occupation-related exposure, residential proximity to

5   locations with greater exposure intensity).

6   The section will consider options for using data related to susceptible populations to impact

7   dose-response analysis. In particular, an attempt will be made to highlight when it might be

8   possible to develop separate risk estimates for a specific population or lifestage or to determine

9   whether evidence is available to select a data-derived UF.

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

1   The previous sections of this protocol describe how systematic review principles are
2   applied to evaluate study quality (potential bias and sensitivity) and reach evidence integration
3   judgments on health outcomes (or hazard identification) associated with exposure to the chemical
4   of interest. Selection of specific data sets for dose-response assessment and performance of the
5   dose-response assessment is conducted after hazard identification is complete and involves
6   database- and chemical-specific biological judgments. Several EPA guidance and support
7   documents detail data requirements and other considerations for dose-response modeling,
8   especially EPA's *Benchmark Dose Technical Guidance* (U.S. EPA, 2012b) and EPA's *Review of the*
9   *Reference Dose and Reference Concentration Processes* (U.S. EPA, 2005, 2002). This section of the
10 protocol provides an overview of considerations for conducting the dose-response assessment,
11 particularly statistical considerations specific to dose-response analysis that support quantitative
12 risk assessment. Importantly, these considerations do not supersede existing EPA guidance.
13 For IRIS assessments, dose-response assessments are typically performed for both
14 noncancer and cancer hazards, and for both oral and inhalation routes of exposure following
15 chronic exposure[17] to the chemical of interest, if supported by existing data. For noncancer
16 hazards, an RfD and an RfC are usually derived. An RfD or an RfC is an estimate, with uncertainty
17 spanning perhaps an order of magnitude, of an exposure to the human population (including
18 susceptible populations or lifestages) likely to be without an appreciable risk of deleterious health
19 effects over a lifetime (U.S. EPA, 2002). Reference values are not predictive risk values; that is, they
20 provide no information about risks at higher or lower exposure levels.
21 As discussed in Section 2 ("Scoping and Initial Problem Formulation Summary") of this
22 assessment, IRIS will conduct the assessment with a goal of developing oral and inhalation
23 reference values for noncancer toxicity from exposure to complex PCB mixtures. The derivation of
24 noncancer reference values might also depend on the nature of the hazard conclusions. Specifically,
25 for noncancer outcomes, dose-response assessment generally will be conducted when the evidence
26 integration judgments indicate there is "sufficient evidence for hazard", with preference given to
27 health effects with stronger evidence scenarios within that category (Section 10.2), and quantitative
28 analyses generally will not be attempted for "insufficient evidence."

---

[17] Dose-response assessments can also be conducted for shorter durations, particularly if the evidence base
for a chemical indicates risks associated with shorter exposures to the chemical (U.S. EPA, 2002).

11.1. SELECTING STUDIES FOR DOSE-RESPONSE ASSESSMENT

1   The dose-response assessment begins with a review of the important health effects

2 highlighted in the hazard identification step (see Section 10), particularly among the studies of

3 highest quality and studies that exemplify the attributes summarized in Table 24. This review also

4 considers whether opportunities for quantitative evidence integration exist. Examples of

5 quantitative integration, from simplest to more complex, include (1) combining results for an

6 outcome across sex (within a study); (2) characterizing overall toxicity, as in combining effects that

7 comprise a syndrome, or occur on a continuum (e.g., precursors and overt toxicity, benign tumors

8 that progress to malignant tumors); and (3) conducting a meta-analysis or metaregression of all

9 studies addressing a category of important health effects.

10   Some studies used qualitatively for hazard identification might or might not be useful

11 quantitatively for dose-response assessment due to such factors as the lack of quantitative

12 measures of exposure or lack of variability measures for response data. If the needed information

13 cannot be located (see Section 7), semiquantitative analysis might be feasible (e.g., using

14 NOAEL/LOAEL). Studies of low sensitivity might be less useful if they fail to detect a true effect or

15 yield PODs with wide confidence limits, but such studies would be considered for inclusion in a

16 meta-analysis.

17   Among the studies that support the hazard conclusions, those most useful for

18 dose-response analysis generally have at least one exposure level in the region of the

19 dose-response curve near the BMR (the response level to be used for deriving toxicity values), to

20 minimize low-dose extrapolation, and more exposure levels and larger sample sizes overall (U.S.

21 EPA, 2012b). These attributes support a more complete characterization of the shape of the

22 exposure-response curve and decrease the uncertainty in the associated exposure-response metric

23 (e.g., RfC) by reducing statistical uncertainty in the POD and minimizing the need for low-dose

24 extrapolation. In addition to these more general considerations, specific issues that could impact

25 the feasibility of dose-response modeling for individual data sets are described in more detail in the

26 *Benchmark Dose Technical Guidance* (U.S. EPA, 2012b).

Table 24. Attributes used to evaluate studies for derivation of toxicity values

Study attributes		Considerations	
		Human studies	**Animal studies**
Study confidence		*High* or *medium* confidence studies are highly preferred over *low* confidence studies. The available *high* and *medium* confidence studies are further differentiated based on the study attributes below and on a reconsideration of the specific limitations identified and their potential impact on dose-response analyses.	
Rationale for choice of species		Human data are preferred over animal data to eliminate interspecies extrapolation uncertainties (e.g., in toxicodynamics, relevance of specific health outcomes to humans).	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 particular 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, inhalation).	Studies by a route of administration relevant to human environmental exposure are preferred. A validated pharmacokinetic model can also be used to extrapolate across exposure routes.
	Exposure durations	When developing a chronic toxicity value, chronic or subchronic studies are preferred over studies of acute exposure durations. Exceptions exist, such as when a susceptible population or lifestage is more sensitive in a particular time window (e.g., developmental exposure).	
	Exposure levels	Exposures near the range of typical environmental human exposures are preferred. Studies with a broad exposure range and multiple exposure levels are preferred to the extent that they can provide information about the shape of the exposure-response relationship [see the EPA *Benchmark Dose Technical Guidance*; (U.S. EPA, 2012b)] 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 development of a causal effect are preferred. Exposure assessment methods that provide measurements at the level of the individual and that reduce measurement error are preferred. Measurements of exposure should not be influenced by knowledge of health outcome status.	Studies providing actual measurements of exposure (e.g., analytical inhalation concentrations vs. 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.
Measurement of health outcome(s)		Studies that can reliably distinguish the presence or absence (or degree of severity) of the outcome are preferred. Outcome ascertainment methods using generally accepted or standardized approaches are preferred.	
		Studies with individual data are preferred in general. Examples include: to characterize experimental variability more realistically, to characterize overall incidence of individuals affected by related outcomes.	

Study attributes	Considerations	
	Human studies	**Animal studies**
	Among several relevant health outcomes, preference is generally given to those with greater biological significance.	
Study size and design	Preference is given to studies using designs reasonably expected to have power to detect responses of suitable magnitude.[b] This does not mean that studies with substantial responses but low power would be ignored, but that they will be interpreted in light of a confidence interval or variance for the response. Studies that address changes in the number at risk (through decreased survival, loss to follow-up) are preferred.	

[a] An exposure or other variable associated with both exposure and outcome but 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).

11.2. CONDUCTING DOSE-RESPONSE ASSESSMENTS

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

4　　1)　Within the observed dose range, the preferred approach is to use dose-response modeling
5　　　　to incorporate as much of the data set as possible into the analysis. This modeling yields a
6　　　　POD, an exposure level ideally near the lower end of the range of observation, without
7　　　　significant extrapolation to lower exposure levels. See Section 11.2.1 for more details.

