EPA/635/R-23/030
IRIS Assessment Protocol
www.epa.gov/iris

Protocol for the Naphthalene IRIS Assessment

(Preliminary Assessment Materials)

[CASRN 91-20-3]

March 2023

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. This information is distributed solely for review purposes under applicable information quality guidelines. It has not been formally disseminated by the Environmental Protection Agency. It does not represent and should not be construed to represent any Agency determination or policy. 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
BMDL	benchmark dose lower confidence limit
$BW^{3/4}$	body-weight scaling to the 3/4 power
CAA	Clean Air Act
CAS	Chemical Abstracts Service
CASRN	Chemical Abstracts Service registry number
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CI	confidence interval
CPHEA	Center for Public Health and Environmental Assessment
COI	conflict of interest
EPA	Environmental Protection Agency
GLP	good laboratory practices
HAP	hazardous air pollutant
HAWC	Health Assessment Workspace Collaborative
HEC	human equivalent concentration
HERO	Health and Environmental Research Online
IRIS	Integrated Risk Information System
IUR	inhalation unit risk
LOAEL	lowest-observed-adverse-effect level
LOEL	lowest-observed-effect level
MeSH	Medical Subject Headings
MOA	mode of action
NMD	normalized mean difference
NOEL	no-observed-effect level
NTP	National Toxicology Program
OAR	Office of Air and Radiation
OECD	Organization for Economic Co-operation and Development
OLEM	Office of Land and Emergency Management
ORD	Office of Research and Development
OSF	oral slope factor
PBPK	physiologically based pharmacokinetic
PECO	populations, exposures, comparators, and outcomes
PK	pharmacokinetic
POD	point of departure
RfC	reference concentration
RfD	oral reference dose
ROBINS-I	Risk of Bias in Nonrandomized Studies of Interventions
UF	uncertainty factor

AUTHORS | CONTRIBUTORS | REVIEWERS

Assessment Team

Ingrid L. Druwe, Ph.D. (co-Assessment Manager)	U.S. EPA/ORD/CPHEA/CPAD
Erin Yost, Ph.D. (co-Assessment Manager)	
Michelle Angrish, Ph.D.	
Bevin Blake, Ph.D.	
J. Allen Davis, M.S.P.H	
Dustin Kapraun, Ph.D.	
Martha Powers, M.P.H., Ph.D.	
Paul Schlosser, Ph.D.	
Rachel M. Shaffer, M.P.H., Ph.D.	

Executive Direction

Wayne Cascio	CPHEA Center Director
V. Kay Holt	CPHEA Deputy Center Director
Samantha Jones	CPHEA Associate Director
Kristina Thayer	CPAD Division Director
Andrew Kraft	CPAD Associate Division Director, IRIS PFAS Team Lead
Paul White	CPAD Senior Science Advisor
Ravi Subramaniam	CPAD Senior Science Advisor (Acting)
Garland Waleko	CPHEA/CPAD/Toxic Effects Assessment (DC) Branch Chief (Acting)
Janice Lee	CPHEA/CPAD/Toxic Effects Assessment (RTP) Branch Chief
Glenn Rice	CPHEA/CPAD/Science Assessment Methods Branch Chief
Viktor Morozov	CPHEA/CPAD/Quantitative Assessment Branch Chief

Contributors and Production Team

Maureen Johnson	U.S. EPA/ORD/CPHEA
Ryan Jones	
Dahnish Shams	
Vicki Soto	
Jessica Soto-Hernandez	
Samuel Thacker	
Garland Waleko	
Channa Keshava	Former Chemical manager, U.S. EPA/ORD/CPHEA
Suryanarayana Vulimiri	
Audrey Galizia	
Amanda Persad	
Rebecca Schaefer	Oak Ridge Associated Universities (ORAU) Contractor
Brittany Schulz	Oak Ridge Associated Universities (ORAU) Contractor
John Bucher	Sole Source Contractor, U.S. EPA

1. INTRODUCTION

The Integrated Risk Information System (IRIS) Program is undertaking a reassessment of the health effects of naphthalene. IRIS assessments provide high quality, publicly available information on the toxicity of chemicals to which the public might be exposed. These science assessments are not regulations. Science assessments such as these provide a critical part of the scientific foundation for subsequent risk assessment and risk management decisions made by EPA program and regional offices to protect public health. IRIS assessments are also used by states and local health agencies, tribes, other federal agencies, international health organizations, and other external stakeholders.

An IRIS assessment plan (IAP) for naphthalene was released for public comment in July 2018, but the IRIS assessment of naphthalene was subsequently suspended prior to a public meeting on the IAP due to changing priorities within the EPA as formally documented in the *IRIS Program Outlook*–April 2019. Naphthalene was renominated as an IRIS assessment in 2021 as described in *A Message from the IRIS Program*–June 2021. An updated IAP and errata sheet were posted to the EPA website in September 2021 and presented at a public science meeting on November 9, 2021 (https://www.epa.gov/iris/iris-public-science-meeting-nov-2021), to seek input on the problem formulation components of the assessment plan.

The IAP summarizes the IRIS Program's scoping and problem formulation conclusions, specifies the objectives and specific aims of the assessment, provides draft PECO (populations, exposures, comparators, and outcomes) criteria, and identifies key areas of scientific complexity. This protocol document incorporates the updated IAP content, including revisions based on public input and updated scoping needs, and presents the methods for conducting the systematic review and dose-response analysis for the assessment. Whereas the IAP describes what the assessment will cover, chemical-specific protocols describe how the assessment will be conducted (see Figure 1).

The systematic review methods described in this protocol are based on the Office of Research and Development's *ORD Staff Handbook for Developing Integrated Risk Information System (IRIS) Assessments* (referred to as the "IRIS Handbook") (U.S. EPA, 2022). The IRIS Handbook was revised in 2022 to incorporate updates to assessment methodology as recommended in a report by the National Academies of Sciences, Engineering, and Medicine (NASEM) (NASEM, 2021) on the draft IRIS Handbook (U.S. EPA, 2020b). Prior to the suspension of the IRIS assessment of naphthalene, some aspects of the assessment were already underway using methods included in the draft Handbook (i.e., literature search, screening, and study evaluation); and when the assessment was renominated, the assessment team considered the revisions made to the Handbook in response to the NASEM report and concluded that the changes would not fundamentally impact

the previously initiated literature search, screening, and overall study evaluation ratings. Therefore, for this assessment, studies will continue to be evaluated using the previously established methodology described in the draft IRIS Handbook (U.S. EPA, 2020b). This is consistent with a 2011 NASEM recommendation not to delay releasing assessments until systematic review methods are finalized (NRC, 2011). The study evaluation methods described in this protocol have been previously presented to NASEM and were positively received (NASEM, 2018); the refinements recommended by NASEM (2021), and reflected in the final IRIS Handbook are generally aimed at clarifying the IRIS study evaluation method but do not request a major overhaul of the study evaluation methods[1]. (U.S. EPA, 2022; NASEM, 2021; U.S. EPA, 2020b; NASEM, 2018)

The IRIS Program posts assessment protocols on its website. Public input received is considered during preparation of the draft assessment.

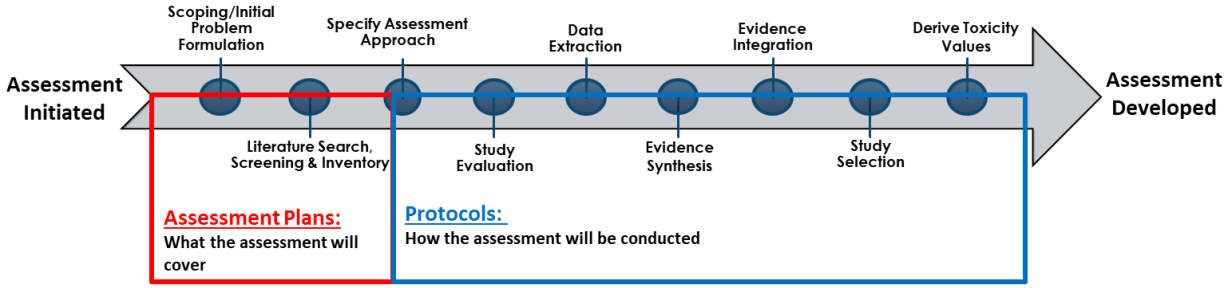


Figure 1-1. IRIS systematic review problem formulation and method documents.

[1] The major study evaluation refinements recommended by NASEM (2021) include (1) clarifications to the procedure for evaluating studies for sensitivity and (2) standardizing the procedure for evaluating reporting quality between human and animal studies.

2. SCOPING AND INITIAL PROBLEM FORMULATION SUMMARY

Section 2.1 provides a brief overview of aspects of the human exposure characteristics of naphthalene that might provide useful context for this protocol. This overview is not intended to provide a comprehensive description of the available information on these topics and is not recommended for use in decision-making. The reader is encouraged to refer to the source materials cited below, more recent publications on these topics, and authoritative reviews or assessments focused on these topics.

2.1. BACKGROUND

Naphthalene is a polycyclic aromatic hydrocarbon that is a white crystalline solid with an aromatic odor. It is soluble in organic solvents and stable in closed containers under normal temperatures and pressures (NTP, 2011). Naphthalene is naturally occurring and is most abundantly found in coal tar, coal, and petroleum (ToxNet Hazardous Substances Data Bank, 2017; ATSDR, 2005). The release of naphthalene also could occur because of its manufacture or use in the chemical industry. In the United States, naphthalene is considered a high production volume (HPV) chemical, though domestic production of naphthalene has decreased significantly from a peak of 900 million pounds in 1968 to an aggregate volume of 100–250 million pounds in 2015 (U.S. EPA, 2016). Naphthalene is also present in jet fuels, such as jet propulsion fuel 8 (JP-8) (ATSDR, 2013). Naphthalene is mainly used in the manufacture of dyes, surfactants, leather tanning agents, dispersants, pesticides, resins, solvents, and chemical intermediates (ATSDR, 2005). Major consumer products containing naphthalene include moth repellents, in the form of mothballs or crystals, and toilet deodorant blocks (ATSDR, 2005). Naphthalene is used as fragrance in non-food-use pesticide products, while naphthalene derivatives are also used as inert ingredients in non-food use pesticide products regulated by EPA (U.S. EPA, 2015a, 2012c). Lastly, naphthalene is also a constituent of tobacco smoke (ATSDR, 2005).

The general public can be exposed to naphthalene via inhalation, ingestion, and dermal routes. Inhalation is generally considered to be the predominant route of exposure (ToxNet Hazardous Substances Data Bank, 2017). Naphthalene is emitted into the atmosphere by industrial facilities, open burning and mobile sources. Naphthalene is a component of fuel oil and gasoline and is produced as a combustion by-product in vehicle exhaust. Exposure to naphthalene may also come from contact with contaminated land and water resulting from spills during storage, transportation and disposal of fuel oil, coal tar, etc. (CalEPA, 2004; IARC, 2002). Because tobacco smoke and numerous consumer products contain and release naphthalene, naphthalene is a

contaminant of indoor air (CalEPA, 2004; IARC, 2002). For nonsmokers exposed to environmental tobacco smoke in their residences, the naphthalene intake rate is estimated to be 1 to 3 μg day^{-1} (Jia and Batterman, 2010; Nazaroff and Singer, 2004). An estimate of the average total intake rate of naphthalene via inhalation in ambient and indoor air is 19 μg day-1 (Jia and Batterman, 2010; Howard, 1989). Children can receive additional exposure to naphthalene through ingestion of soil or food contaminated with naphthalene or through accidental ingestion of household products containing naphthalene, such as mothballs and deodorant blocks (ATSDR, 2005), that are sometimes mistaken for candy. Occupational exposure to naphthalene occurs through inhalation and dermal contact by workers in facilities where naphthalene is produced or used, such as mothball manufacturing plants and creosote-impregnation facilities. High exposures to naphthalene have also been suggested to occur in forest firefighters (Robinson et al., 2008).

Naphthalene is readily absorbed into the systemic circulation following oral, dermal, or inhalation exposure and distributed by the blood throughout the body. It can be transferred to the developing fetus of pregnant women (Anziulewicz et al., 1959; Zinkham and Childs, 1958, 1957) and has been detected in human breast milk (Cok et al., 2012; Tsang et al., 2011; Pellizzari et al., 1982) and umbilical cord serum (Tsang et al., 2011). Naphthalene is rapidly metabolized into a wide array of metabolites, including reactive epoxide and quinone intermediates that may interact with cellular macromolecules such as proteins and DNA. Two major metabolic pathways for naphthalene have been identified: (1) a cytochrome P450 (CYP)-dependent pathway and (2) a glutathione (GSH)-conjugation-dependent pathway. Metabolites pertaining to both major pathways have been identified in the blood and urine of occupationally-exposed individuals and in experimentally-exposed animals (ATSDR, 2005; CalEPA, 2004; IARC, 2002). The naphthalene metabolites 1-naphthol and 2-naphthol have been widely detected in the urine of the U.S. general population, including in children aged 6-19 years old (CDC, 2022).

A summary of existing human health reference values for naphthalene (surveyed in August 2022 using methods described in Appendix A) is provided in Figure 1 (inhalation) and Figure 2 (oral). See Appendix Tables A-2 (inhalation reference values) and A-3 (oral reference values) for a tabular summary, including derivation details of the displayed values; values with no derivation details are listed in Table A-4.

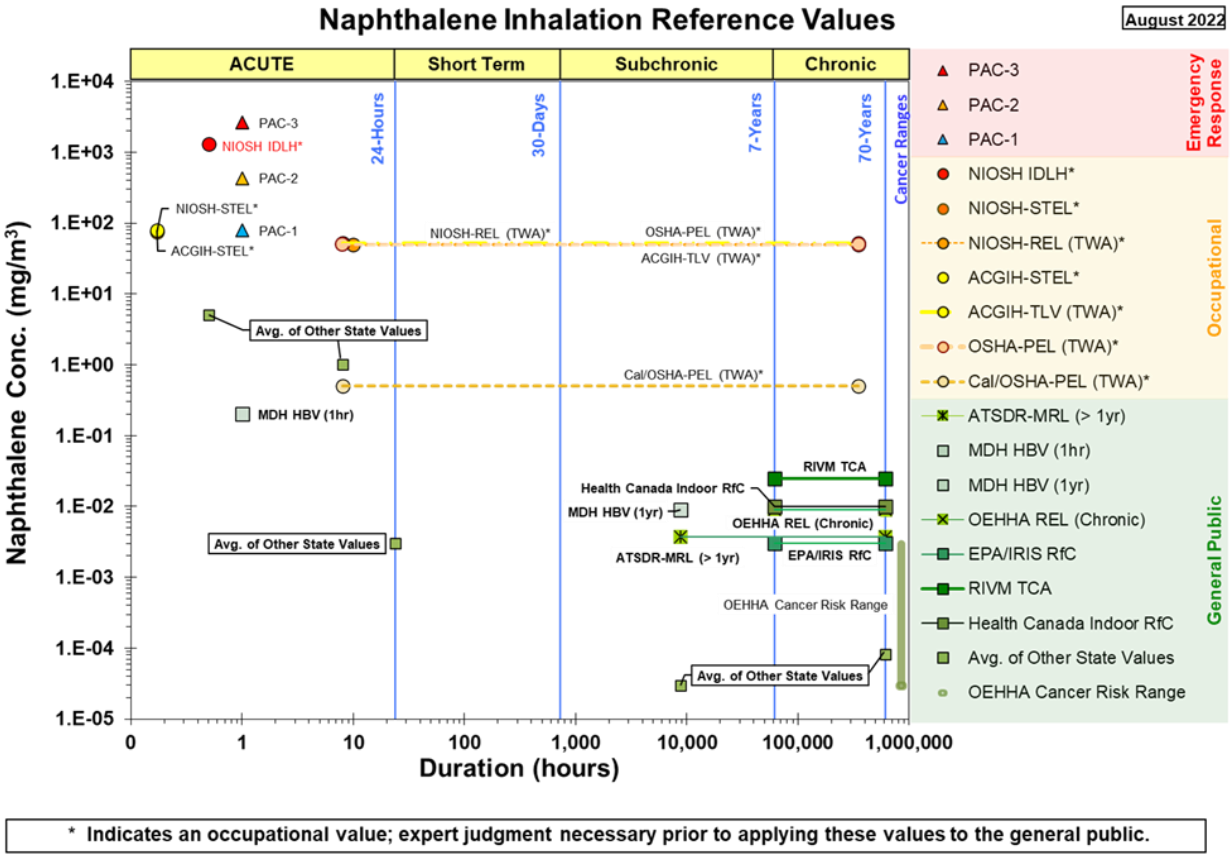


Figure 2-1. Available health effect reference values for inhalation exposure to naphthalene. See Appendix Table A1 for a tabular summary, including information on how each value was derived. Categories for the reference values based on their intended purpose are shown in the legend – red for Emergency Response, gold for Occupational, and green for values applicable to the General Public. OEHHA cancer risk range: range associated with a 10^{-6} - 10^{-4} cancer risk calculated based on the OEHHA cancer slope factor. Abbreviations: ACGIH = American Conference of Governmental Industrial Hygienists; ATSDR = Agency for Toxic Substances and Disease Registry; HBV = Health-Based Value; IDLH = Immediately Dangerous to Life and Health; IRIS = Integrated Risk Information System; MDH = Minnesota Department of Health; MRL = Minimal Risk Level; NIOSH = National Institute for Occupational Safety and Health; OEHHA = California Environmental Protection Agency's Office of Environmental Health Hazard Assessment; OSHA = Occupational Safety and Health Administration; PAC = Protective Action Criteria; PEL = Permissible Exposure Limit; REL = Recommended Exposure Limit (NIOSH) or Reference Exposure Level (California); RfC = Reference Concentration; RIVM = Rijksinstituut voor Volksgezondheid en Milieu, The Netherlands Institute for Public Health and the Environment; STEL = Short-term Exposure Limit; TCA = Tolerable Concentration; TLV = Threshold Limit Value; TWA = Time-weighted average.

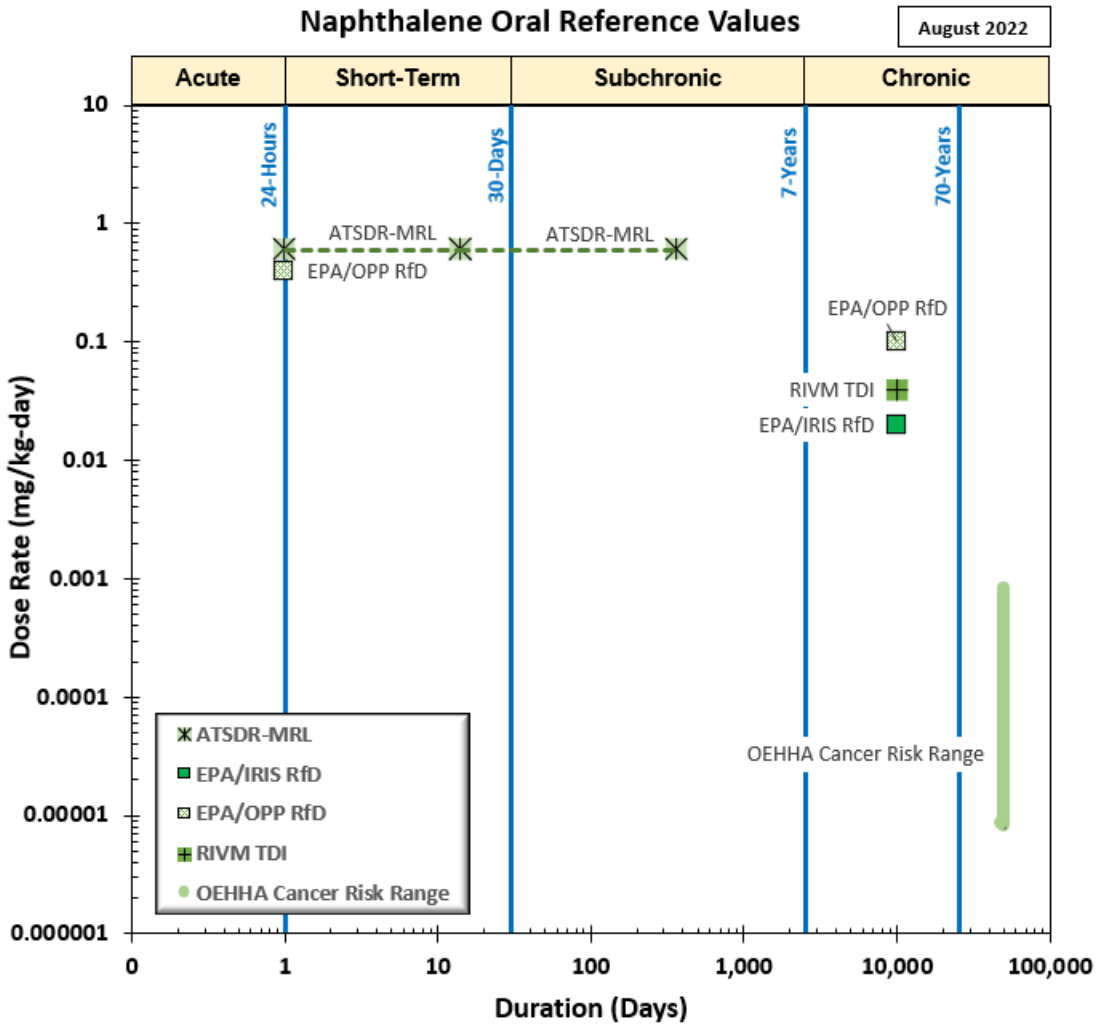


Figure 2-2. Available health effect reference values for oral exposure to naphthalene. See Appendix Table A2 for a tabular summary, including information on how each value was derived. All values in this figure are intended for application in the general public. OEHHA cancer risk range: range associated with a 10^{-6} - 10^{-4} cancer risk calculated based on the OEHHA cancer slope factor. Abbreviations: ATSDR = Agency for Toxic Substances and Disease Registry; IRIS = Integrated Risk Information System; MRL = Minimal Risk Level; OPP = Office of Pesticide Programs; RfD = Reference Dose; RIVM = Rijksinstituut voor Volksgezondheid en Milieu, The Netherlands Institute for Public Health and the Environment; TDI = Tolerable Daily Intake.