8　　2)　Derivation of reference values nearly always involves extrapolation to exposures lower than
9　　　　the POD and is described in more detail in Section 11.2.3.

10　　　　When sufficient and appropriate human data and laboratory animal data are both available
11　for the same outcome, human data are generally preferred for the dose-response assessment
12　because their use eliminates the need to perform interspecies extrapolations.
13　　　　For reference values, IRIS assessments typically derive a candidate value from each suitable
14　data set, whether for human or animal (see Section 11.1). Evaluating these candidate values
15　grouped within a particular organ/system yields a single organ-/system-specific value for each
16　organ/system under consideration. Next, evaluation of these organ-/system-specific values results
17　in the selection of a single overall reference value to cover all health outcomes across all
18　organs/systems. Although this overall reference value is the focus of the assessment, the
19　organ-/system-specific values can be useful for subsequent cumulative risk assessments that
20　consider the combined effect of multiple agents acting on a common organ/system.

11.2.1. Dose-response Analysis in the Range of Observation

21　　　　For conducting a dose-response assessment, toxicodynamic ("biologically based") modeling
22　can be used when data are sufficient to ascertain the MOA and quantitatively support model
23　parameters that represent rates and other quantities associated with the key precursor events of
24　the MOA. Toxicodynamic modeling is potentially the most comprehensive way to account for the
25　biological processes involved in a response. Such models seek to reflect the sequence of key
26　precursor events that lead to a response. Toxicodynamic models can contribute to dose-response
27　assessment by revealing and describing nonlinear relationships between internal dose and
28　response. Such models can provide a useful approach for analysis in the range of observation,
29　provided the purpose of the assessment justifies the effort involved.
30　　　　When a toxicodynamic model is not available for dose-response assessment or when the
31　purpose of the assessment does not warrant developing such a model, empirical modeling will be
32　used to fit the data (on the apical outcome or a key precursor event) in the range of observation.
33　For this purpose, EPA has developed a standard set of models (http://www.epa.gov/ncea/bmds)
34　that can be applied to typical data sets, including those that are linear and nonlinear. When

1   alternative models with significant biological support are available, the decision maker can be

2   informed by the presentation of these alternatives along with the models' strengths and

3   uncertainties. EPA has developed guidance on modeling dose-response data, assessing model fit,

4   selecting suitable models, and reporting modeling results [see the EPA *Benchmark Dose Technical*

5   *Guidance* (U.S. EPA, 2012b)]. Additional judgment or alternative analyses are used if the procedure

6   fails to yield reliable results, for example, if the fit is poor, modeling might be restricted to the lower

7   doses, especially when competing toxicity occurs at higher doses.

8   For each modeled response, a POD from the observed data will be estimated to mark the

9   beginning of extrapolation to lower doses. The POD is an estimated dose (expressed in

10   human-equivalent terms) near the lower end of the observed range without significant

11   extrapolation to lower doses. The POD is used as the starting point for subsequent extrapolations

12   and analyses. For nonlinear extrapolation, the POD is used in calculating an RfD or RfC.

13   Due to the biopersistent nature of many PCB congeners, the relationship between the

14   exposure rate (mg/kg-day administered, absorbed, or inhaled) and the concentration of the

15   congener in the body is a complex function of the exposure rate, exposure duration, and lifestage(s)

16   over which the exposure occurs. Thus, interpretation and comparison of studies with different

17   exposure designs is facilitated by using a pharmacokinetic model that tracks the accumulation of

18   each congener over time and accounts for transfer to offspring during gestation and via lactation

19   for studies that include developmental exposures. Except when comparing studies with otherwise

20   identical exposure designs, comparisons and analyses described in the assessment, including POD

21   estimations, will be based on measures of internal dose averaged over the exposure duration or

22   over the critical window of exposure for the health effect of interest (if known).

23   The response level at which the POD is calculated is guided by the severity of the endpoint.

24   If linear extrapolation is used, selection of a response level corresponding to the POD is not highly

25   influential, so standard values near the low end of the observable range are generally used (for

26   example, 10% extra risk for experimental animal histopathology data, 1% for epidemiological

27   data). Nonlinear approaches account for both statistical and biological considerations. For

28   dichotomous data, a response level of 10% extra risk is generally used for minimally adverse

29   effects, 5% or lower for more severe effects. For continuous data, a response level is ideally based

30   on an established definition of biological significance. In the absence of such definition, one control

31   standard deviation from the control mean is often used for minimally adverse effects, and one-half

32   standard deviation for more severe effects. The POD is the 95% lower bound on the dose

33   associated with the selected response level.

34   EPA has developed standard approaches for determining the relevant dose to be used in the

35   dose-response modeling in the absence of appropriate pharmacokinetic modeling. These standard

36   approaches also facilitate comparison across exposure patterns and species:

37   •   Intermittent study exposures are standardized to a daily average over the duration of

38   exposure. For chronic effects, daily exposures are averaged over the lifespan. Exposures

1   during a critical period, however, are not averaged over a longer duration (U.S. EPA, 2005,
2   1991).

3   • Doses are standardized to equivalent human terms to facilitate comparison of results from
4   different species. Oral doses are scaled allometrically using mg/kg$^{3/4}$-day as the equivalent
5   dose metric across species. Allometric scaling pertains to equivalence across species, not
6   across lifestages, and is not used to scale doses from adult humans or mature animals to
7   infants or children (U.S. EPA, 2011a, 2005). Inhalation exposures are scaled using
8   dosimetry models that apply species-specific physiological and anatomical factors and
9   consider whether the effect occurs at the site of first contact or after systemic circulation
10   (U.S. EPA, 2012a, 1994b).

11   • Converting doses across exposure routes can be informative. If this is done, the assessment
12   describes the underlying data, algorithms, and assumptions (U.S. EPA, 2005).

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

15   These standard approaches will be augmented through the use of pharmacokinetic

16 modeling because of the indirect relationship between exposure and internal concentration or

17 dose, as briefly described above. In particular:

18   • Standardization of intermittent exposures will be conducted by determining the internal
19   concentration averaged over the appropriate time period, or area-under-the-concentration
20   curve for effects that are presumed to result from cumulative damage.

21   • The human equivalent internal dose (concentration) for a given response level is assumed
22   identical to the animal internal dose for exposure over the biologically equivalent period
23   (e.g., human vs. animal gestation). Because data exist to determine the half-lives of
24   congeners in humans independent of animal species, those congener-specific values will be
25   used rather than those derived from allometric scaling. Equivalent human inhalation
26   exposures likewise will be estimated using a human version of a pharmacokinetic model
27   that includes inhalation uptake and exhalation.

28   • Route-to-route extrapolation likewise will be conducted using a pharmacokinetic model
29   capable of describing oral and inhalation (and possibly dermal) exposure.

30   • The pharmacokinetic modeling might use either study-specific intake or body weight or
31   recommended standard values.

11.2.2. Extrapolation: Slope Factors and Unit Risks

32   A cancer assessment is not included in the scope of the current assessment for PCBs.

33 Accordingly, this assessment will not derive an oral slope factor or inhalation unit risk.

11.2.3. Extrapolation: Reference Values

34   Reference value derivation is EPA's most frequently used type of nonlinear extrapolation

35 method. It is most commonly used for noncancer effects.

1   For each data set selected for reference value derivation, reference values are estimated by

2   applying relevant adjustments to the PODs to account for the conditions of the reference value

3   definition—for human variation, extrapolation from animals to humans, extrapolation to chronic

4   exposure duration, and extrapolation to a minimal level of risk (if not observed in the data set).

5   Increasingly, data-based adjustments (U.S. EPA, 2014) and Bayesian methods for characterizing

6   population variability (NRC, 2014) might be feasible and might be distinguished from the UF

7   considerations outlined below. The assessment will discuss the scientific bases for applying these

8   data-based adjustments and UFs:

9   •   *Animal-to-human extrapolation*: If animal results are used to make inferences about
10   humans, the reference value derivation incorporates the potential for cross-species
11   differences, which could arise from differences in toxicokinetics or toxicodynamics. If
12   available, a biologically based model that adjusts fully for toxicokinetic and toxicodynamic
13   differences across species could be used. Otherwise, the POD is standardized to equivalent
14   human terms or is based on pharmacokinetic or dosimetry modeling, that might range from
15   detailed chemical-specific to default approaches (U.S. EPA, 2014, 2011a), and a factor of
16          $10^{1/2}$ (rounded to 3) is applied to account for the remaining uncertainty involving
17   toxicokinetic and toxicodynamic differences.

18   •   *Human variation*: The assessment accounts for variation in susceptibility across the human
19   population and the possibility that the available data might not represent individuals who
20   are most susceptible to the effect, by using a data-based adjustment or UF or a combination
21   of the two. When appropriate data or models for the effect or for characterizing the internal
22   dose are available, the potential for data-based adjustments for toxicodynamics or
23   toxicokinetics is considered (U.S. EPA, 2014, 2002).[18],[19] When sufficient data are available,
24   an intraspecies UF either less than or greater than 10-fold might be justified (U.S. EPA,
25   2002). This factor can be reduced if the POD is derived from or adjusted specifically for
26   susceptible individuals [not for a general population that includes both susceptible and
27   nonsusceptible individuals; (U.S. EPA, 2002, 1998, 1996a, 1994b, 1991)]. When the use of
28   such data or modeling is not supported, a UF with a default value of 10 is considered.

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

---

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

[19] 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.

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

5   •   *Database deficiencies*: In addition to the adjustments above, if database deficiencies raise
6   concern that further studies might identify a more sensitive effect, organ system, or
7   lifestage, the assessment can apply a database UF (U.S. EPA, 2002, 1998, 1996a, 1994b,
8   1991). The size of the factor depends on the nature of the database deficiency. For
9   example, EPA typically follows the recommendation that a factor of 10 be applied if both a
10   prenatal toxicity study and a two-generation reproduction study are missing and a factor of
11        $10^{1/2}$ (i.e., 3) if either is missing (U.S. EPA, 2002).

12   The POD for a reference value is divided by the product of these factors. U.S. EPA (2002)

13 recommends that any composite factor that exceeds 3,000 represents excessive uncertainty, and

14 recommends against relying on the associated reference value. The derivation of oral and

15 inhalation reference values for PCBs conducted as part of the current assessment will be performed

16 consistent with EPA guidance summarized above.

12. PROTOCOL HISTORY

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

REFERENCES

Alford-Stevens, AL. (1986). Analyzing PCBs. Environ Sci Technol 20: 1194-1199.
http://dx.doi.org/10.1021/es00154a001

Alford-Stevens, AL; Budde, WL; Bellar, TA. (1985). Interlaboratory study on determination of polychlorinated biphenyls in environmentally contaminated sediments. Anal Chem 57: 2452-2457. http://dx.doi.org/10.1021/ac00290a007

Arnold, DL; Bryce, F; Karpinski, K; Mes, J; Fernie, S; Tryphonas, H; Truelove, J; Mcguire, PF; Burns, D; Tanner, J. R.; Stapley, R; Zawidzka, ZZ; Basford, D. (1993a). Toxicological consequences of Aroclor 1254 ingestion by female rhesus (Macaca mulatta) monkeys. Part 1B. Prebreeding phase: Clinical and analytical laboratory findings. Food Chem Toxicol 31: 811-824. http://dx.doi.org/10.1016/0278-6915(93)90219-O

Arnold, DL; Bryce, F; Stapley, R; Mcguire, PF; Burns, D; Tanner, J. R.; Karpinski, K. (1993b). Toxicological consequences of Aroclor 1254 ingestion by female rhesus (Macaca mulatta) monkeys. Part 1A. Prebreeding phase: clinical health findings. Food Chem Toxicol 31: 799-810. http://dx.doi.org/10.1016/0278-6915(93)90218-N

ATSDR. (2000). Toxicological profile for polychlorinated biphenyls (PCBs) [ATSDR Tox Profile]. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service. http://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=142&tid=26

ATSDR. (2011). Addendum to the toxicological profile for polychlorinated biphenyls. Atlanta, GA.

Barsotti, DA; van Miller, JP. (1984). Accumulation of a commercial polychlorinated biphenyl mixture (Aroclor 1016) in adult rhesus monkeys and their nursing infants. Toxicology 30: 31-44.

Catlin, NR; Collins, BJ; Auerbach, SS; Ferguson, SS; Harnly, JM; Gennings, C; Waidyanatha, S; Rice, GE; Smith-Roe, SL; Witt, KL; Rider, CV. (2018). How similar is similar enough? A sufficient similarity case study with Ginkgo biloba extract. Food Chem Toxicol 118: 328-339. http://dx.doi.org/10.1016/j.fct.2018.05.013

CDC. (2009). Fourth national report on human exposure to environmental chemicals. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. http://www.cdc.gov/exposurereport/

Crissman, JW; Goodman, DG; Hildebrandt, PK; Maronpot, RR; Prater, DA; Riley, JH; Seaman, WJ; Thake, DC. (2004). Best practices guideline: Toxicologic histopathology. Toxicol Pathol 32: 126-131. http://dx.doi.org/10.1080/01926230490268756

Devoto, E; Fiore, BJ; Millikan, R; Anderson, HA; Sheldon, L; Sonzogni, WC; Longnecker, MP. (1997). Correlations among human blood levels of specific PCB congeners and implications for epidemiologic studies. Am J Ind Med 32: 606-613.

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

EFSA. (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

Erickson, MD. (2018). Aroclor misidentification in environmental samples: how do we communicate more effectively between the laboratory and the data user? Environ Sci Pollut Res Int 25: 16291-16299. http://dx.doi.org/10.1007/s11356-017-0160-6

Feder, PI; Ma, ZJ; Bull, RJ; Teuschler, LK; Schenck, KM; Simmons, JE; Rice, G. (2009). Evaluating sufficient similarity for disinfection by-product (DBP) mixtures: multivariate statistical procedures [Review]. J Toxicol Environ Health A 72: 468-481. http://dx.doi.org/10.1080/15287390802608965