2.2. SCOPING SUMMARY

Naphthalene is subject to regulation under several environmental statutes implemented by EPA, including the Clean Water Act (CWA), Clean Air Act (CAA), Federal Fungicide Insecticide and Rodenticide Act (FIFRA), Toxic Substances Control Act (TSCA); Emergency Planning and Community Right-to-Know Act (EPCRA), Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), and the Resource Conservation and Recovery Act (RCRA). Naphthalene

is also listed as a Hazardous Air Pollutant (HAP) by EPA and is a contaminant found at more than 300 National Priority List (Superfund) (U.S. EPA, 2023).

During initial scoping, the IRIS Program met with EPA program and regional offices that had interest in an IRIS assessment for naphthalene to discuss specific assessment needs. Table 2-1 provides a summary of current programmatic interest. Additional programmatic and regional needs and interests will be reviewed and updated as the assessment progresses.

Table 2-1. EPA program interest in reassessment of naphthalene

EPA program	Oral	Inhalation	Statutes/regulations/policies	Anticipated uses/interest
OLEM, Regions	X	X	Comprehensive Environmental Response, Compensation and Liability Act (CERCLA)	Naphthalene toxicological information could be used to make risk determinations for response actions (e.g., short-term removals, long-term remedial response actions) under CERCLA and RCRA.
OCSPP	X	X	Toxic Substances Control Act (TSCA)	Naphthalene toxicological information could be used to inform risk assessment and risk management decisions under TSCA.
OAR	X	X	Clean Air Act (CAA)	Naphthalene is listed as a Hazardous Air Pollutant (HAP) and is also a mobile source air toxic. Naphthalene toxicological information could be used to inform risk assessment and risk management decisions under CAA.

OLEM (Office of Land and Emergency Management)
OCSPP (Office of Chemical Safety and Pollution Prevention)
OAR (Office of Air and Radiation)

2.3. PROBLEM FORMULATION

A 1998 assessment of naphthalene is currently available on the IRIS website at https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=436 (U.S. EPA, 1998b). This assessment includes a review of inhalation studies which provide support for a reference concentration (RfC) of 3×10^{-3} mg/m^3 for noncancer effects based on hyperplasia and metaplasia in respiratory and olfactory epithelium in mice, and a review of oral studies which provide support for a reference dose (RfD) of 2×10^{-2} mg/kg-day for noncancer effects based on decreased body weight in male rats. EPA's 1998 IRIS Toxicological Review of Naphthalene, which was conducted using EPA's 1986 Cancer Guidelines (U.S. EPA, 1986), classified naphthalene as a Group C, possible human carcinogen. This classification was based on inadequate carcinogenicity data in humans exposed to naphthalene via the oral and inhalation routes, and limited evidence of carcinogenicity in animals exposed to naphthalene via inhalation. The 1998 assessment concluded that a genotoxic

mechanism appeared unlikely but hypothesized that the mechanism for tumorigenesis involves oxygenated reactive metabolites produced via the cytochrome P450 monooxygenase system.

Since the posting of the IRIS toxicological review of naphthalene in 1998 and the 2005 release of EPA's cancer guidelines (U.S. EPA, 2005a), new information on naphthalene has become available (see Section 4.5), including bioassay data, potency estimations, and physiologically-based pharmacokinetic (PBPK) models with the potential to assist in performing route-to-route and animal-to-human extrapolations. More specifically, several significant studies on naphthalene toxicity have been published, including a 2-year inhalation study performed by NTP in which naphthalene-exposed rats showed an increased incidence of nasal tumors (NTP, 2000). In addition to this NTP study, numerous studies (>70) have been published which provide mechanistic information that could inform the naphthalene mode of action for cancer or noncancer effects. These include studies that report on the involvement of specific cytochrome P450 subfamilies like CYP2F and CYP2A in the metabolism and possible activation of reactive naphthalene intermediates (Buckpitt et al., 2013; Morris, 2013; Morris and Buckpitt, 2009; Carlson, 2008; Genter et al., 2006; Buckpitt et al., 2002; Su et al., 2000; Lanza et al., 1999; Shultz et al., 1999) that may interact with biological macromolecules such as proteins or DNA. Additionally, a PBPK model for naphthalene was developed using controlled human dermal and inhalation exposures to JP-8, of which naphthalene is a component (Kim et al., 2007). The results of this more recent research will be evaluated using EPA's current cancer guidelines (U.S. EPA, 2005a) and may provide new evidence to better inform naphthalene toxicity values.

2.4. KEY SCIENCE ISSUES

Based on the preliminary literature survey, the following key scientific issues and potential cancer mode-of-action (MOA) hypotheses were identified that warrant evaluation in this assessment.

- Evaluating interspecies differences in metabolism and toxicity: Naphthalene toxicity is typically attributed to protein binding by naphthalene quinone metabolites and/or the participation of naphthalene quinone metabolites in redox cycles leading to oxidative stress and DNA damage (O'Brien, 1991). These quinone intermediates are produced via cytochrome P450 (CYP)-dependent metabolism and may specifically involve the CYP2F subfamily. While much progress has been made in the characterization of CYP2F2, the CYP thought to be primarily involved in naphthalene metabolism in mice, characterizing the relative contribution of P450 oxidizing enzymes to naphthalene metabolism in rats and humans has been more difficult (Buckpitt et al., 2002; Shultz et al., 1999). Recent studies show that, in addition to the CYP2F subfamily, the CYP2A class also plays an important role in naphthalene-induced lung toxicity and may be the more pertinent enzyme in naphthalene metabolism in humans (Li et al., 2017; Su et al., 2000). The rate and extent of metabolism of naphthalene in various tissues and in different animal species, along with anatomical differences in the nasal turbinates between species, will be important considerations in evaluating differences in naphthalene toxicity across species.

- Cancer mode of action: Multiple animal and in vitro studies published since the 1998 IRIS Toxicological Review have provided mechanistic information and postulated the involvement of several biological processes in the development of naphthalene-induced tumor formation. These proposed processes include genotoxicity, cytotoxicity, and sustained regenerative cell proliferation. Among the key events identified by these studies are the depletion of glutathione and the formation of reactive naphthalene quinone metabolites via the cytochrome P450 pathway. These quinone metabolites may lead to oxidative stress and DNA damage. To help inform the analysis and interpretation of the role and biological plausibility of each of these proposed mechanisms occurring in humans and their role in the formation of naphthalene-induced tumors, the supplemental materials identified in the literature search will be reviewed to identify relevant information [e.g., workshops (U.S. EPA, 2014b)] that inform these topics. Differences in enzyme activities between human and rodent tissues exist; therefore, evaluation of the cancer MOA in the context of toxic metabolite formation and the relevance of these toxic metabolites to human cancer hazard will also be evaluated.

3. OVERALL OBJECTIVES AND SPECIFIC AIMS

The overall objective of this assessment is to identify adverse health effects and characterize exposure-response relationships for these effects of naphthalene to derive toxicity values (e.g., reference doses [RfDs], reference concentrations [RfCs], cancer risk estimates) as supported by the available data. This assessment will use systematic review methods to evaluate the epidemiological and toxicological literature for naphthalene, including consideration of relevant mechanistic evidence. The evaluation conducted in this assessment will be consistent with relevant EPA guidelines.[2]

3.1. SPECIFIC AIMS

- Develop a systematic evidence map (SEM) to identify an initial literature inventory of epidemiological studies (i.e., human), toxicological studies (i.e., experimental animal), PBPK models, and supplemental literature pertinent to characterizing the health effects of naphthalene exposure. The PECO criteria used to develop the SEM (referred to "problem formulation PECO") is conducted according to the methods for literature search, screening, and inventory described in Section 4 (Thayer et al., 2022; NASEM, 2021; Wolffe et al., 2019).
    - Epidemiological studies, toxicological studies, and PBPK models are identified for inclusion based on predefined populations, exposure, comparators, and outcomes (PECO) criteria.
    - Supplemental material content includes: mechanistic studies, including in vivo, in vitro, ex vivo, or in silico models; toxicokinetic and *absorption, distribution, metabolism, and excretion* (ADME) studies; studies with routes of exposure other than oral, inhalation, and dermal; case studies; studies that evaluate exposure and health effects associated with the jet fuel JP-8; studies that are in a non-English language; and studies that are abstract-only or did not have the full text available.
- Use the initial literature inventory identified in the SEM to (1) develop assessment PECO criteria that define the subset of studies that will be the focus of the systematic review; (2) define the unit(s) of analysis at the level of endpoint or health outcome for hazard characterization; and (3) identify priority analyses of supplemental material to address the specific aims, uncertainties in hazard characterization, susceptibility, and dose-response analysis.
- Conduct study evaluations (risk of bias and sensitivity) for individual epidemiological and toxicological studies that meet refined assessment PECO criteria.

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

- Conduct a scientific and technical review for PBPK models considered for use in the assessment. If a PBPK or PK model is selected for use, the most reliable dose metric will be applied based on analyses of the available dose metrics and the outcomes to which they are being applied.

- Conduct data extraction (summarizing study methods and results) from epidemiological and animal toxicological studies that meet the refined assessment PECO criteria.

- For each evidence stream, and for each unit of analysis, use a structured framework to develop and describe the strength of evidence across studies and the supporting rationale ("evidence synthesis"). Depending on the specific health endpoint or outcome, mechanistic information and precursor events might be included in a unit of analysis.

- For each health effect category, use a structured framework to develop and describe weight of evidence judgments across evidence streams and the supporting rationale for those judgments ("evidence integration"). The evidence integration analysis presents inferences and conclusions on human relevance of findings in animals, cross-evidence stream coherence, potentially susceptible populations and lifestages, and other critical inferences supported by mechanistic, ADME, or PK/PBPK data (e.g., biological plausibility).

- For each health effect category, summarize evidence synthesis and evidence integration conclusions in an evidence profile table.

- Derive toxicity values (e.g., reference doses [RfDs], reference concentrations [RfCs], cancer risk estimates) as supported by the available data.

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

4. LITERATURE SEARCH, SCREENING, AND INVENTORY

The literature search and screening processes described in this section were used to conduct an SEM and identify an initial literature inventory for naphthalene (Appendix C), using problem formulation PECO criteria (see Section 4.1) and supplemental screening criteria (see Section 4.2) to guide the inclusion of studies. The resulting initial literature inventory was used to develop assessment PECO criteria and identify priority analyses of supplemental material (described in Chapter 5). The initial literature search as well as all subsequent literature search updates are conducted using the processes described in this chapter, and therefore for the purposes of this assessment the literature inventory developed as part of the SEM will be continually updated with new studies as the assessment progresses.

4.1. POPULATIONS, EXPOSURES, COMPARATORS, AND OUTCOMES (PECO) CRITERIA FOR THE SYSTEMATIC EVIDENCE MAP

PECO criteria are used to focus the research question(s), search terms, and inclusion criteria in a systematic review. The PECO criteria used to develop the SEM and identify an initial literature inventory are referred to hereafter as the "problem formulation PECO" (see Table 5-1) and were intentionally broad in order to identify all the available evidence in humans and animal models.

The problem formulation PECO for naphthalene (see Table 4-1) was based on: (1) nomination of the chemical for assessment, (2) discussions with scientists in EPA program and regional offices to determine the scope of the assessment that will best meet Agency needs, and (3) preliminary review of the health effects literature for naphthalene (primarily focusing on reviews and authoritative health assessment documents) to identify the potential major health hazards associated with exposure to naphthalene and key areas of scientific complexity.

Table 4-1. Populations, exposures, comparators, outcomes (PECO) criteria for the systematic evidence map (i.e., problem formulation PECO)

PECO element	Evidence
Populations[a]	**Human:** Any population and lifestage (occupational or general population, including children and other sensitive populations). The following study designs will be considered most informative: controlled exposure, cohort, case-control, cross-sectional, and ecological.
	Animal: Nonhuman mammalian animal species (whole organism) of any lifestage (including preconception, in utero, lactation, peripubertal, and adult stages). Studies of transgenic animals will be tracked as mechanistic studies under "potentially relevant supplemental material."

PECO element	Evidence
Exposures	**Human:** Any exposure to naphthalene (CASRN 91-20-3), including occupational exposures.
	Animal: Any exposure to naphthalene (CASRN 91-20-3) via oral or inhalation route[s]. Studies involving exposures to mixtures will be included only if they include an arm with exposure to naphthalene alone. Other exposure routes, including injection and dermal, will be tracked during title and abstract screening and tagged as "supplemental information."
	Studies describing physiologically-based pharmacokinetic (PBPK) models for naphthalene will be included.
Comparators	**Human:** A comparison or referent population exposed to lower levels (or no exposure/exposure below detection limits) of naphthalene, or exposure to naphthalene for shorter periods of time.
	Animal: A concurrent control group exposed to vehicle-only treatment.
Outcomes	All health outcomes (both cancer and noncancer). In general, endpoints related to clinical diagnostic criteria, disease outcomes, histopathological examination, or other apical/phenotypic outcomes will be prioritized for evidence synthesis over outcomes such as biochemical measures.

4.2. SUPPLEMENTAL SCREENING CRITERIA

During the literature screening process, studies containing information potentially relevant to the specific aims of the assessment are tagged as supplemental material by category. Some studies could emerge as being critically important to the assessment and may need to be evaluated and summarized at the individual study level (e.g., certain cancer MOA or ADME studies), or might be helpful to provide context (e.g., provide hazard evidence from routes or durations of exposure not meeting the refined assessment PECO), or might not be cited at all in the assessment (e.g., individual studies that contribute to a well-established scientific conclusion). Because it is often difficult to assess the impact of individual studies tagged as supplemental material on assessment conclusions at the screening stage, the tagging structure, described in Table 4-2, allows for easy retrieval later in the assessment process.

Table 4-2. Categories of "Potentially Relevant Supplemental Material"

Category (Tag)	Description
Mechanistic	Studies that do not meet PECO criteria but do report measurements related to a health outcome that inform the biological or chemical events associated with phenotypic effects. Experimental design could include in vitro, in vivo (by any route of exposure), ex vivo, and in silico studies in mammalian and nonmammalian model systems. Studies where the chemical is used as a laboratory reagent generally do not need to be tagged (e.g., as a chemical probe used to measure antibody response). *[During screening, especially at the title and abstract (TIAB) level, it may not be readily apparent for studies that meet P, E, and C criteria if the endpoint(s) in a study are best classified as phenotypic or mechanistic with respect to the O criteria. In these cases, the study should be screened as "unclear" during TIAB screening, and a determination made based on full-text review (in consultation with a content expert as needed). Full-text retrieval is performed for studies of transgenic model systems that meet E and C criteria to determine if they include phenotypic information in wildtype animals that meet P and O criteria but is not reported in the abstract.]*
Toxicokinetic (ADME)	Toxicokinetic (ADME) studies are primarily controlled experiments, where defined exposures usually occur by intravenous, oral, inhalation, or dermal routes, and the concentration of particles, a chemical, or its metabolites in blood or serum, other body tissues, or excreta are then measured. These data are used to estimate the amount absorbed (A), distributed (D), metabolized (M), and/or excreted (E). The most informative studies involve measurements over time such that the initial increase and subsequent concentration decline is observed, preferably at multiple exposure levels. Data collected from multiple tissues or excreta at a single time-point also inform distribution. ADME data can also be collected from human subjects who have had environmental or workplace exposures that are not quantified or fully defined. However, to be useful such data must involve either repeated measurements over a time-period when exposure is known (e.g., is zero because previous exposure ended) *or* time- and subject-matched tissue or excreta concentrations (e.g., plasma and urine, or maternal and cord blood). ADME data, especially metabolism and tissue partition coefficient information, can be generated using in vitro model systems. Although in vitro data may not be as definitive as in vivo data, these studies should also be tracked as ADME. For large evidence bases it may be appropriate to separately track the in vitro ADME studies. **Studies describing environmental fate and transport or metabolism in bacteria or model systems not applicable to humans or animals should not be tagged.*
Non-PECO route of exposure	Epidemiological or animal studies that use a non-PECO route of exposure. (e.g., injection, dermal). **This categorization generally does not apply to epidemiological studies where the exposure route may be unclear; such studies advance to full-text review to determine PECO relevance if the route(s) of exposure are plausible.*
PBPK model application	Studies that describe the application of PBPK model(s) for naphthalene but do not develop a novel, whole-organism PBPK model. Examples: pharmacokinetic and toxicological studies that make use of existing PBPK models; cell culture analogs of PBPK models.
Case reports or case series	Case reports of ≤ 3 subjects that describe health outcomes after exposure.
JP-8 health effect studies	Studies that evaluate exposure and health effects associated with the jet fuel JP-8 but do not evaluate the effects of naphthalene as a standalone compound. Human studies that use measures of JP-8 rather than naphthalene alone in regression analyses will be tagged to this category.
Non-English studies	Records that are in a non-English language.
Abstract only or full text not available	Records that do not contain sufficient documentation to support study evaluation and data extraction.

1

4.3. LITERATURE SEARCH STRATEGIES

4.3.1. Core Database Searches

Literature search strategies were developed using key terms and words related to the problem formulation PECO criteria. Standard terms were used to gather information on health outcomes (e.g., toxicity, hematology, teratogen). Terms for specific experimental animal species were also included. Exposure terms were used to capture studies that are not indexed by the chemical name (e.g., moth balls, camphor). Because each database has its own search architecture, the resulting search strategy was tailored to account for each database's unique search functionality.

The following databases were searched:

- PubMed (National Library of Medicine)
- Web of Science (Thomson Reuters)
- Toxline (National Library of Medicine)[3]

Database searches were conducted in February 2013, December 2014, November 2015, January 2017, September 2017, February 2019, January 2021, and January 2022. Searches conducted in January 2017 added terms to the PubMed query looking for information on naphthalene metabolites (1,4-naphthoquinone; 1,2-naphthoquinone; naphthalene 1,2-oxide; and 1,2-dihydroxy-1,2-dihydronaphthalene). Searches were not restricted by publication date and no language restrictions were applied. The detailed search strategies are presented in Appendix B (Table B-1). Literature searches were conducted using EPA's Health and Environmental Research Online (HERO) database.[4]

The database searches will be updated throughout assessment draft development to identify literature published during the course of review. The last full literature search update will be conducted less than 1 year before the planned release of the draft document for public comment. The results returned (i.e., the number of "hits" from each electronic database or other literature source), including the results of any literature search updates, are documented in the literature flow diagrams (see Appendix C), which also reflect the literature screening decisions. The IRIS Program takes extra steps to ensure identification of pertinent studies by encouraging the scientific community and the public to identify additional studies and ongoing research and by considering late breaking studies that would impact the credibility of the conclusions, even during the review

[3] The ToxLine database was migrated to PubMed after the 2019 literature search update, so was not included in subsequent literature search updates.
[4]Health and Environmental Research Online: https://hero.epa.gov/hero/.

process.[5] Studies identified after peer review begins will be considered for inclusion only if they meet the assessment PECO criteria and could fundamentally alter the assessment's primary conclusions.

4.3.2. Targeted Search for PBPK Models

To ensure that PBPK models for naphthalene were not missed by the broad literature search described in the section above, an additional targeted search for PBPK models for naphthalene was conducted in PubMed in August 2022. This search strategy is presented in Appendix B (Table B-2). These studies were screened according to the methods in Section 4.4 by two independent reviewers with expertise in PBPK modeling.

4.3.3. Other Resources Consulted

The literature search strategies described above are designed to be broad, but like any search strategy, studies can be missed [e.g., cases where the specific chemical is not mentioned in title, abstract, or keyword content; ability to capture "gray" literature (studies not reported in the peer-reviewed literature) that is not indexed in the databases listed above]. Thus, in addition to the core database searches, the sources below are used to identify studies that could have been missed (see Appendix B, Table B-3 and B-4 for details):

- Identification of Toxic Substances Control Act Test Submissions (TSCATS) by searching TSCATS 2, TSCATS 1, EPA's Chemical Data Access Tool (CDAT), and Google searches for TSCA recent submissions.
- Manually searching citations from published review articles and national and international health agency documents.
- "Backward" searches (to identify articles cited by included studies, reviews, or prior assessments by other agencies) and "forward" searches (to identify articles that cite those studies).
- References that had been previously added to the HERO project page for the naphthalene assessment during the development of earlier draft materials.
- Searching a combination of Chemical Abstract Service Registry Numbers (CASRNs) and synonyms on chemical assessment-related websites.

High throughput screening information for naphthalene from EPA's ToxCast or Tox21 will not be pursued in this assessment due to quality control (QC) concerns. The analytical QC performed by ToxCast found that naphthalene was present in the sample at the initial timepoint (T0) but was not detectable at a later timepoint (at 4 months T4), indicating that decomposition

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

had occurred at some point during that time period. Considering that naphthalene is volatile/semi-volatile and the ToxCast assays rely on liquid-based cell and cell-free systems, the stability of the chemical in the test system is uncertain and bioactivity results are difficult to interpret. Access to the naphthalene assays and QC for these may be found at the ToxCast Dashboard by following this link (click on "QC Data ID" to view the results): https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID8020913#invitrodb-bioassays-toxcast-tox21.

4.3.4. Non Peer-Reviewed Data

IRIS assessments rely mainly on publicly accessible, peer-reviewed studies. However, it is possible that unpublished data directly relevant to the PECO might be identified during assessment development. In these instances, EPA will try to get permission to make the data publicly available (e.g., in HERO); data that cannot be made publicly available are not used in IRIS assessments. In addition, on rare occasions where unpublished data would be used to support key assessment decisions (e.g., deriving a toxicity value), EPA may obtain external peer review if the owners of the data are willing to have the study details and results made publicly accessible, or if an unpublished report is publicly accessible (or submitted to EPA in a non-confidential manner) (U.S. EPA, 2015b). This independent, contractor driven, peer review would include an evaluation of the study similar to that for peer review of a journal publication. The contractor would identify and typically select three scientists knowledgeable in scientific disciplines relevant to the topic as potential peer reviewers. Persons invited to serve as peer reviewers would be screened for conflict of interest. In most instances, the peer review would be conducted by letter review. The study and its related information, if used in the IRIS assessment, would become publicly available. In the assessment, EPA would acknowledge that the document underwent external peer review managed by the Agency, and the names of the peer reviewers would be identified. In certain cases, IRIS will assess the utility of an analysis of accessible raw data (with descriptive methods) that has undergone rigorous quality assurance/quality control review (e.g., ToxCast/Tox21 data, results of NTP studies not yet published) but that have not yet undergone external peer review.

Unpublished data from personal author communication can supplement a peer-reviewed study as long as the information is made publicly available. If such ancillary information is acquired, it will be documented in the Health Assessment Workspace Collaborative (HAWC) or HERO project page (depending on the nature of the information received).

4.4. LITERATURE SCREENING

This screening strategy was used to identify an initial literature inventory (described in Appendix C) and will be used in subsequent literature search updates. The problem formulation PECO criteria described in Section 4.1 are used to determine inclusion or exclusion of a reference as a primary source of health effects data or a published PBPK model. In addition to the inclusion of

studies that meet the problem formulation PECO criteria, studies containing supplemental material that is potentially relevant to the specific aims are tracked during the screening process using the categories described in Section 4.2. Although not considered to directly meet PECO criteria, these studies are not strictly excluded unless otherwise specified. Unlike studies that meet PECO criteria, supplemental studies may not be subject to systematic review unless specifically defined questions are identified that focus the mechanistic (or other) analysis to inform the specific aims (see Section 3.1).

4.4.1. Title and Abstract-Level Screening

Following a pilot phase to calibrate screening guidance, two screeners independently conduct a title and abstract screen of the search results to identify records that appeared to meet the problem formulation PECO criteria. For literature searches conducted through November 2015, all identified records were first electronically screened with a set of terms intended to prioritize "on-topic" references for title and abstract review (see Appendix B for a description of electronic screening methods and the list of inclusion terms). Title/abstract screening was then performed manually on all records prioritized by the electronic screen. For literature searches conducted after November 2015, no electronic screen was performed due to the smaller number of records identified, and title/abstract screening was performed on all records.

The software platforms used for screening the literature for naphthalene changed over time, reflecting the technology that was available at the time of each literature search. In all cases, screening was performed manually (machine learning functionality was not applied), and therefore EPA does not anticipate that screening results are affected by the type of software used. The software platforms used for title/abstract screening are EndNote (for literature searches conducted between 2013 and 2017), SWIFT-Active Screener software (for literature search conducted in 2019) (https://swift.sciome.com/activescreener), or DistillerSR (for literature searches conducted in 2021 and thereafter) (https://www.evidencepartners.com/products/distillersr-systematic-review-software/).

For citations with no abstract, articles are screened based on all or some of the following: title relevance, page numbers (articles two pages in length or less may be assumed to be conference reports, editorials, or letters), and PubMed MeSH (Medical Subject Headings, e.g., a study might not be considered further if there are no human health- or biology-related MeSH terms). Screening conflicts are resolved by discussion among the primary screeners with consultation by a third reviewer or technical advisor (if needed) to resolve any remaining disagreements. Eligibility status of non-English studies is assessed using the same approach with online translation tools or engagement with a native speaker. Non-English studies were tracked during screening and tagged as supplemental for possible further evaluation.

4.4.2. Full-Text Level Screening

Records that are not excluded based on the title and abstract are advanced to full-text review. Full-text copies of these potentially relevant records are retrieved, stored in the HERO database, and independently assessed by two screeners to confirm eligibility according to the problem formulation PECO criteria. Screening conflicts are resolved by discussion among the primary screeners with consultation by a third reviewer or technical advisor (as needed to resolve any remaining disagreements). Studies that advance to full-text review can also be tagged as "potentially relevant supplemental material." Approaches for language translation include use of an online translation tool, an engagement of a native speaker from within EPA, or use of fee-based translation services. Fee-based translation services for non-English studies are typically reserved for studies that are anticipated as being useful for toxicity value derivation.

4.4.3. Multiple Publications of the Same Data

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

4.4.4. Literature Screening Results

The results of this screening process are posted on the project page for this assessment in the HERO database (https://hero.epa.gov/hero/index.cfm/project/page/project_id/367) and studies have been "tagged" with appropriate category descriptors (e.g., included, excluded, potentially relevant supplemental material). The literature inventory of studies meeting problem formulation PECO criteria is shown in Appendix C (see Section 4.5 for details on how literature inventories are created).

4.5. LITERATURE INVENTORY

During title/abstract or full text level screening, studies are categorized by evidence type (human or animal) or category of supplemental information (e.g., mechanistic, ADME). Next, study design details for studies that meet the problem formulation PECO criteria are summarized as described in Section 4.5.1. A more granular tagging of supplemental material may also be conducted as described in Section 4.5.2. The results of this categorization and tagging are referred to as the

literature inventory and is the key analysis output of the SEM. The literature inventory of studies meeting the problem formulation PECO criteria is shown in the SEM described in Appendix C.

4.5.1. Studies that Meet the Problem Formulation PECO Criteria

During full text screening, all human and animal studies that met the problem formulation PECO criteria are briefly summarized to facilitate subsequent review by subject matter experts. For animal studies, the following information is captured: study type [acute (<24 hours), short term (1–30 days), subchronic (30–90 days), chronic (>90 days), reproductive, developmental], duration and timing of treatment, route, species, strain, sex, dose or concentration levels tested, dose or concentration units, health system and specific endpoints assessed, and a brief summary of findings at the health system level based on author-reported statistical significance. For human studies, the following information is summarized: population type (e.g., general population-adult, occupational, pregnant women, infants and children), study type (e.g., controlled trial, cross-sectional, cohort, case-control), short free text description of study population, sex, major route of exposure (if known), description of how exposure was assessed, health system and specific outcome assessed, and a summary of findings at the health system level based on author-reported statistical significance (null or an indication of any associations found and a description of how the exposure was quantified in the analysis). Studies are extracted into Excel by one team member and checked by at least one other team member. These study summaries are referred to as literature inventories and are presented using Tableau visualization software (https://www.tableau.com/).

All PBPK models identified in the literature search are reviewed by subject matter experts and are summarized in Appendix C of this protocol in both descriptive text and in a tabular format.

4.5.2. Organizational Approach for Supplemental Material

Inventories may also be created for other categories of studies that were tagged as "potentially relevant supplemental material" during screening, including mechanistic studies (e.g., in vitro or in silico models), ADME studies, and other studies that do not meet the specific PECO criteria but that may still be relevant to the research question(s). Here, the objective is to create an inventory of studies that can be tracked and further summarized as needed—for example, by model system, key characteristic [e.g., of carcinogens; Smith et al. (2016)] mechanistic endpoint, or key event—to support analyses of critical questions that arise at various stages of the systematic review. See Section 5.3 for a description how the inventory and analysis of supplemental material will be approached. Any inventories of potentially relevant supplemental material created for this assessment will be made publicly available.

5. SPECIFY ASSESSMENT APPROACH

The primary purpose of this step is to provide further specification to the assessment methods based on characterization of the extent and nature of the evidence identified from the literature inventory. This includes refinements to PECO criteria and defining the unit(s) of analysis for health endpoints/outcomes during evidence synthesis, and presenting analysis approaches for mechanistic, ADME, and other types of supplemental material content. A unit of analysis is an outcome or group of related outcomes within a health effect category that are considered together during evidence synthesis (see Section 8).

5.1. REFINEMENTS TO PECO CRITERIA

Refinements to the problem formulation PECO criteria were made based on the creation of initial literature inventories by subject matter experts, which are presented in Appendix C. The assessment PECO criteria (see Table 5-1) reflect the subset of studies that will be the focus of the systematic review and will move forward for study evaluation and evidence synthesis.

The systematic review will focus on the health outcome categories identified in the literature inventory, that appear to have sufficient information available to support hazard identification, i.e., respiratory system (nasal and pulmonary), hematological, immune system, reproductive system, developmental, and cancer. Ocular effects such as cataracts were not included in the assessment PECO because they are reported to occur at higher naphthalene exposure levels compared to other types of health outcomes (Yost et al., 2021) and therefore are not likely to drive the derivation of toxicity values. Other health outcome categories identified in the initial literature inventory were not included in the assessment PECO because they do not appear to have enough information to support hazard identification. For instance, although an association between naphthalene and severe neonatal jaundice was identified in a cross-sectional study (Familusi and Dawodu, 1985), this is thought to be a secondary effect of hemolytic anemia and therefore hepatic effects were not included in the assessment PECO. Cardiometabolic effects including obesity, hypertension, and metabolic syndrome were identified in two cross-sectional studies that evaluated association with naphthalene metabolites in urine (Ranjbar et al., 2015; Scinicariello and Buser, 2014) but these observations were considered too limited to support hazard identification. Evidence for other health outcome categories such as renal/urinary and endocrine/exocrine was largely null based on the available studies. Therefore, unless additional evidence becomes available, studies that do not report on any of the health outcome categories listed in the assessment PECO will not be included in the systematic review and will not undergo study evaluation.

Among the available animal studies, literature screening indicated that there were generally sufficient numbers of multi-dose chronic, subchronic, or developmental exposure studies available

to inform weight of evidence and dose-response analysis for each of the major health effect categories being considered for systematic review. Because longer duration studies are preferred for dose-response assessment to inform lifetime toxicity values, it was decided for the purposes of this assessment that non-developmental studies with exposures < 30 days in duration will only be included in the systematic review for a given health effect if longer duration studies are not available or if they contribute critical information to the weight of evidence or dose-response analysis. An iterative approach will be applied when determining which acute and short-term duration studies will be included in the systematic review. For instance, the 1- and 5-day inhalation studies by Dodd et al. (2010) will be included in the systematic review because they provide information on the concentration- and time-dependent development of nasal and olfactory necrosis in rats exposed to naphthalene, which is anticipated to be useful for dose-response analysis. Likewise, the 14-day oral study by Shopp et al. (1984) will be included along with the 90-day study from the same report to demonstrate dose- and time-dependent responses. All studies exposing animals during developmental life stages (e.g., gestational exposure studies) will be included regardless of exposure duration, as short-term exposures may coincide with windows of susceptibility. Studies with exposure durations < 30 days that do not meet these criteria will not be included in the systematic review and will not undergo study evaluation.

Table 5-1. Refined assessment PECO criteria for naphthalene

PECO element	Evidence
Populations[a]	**Human:** Any population and lifestage (occupational or general population, including children and other sensitive populations). The following study designs will be considered most informative: controlled exposure, cohort, case-control, cross-sectional, and ecological.
	Animal: Nonhuman mammalian animal species (whole organism) of any lifestage (including preconception, in utero, lactation, peripubertal, and adult stages). Studies of transgenic animals will be tracked as mechanistic studies under "potentially relevant supplemental material."
Exposures	**Human:** Any exposure to naphthalene (CASRN 91-20-3), including occupational exposures.
	Animal: Any exposure to naphthalene (CASRN 91-20-3) via oral or inhalation, route[s] for ≥30 days. Non-developmental studies with exposures < 30 days in duration will only be included in the systematic review for a given health effect if longer duration studies are not available or if they contribute critical information to the weight of evidence or dose-response analysis. Studies exposing animals during developmental lifestages (e.g., gestational exposure) will be included regardless of exposure duration. Studies involving exposures to mixtures will be included only if they include an arm with exposure to naphthalene alone. Other exposure routes, including injection and dermal, will be tracked during title and abstract screening and tagged as "supplemental information."
	Studies describing physiologically-based pharmacokinetic (PBPK) models for naphthalene will be included.

PECO element	Evidence
Comparators	**Human:** A comparison or referent population exposed to lower levels (or no exposure/exposure below detection limits) of naphthalene, or exposure to naphthalene for shorter periods of time.
	Animal: A concurrent control group exposed to vehicle-only treatment.
Outcomes	Health outcomes: respiratory system, hematological, immune system, reproductive system, developmental, and cancer. In general, endpoints related to clinical diagnostic criteria, disease outcomes, histopathological examination, or other apical/phenotypic outcomes will be prioritized for evidence synthesis over outcomes such as biochemical measures.

5.2. UNITS OF ANALYSIS FOR DEVELOPING EVIDENCE SYNTHESIS AND INTEGRATION JUDGMENTS FOR HEALTH EFFECT CATEGORIES

The planned units of analysis based on outcomes identified in the assessment PECO criteria are summarized in Table 5-2. General considerations for defining the units of analysis are presented in the IRIS Handbook (U.S. EPA, 2022). Each unit of analysis is initially synthesized and judged separately within an evidence stream (see Section 8.1). Evidence integration judgments focus on the stronger within evidence stream synthesis conclusions when multiple units of analysis are synthesized. The evidence synthesis judgments are used alongside other key considerations (i.e., human relevance of findings in animal evidence, coherence across evidence streams, information on susceptible populations or lifestages, and other critical inferences that draw on mechanistic evidence) to draw an overall evidence integration judgment for each health effect category or more granular health outcome grouping (see Section 8.2). As new evidence to inform potential naphthalene-associated health hazards become available, the assessment team will consider updates to the units of analysis as appropriate.

Table 5-2. Health effect categories and human and animal evidence unit of analysis endpoint groupings for which evidence integration judgments will be developed for naphthalene

Health Effect Categories for Evidence Integration	Units of Analysis for Evidence Synthesis That Inform Evidence Integration (Each bullet represents a unit of analysis)	
	Human Evidence	**Animal Evidence**
Respiratory	• Any noncancer respiratory outcomes	• Pulmonary lesions • Nasal/olfactory lesions • Lung weight
Hematological	• Hematological evaluations of red blood cells, platelets, and clotting factors	• Hematological evaluations of red blood cells, platelets, and clotting factors
Immune	• Functional immune measures of sensitization or allergic response (asthma, dermal and nasal allergic measures)	• Functional immunotoxicity battery • Leukocyte counts • Thymus and spleen weights

Health Effect Categories for Evidence Integration	Units of Analysis for Evidence Synthesis That Inform Evidence Integration (Each bullet represents a unit of analysis)	
	Human Evidence	**Animal Evidence**
	• Observable immune measures of sensitization or allergic response (e.g., leukocyte counts, cytokine secretion) • Immunosuppression	• Histopathology of lymph nodes, thymus, and spleen
Reproductive	• Sperm/semen parameters • Reproductive hormones • Preterm birth	• Pregnancy outcomes (pregnant at sacrifice/premature delivery, maternal body weight) • Gonad weights • Histopathology of male and female reproductive organs
Developmental	• Fetal growth (e.g., birth weight, birth length) • Neurodevelopment *Maternal-fetal parameters described in the analysis of reproductive outcomes (preterm birth, cord blood hormone levels) may also be used to support the analysis of developmental outcomes.	• Fetal viability (live and dead fetuses, implantations, resorptions) • Fetal body weight • Fetal structural alterations • Postnatal growth and viability *An analysis of dam health (e.g., weight gain, food consumption) is also conducted to support conclusions of specificity of the effects as being developmental (versus derivative of maternal toxicity). Exposure during pregnancy can affect both the mother and the fetus, and it is frequently not possible to determine whether effects on the fetus are in response to or separate from maternal toxicity in studies that report both. The maternal endpoints in animal toxicology studies described in this section (maternal body weight gain and gestation length) must therefore be considered in conjunction with the fetal endpoints (survival, growth, and structural alterations)
Carcinogenicity	• Lung cancer	• Pulmonary tumors or precancerous lesions • Nasal tumors or precancerous lesions

5.3. CONSIDERATION OF SUPPLEMENTAL MATERIAL

5.3.1. Toxicokinetic (ADME) Information

Naphthalene toxicity is related to the production of reactive metabolites in the body (naphthalene 1,2-oxide; 1,2-naphthoquinone; and 1,4-naphthoquinone). The analysis of interspecies differences that could affect the formation and elimination of these toxic metabolites was identified as a key science issue during problem formulation (Section 2.4). The studies identified as "Toxicokinetic (ADME)" in the literature search will be reviewed and synthesized with focus on interspecies differences, such as CYP enzyme activity, that could affect the biological plausibility of these toxic metabolites being formed in humans.

5.3.2. Mechanistic Information

The analysis of biological processes underlying naphthalene-induced tumor formation was identified as a key science issue during problem formulation (see Section 2.4). Studies tagged as containing mechanistic information will be inventoried to identify and organize data that can be used to support the analysis of cancer MOA in the context of toxic naphthalene metabolite formation.

5.3.3. Case Studies

Human case studies exist for naphthalene that may provide relevant supporting information for hazard identification. For instance, case reports have documented laryngeal cancer among workers in a German naphthalene purification plant (Wolf, 1978, 1976) and colorectal cancer among Nigerian patients with a history of taking a naphthalene-containing indigenous treatment (Ajao et al., 1988). Hemolytic anemia has been frequently documented in case reports of individuals exposed to naphthalene, particularly among children who have ingested mothballs and in infants whose clothing or bedding was stored in mothballs (ATSDR, 2005). The case reports identified in the literature search for naphthalene will be inventoried to capture information on the study populations and the types of health effects observed and may be used to supplement the human evidence syntheses.

6. STUDY EVALUATION (RISK OF BIAS AND SENSITIVITY)

The general approach for evaluating primary health effect studies that meet assessment PECO criteria is described in Section 6.1. Instructional and informational materials for study evaluations are available at https://hawcprd.epa.gov/assessment/100000039/. The approach is conceptually the same for epidemiology, controlled human exposure, animal toxicology, and in vitro studies but the application specifics differ; thus, they are described separately in Sections 6.2, 6.3 and 6.5, respectively. Any physiologically based PBPK models used in the assessment are evaluated using methods described in the Quality Assurance Project Plan for PBPK models (U.S. EPA, 2018d), which is summarized in Section 6.4.