Foster, P; Gray, LE, Jr. (2013). Toxic responses of the reproductive system. In CD Klaassen (Ed.),
  (8th ed., pp. 861-906). New York, NY: McGraw-Hill Education.
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
Harrad, S; Ibarra, C; Robson, M; Melymuk, L; Zhang, X; Diamond, M; Douwes, J. (2009).
  Polychlorinated biphenyls in domestic dust from Canada, New Zealand, United Kingdom
  and United States: Implications for human exposure. Chemosphere 76: 232-238.
  http://dx.doi.org/10.1016/j.chemosphere.2009.03.020
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.
IARC. (2015). Polychlorinated and polybrominated biphenyls: Volume 107. In IARC Monographs on
  the Evaluation of Carcinogenic Risks to Humans. Geneva, Switzerland: World Health
  Organization. http://monographs.iarc.fr/ENG/Monographs/vol107/index.php
ICF. (2018). DRAGON: An online tool for systematic review [Website].
  https://www.icf.com/resources/solutions-and-apps/dragon-online-tool-systematic-review
Kodavanti, PRS; Kannan, N; Yamashita, N; Derr-Yellin, EC; Ward, TR; Burgin, DE; Tilson, HA;
  Birnbaum, LS. (2001). Differential effects of two lots of Aroclor 1254: Congener-specific
  analysis and neurochemical end points. Environ Health Perspect 109: 1153-1161.
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
Lehmann, GM; Christensen, K; Maddaloni, M; Phillips, LJ. (2015). Evaluating health risks from
  inhaled polychlorinated biphenyls: research needs for addressing uncertainty. Environ
  Health Perspect 123: 109-113. http://dx.doi.org/10.1289/ehp.1408564
Levin, ED; Schantz, SL; Bowman, RE. (1988). Delayed spatial alternation deficits resulting from
  perinatal PCB exposure in monkeys. Arch Toxicol 62: 267-273.
  http://dx.doi.org/10.1007/BF00332486
Marek, RF; Thorne, PS; Herkert, NJ; Awad, AM; Hornbuckle, KC. (2017). Airborne PCBs and OH-PCBs
  inside and outside urban and rural U.S. schools. Environ Sci Technol 51: 7853-7860.
  http://dx.doi.org/10.1021/acs.est.7b01910
Marshall, S; Gennings, C; Teuschler, LK; Stork, LG; Tornero-Velez, R; Crofton, KM; Rice, GE. (2013).
  An empirical approach to sufficient similarity: combining exposure data and mixtures
  toxicology data. Risk Anal 33: 1582-1595. http://dx.doi.org/10.1111/risa.12015
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
NIOSH. (2019a). NIOSH pocket guide to chemical hazards. RTECS. Chlorodiphenyl (42% chlorine).
  http://www.cdc.gov/niosh/npg/npgd0125.html
NIOSH. (2019b). NIOSH pocket guide to chemical hazards. RTECS. Chlorodiphenyl (54% chlorine).
  http://www.cdc.gov/niosh/npg/npgd0126.html

1 NRC. (2014). Review of EPA's Integrated Risk Information System (IRIS) process. Washington, DC:
2   The National Academies Press. http://www.nap.edu/catalog.php?record_id=18764
3 NTP. (2011). Specifications for the conduct of studies to evaluate the reproductive and
4   developmental toxicity of chemical, biological and physical agents in laboratory animals.
5   https://ntp.niehs.nih.gov/ntp/test_info/finalntp_reprospecsmay2011_508.pdf
6 NTP. (2015). Handbook for conducting a literature-based health assessment using OHAT approach
7   for systematic review and evidence integration. U.S. Dept. of Health and Human Services,
8   National Toxicology Program.
9   https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf
10 O'Brien, KM; Upson, K; Cook, NR; Weinberg, CR. (2016). Environmental chemicals in urine and
11   blood: Improving methods for creatinine and lipid adjustment. Environ Health Perspect
12   124: 220-227. http://dx.doi.org/10.1289/ehp.1509693
13 OSHA. (2018). Chemical sampling information, chlorodiphenyl (54% Cl).
14   https://www.osha.gov/chemicaldata/chemResult.html?recNo=121
15 OSHA. (2019). Chemical sampling information, chlorodiphenyl (42% Cl).
16   https://www.osha.gov/chemicaldata/chemResult.html?recNo=392
17 Rice, DC. (1997). Effect of postnatal exposure to a PCB mixture in monkeys on multiple fixed
18   interval-fixed ratio performance. Neurotoxicol Teratol 19: 429-434.
19 Rice, DC. (1998). Effects of postnatal exposure of monkeys to a PCB mixture on spatial
20   discrimination reversal and DRL performance. Neurotoxicol Teratol 20: 391-400.
21 Rice, DC. (1999). Effect of exposure to 3,3',4,4',5-pentachlorobiphenyl (PCB 126) throughout
22   gestation and lactation on development and spatial delayed alternation performance in rats.
23   Neurotoxicol Teratol 21: 59-69. http://dx.doi.org/10.1016/S0892-0362(98)00031-2
24 Rice, DC; Hayward, S. (1997). Effects of postnatal exposure to a PCB mixture in monkeys on
25   nonspatial discrimination reversal and delayed alternation performance. Neurotoxicology
26   18: 479-494.
27 Rice, DC; Hayward, S. (1999). Effects of exposure to 3,3',4,4',5-pentachlorobiphenyl (PCB 126)
28   throughout gestation and lactation on behavior (concurrent random interval-random
29   interval and progressive ratio performance) in rats. Neurotoxicol Teratol 21: 679-687.
30   http://dx.doi.org/10.1016/S0892-0362(99)00021-5
31 Rice, GE; Ingvar, E; Feder, EI; Gennings, C. (2018). Assessing human health risks using information
32   on whole mixtures. In C Rider; J Simmons (Eds.), (pp. 421-463). Cham: Springer.
33   http://dx.doi.org/10.1007/978-3-319-56234-6_15
34 Schantz, SL; Levin, ED; Bowman, RE. (1991). Long-term neurobehavioral effects of perinatal
35   polychlorinated biphenyl (PCB) exposure in monkeys. Environ Toxicol Chem 10: 747-756.
36   http://dx.doi.org/10.1002/etc.5620100606
37 Schantz, SL; Levin, ED; Bowman, RE; Heironimus, MP; Laughlin, NK. (1989). Effects of perinatal PCB
38   exposure on discrimination-reversal learning in monkeys. Neurotoxicol Teratol 11: 243-
39   250. http://dx.doi.org/10.1016/0892-0362(89)90066-4
40 Schisterman, EF; Whitcomb, BW; Louis, GM; Louis, TA. (2005). Lipid adjustment in the analysis of
41   environmental contaminants and human health risks. Environ Health Perspect 113: 853-
42   857. http://dx.doi.org/10.1289/ehp.7640
43 Schünemann, H; Hill, S; Guyatt, G; Akl, EA; Ahmed, F. (2011). The GRADE approach and Bradford
44   Hill's criteria for causation. J Epidemiol Community Health 65: 392-395.
45   http://dx.doi.org/10.1136/jech.2010.119933
46 Smith, MT; Guyton, KZ; Gibbons, CF; Fritz, JM; Portier, CJ; Rusyn, I; DeMarini, DM; Caldwell, JC;
47   Kavlock, RJ; Lambert, PF; Hecht, SS; Bucher, JR; Stewart, BW; Baan, RA; Cogliano, VJ; Straif,
48   K. (2016). Key characteristics of carcinogens as a basis for organizing data on mechanisms
49   of carcinogenesis [Review]. Environ Health Perspect 124: 713-721.
50   http://dx.doi.org/10.1289/ehp.1509912

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.