6.1. STUDY EVALUATION OVERVIEW FOR HEALTH EFFECT STUDIES

The IRIS Program uses a domain-based approach to evaluate studies. Key concerns for the review of epidemiology and animal toxicology studies are potential bias (factors that affect the magnitude or direction of an effect in either direction) and insensitivity (factors that limit the ability of a study to detect a true effect; low sensitivity is a bias toward the null when an effect exists). The study evaluations are aimed at discerning the expected magnitude of any identified limitations (focusing on limitations that could substantively change a result), considering the expected direction of the bias. The study evaluation approach is designed to address a range of study designs, health effects, and chemicals. The general approach for reaching an overall judgment regarding confidence in the reliability of the results is illustrated in Figure 6-1.

(a) Individual evaluation domains

Epidemiology	Animal	In vitro
• Exposure measurement • Outcome ascertainment • Participant selection • Confounding • Analysis • Selective reporting • Sensitivity	• Reporting quality • Allocation • Observational bias/blinding • Confounding • Selective reporting and attrition • Chemical administration and characterization • Exposure timing, frequency, and duration • Endpoint sensitivity and specificity • Results presentation	• Reporting quality • Observational bias/blinding • Variable control • Specificity • Selective reporting • Chemical administration and characterization • Exposure timing, frequency, and duration • Endpoint sensitivity • Results presentation and analysis

(b) Domain level judgments and overall study rating

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 are 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 they are unlikely to have a significant impact on results.
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 uninterpretable but may be used to highlight possible research gaps.

Figure 6-1. Overview of Integrated Risk Information System (IRIS) study evaluation approach. (a) individual evaluation domains organized by evidence type, and (b) individual evaluation domains, judgments, and definitions for overall ratings (i.e., domain and overall judgments are performed on an outcome-specific basis).

To calibrate the assessment specific considerations, the study evaluation process includes a pilot phase to assess and refine the evaluation process. Following this pilot, at least two reviewers independently evaluate studies to identify characteristics that bear on the informativeness (i.e., validity and sensitivity) of the results. The independent reviewers use structured web-forms for study evaluation housed within EPA's version of HAWC (https://hawc.prd.epa.gov) to record separate judgments for each domain and the overall study for each outcome and unit of analysis, to reach consensus between reviewers, and when necessary, resolve differences by discussion between the reviewers or consultation with additional independent reviewers. As reviewers examine a group of studies, additional chemical specific knowledge or methodological concerns could emerge, and a second pass of all pertinent studies might become necessary.

In general, considerations for reviewing a study with regard to its conduct for specific health outcomes are based on considerations presented in the IRIS Handbook (U.S. EPA, 2022) and use of existing guideline documents when available, including EPA guidelines for carcinogenicity, neurotoxicity, reproductive toxicity, and developmental toxicity (U.S. EPA, 2005a, 1998a, 1996, 1991).

Authors might be queried to obtain critical information, particularly that involving missing key study design, results information, or additional analyses that could address potential study limitations. During study evaluation, the decision on whether to seek missing information focuses on information that could result in a reevaluation of the overall study confidence for an outcome. Outreach to study authors is documented in HAWC and considered unsuccessful if researchers do not respond to an email or phone request within one month. Only information or data that can be made publicly available (e.g., within HAWC or HERO) will be considered.

When evaluating studies that examine more than one outcome, the evaluation process is explicitly conducted at the individual outcome level within the study. Thus, the same study may have different outcome domain judgments for different outcomes. These measures could still be grouped for evidence synthesis.

During review, for each evaluation domain, reviewers reach a consensus judgment of *good*, *adequate*, *deficient*, *not reported*, or *critically deficient*. If a consensus is not reached, a third reviewer performs conflict resolution. It is important to emphasize that evaluations are performed in the context of the study's utility for identifying individual hazards. Limitations specific to the usability of the study for dose-response analysis are useful to note and applicable to selecting studies for that purpose (see Section 9), but they do not contribute to the study confidence classifications. These four categories are applied to each evaluation domain for each outcome considered within a study, as follows:

- *Good* represents a judgment that the study was conducted appropriately in relation to the evaluation domain, and any minor deficiencies noted are not expected to influence the study results or interpretation of the study findings.

- *Adequat*e indicates a judgment that methodological limitations related to the evaluation domain are (or are likely to be) present, but those limitations are unlikely to be severe or to notably impact the study results or interpretation of the study findings.

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

- *Not reported* indicates the information necessary to evaluate the domain question was not available in the study. Depending on the expected impact, the domain may be interpreted as *adequate* or *deficient* for the purposes of the study confidence rating.

- *Critically deficient* reflects a judgment that the study conduct relating to the evaluation domain introduced a serious flaw that is interpreted to be the primary driver of any observed effect(s) or makes the study uninterpretable. Studies with *critically deficient* judgments in any evaluation domain are almost always classified as overall *uninformative* for the relevant outcome(s).

Once the evaluation domains are rated, the identified strengths and limitations are considered collectively to reach a study confidence classification of *high*, *medium*, or *low* confidence, or *uninformative* for each specific health outcome(s). This classification is based on the reviewer judgments across the evaluation domains and considers the likely impact that the noted deficiencies in bias and sensitivity have on the outcome-specific results. There are no pre-defined weights for the domains, and the reviewers are responsible for applying expert judgment to make this determination. The study confidence classifications, which reflect a consensus judgment between reviewers, are defined as follows:

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

- *Medium* confidence: Possible deficiencies or concerns were identified, but the limitations are unlikely to have a significant impact on the study results or their interpretation. Generally, *medium* confidence studies include *adequate* or *good* judgments across most domains, with the impact of any identified limitation not being judged as severe.

- *Low* confidence: Deficiencies or concerns are identified, and the potential for bias or inadequate sensitivity is expected to have a significant impact on the study results or their interpretation. Typically, *low* confidence studies have a *deficient* evaluation for one or more domains, although some *medium* confidence studies might have a *deficient* rating in domain(s) considered to have less influence on the magnitude or direction of effect estimates. *Low* confidence results are given less weight compared to *high* or m*edium* confidence results during evidence synthesis and integration (see Sections 7 and 8) and are generally not used as the primary sources of information for hazard identification or derivation of toxicity values unless they are the only studies available (in which case, this significant uncertainty would be emphasized during dose-response analysis). Studies rated *low* confidence only because of sensitivity concerns are asterisked or otherwise noted because they often require additional consideration during evidence synthesis. Effects

observed in studies that are biased toward the null may increase confidence in the results, assuming the study is otherwise well conducted (see Section 8).

- *Uninformative*: Serious flaw(s) are judged to make the study results uninterpretable for use in the assessment. Studies with *critically deficient* judgments in any evaluation domain are almost always rated *uninformative*. Studies with multiple *deficient* judgments across domains may also be considered *uninformative*. Given that the findings of interest are considered uninterpretable based on the identified flaws (see above definition of *critically deficient*) and do not provide information of use to assessment interpretations, these studies have no impact on evidence synthesis or integration judgments and are not useable for dose-response analyses but may be used to highlight research gaps.

As previously noted, study evaluation determinations reached by each reviewer and the consensus judgment between reviewers are recorded in HAWC. Final study evaluations housed in HAWC are made available when the draft is publicly released. The study confidence classifications and their rationales are carried forward and considered as part of evidence synthesis (see Section 11) to help interpret the results across studies.

6.2. EPIDEMIOLOGY STUDY EVALUATION

Evaluation of epidemiology studies of health effects to assess risk of bias and study sensitivity are conducted for the following domains: exposure measurement, outcome ascertainment, participant selection, potential confounding, analysis, study sensitivity, and selective reporting. Bias can result in false positives and negatives (i.e., Types I and II errors), whereas study sensitivity is typically concerned with identifying the latter.

The principles and framework used for evaluating epidemiology studies are adapted from the principles in the Cochrane Risk of Bias in Nonrandomized Studies of Interventions [ROBINS-I; Sterne et al. (2016)] but modified to address environmental and occupational exposures. The types of information that may be the focus of those criteria are listed in Table 6-1. Core and prompting questions, presented in Table 6-2, are used to collect information to guide evaluation of each domain. Core questions represent key concepts while the prompting questions help the reviewer focus on relevant details under each key domain. Exposure- and outcome-specific criteria to use during study evaluation are developed using the core and prompting questions and refined during a pilot phase with engagement from topic-specific experts. The protocol may also be adjusted in the early phases of the study evaluation process if corrections are identified based on initial literature reviews. Exposure domain considerations specific to naphthalene are presented in Sections 6.2.1 to 6.2.2.

Table 6-1. Information relevant to evaluation domains for epidemiology studies

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

1

Table 6-2. 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? • Is the exposure measurement likely to be affected by a knowledge of the outcome? • Is the exposure measurement likely to be affected by the presence of the outcome (i.e., reverse causality)? For case-control studies of occupational exposures: • Is exposure based on a comprehensive job history describing tasks, setting, time period, and use of specific materials? For biomarkers of exposure, general population: • Is a standard assay used? What are the intra- and inter-assay coefficients of variation? Is the assay likely to be affected by contamination? Are values less than the limit of detection dealt with adequately? • What exposure time period is reflected by the biomarker? If the half-life is short, what is the correlation between serial measurements of exposure?	Is the degree of exposure misclassification likely to vary by exposure level? If the correlation between exposure measurements is moderate, is there an adequate statistical approach to ameliorate variability in measurements? If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations require customization to the exposure and outcome (relevant timing of exposure). **Good** • Valid exposure assessment methods used, which represent the etiologically relevant time period of interest. • Exposure misclassification is expected to be minimal. **Adequate** • Valid exposure assessment methods used, which represent the etiologically relevant time period of interest. • Exposure misclassification may exist but is not expected to greatly change the effect estimate. **Deficient** • Valid exposure assessment methods used, which represent the etiologically relevant time period of interest. Specific knowledge about the exposure and outcome raise concerns about reverse causality, but there is uncertainty about whether it is influencing the effect estimate. • Exposed groups are expected to contain a notable proportion of unexposed or minimally exposed individuals, the method did not capture important temporal or spatial variation, or there is other evidence of exposure misclassification that would be expected to notably change the effect estimate. **Critically deficient** • Exposure measurement does not characterize the etiologically relevant time period of exposure or is not valid. • There is evidence that reverse causality is very likely to account for the observed association. • Exposure measurement was not independent of outcome status.

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

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

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

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
Continued:	Continued:	Continued:	Continued: **Deficient** • Does not include variables in the models that are likely to be influential colliders or intermediates on the causal pathway. And any of the following: • The potential for bias to explain some of the results is high based on an inability to rule out residual confounding, such as a lack of demonstration that key confounders of the exposure-outcome relationships are considered; • Descriptive information on key confounders (e.g., their relationship relative to the outcomes and exposure levels) is not presented; or • Strategy of evaluating confounding is unclear or is not recommended (e.g., only based on statistical significance criteria or stepwise regression [forward or backward elimination]). **Critically deficient** • Includes variables in the models that are colliders and/or intermediates in the causal pathway, indicating that substantial bias is likely from this adjustment, or • Confounding is likely present and not accounted for, indicating that all of the results are most likely due to bias. ○ Presenting a progression of model results with adjustments for different potential confounders, if warranted.

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

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
Continued:	Continued:	Continued:	Continued: **Deficient** • Does not conduct analysis using optimal characterization of the outcome variable. • Descriptive information about exposure levels is not provided (where applicable). • Effect estimate and *p*-value are presented, without standard error or confidence interval. • Results are presented as statistically "significant"/"not significant." **Critically deficient** • Results of analyses of effect modification are 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.

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
Selective reporting Is there reason to be concerned about selective reporting?	• Are results provided for all the primary analyses described in the methods section? • Is there appropriate justification for restricting the amount and type of results that are shown? • Are only statistically significant results presented?	If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations generally do not require customization and 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 are reported for all primary analyses. **Deficient** • Concerns are raised based on previous publications, a methods paper, or a registered protocol indicating that analyses are planned or conducted that are not reported, or that hypotheses originally considered to be secondary are represented as primary in the reviewed paper. • Only subgroup analyses are reported, suggesting that results for the entire group are omitted. • Only statistically significant results are reported.

Sensitivity Is there a concern that sensitivity of the study is not adequate to detect an effect?	• Is the exposure range adequate to detect associations and exposure-response relationships? • Was the appropriate population included? • Was the length of follow-up adequate? Is the time/age of outcome ascertainment optimal given the interval of exposure and the health outcome? • Are there other aspects related to risk of bias or otherwise that raise concerns about sensitivity?		These considerations may require customization to the exposure and outcome. Depending on the needs of the assessment, there may be fewer than four rating levels. Some study features that affect study sensitivity may have already been included in the other evaluation domains; these should be noted in this domain again, along with any features that have not been addressed elsewhere so that the rating provides an overall summary of factors that may impact sensitivity. When determining the overall study confidence rating, the evaluator should be conscious that a limitation could contribute to multiple domains and not double-penalize the study. Some considerations include: **Good** • The range of exposure levels provides sufficient variability in exposure distribution and/or sufficient range or contrasts (e.g., across groups or exposure categories) to detect associations or exposure-response relationships that may be present. • The population was exposed to levels expected to have an impact on response. • The study population was at risk of developing the outcomes of interest (e.g., ages, life stage, sex). • The timing of outcome ascertainment was appropriate given expected latency for outcome development (i.e., adequate follow-up interval). • There was evidence of sufficient statistical power (which may include formal power calculations) to observe an effect if one exists. • No other concerns raised regarding study sensitivity (e.g., no evidence that results would be attenuated enough to preclude detection of an adverse health effect). **Adequate** • Same considerations as *good*, except: ○ Issues are identified that could reduce sensitivity, but they are unlikely to impact the overall findings of the study. **Deficient** • Concerns were raised about the issues described for *good* that are expected to notably decrease the sensitivity of the study to detect associations for the outcome (i.e., reasonably high likelihood of a false null result).

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
			• Note: *Deficient* sensitivity indicates that null findings should be interpreted with caution and may not represent a lack of association. **Critically deficient** • Severe concerns were raised about the sensitivity of the study such that any observed association is uninterpretable (e.g., exposure gradients/contrasts that precluded an ability to distinguish exposure levels between study participants).

1

For evaluation of the exposure measures domain, studies in which human exposure is quantified in the air or in urinary biomarker measurements will be preferred. Studies where naphthalene exposure is inferred but not confirmed by quantitative measurements will be given lower preference. Studies that only use measurements of JP-8 jet fuel rather than naphthalene alone in regression analyses will be marked as potentially relevant supplemental material, given the concerns with confounding due to the diverse components of the jet fuel.

6.2.1. Air monitoring or modeling

Naphthalene can exist in both the vapor and particulate phases, but more than 95% is anticipated to occur in the vapor phase (Lai et al., 2009; Eiguren-Fernandez et al., 2004; Fang et al., 2004; Harrison et al., 1996). The half-life of naphthalene in the atmosphere is less than 1 day (ATSDR, 2005); specific data about the half-life in indoor environments were not identified but would depend on concentrations of hydroxyl radicals present (ATSDR, 2005). Naphthalene concentrations may be higher in indoor air than outdoor air due to certain exposure sources, such as mothballs or paint (WHO, 2010; ATSDR, 2005). In these situations where indoor sources are expected to dominate, measurement of naphthalene concentrations in indoor air is preferred over outdoor air estimates alone. In general, however, due to the relevance of both indoor and outdoor sources, individual-level exposure assessments for health effects studies ideally would capture contributions from time at home, school or work, and in-transit. For this reason, individual-level or time-weighted summaries are preferred over area-level monitoring that does not incorporate individual movement/behaviors and the potential contribution of multiple sources.

The effectiveness of air monitoring for naphthalene depends on the approach (active vs. passive) and the sorbent utilized. Passive sampling approaches require long sampling times in situations with low PAH concentrations and low sensitivity of analytical methods. With regard to sorbent, the U.S. EPA Compendium Method TO-13A for PAHs (including naphthalene) in ambient air allows for either a polyurethane (PUF) or XAD-2 adsorbent cartridge (U.S. EPA, 1999). However, PUF has a lower recovery efficiency for naphthalene and may result in an underestimate of airborne concentrations, particularly with passive sampling (Strandberg et al., 2018; Chuang et al., 1987). Therefore, XAD-2 active sampling is the preferred method for naphthalene sampling (Piñeiro et al., 2021). Additional methods are described in (EIC, 2015).

The time frame represented by the exposure estimates should correspond to the period in which the health outcomes were expected to have developed. Indoor exposure assessments representing a period of week(s) in more than one season could reasonably characterize average exposure over the previous year and would be relevant to immune-related or other symptoms (e.g., asthma, wheezing illness, allergy symptoms, sensory irritation) occurring over the previous several weeks to a year. Daily sampling is best, but periodic sampling on a less than daily basis could be sufficient depending on the variability in air concentrations. Shorter duration monitoring could be relevant for acute outcomes. Developmental outcomes should be evaluated in relation to the relevant critical exposure periods during pregnancy if they are known. Exposure measurements

with shorter time frames are less informative for studying the prevalence or incidence of chronic disease, such as physician-diagnosed asthma, cardiovascular disease, or cancer.

There has been limited use of modeling (e.g., land use regression (LUR)) to assess exposure to naphthalene (Lu et al., 2019). Primary concerns with these approaches are that they only capture potential outdoor sources of exposure and there is uncertainty regarding their validity or reliability given the lack of a robust literature base. As such, decisions regarding the appropriateness of modeling approaches will be made on a case-by-case basis based on the description of model development and how adequately the model characterizes spatial variation in the community.

6.2.2. Biomarker assessment

Urinary

When biomarkers of exposure are used to identify the presence of naphthalene, monohydroxylated metabolites of naphthalene are preferred. Alternative metabolites, such as dihydroxy urinary metabolites of naphthalene, are more challenging to quantify and analyze with current capabilities (Klotz et al., 2011) but may be more reliable in the future. With regard to monohydroxylated metabolites, studies measuring 2-naphthol are preferred versus studies measuring 1-naphthol only. 1-Naphthol is a metabolite of both naphthalene and the pesticide carbaryl (one of the most commonly used insecticides in home and garden settings, with widespread low-level exposure expected across the population), and therefore is a less specific biomarker of naphthalene exposure compared to 2-naphthol. Measurement of 1-naphthol may be appropriate if the study uses approaches to distinguish between source (e.g., naphthalene vs. carbaryl) (Meeker et al., 2007). Naphthalene metabolites measured in urine may reflect internal dose and can be utilized as sensitive biomarkers of exposure if specific metabolites are measured in relation to etiologically relevant periods. However, because the half-life of naphthalene in the body is short [4 hours (ATSDR, 2005)] and the metabolites are excreted rapidly, there are temporal variations in urinary metabolite levels relative to the timeframe of exposure. A single spot urine sample therefore may not be a reliable surrogate for longer-term exposure. This question of reproducibility of biomarker measures over time has been discussed for other environmental exposures, such as phthalates (Radke et al., 2018; Johns et al., 2015). The intraclass correlation coefficient (ICC), a measure of reliability, for naphthalene metabolites in urine has been reported in a variety of populations and in a variety of settings as approximately 0.3-0.7 (Zhu et al., 2021; Cathey et al., 2018; Dobraca et al., 2018; Yang et al., 2017a; Wheeler et al., 2014; Li et al., 2010), though poorer reproducibility has also been reported (Yang et al., 2017b). While 2-naphthol is a more specific marker of naphthalene exposure, it sometimes – but not always – has a lower ICC than 1-naphthol in a sample of examined studies. If results are available for both metabolites, consistent patterns across both would provide more confidence in drawing conclusions. Overall, use of a single spot sample to reflect longer term exposure is likely to induce non-differential exposure misclassification into the analysis (which, in most cases, would produce bias towards the

null). Use of pooled samples over multiple days is preferred over a spot sample from a single day (Perrier et al., 2016).

Overall, judgement of the adequacy of a spot urine sample depends in part upon whether the exposure source is expected to be consistent over time and whether the sample falls within the etiologically relevant time period. There is more concern regarding the appropriateness of spot samples for chronic compared to acute outcomes. General guidelines are provided in the table below.

Nonurinary

Most studies evaluating PAH exposure, such as naphthalene, measure the concentration of PAH metabolites in urine, as PAHs are metabolized rapidly in the body (Yin et al., 2017). Other potential biomarkers of exposure include umbilical cord blood, breast milk, and placenta tissue; however, there is currently limited information on the usefulness of these measures as exposure biomarkers for naphthalene in epidemiological research (Powers, 2022). Combined with the short elimination half-life of naphthalene in the body, biomarkers other than urine will be rated as *critically deficient*.

Additionally, some studies have used unmetabolized PAHs to measure body burden (De Craemer et al., 2016). Because of the short half-life of PAH parent compounds, the appropriate quantification approach is to measure metabolites. Therefore, studies attempting to quantify naphthalene burden through assessment of the parent compound only will be rated *critically deficient.*

Table 6-3. Evaluation of exposure biomarkers in general population studies of naphthalene (adapted from Phthalates SR protocol) (Radke et al., 2018)

Level	Criteria	
	Biomarkers	**Air**
Good	• Two or more urine samples within the etiologically relevant period [i.e., temporality is established, and sufficient latency occurred before disease onset] for development of the outcome based on current biological understanding) **and** • Measurement of 2-naphthol metabolites in urine **and** • Discussion of laboratory QC procedures **or** no discussion of laboratory QC procedures but analysis	• Integrated personal measurements or time-weighted summary concentrations incorporating concentrations in residence and school/workplace **and** • Appropriate and validated methods used for sampling (e.g., NMAM 5528, TO13A, XAD) and analysis (e.g., GC/MS, HPLC). Sampling details provided (e.g., type of samplers, placement of samplers, sampling periods, status of activities in structures, chemical analysis methods (or citation provided). Validation with paired tests to ensure consistency. Calibration of

Level	Criteria	
	Biomarkers	**Air**
	by an experienced laboratory (e.g., Centers for Disease Control and Prevention [CDC])	automated instruments if relevant. Sufficient samples above the LOD **and** • Time-frame of measurements appropriate to development of health outcome
Adequate	• One urine sample within etiologically relevant period for development of outcome **and** • Measurement of 2-naphthol metabolites in urine or measurement of 1-naphthol with methods to distinguish original source **and** Evidence that exposure was consistently assessed using methods described in Good, but there were some concerns about quality control measures or other potential for nondifferential misclassification	• Area measurements in home, average of measurements in 1 or more rooms; over multiple seasons if estimating annual average **and** • Appropriate and validated methods used (e.g., NMAM 5528, XAD) and analysis (e.g., GC/MS, HPLC). Sufficient samples above the LOD. Sampling details provide adequate level of confidence in approach, though less detailed provided than for "Good" above **and** • Time-frame of measurements appropriate to development of health outcome **Or** • Average estimates based on land use regression models developed for location where study was conducted including description of model development and sufficient information about how the model adequately characterizes spatial variation in the community. Potentially other methods besides LUR might fall into this category if detailed validation information was provided to ensure model adequately characterizes spatial variation **and** • Time-frame of modeling relevant to the development of health outcome
Deficient	• One urine sample; sample collection may be outside the etiologically relevant period and/or there is some concern for reverse causation	*For monitoring:* • Monitoring with PUF adsorbent cartridge, or an approach that may not be fully appropriate or validated

Level	Criteria	
	Biomarkers	**Air**
	Or Measurement of 1-naphthol metabolites in urine without methods to account for original source **or** Concerns with QC/QA	• Monitoring of outdoor air concentrations only if indoor sources are expected to dominate • Area measurements in home obtained on one occasion but study is estimating annual average **Or** *For modeling:* • Average estimates based on land use regression models developed for location where study was conducted, but some uncertainties remain regarding how the model was developed or how the model adequately characterizes spatial variation in the community due to what was known about sources • Estimates based on other modeling approaches (e.g., NATA, CMAQ) with more limited ability to accurately capture spatial/temporal variation **Or** • Use of questionnaires or observations of sources in the home by trained study personnel
Critically Deficient	• Biomarker measured in tissue other than urine **or** • Clear concern for reverse causation would make the results uninterpretable	• No explanation or insufficient detail provided about air monitoring or modeling methods • Air monitoring or modeling occurred outside of a relevant window for health outcome of interest • Use of air monitoring approach that has not been validated for naphthalene or does not sufficiently capture spatial/temporal variation • Technical issues during monitoring (e.g., inconsistency during sampling, pump faults from overloading)

6.3. EXPERIMENTAL ANIMAL STUDY EVALUATION

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

Table 6-3 provides the standard domains and core questions along with some basic considerations for guiding the evaluation. Each domain receives a consensus judgment of *Good, Adequate, Deficient, Not Reported,* or *Critically Deficient* (as described in Section 6.1) accompanied by a rationale for the judgment. Once all domains are rated, an overall confidence classification of *High, Medium,* or *Low* confidence or *Uninformative* is assigned (as described in Section 6.1). The rationale for the classification, including a brief description of any identified strengths and/or limitations from the domains and their potential impact on the overall confidence determination, should be documented clearly and consistently. This rationale should, to the extent possible, reflect an interpretation of the potential influence on the results (including the direction and/or magnitude of influence).

Table 6-4. Questions to guide the development of criteria for each domain in experimental animal toxicology studies

Evaluation type	Domain name – core question	Prompting questions	Basic 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?
	
	Note: Reviewers should reach out to authors to obtain missing information when studies are considered key for hazard evaluation and/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 life stage 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 by assessment teams, although in some instances the **important information** may be refined depending on the endpoints/outcomes of interest or the chemical under investigation.
			
			A judgment and rationale for this domain should be given for the study. Typically, these will not change regardless of the endpoints/outcomes investigated by the study. **In the rationale, reviewers should indicate whether the study adhered to GLP, 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 any pieces of **critical information**. Studies that are Critically Deficient for reporting are Uninformative for the overall rating and not considered further for evidence synthesis and integration.

Evaluation type		Domain name – core question	Prompting questions	Basic 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)? • Is the allocation method described? • Aside from randomization, were any steps taken to balance variables across experimental groups during allocation?	These considerations typically do not need to be refined by assessment teams. A judgment and rationale for this domain should 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: 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 type		Domain name – core question	Prompting questions	Basic considerations

Risk of bias	**Selection and performance bias**	**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 by the assessment teams.
				
				Note: It can be useful for teams to identify highly subjective measures of endpoints/outcomes where observational bias may strongly influence results prior to performing evaluations.
				
				A judgment and rationale for this domain should 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).
				• *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 could have impacted results.

Evaluation type		Domain name – core question	Prompting questions	Basic considerations
Risk of bias	**Confounding/variable control**	**Confounding** Are variables with the potential to confound or modify results controlled 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 may need to be refined by assessment teams, as the specific variables of concern can vary by experiment or chemical. A judgment and rationale for this domain should be given for each cohort or experiment in the study, noting when the potential for confounding is restricted to specific endpoints/outcomes. • *Good*: Outside of the exposure of interest, variables that are likely to confound or modify results appear to be controlled and consistent across experimental groups. • *Adequate*: Some concern that variables that were 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 type		Domain name – core question	Prompting questions	Basic considerations
Risk of bias	**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 under core question)? *Attrition bias:* • Are all animals accounted for in the results? • If there are discrepancies, do authors provide an explanation (e.g., death or unscheduled sacrifice during the study)? • If omitted results and/or attrition are unexplained, what is the expected impact on the interpretation of the results?	These considerations typically do not need to be refined by assessment teams. A judgment and rationale for this domain should 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 timepoints. Data not reported in the primary article is 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 and/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 and/or high animal attrition; omissions and/or attrition are not explained and may significantly impact the interpretation of the results. • *Critically Deficient*: Extensive results omission and/or animal attrition are identified and prevent comparison of results across treatment groups.

Sensitivity	**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 and/or composition (e.g., identity and percent distribution of different isomers) of the chemical? If not, can the purity and/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?
			
			  ○ For inhalation studies: Were target concentrations confirmed using reliable analytical measurements in chamber air?
			
			  ○ For oral studies: If necessary, based on consideration of chemical-specific knowledge (e.g., instability in solution; volatility) and/or exposure design (e.g., the frequency and duration of exposure), were chemical concentrations in the dosing solutions or diet analytically confirmed?
			
			• Are there concerns about the methods used to administer the chemical (e.g., inhalation chamber type, gavage volume, etc.)?
				It is essential that these criteria are considered, and potentially refined, by assessment teams, as the specific variables of concern can vary by chemical.
				
				A judgment and rationale for this domain should 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).
				
				• *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 and/or life stage 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).

Evaluation type		Domain name – core question	Prompting questions	Basic considerations
Sensitivity	**Exposure methods sensitivity**	**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 by assessment teams. A judgment and rationale for this domain should 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 and/or frequency of the exposure is not sensitive and did not include most of the critical window of sensitivity (if known). These limitations are expected to bias the results towards the null. • *Critically Deficient*: The exposure design was not sensitive and is expected to strongly bias the results towards the null. The rationale should indicate the specific concern(s).

Evaluation type		Domain name – core question	Prompting questions	Basic considerations
Sensitivity	**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.*	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 (see note)? • 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 by assessment teams. A judgment and rationale for this domain should 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, etc.) • 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 non-exposure before testing)

Evaluation type		Domain name – core question	Prompting questions	Basic considerations
Sensitivity	**Outcome measures and results display**	**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 a way that is inappropriate or misleading?	Considerations for this domain are highly variable depending on the outcomes of interest and must be refined by assessment teams. A judgment and rationale for this domain should be given for each endpoint/outcome or group of endpoints/outcomes investigated in the study. Examples of potential concerns include: • Nonpreferred presentation, such as developmental toxicity data averaged across pups in a treatment group, when litter responses 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 type	Domain name – core question	Prompting questions	Basic 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 should mark studies that are rated lower than high confidence only due to low sensitivity (i.e., bias towards the null) for additional consideration during evidence synthesis. If the study is otherwise well-conducted and an effect is observed, the confidence may be increased.*	For each endpoint/outcome or grouping of 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 should be given for each endpoint/outcome or group of endpoints/outcomes investigated in the study. • *High Confidence*: No notable concerns are identified (e.g., most or all domains rated Good). • *Medium Confidence*: Some concerns are identified but expected to have minimal impact on the interpretation of the results (e.g., most domains rated Adequate or Good; may include studies with Deficient ratings if concerns are not expected to strongly impact the magnitude or direction of the results). Any important concerns should be carried forward to evidence synthesis. • *Low Confidence*: Identified concerns are expected to significantly impact on the study results or their interpretation (e.g., generally, Deficient ratings for one or more domains). The concerns leading to this confidence judgment must be carried forward to evidence synthesis (see note). • *Uninformative*: Serious flaw(s) that make the study results unusable for informing hazard identification (e.g., generally, a Critically Deficient rating in any domain; many Deficient ratings). Uninformative studies are not considered further in the synthesis and integration of evidence.

6.4. PHYSIOLOGICALLY BASED PHARMACOKINETIC (PBPK) MODEL EVALUATION

PBPK (or classical pharmacokinetic [PK])[6] models should be used in an assessment when an applicable one exists and no equal or better alternative for dosimetric extrapolation is available. Any models used should represent current scientific knowledge and accurately translate the science into computational code in a reproducible, transparent manner. For a specific target organ/tissue, it may be possible to employ or adapt an existing PBPK model or develop a new PBPK model or an alternate quantitative approach. Data for PBPK models could come from studies across various species and may be in vitro or in vivo in design.

Existing naphthalene PBPK models were identified through a literature search and are summarized in Appendix C. Of these, the model of Campbell et al. (2014) is the penultimate model in its lineage and it explicitly describes dosimetry for specific regions in the upper respiratory tract, which is a feature that distinguishes it from all previous models. Kapraun et al. (2020) extended the model of Campbell et al. (2014) by incorporating a skin route of exposure and demonstrated that their model could be used to reproduce human pharmacokinetic data; they also performed quality assurance procedures (U.S. EPA, 2018d) for their model. This most recently published naphthalene PBPK model (Kapraun et al., 2020) is therefore the clear choice for use in this assessment.

EPA has evaluated the Kapraun et al. (2020) model in accordance with criteria outlined by U.S. EPA (2018d). Judgments on the suitability of a model are separated into two categories: scientific and technical (see Table 6-5). The scientific criteria focus on whether the biology, chemistry, and other information available for chemical MOA(s) are justified (i.e., preferably with citations to support use) and represented by the model structure and equations. The scientific criteria are judged based on information presented in the publication or report that describes the model and do not require evaluation of the computer code. Preliminary technical criteria include availability of the computer code and completeness of parameter listing and documentation. Studies that meet the preliminary scientific and technical criteria are then subjected to an in-depth technical evaluation, which includes a thorough review and testing of the computational code. The in-depth technical and scientific analyses focus on the accurate implementation of the conceptual model in the computational code, use of scientifically supported and biologically consistent parameters in the model, and reproducibility of model results reported in journal publications and

[6] Note that the terms "pharmacokinetic" (adjective) and "pharmacokinetics" (noun), which are both abbreviated as "PK," are used in this document when discussing absorption, distribution, metabolism, and excretion (ADME) of a substance by an organism or any related quantities, experiments, or models. The terms "toxicokinetic" and "toxicokinetics," which are both abbreviated as "TK," are frequently used as synonyms for "pharmacokinetic" and "pharmacokinetics" in the literature, but the latter terms are used preferentially here for document-wide consistency. Also, PBPK models are sometimes described as "physiologically based toxicokinetic models" (abbreviated "PBTK models") or even as "physiologically based kinetic models" (abbreviated "PBK models") in the literature, but in this document the term "PBPK model" is used preferentially for purposes of consistency.

other documents. This approach stresses (1) clarity in the documentation of model purpose, structure, and biological characterization; (2) validation of mathematical descriptions, parameter values, and computer implementation; and (3) evaluation of each plausible dose metric. The in-depth analysis is used to evaluate the potential value and cost of developing a new model or substantially revising an existing one.

Table 6-5. 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., $BW^{3/4}$ scaling).
	• Model parameterization for critical life stages or windows of susceptibility. Evaluation of these criteria should also consider the model's fidelity vs. default approaches and possible use of an intraspecies uncertainty factor (UF) in conjunction with the model to account for variations in sensitivity between life stages.
	• Predictive power of model-based dose metric vs. default approach, based on exposure.
	  ○ Specifically, model-based metrics may correlate better than the applied doses with animal/human dose-response data.
	  ○ The degree of certainty in model predictions vs. default is also a factor. For example, while target tissue metrics are generally considered better than blood concentration metrics, lack of data to validate tissue predictions when blood data are available may lead to choosing the latter.

	Principle of parsimony • Model complexity or biological scale, including number and parameterization of (sub)compartments (e.g., tissue or subcellular levels) should be commensurate with data available to identify parameters.
	Model describes existing PK data reasonably well, both in "shape" (matches curvature, inflection points, peak concentration time, etc.) and quantitatively (e.g., within factor of 2−3).
	Model equations are consistent with biochemical understanding and biological plausibility.
Initial technical	Well-documented model code is readily available to EPA and the 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).

Category	Specific criteria
	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 may decide to independently conduct this additional work before using the model in the assessment. • A sound explanation should be provided when sensitivity of the dose metric to model parameters differs from what is reasonably expected based on experience.

6.5. IN VITRO STUDY EVALUATION

As described in Section 4.4, the initial literature screening identifies sets of other potentially informative studies, including mechanistic studies, as "potentially relevant supplemental information." Mechanistic information includes any experimental measurement related to a health outcome that informs the biological or chemical events associated with phenotypic effects; these measurements can improve understanding of the mechanisms involved in the biological effects following exposure to a chemical but are not generally considered by themselves adverse outcomes. Mechanistic data are reported in a diverse array of observational and experimental studies across species, model systems, and exposure paradigms, including in vitro, in vivo (by various routes of exposure), ex vivo, and in silico studies. Section 5.3.2 outlines an approach for the consideration of information from mechanistic studies where the specific analytical approach is targeted to the assessment needs depending on the extent and nature of the human and animal evidence.

Individual study-level evaluations of mechanistic endpoints are not typically pursued. This is because each identified study that reported mechanistic information would need to undergo a full evaluation of risk of bias and sensitivity before the relevant toxicity pathways are identified and the needs of the assessment are better understood, which would not be an effective use of time and resources. For some chemical assessments, however, it may be necessary to identify assay-specific considerations for study endpoint evaluations on a case-by-case basis to provide a more detailed summary and evaluation for the most relevant individual studies. This may be done, for example, when the scientific understanding of a critical mechanistic event or MOA is less established or lacks scientific consensus, the reported findings on a mechanistic endpoint are conflicting, the available mechanistic evidence addresses a complex and influential aspect of the assessment, or in vitro or in silico data make up the bulk of the evidence base and there is little or no evidence from epidemiological studies or animal bioassays.

If a subset of individual mechanistic studies is identified for evaluation, the study evaluation considerations will differ depending on the type of endpoints, study designs, and model systems or populations evaluated. It should be noted that because the evaluation process is outcome specific, overall confidence classifications for human or animal studies that have already been determined will not automatically apply to mechanistic endpoints if reported in the same study; a separate evaluation of the mechanistic endpoints should be performed as the utility of a study may vary for

the different outcomes reported. Developing specific considerations requires a familiarity with the studies to be evaluated and cannot be conducted in the absence of knowledge of the relevant study designs, measurements, and analytic issues. Knowledge of issues related to the hazards and the outcomes identified in the revised evaluation plan is also important for developing specific evaluation considerations. One challenge is that novel methodologies for studying mechanistic evidence are continuously being developed and implemented and often no "standard practices" exist.

The evaluation of mechanistic studies applies similar principles as those described above for the evaluation of epidemiological and experimental animal studies. Table 6- provides the standard domains and core questions for the evaluation of studies conducted in in vitro test systems, along with some basic considerations for guiding the evaluation. The evaluation process focuses on assessing aspects of the study design and conduct through three broad types of evaluations: reporting quality, risk of bias, and study sensitivity. Some domain considerations are tailored to the chemical and to the assay(s) or endpoint(s) being evaluated. Assessment teams work with subject matter experts to develop specific considerations. These specific considerations are determined prior to performing study evaluation, although they may be refined as the study evaluation proceeds (e.g., during pilot testing). Assessment- or assay-specific considerations are documented and made publicly available in the assessment.

Table 6-6. Domains, questions, and general considerations to guide the evaluation of in vitro studies

Domain and core question	Prompting questions	General considerations
Observational bias/blinding Did the study implement measures, where possible, to reduce observational bias? Considerations will vary depending on the specific assay/model system being used and may not be applicable to some analyses.	For each assay or endpoint in a study: Did the study report steps taken to minimize observational bias during analysis (e.g., blinding/coding of slides or plates for analysis; collection of data from randomly selected fields; positive controls that are not immediately identifiable)? If not, did the study use a design or approach for which such procedures can be inferred, or which would not be possible to implement? Were the assays evaluated using automated approaches (e.g., microplate readers) that reduce concern for observational bias? 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 by the assessment teams. Prior to performing evaluations, teams should consider the specific assay to identify highly subjective measures of endpoints where observational bias may strongly influence results. A judgment and rationale for this domain should be given for each assay or endpoint or group of endpoints investigated in the study. **Good:** Measures to reduce observational bias were described (e.g., specific mention of blinding and/or coding of slides for analysis), or observational bias is not a concern because of use of automated/computer driven systems and/or standard laboratory kits. **Not reported, interpreted as adequate:** Measures to reduce observational bias were not described, but the potential concern for bias was mitigated because protocol cited includes a description of requirements for blinding/coding, or the impact on results is expected to be minor because the specific measurement is more objective. **Not reported, interpreted as *deficient***: No protocol cited; the potential impact on the results is major because the endpoint measures are highly subjective (e.g., counting plaques or live vs. dead cells). **Critically deficient:** Strong evidence for observational bias that could have impacted the results.
Variable Control	For each study:	These considerations will need to be refined by assessment teams as the specific variables of concern can vary by the experimental test system and chemical.