Thomas, K; Xue, J; Williams, R; Jones, P; Whitaker, D. (2012). Polychlorinated biphenyls (PCBs) in school buildings: Sources, environmental levels, and exposure [EPA Report]. (EPA/600/R-12/051). U.S. Environmental Protection Agency. https://www.epa.gov/sites/production/files/2015-08/documents/pcb_epa600r12051_final.pdf

ToxNet Hazardous Substances Data, B. (2011). HSDB: Polychlorinated biphenyls [Website]. https://toxnet.nlm.nih.gov/cgi-bin/sis/search/a?dbs+hsdb:@term+@DOCNO+3945

Tryphonas, H; Hayward, S; O'Grady, L; Loo, JC; Arnold, DL; Bryce, F; Zawidzka, ZZ. (1989). Immunotoxicity studies of PCB (Aroclor 1254) in the adult rhesus (Macaca mulatta) monkey--preliminary report. International Journal of Immunopharmacology 11: 199-206. http://dx.doi.org/10.1016/0192-0561(89)90072-6

Tryphonas, H; Luster, MI; Schiffman, G; Dawson, LL; Hodgen, M; Germolec, D; Hayward, S; Bryce, F; Loo, JC; Mandy, F. (1991a). Effect of chronic exposure of PCB (Aroclor 1254) on specific and nonspecific immune parameters in the rhesus (Macaca mulatta) monkey. Fundam Appl Toxicol 16: 773-786. http://dx.doi.org/10.1093/toxsci/16.4.773

Tryphonas, H; Luster, MI; White, KL, Jr.; Naylor, PH; Erdos, MR; Burleson, GR; Germolec, D; Hodgen, M; Hayward, S; Arnold, DL. (1991b). Effects of PCB (Aroclor 1254) on non-specific immune parameters in rhesus (Macaca mulatta) monkeys. International Journal of Immunopharmacology 13: 639-648. http://dx.doi.org/10.1016/0192-0561(91)90176-8

U.S. EPA. (1988). 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. (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. (1993). IRIS chemical assessment summary for aroclor 1016 (CASRN 12674-11-2). https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=462

U.S. EPA. (1994a). IRIS chemical assessment summary for aroclor 1254 (CASRN 11097-69-1). https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=389

U.S. EPA. (1994b). 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. (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. (1996b). PCBs: Cancer dose-response assessment and application to environmental mixtures [EPA Report]. (EPA/600/P-96/001). Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=12486

1   U.S. EPA. (1998). Guidelines for neurotoxicity risk assessment [EPA Report] (pp. 1-89).
2   (EPA/630/R-95/001F). Washington, DC: U.S. Environmental Protection Agency, Risk
3   Assessment Forum. http://www.epa.gov/risk/guidelines-neurotoxicity-risk-assessment
4   U.S. EPA. (2000). Supplementary guidance for conducting health risk assessment of chemical
5   mixtures (pp. 1-209). (EPA630R00002). Washington, DC: U.S. Environmental Protection
6   Agency, Risk Assessment Forum.
7   http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=20533
8   U.S. EPA. (2002). A review of the reference dose and reference concentration processes.
9   (EPA/630/P-02/002F). Washington, DC: U.S. Environmental Protection Agency, Risk
10   Assessment Forum. https://www.epa.gov/sites/production/files/2014-12/documents/rfd-
11   final.pdf
12 U.S. EPA. (2004). Toxicological review of boron and compounds. In support of summary
13   information on the Integrated Risk Information System (IRIS) [EPA Report].
14   (EPA/635/04/052). Washington, DC: U.S. Environmental Protection Agency, IRIS.
15   http://nepis.epa.gov/exe/ZyPURL.cgi?Dockey=P1006CK9.txt
16 U.S. EPA. (2005). Guidelines for carcinogen risk assessment [EPA Report] (pp. 1-166). (EPA/630/P-
17   03/001F). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum.
18   https://www.epa.gov/sites/production/files/2013-
19   09/documents/cancer_guidelines_final_3-25-05.pdf
20 U.S. EPA. (2011a). Recommended use of body weight 3/4 as the default method in derivation of the
21   oral reference dose (pp. 1-50). (EPA/100/R-11/0001). Washington, DC: U.S. Environmental
22   Protection Agency, Risk Assessment Forum, Office of the Science Advisor.
23   https://www.epa.gov/risk/recommended-use-body-weight-34-default-method-derivation-
24   oral-reference-dose
25 U.S. EPA. (2011b). Toxicological review of trichloroethylene (CASRN 79-01-6) in support of
26   summary information on the Integrated Risk Information System (IRIS) [EPA Report].
27   (EPA/635/R-09/011F). Washington, DC.
28   https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0199tr/0199tr.p
29   df
30 U.S. EPA. (2012a). Advances in inhalation gas dosimetry for derivation of a reference concentration
31   (RfC) and use in risk assessment (pp. 1-140). (EPA/600/R-12/044). Washington, DC.
32   https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=244650&CFID=50524762&CFTOK
33   EN=17139189
34 U.S. EPA. (2012b). Benchmark dose technical guidance. (EPA/100/R-12/001). Washington, DC: U.S.
35   Environmental Protection Agency, Risk Assessment Forum.
36   https://www.epa.gov/risk/benchmark-dose-technical-guidance
37 U.S. EPA. (2014). Guidance for applying quantitative data to develop data-derived extrapolation
38   factors for interspecies and intraspecies extrapolation [EPA Report]. (EPA/100/R-
39   14/002F). Washington, DC: Risk Assessment Forum, Office of the Science Advisor.
40   https://www.epa.gov/sites/production/files/2015-01/documents/ddef-final.pdf
41 U.S. EPA. (2015a). Science policy council peer review handbook [EPA Report] (4th ed.).
42   (EPA/100/B-15/001). Washington, DC: U.S. Environmental Protection Agency, Science
43   Policy Council. https://www.epa.gov/osa/peer-review-handbook-4th-edition-2015
44 U.S. EPA. (2015b). Scoping and problem formulation for the toxicological review of polychlorinated
45   biphenyls (PCBS): Effects other than cancer [CASRN 1336-36-3] (Draft). (EPA/635/R-
46   14/198). Washington, DC. http://ofmpub.epa.gov/eims/eimscomm.getfile?
47   p_download_id=524917
48 U.S. EPA. (2017). Guidance to assist interested persons in developing and submitting draft risk
49   evaluations under the Toxic Substances Control Act. (EPA/740/R17/001). Washington, DC:
50   U.S Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention.

1   https://www.epa.gov/sites/production/files/2017-
2   06/documents/tsca_ra_guidance_final.pdf
3 U.S. EPA. (2018). An umbrella Quality Assurance Project Plan (QAPP) for PBPK models [EPA
4   Report]. (ORD QAPP ID No: B-0030740-QP-1-1). Research Triangle Park, NC.
5 Varghese, A; Cawley, M; Hong, T. (2017). Supervised clustering for automated document
6   classification and prioritization: A case study using toxicological abstracts. Environ Syst
7   Decis 38: 398-414. http://dx.doi.org/10.1007/s10669-017-9670-5
8 Varghese, A; Hong, T; Hunter, C; Agyeman-Badu, G; Cawley, M. (2019). Active learning in automated
9   text classification: a case study exploring bias in predicted model performance metrics.
10   Environ Syst Decis1-12. http://dx.doi.org/10.1007/s10669-019-09717-3
11 Vesterinen, HM; Sena, ES; Egan, KJ; Hirst, TC; Churolov, L; Currie, GL; Antonic, A; Howells, DW;
12   Macleod, MR. (2014). Meta-analysis of data from animal studies: a practical guide. J
13   Neurosci Methods 221: 92-102. http://dx.doi.org/10.1016/j.jneumeth.2013.09.010
14 Vorkamp, K. (2015). An overlooked environmental issue? A review of the inadvertent formation of
15   PCB-11 and other PCB congeners and their occurrence in consumer products and in the
16   environment [Review]. Sci Total Environ 541: 1463-1476.
17   http://dx.doi.org/10.1016/j.scitotenv.2015.10.019
18 WHO. (2003). Polychlorinated biphenyls: Human health aspects. (Concise International Chemical
19   Assessment Document 55). Geneva.
20 Wolff, MS; Camann, D; Gammon, M; Stellman, SD. (1997). Proposed PCB congener groupings for
21   epidemiological studies [Letter]. Environ Health Perspect 105: 13-14.
22 Wolff, MS; Toniolo, PG. (1995). Environmental organochlorine exposure as a potential etiologic
23   factor in breast cancer. Environ Health Perspect 103: 141-145.
24 Xue, J; Liu, SV; Zartarian, VG; Geller, AM; Schultz, BD. (2014). Analysis of NHANES measured blood
25   PCBs in the general US population and application of SHEDS model to identify key exposure
26   factors. J Expo Sci Environ Epidemiol 24: 615-621. http://dx.doi.org/10.1038/jes.2013.91