Domain and core question	Prompting questions	General considerations
Are all introduced variables with the potential to affect the results of interest controlled for and consistent across experimental groups?	Are there any known or presumed differences across treatment groups (e.g., co-exposures, culture conditions, cell passages, variations in reagent production lots, mycoplasma infections) that could bias the results? If differences are identified, to what extent are they expected to impact the results? Did the study address feature inherent to the physico-chemical properties of the test substance(s) that have the potential to bias the results away from the null? For example, could the test article interfere with a given assay (e.g., auto-fluoresces or inhibits enzymatic processes necessary for assay signals), potentially leading to an erroneous positive signal? *(Note that concerns related to dose are addressed in chemical administration and characterization.)* Are there known variations in cellular signaling unique to the model system that could influence the possibility of detecting the effect(s) of interest? Are there concerns regarding the negative (untreated and/or vehicle) controls used? Were negative controls run concurrently?	A judgment and rationale for this domain should be given for each experiment in the study, noting when the potential to affect results is restricted to specific assays or endpoints. Good: Outside of the exposure of interest, variables or features of the test system and/or chemical properties that are likely to impact results appear to be controlled for and consistent across experimental groups. **Adequate:** Some concern that variables or features of the test system and/or chemical properties that are likely to modify or interfere with results were uncontrolled or inconsistent across groups but are expected to have a minimal impact on the results. Deficient: Notable concern that important study variables and/or features of the test system lacked specificity or were uncontrolled or inconsistent across groups and are expected to substantially impact the results. **Critically deficient:** Features of the test system are known to be nonspecific for this endpoint, and/or influential study variables were presumed to be uncontrolled or inconsistent across groups and are expected to be a primary driver of the results.
Selective Reporting Did the study present results, quantitatively or qualitatively, for all prespecified assays or	For each study: Are results presented for all endpoints/outcomes described in the methods?	These considerations typically do not need to be refined by assessment teams. A judgment and rationale for this domain should be given for each assay or endpoint in the study.

Domain and core question	Prompting questions	General considerations
endpoints and replicates described in the methods? *Note: The appropriateness of the analysis or results presentation is considered under results presentation.*	Did the study clearly indicate the number of replicate experiments performed? Were the replicates technical (from the same sample) or independent (from separate, distinct exposures)? If unexplained results omissions are identified, what is the expected impact on the interpretation of the results?	**Good:** Quantitative or qualitative results were reported for all prespecified assays or endpoints (explicitly stated or inferred), exposure groups and evaluation timepoints. Data not reported in the primary article is available from supplemental material. If results omissions 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 assays or endpoints (explicitly stated or inferred), exposure groups and evaluation timepoints. Omissions 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 assays or endpoints (explicitly stated or inferred), exposure groups and evaluation timepoints; omissions are not explained and may significantly impact the interpretation of the results. **Critically Deficient:** Extensive results omissions are identified, preventing comparisons of results across treatment groups.
Chemical administration and characterization Did the study adequately characterize exposure to the chemical of interest and the exposure administration methods?	For each study: Are there concerns regarding the purity and/or composition (e.g., identity and percent distribution of different isomers) of the test material/chemical? If so, can the purity and/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	It is essential that these criteria are considered, and potentially refined, by assessment teams, as the specific variables of concern can vary by chemical (e.g., stability may be an issue for one chemical but not another). A judgment and rationale for this domain should be given for each 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.

Domain and core question	Prompting questions	General considerations
	performed? If not, is this a significant concern for this substance? Are there concerns about the stability of the test chemical in the vehicle and/or culture media (e.g., pH, solubility, volatility, adhesion to plastics) that were not corrected for, leading to potential bias away from the null (e.g., observed precipitate formation at high concentrations) or toward the null (e.g., enclosed chambers not used for testing volatile chemicals)? Are there concerns about the preparation or storage conditions of the test substance? Are there concerns about the methods used to administer the chemical?	**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). **Deficient:** Uncertainties in the exposure characterization are identified and expected to substantially impact the results (e.g., the source and purity of the test article are not reported, and no independent verification of the test article was conducted; levels of impurities are substantial or concerning; deficient administration methods were used). **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).

Endpoint measurement Are the selected protocols, procedures, and test systems adequately described and appropriate for evaluating the endpoint(s) of interest? *Notes:* *Considerations related to adjustments or corrections to endpoint measurements are addressed under results presentation.* *Considerations related to the sensitivity of the animal model and timing of endpoint measurement are evaluated under sensitivity.*	For each endpoint or grouping of endpoints in a study: Are the evaluation methods and test systems adequately described and appropriate? Are there concerns regarding the methodology selected (e.g., accepted guidelines, established criteria) for endpoint evaluation? Are there concerns about the specificity of the experimental design? Did the study address feature inherent to the test system or experiment that have the potential to lead to bias away from the null? Are there serious concerns about the number of replicates or sample size in the study? Are appropriate control groups for the study/assay type included? Was there a need for the assay to include specific controls to reduce potential sources of underlying bias? Did the test compound induce cytotoxicity (known, or expected based on other studies of similar design) to a degree that is expected to affect interpretation of results?	Considerations for this domain are highly variable depending on the assay or endpoint(s) of interest and must be refined by assessment teams. A judgment and rationale for this domain should be given for each assay or endpoint or group of endpoints investigated in the study. Some considerations include the following: **Good**: Adequate description of methods and test system. Use of generally accepted and reliable endpoint methods that are consistent with accepted guidelines or established criteria for the assay(s)/endpoint(s) of interest. Sample sizes are generally considered adequate for the assay or protocol of interest and there are no notable concerns about sampling in the context of the endpoint protocol. Includes appropriate control groups (e.g., use of loading controls) and any use of nonconcurrent or historical control data (e.g., for comparison to background levels in negative controls) is justified (e.g., authors or evaluators considered the similarity between current cell cultures and laboratory conditions to historical controls). Ratings of Adequate, Deficient, and Critically Deficient are generally defined as follows: **Adequate**: Issues are identified that may affect endpoint measurement but are considered unlikely to substantially impact the overall findings or the ability to reliably interpret those findings. **Deficient**: Concerns are raised that are expected to notably affect endpoint measurement and reduce the reliability of the study findings.

Domain and core question	Prompting questions	General considerations
		Critically deficient: Severe concerns are raised about endpoint measurement and any findings are likely to be largely explained by these limitations. The following specific examples of relevant concerns are typically associated with a Deficient rating, but Adequate or Critically Deficient might be applied depending on the expected impact of limitations on the reliability and interpretation of the results: Study report lacks important details that are necessary to evaluate the appropriateness of the study design (e.g., description of the assays or protocols; information on the cell line, passage number). Selection of protocols that are nonpreferred or lack specificity for investigating the endpoint of interest. This includes omission of additional experimental criteria (e.g., inclusion of a positive control or dosing up to levels causing minimal toxicity) when required by specific testing guidelines/protocols. * Cytotoxicity is observed or expected based on findings from similarly designed studies and may mask interpretation of outcome(s) of interest. Sample sizes are smaller than is generally considered adequate for the assay or protocol of interest. Inadequate sampling can also be raised within the context of the endpoint protocol (e.g., in a pathology study, bias that is introduced by only sampling a single tissue depth or an inadequate number of slides per animal).** Controls are not included or considered inappropriate. *These limitations typically also raise a concern for insensitivity. **Sample size alone is not a reason to conclude an individual study is critically deficient.

Results presentation Are the results presented and compared in a way that is appropriate and transparent and makes the data usable?	For each assay/endpoint or grouping of endpoints in a study: Does the level of detail allow for an informed interpretation of the results? If applicable, was the assay signal normalized to account for non-biological differences across replicates and exposure groups? Are the data compared or presented in a way that is inappropriate or misleading (e.g., presenting western blot images without including numerical values for densitometry analysis, or vice versa)? Flag potentially inappropriate statistical comparisons for further review.	Considerations for this domain are highly variable depending on the endpoints of interest and must be refined by assessment teams. A judgment and rationale for this domain should be given for each assay or endpoint or group of endpoints investigated in the study. Some considerations include the following: **Good**: No concerns with how the data are presented. Results are quantified or otherwise presented in a manner that allows for an independent consideration of the data (assessments do not rely on author interpretations). No concerns with completeness of the results reporting.* Ratings of **Adequate**, **Deficient**, and **Critically Deficient** are generally defined as follows: **Adequate:** Concerns are identified that may affect results presentation but are considered unlikely to substantially impact the overall findings or the ability to reliably interpret those findings. **Deficient:** Concerns with results presentation are identified and expected to substantially impact results interpretation and reduce the reliability of the study findings. **Critically deficient:** Severe concerns about results presentation were identified and study findings are likely to be largely explained by these limitations. The following specific examples of relevant concerns are typically associated with a **Deficient** rating but **Adequate** or **Critically Deficient** might be applied depending on expected impact of limitations on the reliability and interpretation of the results: Nonpreferred presentation of data (e.g., averaging technical replicates rather than independent replicates).

Domain and core question	Prompting questions	General considerations
		Failure to present quantitative results. Pooling data when responses are known or expected to differ substantially (e.g., across cell types or passage number). Incomplete presentation of the data* (e.g., presentation of mean without variance data; concurrent control data are not presented; failure to report or address overt cytotoxicity). *Failure to describe *any* findings for assessed outcomes (i.e., report lacks any qualitative or quantitative description of the results in tables, figures, or text) will result in a critically deficient rating for the outcome(s) of interest for Results Presentation; overall completeness of reporting at the study level is addressed under Selective Reporting.

Sensitivity Are there concerns that sensitivity in the study is not adequate to detect an effect?	Was the exposure period, timing (i.e., cell passage number, insufficient culture maturity for the adequate expression of mature cell markers; insufficient treatment and/or measurement duration for the production of protein above the level of detection), frequency, and duration of exposure sensitive for the assay/model system of interest, particularly in the absence of a positive control? Assay-specific considerations regarding sensitivity, specificity, and validity of the selection of the test methods will be described here (e.g., metabolic competency, antibody specificity) (some of these external considerations may have been applied during prioritization of studies for evaluation). Are there aspects related to risk of bias domains that raise concerns about insensitivity (e.g., selection of protocols or methods that are known to be insensitive or nonspecific for the outcome(s) of interest)? Are there concerns regarding the need for positive controls (e.g., concerns that the effects of interest may be inhibited or otherwise poorly manifest in the test system, for example due to differences from in vivo biology)? If used, was the selected positive test substance (and dose) reasonable and appropriate and was the intended positive response induced?	Are there concerns regarding the need for positive controls (e.g., concerns that the effects of interest may be inhibited or otherwise poorly manifest in the test system, for example due to differences from in vivo biology)? If used, was the selected positive test substance (and dose) reasonable and appropriate and was the intended positive response induced? Considerations for this domain are highly variable depending on the specific assay/model system used or endpoint(s) of interest and must be refined by assessment teams. Some study design features that affect study sensitivity may have already been included in the other evaluation domains; these should be noted in this domain, along with any features that have not been addressed elsewhere. Some considerations include: **Good** The experimental design (considering exposure period, timing, frequency, and duration) is appropriate and sensitive for evaluating the outcome(s) of interest. The selected test system is appropriate and sensitive for evaluating the outcome(s) of interest (e.g., cell line/cell type is appropriate and routinely used for the selected assay). No significant concerns with the ability of the experimental design to detect the specific outcome(s) of interest (e.g., study designed to address known endpoint variability that is unrelated to treatment, such as doubling time or confluency). Timing of endpoint measurement in relation to the chemical exposure is appropriate and sensitive (e.g., cultures adequately express mature cell markers). Potential sources of bias towards the null are not a substantial concern.

Domain and core question	Prompting questions	General considerations
		Adequate Potential issues are identified related to the considerations described for *Good* that could reduce sensitivity, but they are unlikely to impact the overall findings of the study. **Deficient** Concerns were raised about the considerations described for *Good* that are expected to notably decrease the sensitivity of the study to detect a response in the exposed group(s). **Critically deficient** Severe concerns were raised about the sensitivity of the study and experimental design such that any observed associations are likely to be explained by bias. The rationale should indicate the specific concern(s).
Overall confidence Considering the identified strengths and limitations, what is the overall confidence rating for the assay(s) or endpoint(s) of interest?	For each assay or endpoint or grouping of endpoints in a study: • Were concerns (i.e., limitations or uncertainties) related to the 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 should be given for each assay or endpoint, or group of endpoints investigated in the study. Confidence rating definitions are described above (see Section 4.1).

7. DATA EXTRACTION OF STUDY METHODS AND RESULTS

The process of summarizing study methods and results is referred to as data extraction. All epidemiology and experimental animal studies meeting the problem formulation PECO criteria after full-text review are briefly summarized in literature inventories and visualized using Tableau software (see Section 4.5 for a description of the information captured in the literature inventory). For this assessment, for all studies that met the refined assessment PECO criteria in Table 5-1, HAWC is used for full extraction of study methods and results. For animal studies, compared to the literature inventory forms used to describe studies that meet initial PECO criteria, full data extraction in HAWC includes summarizing more details of study design (e.g., diet, chemical purity) and gathering effect size information. Instructions on how to conduct data extraction in HAWC are available at https://hawcproject.org/resources/. Over 100 distinct extraction fields are collected for each animal study and endpoint (for list of data extraction fields, see Downloads > Animal Bioassay Data > Complete Export at the HAWC Naphthalene Project https://hawc.epa.gov/assessment/100500288/). An additional resource used to implement use of a consistent vocabulary to summarize endpoints assessed in animal studies is available in the HAWC project "IRIS PPRTV SEM Template Figures and Resources" (see "Attachments", then select the "Environmental Health Vocabulary (EHV) — a recommended terminology for outcomes/endpoints" file).

All findings are considered for extraction, regardless of statistical significance. The level of extraction for specific outcomes within a study could differ (i.e., narrative only if the finding was qualitative). For quality control, studies are summarized by one member of the evaluation team and independently verified by at least one other member. Discrepancies are resolved by discussion or consultation within the evaluation team. Data extraction results are presented via figures, tables, or interactive web-based graphics in the assessment. The information is also made available for download in Excel format when the draft is publicly released.

For non-English studies online translation tools (e.g., Google translator) or engagement with a native speaker will be considered for use in summarizing studies at the level of the literature inventory. Fee-based translation services for non-English studies are typically reserved for studies considered potentially informative for dose response, a consideration that occurs after preparation of the initial literature inventory during draft assessment development. Digital rulers, such as WebPlotDigitizer (http://arohatgi.info/WebPlotDigitizer/), are used to extract numerical information from figures, and their use is be documented during extraction. For studies that evaluate endpoints at multiple time points (e.g., 7 days, 3 weeks, 3 months) data are generally

summarized for the longest duration in the study report, but other durations may be summarized if they provide important contextual information for hazard characterization (e.g., an effect was present at an interim time point but did not appear to persist or the magnitude of the effect diminished). A free text field is available in HAWC to describe cases when the approach for summarizing results requires explanation.

Author queries may be conducted for studies considered for dose-response analysis to facilitate quantitative analysis (e.g., information on variability or availability of individual animal data). Outreach to study authors or designated contact persons is documented and considered unsuccessful if researchers do not respond to email or phone requests within 1 month of initial attempt(s) to contact. Only information or data that can be made publicly available (e.g., within HAWC or HERO) will be considered.

In some cases, EPA may conduct its own statistical analysis of human and animal toxicology data (assuming the data are amenable to doing so and the study is otherwise well-conducted) during evidence synthesis.

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

8. EVIDENCE SYNTHESIS AND INTEGRATION

Evidence synthesis[7] is a within-stream analysis, conducted separately for human, animal, and mechanistic evidence. Findings from human and animal evidence for each unit of analysis are separately judged to reach an expression of certainty in the evidence for a hazard (*robust, moderate, slight, indeterminate,* or *compelling evidence of no effect*). Within-stream evidence synthesis conclusions directly inform the integration across the evidence streams to draw overall conclusions for each of the assessed health effect categories (*evidence demonstrates, evidence indicates, evidence suggests, evidence inadequate,* or *strong evidence supports no effect*). A structured framework approach is used to guide both evidence synthesis and integration. While there are circumstances where specific mechanistic evidence (typically biological precursors) is included in the unit of analysis for human or animal evidence synthesis, in most cases mechanistic findings are presented separately from the human and animal evidence and used to inform conclusions on (1) the coherence, directness of outcome measures, and biological significance of findings within the animal or human evidence streams during evidence synthesis and, (2) evidence integration judgments on the human relevance of findings in animals, coherence across evidence streams ("cross-stream coherence"), information on susceptible populations or lifestages, understanding of biological plausibility and MOA, and possibly other critical inferences (e.g., read-across analyses). The structured framework also accommodates consideration of supplemental information (e.g., ADME, non-PECO route of exposure) that can inform evidence synthesis and integration judgments.

- Evidence synthesis: A summary of findings and judgment(s) regarding the certainty in the evidence for hazard for each unit of analysis from the human and animal studies are made in parallel, but separately. A unit of analysis is an outcome or group of related outcomes within a health effect category that are considered together during evidence synthesis. These judgments can incorporate mechanistic and other supplemental evidence when the unit of analysis is defined as such (see Section 3). The units of analysis can also include or be framed to focus on precursor events (e.g., biomarkers). In addition, this can include an evaluation of coherence across units of analysis within an evidence stream. At this stage, the animal evidence judgment(s) does not yet consider the human relevance of that evidence.

- Evidence integration: The animal and human evidence judgments are combined to draw an overall evidence integration judgment(s) that incorporates inferences drawn based on information on the human relevance of the animal evidence, coherence across evidence

[7] The phrases "evidence synthesis" and "evidence integration" used here are analogous to the phrases "strength of evidence" and "weight of evidence," respectively, used in some other assessment processes (EFSA, 2017; U.S. EPA, 2017; NRC, 2014; U.S. EPA, 2005a).

streams, potential susceptibility, understanding of biological plausibility and MOA, and other critical inferences informed by mechanistic, ADME, or other supplemental data.

Evidence synthesis and integration judgments are expressed both narratively in the assessment and summarized in tabular format in evidence profile tables (see Table 8-1). Key findings and analyses of mechanistic and other supplemental content are also summarized in narrative and tabular format to inform evidence synthesis and integration judgments (see Table 8-2). In brief, after synthesis a certainty in the evidence judgment is drawn for each unit of analysis summarized as *robust*, *moderate*, *slight*, *indeterminate,* or *compelling evidence of no effect* (see Section 8.1). Next, these judgments are used to inform evidence integration judgments summarized as ***evidence demonstrates, evidence indicates, evidence suggests, evidence inadequate,*** or ***strong evidence supports no effect***) (see Section 8.2). These summary judgments are included as part of the evidence synthesis and integration narratives. When multiple units of analysis are synthesized, the main evidence integration judgments typically focus on the unit of analysis with the strongest evidence synthesis judgments, although exceptions may occur.[8] Health outcomes or endpoints where the unit of analysis is considered to present *slight, indeterminant* or *compelling evidence of no effect* can inform the evidence integration hazard judgement but would typically not be used as the basis for deriving a toxicity value. Structured evidence profile tables are used to summarize these analyses and foster consistency within and across assessments. Instructions for using HAWC to create these tables are available at the HAWC project "IRIS PPRTV SEM Template Figures and Resources" (see "Attachments," then select the "Creating Evidence Profile Tables in HAWC").

[8]In some cases, it may be appropriate to draw multiple evidence integration judgments within a given health effect category. This is generally dependent on data availability (i.e., more narrowly defined categories may be possible with more evidence) and the ability to integrate the different evidence streams at the level of these more granular categories. More granular categories will generally be organized by pre-defined manifestations of potential toxicity. For example, within the health effect category of immune effects, separate and different evidence integration judgments might be appropriate for immusuppression, immunostimulation, and sensitization and allergic response (i.e., the three types of immunotoxicity described in the WHO guidance [2012]). Likewise, within the category of developmental effects, it may be appropriate to draw separate judgments for potential effects on fetal death, structural abnormality, altered growth, and functional deficits [i.e., the four manifestations of developmental toxicity described in EPA guidelines (U.S. EPA, 1991)]. These separate judgments are particularly important when the evidence supports that the different manifestations might be based on different toxicological mechanisms. As described for the evidence synthesis judgments, the strongest evidence integration judgment will typically be used to reflect certainty in the broader health effect category.

Table 8-1. Generalized evidence profile table to show the relationship between evidence synthesis and evidence integration to reach judgment of the evidence for hazard

Evidence Synthesis Judgments (note that many factors and judgments require elaboration or evidence-based justification; see IRIS Handbook for details)					Evidence Integration (Weight of Evidence) Judgment(s)
Studies	**Summary of key findings**	**Factors that increase certainty** (Applied to each unit of analysis)	**Factors that decrease certainty** (Applied to each unit of analysis)	**Evidence Synthesis Judgment(s)**	Describe overall evidence integration judgment(s): ⊕⊕⊕ Evidence demonstrates ⊕⊕⊙ Evidence indicates (likely) ⊕⊙⊙ Evidence suggests ⊙⊙⊙ Evidence inadequate – – – Strong evidence supports no effect Highlight the primary supporting evidence for each integration judgment* Present inferences and conclusions on: • Human relevance of findings in animals* • Cross-stream coherence* • Potential susceptibility* • Understanding of biological plausibility and MOA* • Other critical inferences
Evidence from human studies					
Unit of analysis #1 Studies considered and study confidence	Description of the primary results	• All/Mostly *medium* or *high* confidence studies • Consistency • Dose-response gradient • Large or concerning magnitude of effect • Coherence*	• All/Mostly *low* confidence studies • Unexplained inconsistency • Imprecision • Concerns about biological significance* • Indirect outcome measures* • Lack of expected coherence*	**Judgment reached for each unit of analysis*** ⊕⊕⊕ *Robust* ⊕⊕⊙ *Moderate* ⊕⊙⊙ *Slight* ⊙⊙⊙ *Indeterminate* – – – *Compelling evidence of no effect*	
Evidence from animal studies					
Unit of analysis #1 Studies considered and study confidence	Description of the primary results	• All/Mostly *medium* or *high* confidence studies • Consistency • Dose-response gradient • Large or concerning magnitude of effect • Coherence*	• All/Mostly l*ow* confidence studies • Unexplained inconsistency • Imprecision • Concerns about biological significance* • Indirect outcome measures* • Lack of expected coherence*	**Judgment reached for each unit of analysis** ⊕⊕⊕ *Robust* ⊕⊕⊙ *Moderate* ⊕⊙⊙ *Slight* ⊙⊙⊙ *Indeterminate* – – – *Compelling evidence of no effect*	

*Can be informed by key findings from the mechanistic analyses (see Table 8-2)

Table 8-2. Generalized evidence profile table to show the key findings and supporting rationale from mechanistic analyses

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

AOP = Adverse Outcome Pathway.

8.1. EVIDENCE SYNTHESIS

IRIS assessments synthesize the evidence separately for each unit of analysis by focusing on factors that increase or decrease certainty in the reported findings (see Table 8-1). These factors are adapted from considerations for causality introduced by Austin Bradford Hill (Hill, 1965) with some expansion and adaptation of how they are applied to facilitate transparent application to chemical assessments that consider multiple streams of evidence. Specifically, the factors considered are confidence in study findings (risk of bias and sensitivity), consistency across studies or experiments, dose/exposure response gradient, strength (effect magnitude) of the association, directness of outcome or endpoint measures, and coherence [Table 8-3; see additional discussion in U.S. EPA (2005a), U.S. EPA (1994), and U.S. EPA (2020b)]. These factors are similar to the domains considered in the GRADE Quality of Evidence framework (Schünemann et al., 2013). Each of the considered factors and the certainty of evidence judgments require elaboration or evidence-based justification in the synthesis narrative. Analysis of evidence synthesis considerations is qualitative (i.e., numerical scores are not developed, summed, or subtracted).

Biological understanding (e.g., knowledge of how an effect manifests or progresses) or mechanistic inference (e.g., dependency on a conserved key event across outcomes) can be used to define which related outcomes are considered as a unit of analysis. The units of analysis may also include predefined categories of mechanistic evidence (typically precursor events). When mechanistic evidence is included in the units of analysis, it is evaluated against all evidence synthesis factors. Mechanistic and other supplemental evidence not included in the units of analysis can be analyzed to inform select evidence synthesis factors (i.e., coherence, directness of outcome measures, or biological significance) within the animal and human evidence synthesis. Additional mechanistic evaluations (e.g., biological plausibility) are considered as part of across stream evidence integration (see Section 8.2).

Five levels of certainty in the evidence for a hazard are used to summarize evidence synthesis judgments: *robust* (⊕⊕⊕, very little uncertainty exists), *moderate* (⊕⊕⊙, some uncertainty exists), *slight* (⊕⊙⊙, large uncertainty exists), *indeterminate* (⊙⊙⊙), or *compelling evidence of no effect* (– – –, little to no uncertainty exists for lack of hazard) (see Tables 8.4 and 8.5 for descriptions). Conceptually, before the evidence synthesis framework is applied, certainty in the evidence is neutral (i.e., functionally equivalent to indeterminate). Next, the level of certainty regarding the evidence for (or against) hazard is increased or decreased depending on interpretations using the factors described in Table 8.3. Level of certainty analyses are conducted for each unit of analysis within an evidence stream. Observations that increase certainty are having an evidence base exhibiting a signal of an effect on the health outcome based on evaluation of consistency across studies or experiments, the presence of a dose or exposure-response gradient, observing a large or concerning magnitude of effect, and coherent findings for closely related endpoints (can include mechanistic endpoints). These patterns are more compelling when

observed among high or medium confidence studies. Observations that decrease certainty are having an evidence base of mostly low confidence studies, unexplained inconsistency, imprecision, concerns about biological significance, indirect measures of outcomes, and lack of expected coherence. Study sensitivity considerations can be expressed as a factor that can either increase or decrease certainty in the evidence, depending on whether an association is observed. An evidence base of mostly null findings where insensitivity is a serious concern decreases certainty that the evidence is sufficient to support a lack of health effect or association. Conversely, there may be an increase in the evidence certainty in cases where an association is observed although the expected impact of study sensitivity is towards the null.

Table 8-3. Considerations that inform evaluations and judgments of the strength of the evidence for hazard

Consideration	Increased evidence certainty (of the human or animal evidence for hazard[a])	Decreased evidence certainty (of the human or animal evidence for hazard[a])
Risk of bias & sensitivity (across studies)	• An evidence base of mostly (or all) *high* or *medium* confidence studies is interpreted as being only minimally affected by bias and insensitivity. • This factor should not be used if no other factors would increase or decrease the confidence for a given unit of analysis. • In addition, consideration of risk of bias and sensitivity should inform how other factors are evaluated, i.e., can inconsistency be potentially explained by variation in confidence judgments?	• An evidence base of mostly (or all) *low* confidence studies decreases certainty. An exception to this is an evidence base of studies in which the issues resulting in *low* confidence are related to insensitivity. This may increase evidence certainty in cases where an association is identified because the expected impact of study insensitivity is towards the null. • An evidence base of mostly null findings where insensitivity is a serious concern decreases certainty that the evidence is sufficient to support a lack of health effect or association. • Decisions to increase certainty for other considerations in this table should generally not be made if there are serious concerns for risk of bias.
Consistency	• Similarity of findings for a given outcome (e.g., of a similar direction) across independent studies or experiments, especially when *medium* or *high* confidence, increases certainty. The increase in certainty is larger when consistency is observed across populations (e.g., geographical location) or exposure scenarios in human studies, and across laboratories, species, or exposure scenarios (e.g., route, timing) in animal studies. When seemingly inconsistent findings are identified, patterns should be further analyzed to discern if the inconsistencies can potentially be explained based on study confidence, dose or exposure levels, population, or experimental model differences, etc. This factor is typically given the most attention during evidence synthesis.	• Unexplained inconsistency [i.e., conflicting evidence; see (U.S. EPA, 2005a)] decreases certainty. Generally, certainty should not be decreased if discrepant findings can be reasonably explained by considerations such as study confidence conclusions (including sensitivity); variation in population or species, sex, or lifestage (including understanding of differences in pharmacokinetics); or exposure patterns (e.g., intermittent versus continuous), levels (*low* versus *high*), or duration. Similar to current recommendations in the Cochrane Handbook [(Higgins et al., 2022), see Section 7.8.6], clear conflicts of interest (COI) related to funding source can be considered as a factor to explain apparent inconsistency. For small evidence bases, it may be hard to assess consistency. An evidence base of a single or a few studies where consistency cannot be accurately assessed does not, on its own, increase or decrease evidence certainty. Similarly, a reasonable explanation for inconsistency does not necessarily result in an increase in evidence certainty.
Effect magnitude and imprecision	• Evidence of a large or concerning magnitude of effect can increase certainty (generally only when observed in *medium* or *high* confidence studies). • Judgments on effect magnitude and imprecision consider the rarity and severity of the effect.	• Certainty may be decreased if the findings are considered not likely to be biologically significant. Effects that are small in magnitude might not be considered to be biologically significant (adverse[b]) based on information such as historical responses and variability. However, effects that appear to be of small magnitude may be meaningful at the population level (e.g., IQ shifts); in such cases, certainty would not be decreased. • Certainty may also be decreased for imprecision, particularly if there are only a few studies available to evaluate consistency in effect magnitude across studies.

Consideration	Increased evidence certainty (of the human or animal evidence for hazard[a])	Decreased evidence certainty (of the human or animal evidence for hazard[a])
Dose-response	• Evidence of dose-response or exposure-response in *high* or *medium* confidence studies increases certainty. Dose-response may be demonstrated across studies or within studies, and it can be dose- or duration-dependent. It may also not be a monotonic dose-response (monotonicity should not necessarily be expected as different outcomes may be expected at low vs. high doses or long vs. short durations due to factors such as activation of different mechanistic pathways, systemic toxicity at high doses, or tolerance/acclimation). Sometimes, grouping studies by level of exposure is helpful to identify the dose-response pattern. • Decreases in a response (e.g., symptoms of current asthma) after a documented cessation of exposure also may increase certainty in a relationship between exposure and outcome (this is primarily applicable to epidemiology studies because of their observational nature).	• A lack of dose-response when expected based on biological understanding can decrease certainty in the evidence. If the data are not adequate to evaluate a dose-response pattern, however, then certainty is neither increased nor decreased. • In some cases, duration-dependent patterns in the dose-response can decrease evidence certainty. Such patterns are generally only observable in experimental studies. Specifically, the magnitude of effects at a given exposure level might decrease with longer exposures (e.g., due to tolerance or acclimation). Or, effects might rapidly resolve under certain experimental conditions (e.g., reversibility after removal of exposure). As many reversible and short-lived effects can be of high concern, decisions about whether such patterns decrease evidence certainty depend on considering the pharmacokinetics of the chemical and the conditions of exposure [see (U.S. EPA, 1998a)], endpoint severity, judgments regarding the potential for delayed or secondary effects, the underlying mechanism(s) involved, as well as the exposure context focus of the assessment (e.g., addressing intermittent or short-term exposures).
Directness of outcome/endpoint measures	• Not applicable	• If the evidence base primarily includes outcomes or endpoints that are indirect measures (e.g., biomarkers) of the unit of analysis, certainty (for that unit of analysis) is typically decreased. Judgments to decrease certainty based on indirectness should focus on findings that have an unclear linkage to an apical or clinical (adverse[b]) outcome. Scenarios where the magnitude of the response is not considered to reflect a biologically meaningful level of change (i.e., biological significance; see 'effect magnitude and imprecision' row above) are not considered under indirectness. • Related to indirectness, certainty in the evidence may be decreased when the findings are determined to be nonspecific to the hazard under evaluation. This consideration is generally only applicable to animal evidence and the most common example is effects only with exposures (level, duration) shown to cause excessive toxicity in that species and lifestage (including consideration of maternal toxicity in developmental evaluations). This does not apply when an effect is viewed as secondary to other changes (e.g., effects on pulmonary function because of disrupted immune responses).

Consideration	Increased evidence certainty (of the human or animal evidence for hazard[a])	Decreased evidence certainty (of the human or animal evidence for hazard[a])
Coherence	• Biologically related findings within or across studies, within an organ system or across populations (e.g., sex), increase strength (generally only when observed in *medium* or *high* confidence studies). Certainty is further increased 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. • Coherence across findings within a unit of analysis (e.g., consistent changes in disease markers and biological precursors in exposed humans) can increase certainty in the evidence for an effect. • Coherence within or across biologically related units of analysis can also increase strength for a given (or multiple) unit(s) of analysis. This considers certainty in the biological relationships between the endpoints being compared, and the sensitivity and specificity of the measures used. • Mechanistic support for, or biological understanding of, the relatedness between different endpoints within (or across different) units of analysis, can inform an understanding of coherence.	• An observed lack of expected coherent changes (e.g., in well-established biological relationships) within or across biologically related units of analysis typically decrease evidence strength. This includes mechanistic changes when included in the unit of analysis. However, as described for decisions to increase strength, certainty in 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.
Other factors	• Unusual scenarios that cannot be addressed by the considerations above, e.g., read across inferences supporting the adversity of observed changes.	• Unusual scenarios that cannot be addressed by the considerations above, e.g., strong evidence of publication bias.[c]

[a]While the focus is on identifying potential adverse human health effects (hazards) of exposure, these factors can also be used to increase or decrease certainty in the evidence supporting lack of an effect (e.g., leading to a judgment of compelling evidence of no effect). The latter application is not explicitly outlined here.

[b]Within this framework, evidence synthesis judgments reflect an interpretation of the evidence for) a hazard; thus, consideration of the adversity of the findings is an explicit aspect of the analyses. To better define how adversity is evaluated, the consideration of adversity is broken into the two, sometimes related, considerations of the indirectness of the outcome measures and the interpreted biological significance of the effect magnitude.

[c]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). This may make the available evidence base unrepresentative. However, publication bias can be difficult to evaluate (NTP, 2019) and should not be used as a factor that decreases certainty unless there is strong evidence.

A structured framework approach is used to draw evidence synthesis judgments for human and animal evidence. Tables 8-3 and 8-4 (for human and animal evidence, respectively) provide the example-based criteria that guide how to draw the strength of evidence judgments for each unit of analysis within a health effect category and the terms used to summarize those judgments. These terms are applied to human and animal evidence separately. The terms *robust* and *moderate* are characterizations for judgments that the evidence (across studies) supports that the effect(s) results from the exposure being assessed. These two terms are differentiated by the quality and amount of information available to rule out alternative explanations for the results. For example, repeated observations of effects by independent studies or experiments examining various aspects of exposure or response (e.g., different exposure settings, dose levels or patterns, populations or species, biologically related endpoints) result in a stronger certainty of evidence judgment. The term *slight* indicates situations in which there is some evidence supporting an association within the evidence stream, but substantial uncertainties in the data exist to prevent judgments that the effect(s) can be reliably attributed to the exposure being assessed. *Indeterminate* reflects judgments for a wide variety of evidence scenarios, including when no studies are available or when the evidence from studies of similar confidence has a high degree of unexplained inconsistency. *Compelling evidence of no effect* represents a rare situation in which extensive evidence across a range of populations and exposures has demonstrated that no effects are likely to be attributable to the exposure being assessed. This category is applied at the health effect level (e.g., hepatic effects) rather than more granular units of analysis level to avoid giving the impression of confidence in lack of a health effect when aspects of potential toxicity have not been adequately examined. Reaching this judgment is infrequent because it requires both a high degree of confidence in the conduct of individual studies, including consideration of study sensitivity, as well as comprehensive assessments of outcomes and lifestages of exposure that adequately address concern for the hazard under evaluation.

Table 8-4. Framework for strength of evidence judgments from studies in humans

Strength of evidence judgment	Description
Robust (⊕⊕⊕) …evidence in human studies *(strong signal of effect with very little uncertainty)*	A set of high or medium confidence independent studies (e.g., in different populations) reporting an association between the exposure and the health outcome(s), 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; the findings are considered adverse (i.e., biologically significant and without notable concern for indirectness); and an exposure response gradient is demonstrated. Additional supporting evidence, such as associations with biologically related endpoints in human studies (coherence) or large estimates of risk or severity of the response, can increase certainty, but are not required. Supplemental evidence included in the unit of analysis (e.g., mechanistic studies in exposed humans or human cells) may raise the strength of evidence to *robust* for a set of studies that otherwise would be described as *moderate*. Such evidence not included in the unit of analysis can also inform evaluations of the coherence of the human evidence, the directness of the outcome measures, and the biological significance of the findings. Causality is inferred for a human evidence base of *robust*.
Moderate (⊕⊕⊙) …evidence in human studies *(signal of effect with some uncertainty)*	A set of evidence that does not reach the degree of certainty required for *robust*, but which includes at least one *high* or *medium* confidence study reporting an association and additional information increasing the strength of evidence. For multiple studies, there is primarily consistent evidence of an association with reasonable support for adversity, but there may be some uncertainty due to potential chance, bias, or confounding or because of the indirectness of some measures. For a single study, there is a large magnitude or severity of the effect, or a dose-response gradient, or other supporting evidence, and there are no serious residual methodological uncertainties. Supporting evidence could include associations with related endpoints, including mechanistic evidence from exposed humans when included within the unit of analysis. When available and included in the unit of analysis, mechanistic data in humans that address the above considerations may raise the strength of evidence to *moderate* for a set of studies that otherwise would be described as *slight*. In exceptional cases, biological support from mechanistic evidence in exposed humans may support raising the strength of evidence to *moderate* for evidence that would otherwise be described as *indeterminate*.
Slight (⊕⊙⊙) …evidence in human studies *(signal of effect with large amount of uncertainty)*	One or more studies reporting an association between exposure and the health outcome, but considerable uncertainty exists and supporting coherent evidence is sparse. In general, the evidence is limited to a set of consistent *low* confidence studies, or higher confidence studies with significant unexplained heterogeneity or other serious residual uncertainties. It also applies when one *medium* or *high* confidence study is available without additional information strengthening the likelihood of a causal association (e.g., coherent findings within the same study or from other studies). This category serves primarily to encourage additional study where evidence does exist that might provide some support for an association, but for which the evidence does not reach the degree of confidence required for *moderate*.

Strength of evidence judgment	Description
Indeterminate (⊙⊙⊙) …evidence in human studies *(signal cannot be determined for or against an effect)*	No studies available in humans or situations when the evidence is inconsistent and primarily of *low* confidence. In addition, this may include situations where higher confidence studies exist, but there are major concerns with the evidence base such as unexplained inconsistency, a lack of expected coherence from a stronger set of studies, very small effect magnitude (i.e., major concerns about biological significance), or uncertainties or methodological limitations that result in an inability to discern effects from exposure. It also applies for a single *low* confidence study in the absence of factors that increase certainty. A set of largely null studies could be concluded to be *indeterminate* if the evidence does not reach the level required for *Compelling evidence of no effect*.
Compelling evidence of no effect (– – –) …in human studies *(strong signal for lack of an effect with little uncertainty)*	A set of *high* confidence studies examining a reasonable spectrum of endpoints showing null results (for example, an odds ratio of 1.0), ruling out alternative explanations including chance, bias, and confounding) with reasonable confidence. Each of the studies should have used an optimal outcome and exposure assessment and adequate sample size (specifically for higher exposure groups and for susceptible populations). The set as a whole should include diverse sampling (across sexes [if applicable] and different populations) and include the full range of levels of exposures that human beings are known to encounter, an evaluation of an exposure response gradient, and an examination of at-risk populations and lifestages. Mechanistic data in humans that address the above considerations or that provide information supporting the lack of an association between exposure and effect with reasonable confidence may provide additional support for this judgment.