27

<div style="background-color:#d9e5c0; padding:10px;">

APPENDICES

</div>

APPENDIX A. ELECTRONIC DATABASE SEARCH STRATEGIES

Table A-1. Database search strategy

Search	Search strategy	Date and results
Pub Med		
Chemical terms	((((pcb[tw] OR pcb's OR "pcbs") NOT ("printed circuit board" OR "printed circuit boards")) OR "polychlorinated biphenyl" OR "polychlorinated biphenyls" OR "aroclor" OR "aroclors" OR "arochlor" OR "arochlors" OR "chlophen" OR "chlophens" OR "chlorinated biphenyl" OR "chlorinated biphenyls" OR "chlorinated diphenyl" OR "chloro biphenyl" OR "chloro biphenyls" OR clophen OR "clophens" OR fenclor OR "fenclors" OR inerteen OR "inerteens" OR kanechlor OR "kanechlors" OR phenochlor OR "phenochlors" OR phenoclor OR "phenoclors" OR "polychlorobiphenyl" OR polychlorobiphenyls OR "pyralene" OR "pyranol" OR "sovol"[tw] OR "sovols" OR therminol OR "therminols" OR "polychloro biphenyl" OR "polychloro biphenyls" OR "polychlorodiphenyls" OR "polychlorinated diphenyls" OR delor[tw] OR delors[tw] OR chlorofen OR "chlorofens" OR monochlorobiphenyl OR monochlorobiphenyls OR chlorobiphenyl OR "chlorobiphenyls" OR chlorodiphenyl OR "chlorodiphenyls" OR monochlorodiphenyl OR "monochlorodiphenyls" OR dichlorobiphenyl OR dichlorobiphenyls OR dichlorodiphenyl OR "dichlorodiphenyls" OR "bichlorobiphenyl" OR "bichlorobiphenyls" OR trichlorobiphenyl OR trichlorobiphenyls OR trichlorodiphenyl OR "trichlorodiphenyls" OR tetrachlorobiphenyl OR tetrachlorobiphenyls OR tetrachlorodiphenyl OR "tetrachlorodiphenyls" OR "tetrachloro biphenyl" OR pentachlorobiphenyl OR pentachlorobiphenyls OR pentachlorodiphenyl OR "pentachlorodiphenyls" OR "pentachloro biphenyl" OR hexachlorobiphenyl OR hexachlorobiphenyls OR "hexachloro biphenyl" OR "hexachloro biphenyls" OR heptachlorobiphenyl OR heptachlorobiphenyls OR octachlorobiphenyl OR octachlorobiphenyls OR nonachlorobiphenyl OR nonachlorobiphenyls OR decachlorobiphenyl OR decachlorobiphenyls)) or "capacitor manufacturing workers" or "Yu-Cheng" or Yucheng or Yusho or "polychlorinated-biphenyls" or "Lake Michigan fish" or "North Carolina Breast Milk and Formula Project" or "Great Lakes fish" or "Lake Ontario fish" or "European Background PCB Study" or "Great Lakes Consortium" or "New York State Angler Cohort" or "Lake Michigan Aging Population Study" or "Michigan Anglers Cohort" or "Michigan Long-Term PBB Study"	7/29/2015: 19,089 8/31/2016: 388
Health effect terms	Not applicable	
Other concepts	Health effect literature was prioritized using supervised clustering and machine learning (DoCTOR). A total of 487 health effect-related publications derived from the 2012 *Toxicological Review of Polychlorinated Biphenyls (PCBs): Effects Other Than Cancer* (EPA/635/R-11/079C) were used as seed studies for clustering and machine learning.	
Web of Science		
Chemical terms	((((TS="pcb" OR TS="pcbs") NOT (TS="printed circuit board" OR TS="printed circuit boards")) OR TS="polychlorinated biphenyl" OR TS="polychlorinated biphenyls" OR TS="aroclor" OR TS="aroclors" OR TS="arochlor" OR TS="arochlors" OR	7/29/15: 35,962 8/31/2016: 1,739

Search	Search strategy	Date and results
	TS="chlophen" OR TS="chlophens" OR TS="chlorinated biphenyl" OR TS="chlorinated biphenyls" OR TS="chlorinated diphenyl" OR TS="chlorinated diphenyls" OR TS="chloro biphenyl" OR TS="chloro biphenyls" OR TS="clophen" OR TS="clophens" OR TS="fenclor" OR TS="fenclors" OR TS="inerteen" OR TS="inerteens" OR TS="kanechlor" OR TS="kanechlors" OR TS="phenochlor" OR TS="phenochlors" OR TS="phenoclor" OR TS="phenoclors" OR TS="polychlorobiphenyl" OR TS="polychlorobiphenyls" OR TS="pyralene" OR TS="pyranol" OR TS="sovol" OR TS="sovols" OR TS="therminol" OR TS="therminols" OR TS="polychloro biphenyl" OR TS="polychloro biphenyls" OR TS="polychlorodiphenyls" OR TS="polychlorinated diphenyl" OR TS="polychlorinated diphenyls" OR TS="delor" OR TS="delors" OR TS="chlorofen" OR TS="chlorofens" OR TS="monochlorobiphenyl" OR TS="monochlorobiphenyls" OR TS="chlorobiphenyl" OR TS="chlorobiphenyls" OR TS="chlorodiphenyl" OR TS="chlorodiphenyls" OR TS="monochlorodiphenyl" OR TS="monochlorodiphenyls" OR TS="monochloro biphenyl" OR TS="monochloro biphenyls" OR TS="dichlorobiphenyl" OR TS="dichlorobiphenyls" OR TS="dichlorodiphenyl" OR TS="dichlorodiphenyls" OR TS="dichloro biphenyl" OR TS="dichloro biphenyls" OR TS="bichlorobiphenyl" OR TS="bichlorobiphenyls" OR TS="trichlorobiphenyl" OR TS="trichlorobiphenyls" OR TS="trichlorodiphenyl" OR TS="trichlorodiphenyls" OR TS="trichloro biphenyl" OR TS="trichloro biphenyls" OR TS="tetrachlorobiphenyl" OR TS="tetrachlorobiphenyls" OR TS="tetrachlorodiphenyl" OR TS="tetrachlorodiphenyls" OR TS="tetrachloro biphenyl" OR TS="tetrachloro biphenyls" OR TS="pentachlorobiphenyl" OR TS="pentachlorobiphenyls" OR TS="pentachlorodiphenyl" OR TS="pentachlorodiphenyls" OR TS="pentachloro biphenyl" OR TS="pentachloro biphenyls" OR TS="hexachlorobiphenyl" OR TS="hexachlorobiphenyls" OR TS="hexachloro biphenyl" OR TS="hexachloro biphenyls" OR TS="heptachlorobiphenyl" OR TS="heptachlorobiphenyls" OR TS="heptachloro biphenyl" OR TS="heptachloro biphenyls" OR TS="octachlorobiphenyl" OR TS="octachlorobiphenyls" OR TS="octachloro biphenyl" OR TS="octachloro biphenyls" OR TS="nonachlorobiphenyl" OR TS="nonachlorobiphenyls" OR TS="nonachloro biphenyl" OR TS="nonachloro biphenyls" OR TS="decachlorobiphenyl" OR TS="decachlorobiphenyls" OR TS="decachloro biphenyl" OR TS="decachloro biphenyls"))	
Health effect terms	Not applicable	
Other concepts	Health effect literature was prioritized using supervised clustering and machine learning (DoCTOR). A total of 487 health effect-related publications cited in the 2012 *Toxicological Review Of Polychlorinated Biphenyls (PCBs): Effects Other Than Cancer* (EPA/635/R-11/079C) were used as seed studies for clustering and machine learning.	
Toxline		
Chemical terms	@OR+(@TERM+@rn+1336-36-3+@TERM+@rn+12767-79-2+@TERM+@rn+64093-59-0+@TERM+@rn+12521-86-7+@TERM+@rn+55945-68-1+@TERM+@rn+62251-11-0+@TERM+@rn+37353-77-8+@TERM+@rn+11137-46-5+@TERM+@rn+27323-18-8+@TERM+@rn+25512-42-9+@TERM+@rn+25323-68-6+@TERM+@rn+26914-33-0+@TERM+@rn+25429-29-2+@TERM+@rn+26601-64-9+@TERM+@rn+28655-71-2+@TERM+@rn+55722-26-4+@TERM+@rn+53742-07-7+@TERM+@rn+2051-24-3)+@NOT+@org+"nih+reporter"	7/29/15: 30,443 8/31/2016: 0
	@OR+(@TERM+@rn+38444-93-8+@TERM+@rn+52663-59-9+@TERM+@rn+36559-22-5+@TERM+@rn+70362-46-8+@TERM+@rn+41464-39-5+@TERM+@rn+70362-45-7+@TERM+@rn+41464-47-5+@TERM+@rn+2437-79-8+@TERM+@rn+70362-47-9+@TERM+@rn+41464-40-8+@TERM+@rn+62796-65-0+@TERM+@rn+68194-04-7+@TERM+@rn+35693-99-3+@TERM+@rn+41464-41-9+@TERM+@rn+15968-05-5+@TERM+@rn+74338-24-2+@TERM+@rn+41464-43-	

Search	Search strategy	Date and results
	1+@TERM+@rn+70424-67-8+@TERM+@rn+41464-49-7+@TERM+@rn+74472-33-6+@TERM+@rn+33025-41-1+@TERM+@rn+33284-53-6+@TERM+@rn+54230-22-7+@TERM+@rn+74472-34-7+@TERM+@rn+52663-58-8+@TERM+@rn+33284-54-7+@TERM+@rn+32598-10-0+@TERM+@rn+73575-53-8+@TERM+@rn+73575-52-7+@TERM+@rn+60233-24-1+@TERM+@rn+32598-11-1+@TERM+@rn+41464-46-4+@TERM+@rn+41464-42-0+@TERM+@rn+74338-23-1+@TERM+@rn+32690-93-0+@TERM+@rn+32598-12-2+@TERM+@rn+70362-48-0+@TERM+@rn+32598-13-3+@TERM+@rn+70362-49-1+@TERM+@rn+41464-48-6+@TERM+@rn+33284-52-5+@TERM+@rn+70362-50-4)+@NOT+@org+"nih+reporter"	
	@OR+(@TERM+@rn+2051-60-7 +@TERM+@rn+2051-61-8+@TERM+@rn+2051-62-9+@TERM+@rn+13029-08-8+@TERM+@rn+16605-91-7+@TERM+@rn+25569-80-6+@TERM+@rn+33284-50-3+@TERM+@rn+34883-43-7+@TERM+@rn+34883-39-1+@TERM+@rn+33146-45-1+@TERM+@rn+2050-67-1+@TERM+@rn+2974-92-7+@TERM+@rn+2974-90-5+@TERM+@rn+34883-41-5+@TERM+@rn+2050-68-2+@TERM+@rn+38444-78-9+@TERM+@rn+37680-66-3+@TERM+@rn+37680-65-2+@TERM+@rn+38444-73-4+@TERM+@rn+38444-84-7+@TERM+@rn+55702-46-0+@TERM+@rn+38444-85-8+@TERM+@rn+55720-44-0+@TERM+@rn+55702-45-9+@TERM+@rn+55712-37-3+@TERM+@rn+38444-81-4+@TERM+@rn+38444-76-7+@TERM+@rn+7012-37-5+@TERM+@rn+15862-07-4+@TERM+@rn+35693-92-6+@TERM+@rn+16606-02-3+@TERM+@rn+38444-77-8+@TERM+@rn+38444-86-9+@TERM+@rn+37680-68-5+@TERM+@rn+37680-69-6+@TERM+@rn+38444-87-0+@TERM+@rn+38444-90-5+@TERM+@rn+53555-66-1+@TERM+@rn+38444-88-1)+@NOT+@org+"nih+reporter"	
	@OR+(@TERM+@rn+38380-07-3+@TERM+@rn+55215-18-4+@TERM+@rn+52663-66-8+@TERM+@rn+61798-70-7+@TERM+@rn+38380-05-1+@TERM+@rn+35694-04-3+@TERM+@rn+52704-70-8+@TERM+@rn+52744-13-5+@TERM+@rn+38411-22-2+@TERM+@rn+35694-06-5+@TERM+@rn+35065-28-2+@TERM+@rn+56030-56-9+@TERM+@rn+59291-64-4+@TERM+@rn+52712-04-6+@TERM+@rn+41411-61-4+@TERM+@rn+68194-15-0+@TERM+@rn+68194-14-9+@TERM+@rn+74472-40-5+@TERM+@rn+51908-16-8+@TERM+@rn+68194-13-8+@TERM+@rn+74472-41-6+@TERM+@rn+38380-04-0+@TERM+@rn+68194-08-1+@TERM+@rn+52663-63-5+@TERM+@rn+68194-09-2+@TERM+@rn+35065-27-1+@TERM+@rn+60145-22-4+@TERM+@rn+33979-03-2+@TERM+@rn+38380-08-4+@TERM+@rn+69782-90-7+@TERM+@rn+74472-42-7+@TERM+@rn+39635-35-3+@TERM+@rn+41411-62-5+@TERM+@rn+74472-43-8+@TERM+@rn+39635-34-2+@TERM+@rn+74472-44-9+@TERM+@rn+74472-45-0+@TERM+@rn+74472-46-1+@TERM+@rn+41411-63-6+@TERM+@rn+52663-72-6+@TERM+@rn+59291-65-5+@TERM+@rn+32774-16-6)+@NOT+@org+"nih+reporter"	
	@OR+(@TERM+@rn+52663-62-4+@TERM+@rn+60145-20-2+@TERM+@rn+52663-60-2+@TERM+@rn+65510-45-4+@TERM+@rn+55312-69-1+@TERM+@rn+38380-02-8+@TERM+@rn+55215-17-3+@TERM+@rn+73575-57-2+@TERM+@rn+68194-07-0+@TERM+@rn+68194-05-8+@TERM+@rn+52663-61-3+@TERM+@rn+73575-56-1+@TERM+@rn+73575-55-0+@TERM+@rn+38379-99-6+@TERM+@rn+73575-54-9+@TERM+@rn+41464-51-1+@TERM+@rn+60233-25-2+@TERM+@rn+38380-01-7+@TERM+@rn+39485-83-1+@TERM+@rn+37680-73-2+@TERM+@rn+68194-06-9+@TERM+@rn+60145-21-3+@TERM+@rn+56558-16-8+@TERM+@rn+32598-14-4+@TERM+@rn+70424-69-0+@TERM+@rn+70424-68-9+@TERM+@rn+70362-41-3+@TERM+@rn+74472-35-8+@TERM+@rn+38380-03-9+@TERM+@rn+39635-32-0+@TERM+@rn+74472-36-9+@TERM+@rn+68194-10-5+@TERM+@rn+74472-37-0+@TERM+@rn+74472-38-1+@TERM+@rn+18259-05-7+@TERM+@rn+68194-11-6+@TERM+@rn+31508-00-6+@TERM+@rn+56558-17-9+@TERM+@rn+68194-12-7+@TERM+@rn+56558-18-0+@TERM+@rn+76842-07-4+@TERM+@rn+65510-44-3+@TERM+@rn+70424-70-3+@TERM+@rn+74472-39-2+@TERM+@rn+57465-28-8+@TERM+@rn+39635-33-1)+@NOT+@org+"nih+reporter"	

APPENDIX B. DATA EXTRACTION FIELDS

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

Field label	Data extraction elements
HUMAN	
Funding	Funding source(s)
	Reporting of conflict of interest by authors
Subjects	Study population name/description
	Dates of study and sampling time frame
	Geography (country, region, state, etc.)
	Demographics (sex, race/ethnicity, age or lifestage at exposure and at outcome assessment)
	Number of subjects (target, enrolled, *n* per group in analysis, and participation/follow-up rates)
	Inclusion/exclusion criteria/recruitment strategy
	Description of reference group
Methods	Study design (e.g., prospective or retrospective cohort, nested case-control study, cross-sectional, population-based case-control study, intervention, case report)
	Length of follow-up
	Health outcome category (e.g., cardiovascular)
	Health outcome (e.g., blood pressure)
	Diagnostic or methods used to measure health outcome
	Confounders or modifying factors and how considered in analysis (e.g., included in final model, considered for inclusion but determined not needed)
	Chemical/Mixture name
	Exposure assessment (e.g., blood, urine, hair, air, drinking water, job classification, residence, administered treatment in controlled study)
	Methodological details for exposure assessment (e.g., analytical method, limit of detection)
	Statistical methods
Results	Exposure levels (e.g., mean, median, measures of variance as presented in paper, such as SD, SEM, 75th/90th/95th percentile, minimum/maximum); range of exposure levels, number of exposed cases

Field label	Data extraction elements
	Statistical findings (e.g., adjusted β, standardized mean difference, adjusted odds ratio, standardized mortality ratio, relative risk) or description of qualitative results. When possible, convert measures of effect to a common metric with associated 95% confidence intervals. Most often, measures of effect for continuous data are expressed as mean difference, standardized mean difference, and percentage control response. Categorical data are typically expressed as odds ratio, relative risk (also called risk ratio), or β values, depending on the metric most commonly reported in the included studies and ability to obtain information for effect conversions from the study or through author query.
	Observations on dose-response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, nonmonotonic)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.
ANIMAL	
Funding	Funding source(s)
	Reporting of conflict of interest by authors
Animal model	Sex
	Species
	Strain
	Source of animals
	Age or lifestage at start of dosing and at health outcome assessment
	Diet and husbandry information (e.g., diet name/source)
Treatment	Chemical/Mixture name and CAS (Chemical Abstracts Service) number
	Source of chemical
	Purity or Lot # of chemical
	Dose levels or concentration (as presented and converted to mg/kg bw-day when possible)
	Other dose-related details, such as whether administered dose level was verified by measurement, information on internal dosimetry
	Vehicle used for exposed animals
	Route of administration (e.g., oral, inhalation, dermal, injection)
	Duration and frequency of dosing (e.g., hours, days, weeks when administration was ended, days per week)
Methods	Study design (e.g., single treatment, acute, subchronic [e.g., 90 days in a rodent], chronic, multigenerational, developmental, other)
	Guideline compliance (i.e., use of EPA, OECD, NTP, or another guideline for study design, conducted under GLP guideline conditions, non-GLP but consistent with guideline study, nonguideline peer reviewed publication)
	Number of animals per group (and dams per group in developmental studies)
	Randomization procedure, allocation concealment, blinding during outcome assessment

Field label	Data extraction elements
	Method to control for litter effects in developmental studies
	Use of negative controls and whether controls were untreated, vehicle-treated, or both
	Report on data from positive controls—was expected response observed?
	Endpoint health category (e.g., reproductive)
	Endpoint (e.g., infertility)
	Diagnostic or method to measure endpoint
	Statistical methods
Results	Measures of effect at each dose or concentration level (e.g., mean, median, frequency, and measures of precision or variance) or description of qualitative results. When possible, convert measures of effect to a common metric with associated 95% confidence intervals. Most often, measures of effect for continuous data will be expressed as mean difference, standardized mean difference, and percentage control response. Categorical data will be expressed as relative risk (also called risk ratio).
	NOAEL, LOAEL, benchmark dose analysis, statistical significance of other dose levels, or other estimates of effect presented in paper. **Note:** The NOAEL and LOAEL are highly influenced by study design, do not give any quantitative information about the relationship between dose and response, and can be subject to author's interpretation (e.g., a statistically significant effect might not be considered biologically important). Also, a NOAEL does not necessarily mean zero response. Ideally, the response rate at specific dose levels is used as the primary measure to characterize the response.
	Observations on dose-response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, nonmonotonic)
	Data on internal concentration, toxicokinetics, or toxicodynamics (when reported)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.
IN VITRO	
Funding	Funding source(s)
	Reporting of conflict of interest by authors (reporting bias)
Cell/tissue model	Cell line, cell type, or tissue
	Source of cells/tissue (and validation of identity)
	Sex of human/animal of origin
	Species
	Strain
Treatment	Chemical name and CAS number
	Concentration levels (as presented and converted to µM when possible)
	Source of chemical
	Purity of chemical

Field label	Data extraction elements
	Vehicle used for experimental/control conditions
	Duration and frequency of dosing (e.g., hours, days, weeks when administration was ended, times per day or week)
Methods	Guideline compliance (i.e., use of EPA, OECD, NTP, or another guideline for study design, conducted under GLP guideline conditions, non-GLP but consistent with guideline study, nonguideline peer reviewed publication)
	Randomization procedure, allocation concealment, blinding during outcome assessment (selection bias)
	Number of replicates per group (information bias)
	Percentage serum/plasma in medium
	Use of negative controls and whether controls were untreated, vehicle-treated, or both
	Report on data from positive controls—was expected response observed? (information bias)
	Endpoint health category (e.g., endocrine)
	Endpoint or assay target (e.g., estrogen receptor binding or activation)
	Name and source of assay kit
	Diagnostic or method to measure endpoint (e.g., reporter gene; information bias)
	Statistical methods (information bias)
Results	No-observed-adverse-effect concentration (NOAEC), lowest-observed-adverse-effect concentration (LOAEC), statistical significance of other concentration levels, AC50, or other estimates of effect presented in paper. **Note:** The NOAEC and LOAEC are highly influenced by study design, do not give any quantitative information about the relationship between dose and response, and can be subject to author's interpretation (e.g., a statistically significant effect might not be considered biologically important). Also, a NOAEC does not necessarily mean zero response.
	Observations on dose-response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, nonmonotonic)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

AC50 = 50% activity concentration; EPA = U.S. Environmental Agency; GLP = Good Laboratory Practice; LOAEC = lowest-observed-adverse-effect concentration; LOAEL = lowest-observed-adverse-effect level; NOAEC = no-observed-adverse-effect concentration; NOAEL = no-observed-adverse-effect level; NTP = National Toxicology Program; OECD = Organisation for Economic Co-operation and Development.