Table 8-5. Framework for strength of evidence judgments from studies in animals

Strength of evidence judgment	Description
Robust (⊕⊕⊕) …evidence in animal studies *(strong signal of effect with very little uncertainty)*	The set of *high* or *medium* confidence, independent experiments (i.e., across laboratories, exposure routes, experimental designs [for example, a subchronic study and a multigenerational study], or species) reporting effects of exposure on the health outcome(s). The set of studies is primarily consistent, with reasonable explanations when results differ (i.e., due to differences in study design, exposure level, animal model, or study confidence), and the findings are considered adverse (i.e., biologically significant and without notable concern for indirectness). At least two of the following additional factors in the set of experiments increase the strength of evidence: coherent effects across multiple related endpoints (within or across biologically related units of analysis and may include mechanistic 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. Mechanistic evidence from animals included in the unit of analysis or used to assess coherence of findings in the animal evidence may raise the strength of evidence to *robust* for a set of studies that otherwise would be described as *moderate.*
Moderate (⊕⊕⊙)	A set of evidence that does not reach the degree of certainty required for *robust*, but which includes at least one *high* or *medium* confidence study and additional information increasing the strength of evidence. For multiple studies or a single study, the evidence is primarily consistent or coherent with

Strength of evidence judgment	Description
…evidence in animal studies *(signal of effect with some uncertainty)*	reasonable support for adversity, but there are notable remaining uncertainties (e.g., difficulty interpreting the findings due to concerns for indirectness of some measures); however, these uncertainties are not sufficient to reduce or discount the level of concern regarding the positive findings and any conflicting findings are from a set of experiments of lower confidence. The set of experiments supporting the effect provide additional information increasing the strength of evidence, such as consistent effects across laboratories or species; coherent effects across multiple related endpoints (may include mechanistic endpoints within the unit of analysis); an unusual magnitude of effect, rarity, age at onset, or severity; a strong dose-response relationship; and/or consistent observations across exposure scenarios (e.g., route, timing, duration), sexes, or animal strains. When available and included in the unit of analysis, mechanistic data in animals that address the above considerations may raise the strength of evidence to *moderate* for a set of studies that otherwise would be described as *slight*. In exceptional cases, strong biological support from mechanistic studies may raise the strength of evidence to *moderate* for evidence that would otherwise be described as *indeterminate*.
Slight (⊕⊙⊙) …evidence in animal studies *(signal of effect with large amount of uncertainty)*	One or more studies reporting an effect on an exposure on the health outcome, but considerable uncertainty exists and supporting coherent evidence is sparse. In general, the evidence is limited to a set of consistent *low* confidence studies, or higher confidence studies with significant unexplained heterogeneity or other serious uncertainties (e.g., concerns about adversity) across studies. It also applies when one *medium* or *high* confidence experiment is available without additional information increasing the strength of evidence (e.g., coherent findings within the same study or from other studies). Biological evidence from mechanistic studies may also be independently interpreted as s*light*. This category serves primarily to encourage additional study where evidence does exist that might provide some support for an association, but for which the evidence does not reach the degree of confidence required for *moderate*.
Indeterminate (⊙⊙⊙) …evidence in animal studies *(signal cannot be determined for or against an effect)*	No studies available in animals or situations when the evidence is inconsistent and primarily of *low* confidence. In addition, this may include situations where higher confidence studies exist, but there are major concerns with the evidence base such as unexplained inconsistency, a lack of expected coherence from a stronger set of studies, very small effect magnitude (i.e., major concerns about biological significance), or uncertainties or methodological limitations that result in an inability to discern effects from exposure. It also applies for a single *low* confidence study in the absence of factors that increase certainty. A set of largely null studies could be concluded to be *indeterminate* if the evidence does not reach the level required for *Compelling evidence of no effect*.
Compelling evidence of no effect (– – –) …in animal studies *(strong signal for lack of an effect*	A set of *high* confidence experiments examining a reasonable spectrum of endpoints that demonstrate a lack of biologically significant effects across multiple species, both sexes, and a broad range of exposure levels. The data are compelling in that the experiments have examined the range of scenarios across which health effects in animals could be observed, and an alternative explanation (e.g., inadequately controlled features of the studies' experimental designs; inadequate sample sizes) for the observed lack of effects is not available. Each of the studies should have used an optimal endpoint and exposure assessment and adequate sample size. The evidence base should represent both sexes and address potentially susceptible populations and lifestages.

Strength of evidence judgment	Description
with little uncertainty)	Mechanistic data in animals that address the above considerations or that provide information supporting the lack of an association between exposure and effect with reasonable confidence may provide additional support for this judgment.

8.2. EVIDENCE INTEGRATION

The phase of evidence integration combines animal and human evidence synthesis judgments while also considering information on the human relevance of findings in animal evidence, coherence across evidence streams ("cross-stream coherence"), information on susceptible populations or lifestages, understanding of biological plausibility and MOA, and possibly other critical inferences (e.g., read-across analyses) that generally draw on mechanistic and other supplemental evidence (see Table 8-6). This analysis culminates in an evidence integration judgment and narrative for each potential health effect category (i.e., each noncancer health effect and specific type of cancer, or broader grouping of related outcomes as defined in the evaluation plan). To the extent it can be characterized prior to conducting dose-response analyses, exposure context is also provided.

Table 8-6. Considerations that inform evidence integration judgments

Judgment	Description
Human relevance of findings	Used to describe and justify the interpretation of the relevance of the animal data to humans. This can include consideration of mechanistic or other supplemental information. When human evidence is lacking or has results that differ from animals, analyses of the mechanisms underlying the animal response in relation to those presumed to operate in humans, and the chemical's pharmacokinetics, can inform the extent to which the animal response is likely to be relevant to humans and potentially strengthen overall confidence in the evidence integration conclusion. Conversely, evidence for a mechanistic pathway that is expected to only occur in animals and not in humans can provide support for a conclusion that the animal evidence for an effect is not relevant to humans. In the absence of chemical-specific evidence informing human relevance, the evidence integration narrative will briefly describe the interpreted comparability of experimental animal organs/systems to humans based on underlying biological similarity (e.g., thyroid signaling processes are well conserved across rodents and humans). Generally, a high-level systems summary should be possible for most encountered effects. In some cases, however, it may be appropriate to use a statement such as, 'without evidence to the contrary, [health effect described in the table] responses in animals are presumed to be relevant to humans.' As noted in EPA guidelines (**U.S. EPA, 2005a**), there needs to be evidence or a biological explanation to support an interpreted lack of human relevance for findings in animals, and site concordance is neither expected nor required.

Cross-stream coherence	Addresses the concordance of findings known to be biologically related across human, animal, and mechanistic studies, considering factors such as exposure timing and levels. Notably, for many health effects (e.g., some nervous system and reproductive effects; cancer), it is not necessary (or expected) that effects manifested in humans are identical to those observed in animals, although this typically provides stronger evidence. For example, tumors in one animal species can be predictive of carcinogenic potential in humans or other species, but not necessarily at the same site. EPA guidelines and other resources (e.g., OECD guidance) are consulted when drawing these inferences. Mechanistic support for, or biological understanding of, the relatedness between different outcomes (and the manner in which they are manifest) in different species can inform an understanding of coherence across evidence streams. Evidence supporting a biologically plausible mechanistic pathway across species adds coherence (see below).
Potential susceptibility Susceptible populations and lifestages	Used to summarize analyses relating to individual and social factors that may increase susceptibility to exposure-related health effects in certain populations or lifestages, or to highlight the lack of such information. These analyses are based on knowledge about the health outcome or organ system affected and focus primarily on the influence of intrinsic biological factors such as race/ethnicity, genetic variability, sex, lifestage, and pre-existing health conditions (which can also have an extrinsic basis). Information on extrinsic factors potentially influencing susceptibility (e.g., proximity to exposure; certain lifestyle factors including subsistence living) are not considered in evidence integration judgments on potential susceptibility; these exposure-focused factors are considered by risk managers after the human health assessment is complete. Evaluation of potential susceptibility can also include consideration of mechanistic and ADME evidence.
Biological plausibility and MOA considerations	Support for the biological plausibility of an association between exposure and the health effect increases evidence strength, particularly when observed across species. This may be provided by data from experimental studies of mechanistic pathways, particularly when support is provided for key events or is conserved across multiple components of the pathway. Mechanisms or biological changes with broad scientific acceptance for their relevance to chemical toxicity or the health effect (e.g., key characteristics, hallmarks of cancer) may be used to organize the chemical-specific evidence and identify key events leading from exposure to the health effect. For each key event and key event relationship, the evidence is considered regarding the consistency of experimental data and the generalizability, or likelihood of similarities (e.g., in presence or function) across species, as well as the strength of the support for the biological mechanism. Mechanistic evidence from well-conducted studies that demonstrates that the health effect is unlikely to occur (i.e., species specific effects, irrelevant exposure conditions) can support a judgment that the effects from animal or human studies are not biologically relevant, which weakens the summary evidence integration judgment. Such a decision depends on an evaluation of the strength of the information supporting vs. opposing biological plausibility, as well as the strength of the health effect-specific findings (e.g., stronger health effect data require more certainty in mechanistic evidence opposing plausibility). Importantly, because understanding biological plausibility is dependent on expert knowledge and canonical scientific knowledge, the lack of such understanding does not provide a rationale to decrease the strength of the evidence for an effect (**NTP, 2015**; **NRC, 2014**). These analyses are typically conducted separately to establish MOA understanding and referenced in the evidence integration judgment. If sufficiently supported, MOA understanding can serve to strengthen (e.g., strong support for mutagenicity) or weaken (e.g., critical dependence on a key event not likely to be operant in humans) evidence integration judgments.

Other critical inferences (optional)	Consideration of other evidence or non-chemical-specific information that informs evidence integration judgments (e.g., read across analyses, ADME understanding used to inform other considerations; judgments on other health effects expected to be linked to the health effect under evaluation; read-across analyses or inferences) may be separately described as "other critical inferences."

Using a structured framework approach, one of five phrases is used to summarize the evidence integration judgment based on the within evidence stream integration of the human and animal evidence, and supplemental (mechanistic) evidence: ***evidence demonstrates, evidence indicates, evidence suggests, evidence is inadequate,*** or ***strong evidence supports no effect*** (see Table 8-7). The five integration judgment levels reflect the differences in the amount and quality of the data that inform the evaluation of whether exposure may cause the health effect(s). As it is assumed that any identified health hazards will only manifest given exposures of a certain type and amount (e.g., a specific route; a minimal duration, periodicity, and level), the evidence integration narrative and summary judgment levels include the generic phrase, "given sufficient exposure conditions." This highlights that, for those assessment-specific health effects identified as potential hazards, the exposure conditions associated with those health effects will be defined (as will the uncertainties in the ability to define those conditions) during dose-response analysis. More than one descriptor can be used when the evidence base is able to support that a chemical's effects differ by exposure level or route (U.S. EPA, 2005a). The analyses and judgments are summarized in the evidence profile table (see Table 8-1).

Table 8-7. Framework for summary evidence integration judgments in the evidence integration narrative

Summary evidence integration judgment[a] in narrative	Evidence integration judgment level	Explanation and example scenarios[b]
The currently available **evidence demonstrates** that [chemical] causes [health effect] in humans[c] given sufficient exposure conditions. This conclusion is based on studies of [humans or animals] that assessed [exposure or dose] levels of [range of concentrations or specific cutoff level concentration[d]].	**Evidence demonstrates**	A strong evidence base demonstrating that [chemical] exposure causes [health effect] in humans. • This conclusion level is used if there is *robust* human evidence supporting an effect. • This conclusion level could also be used with *moderate* human evidence and *robust* animal evidence if there is strong mechanistic evidence that MOAs and key precursors identified in animals are anticipated to occur and progress in humans.
The currently available **evidence indicates** that [chemical] likely causes [health effect] in humans given sufficient exposure conditions. This conclusion is based on studies of [humans or animals] that assessed [exposure or dose] levels of [range of concentrations or specific cutoff level concentration].	**Evidence indicates (likely[e])**	An evidence base that indicates that [chemical] exposure likely causes [health effect] in humans, although there may be outstanding questions or limitations that remain, and the evidence is insufficient for the higher conclusion level. • This conclusion level is used if there is *robust* animal evidence supporting an effect and *slight*-to-*indeterminate* human evidence, or with *moderate* human evidence when strong mechanistic evidence is lacking. • This conclusion level could also be used with *moderate* human evidence supporting an effect and *moderate-to-indeterminate* animal evidence, or with *moderate* animal evidence supporting an effect and *moderate-to-indeterminate* human evidence. In these scenarios, any uncertainties in the *moderate* evidence are not sufficient to substantially reduce confidence in the reliability of the evidence, or mechanistic evidence in the *slight* or *indeterminate* evidence base (e.g., precursors) exists to increase confidence in the reliability of the *moderate* evidence.
The currently available **evidence suggests** that [chemical] may cause [health effect] in humans. This conclusion is based on studies of [humans or animals] that assessed [exposure or dose] levels of [range of concentrations or specific cutoff level concentration].	**Evidence suggests**	An evidence base that suggests that [chemical] exposure may cause [health effect] in humans, but there are very few studies that contributed to the evaluation, the evidence is very weak or conflicting, and/or the methodological conduct of the studies is poor. • This conclusion level is used if there is *slight* human evidence and *indeterminate*-to-*slight* animal evidence. • This conclusion level is also used with *slight* animal evidence and *indeterminate*-to-*slight* human evidence.

Summary evidence integration judgment[a] in narrative	Evidence integration judgment level	Explanation and example scenarios[b]
		• This conclusion level could also be used with *moderate* human evidence and *slight* or *indeterminate* animal evidence, or with *moderate* animal evidence and *slight* or *indeterminate* human evidence. In these scenarios, there are outstanding issues or uncertainties regarding the *moderate* evidence (i.e., the synthesis judgment was borderline with *slight*), or mechanistic evidence in the *slight* or *indeterminate* evidence base (e.g., null results in well-conducted evaluations of precursors) exists to decrease confidence in the reliability of the *moderate* evidence. • Exceptionally, when there is general scientific understanding of mechanistic events that result in a health effect, this conclusion level could also be used if there is strong mechanistic evidence that is sufficient to highlight potential human toxicity[f]—in the absence of informative conventional studies in humans or in animals (i.e., *indeterminate* evidence in both).
The currently available **evidence is inadequate** to assess whether [chemical] may cause [health effect] in humans.	**Evidence inadequate**	This conveys either a lack of information or an inability to interpret the available evidence for [health effect]. On an assessment-specific basis, a single use of this "inadequate" conclusion level might be used to characterize the evidence for multiple health effect categories (i.e., all health effects that were examined and did not support other conclusion levels).[g] • This conclusion level is used if there is *indeterminate* human and animal evidence. • This conclusion level is also used with *slight* animal evidence and *compelling evidence of no effect* human evidence. • This conclusion level could also be used with *slight*-to-*robust* animal evidence and *indeterminate* human evidence if strong mechanistic information indicated that the animal evidence is unlikely to be relevant to humans. • A conclusion of **inadequate** is not a determination that the agent does not cause the indicated health effect(s). It simply indicates that the available evidence is insufficient to reach conclusions.
Strong evidence supports no effect in humans. This conclusion is based on studies of [humans or animals] that assessed [exposure or dose] levels of [range of concentrations].	**Strong evidence supports no effect**	This represents a situation in which extensive evidence across a range of populations and exposure levels has identified no effects/associations. This scenario requires a *high* degree of confidence in the conduct of individual studies, including consideration of study sensitivity, and comprehensive assessments of the endpoints and lifestages of exposure relevant to the heath effect of interest.

Summary evidence integration judgment[a] in narrative	Evidence integration judgment level	Explanation and example scenarios[b]
		• This conclusion level is used if there is compelling evidence of no effect in human studies and compelling evidence of no effect to indeterminate in animals. • This conclusion level is also used if there is indeterminate human evidence and compelling evidence of no effect animal evidence in models concluded to be relevant to humans. • This conclusion level could also be used with compelling evidence of no effect in human studies and moderate to robust animal evidence if strong mechanistic information indicated that the animal evidence is unlikely to be relevant to humans.

[a]Evidence integration judgments are typically developed at the level of the health effect when there are sufficient studies on the topic to evaluate the evidence at that level; this should always be the case for "evidence demonstrates" and "strong evidence supports no effect," and typically for "evidence indicates (likely)." However, some databases only allow for evaluations at the category of health effects examined; this will more frequently be the case for conclusion levels of "evidence suggests" and "evidence inadequate." A judgment of "strong evidence supports no effect" is drawn at the health effect level.

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

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

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

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

[f]Scientific understanding of adverse outcome pathways (AOPs) 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 testing) will continue to increase. This may make possible the development of hazard conclusions when there are mechanistic or other relevant data that can be interpreted with a similar level of confidence to positive animal results in the absence of conventional studies in humans or in animals.

[g]Specific narratives for each of these health effects may also be deemed unnecessary.

1

For evaluations of carcinogenicity, consistent with EPA's cancer guidelines (U.S. EPA, 2005a), one of EPA's standardized cancer descriptors is used to describe the overall potential for carcinogenicity within the evidence integration narrative for carcinogenicity. These descriptors are: (1) ***carcinogenic to humans,*** (2) ***likely to be carcinogenic to humans***, (3) ***suggestive evidence of carcinogenic potential,*** (4) ***inadequate information to assess carcinogenic potential, or*** (5) ***not likely to be carcinogenic to humans.*** The standardized cancer descriptors will often align with the evidence integration judgements (i.e., "evidence demonstrates" aligns with "carcinogenic to humans") but not in all cases. For example, the evidence integration judgements are generally used for individual tumor or cancer types and the standardized EPA descriptors are used to characterize overall cancer hazard.

For each type of cancer evaluated (e.g., lung cancer, renal cancer) or sets of related cancer types, an evidence integration narrative and summary judgment level are provided as described above for noncancer health effects. When considering evidence on carcinogenicity across human and animal evidence, site concordance is not required (U.S. EPA, 2005a). If a systematic review of more than one cancer type was conducted, then the strongest evidence integration judgment(s) is used as the basis for selecting the standardized cancer descriptor in accordance with the EPA cancer guidelines (U.S. EPA, 2005a).

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

9.1. OVERVIEW

Selection of specific data sets for dose-response assessment and performance of the dose-response assessment is conducted after hazard identification is complete and involves database- and chemical-specific biological judgments. A number of EPA guidelines and support documents detail data requirements and other considerations for dose-response modeling, especially EPA's *Benchmark Dose Technical Guidance* (U.S. EPA, 2012b), EPA's *Review of the Reference Dose and Reference Concentration Processes* (U.S. EPA, 2005a, 2002), *Guidelines for Carcinogen Risk Assessment* (U.S. EPA, 2005a), and *Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens* (U.S. EPA, 2005b). This section of the protocol provides an overview of considerations for conducting the dose-response assessment, particularly statistical considerations specific to dose-response analysis that support quantitative risk assessment. Importantly, these considerations do not supersede existing EPA guidelines.

For IRIS assessments, dose response- assessments are typically performed for both noncancer and cancer hazards, and for both oral and inhalation routes of exposure following chronic exposure[9] to the chemical of interest, if supported by existing data. For noncancer hazards, an inhalation reference concentration (RfC) or oral reference dose (RfD) will be derived, if possible. A reference value (i.e., RfC or RfD) is an estimate, with uncertainty spanning perhaps an order of magnitude, of an exposure to the human population (including susceptible populations and lifestages) that is likely to be without an appreciable risk of deleterious health effects over a lifetime [(U.S. EPA, 2002) see section 4.2]. In addition to an RfC or RfD, this assessment will attempt to derive organ- or system-specific RfCs (osRfCs) or RfDs (osRfDs) when the data are sufficiently strong (i.e., with rare exception as described below, noncancer conclusions of *evidence demonstrates* or *evidence indicates [likely]*). In addition to chronic RfCs or chronic RfDs, when feasible and if the available data are appropriate for doing so, the assessments will derive a less-than-lifetime toxicity value (a "subchronic" reference value) for noncancer hazards. Both less-than-lifetime and hazard-specific values may be useful to EPA risk assessors within specific decision contexts.

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

When low-dose linear extrapolation for cancer effects is supported, particularly for chemicals with direct mutagenic activity or those for which the data indicate a linear component below the point of departure (POD), an inhalation unit risk (IUR) or oral cancer slope factor (CSF) facilitates estimation of human cancer risks. Low-dose linear extrapolation is also used as a default when the data are insufficient to establish the mode of action (U.S. EPA, 2005a). An IUR is a plausible upper-bound lifetime cancer risk from chronic inhalation of a chemical per unit of air concentration (expressed as ppm or $\mu g/m^3$); a CSF is a plausible upper bound lifetime cancer risk from chronic oral exposure to a chemical.

The derivation of toxicity values depends on the nature of the hazard conclusion. Specifically, EPA generally conducts dose-response assessments and derives cancer values for chemicals that are classified as *carcinogenic* or *likely to be carcinogenic* to humans. When there is *suggestive evidence* of carcinogenic potential to humans, EPA generally would not conduct a dose-response assessment and derive a cancer value. Similarly, for noncancer outcomes, dose-response is conducted based on having stronger evidence of a hazard (generally, "*evidence demonstrates*" and "*evidence indicates [likely]*". When the noncancer outcome is considered *evidence suggests* of potential hazard to humans, EPA generally would not conduct a dose-response assessment and derive a RfC or RfD. Cases where suggestive evidence might be used to develop cancer risk estimates or a noncancer toxicity value include when the evidence base includes a well-conducted study (overall *medium* or *high* confidence for the outcome) and quantitative analyses may be useful for some purposes, (e.g., providing a sense of the magnitude and uncertainty of potential risks, ranking potential hazards, or setting research priorities) (U.S. EPA, 2005a).

9.2. SELECTING STUDIES FOR DOSE-RESPONSE ASSESSMENT

9.2.1. Hazard and MOA Considerations for Dose Response

The assessment presents a summary of hazard identification conclusions to transition to dose response considerations, highlighting (1) information used to inform the selection of outcomes or broader health effect categories for which toxicity values will be derived, (2) whether toxicity values can be derived to protect specific populations or lifestages, (3) how dose response modeling will be informed by pharmacokinetic information, and (4) the identification of biologically based BMR levels (where possible and supported by the data). The pool of outcomes and study-specific endpoints is discussed to identify which categories of effects and study designs are considered the strongest and most appropriate for quantitative assessment of a given health effect, particularly among the studies that exemplify the study attributes summarized in Table 9-1.

Also considered is whether there are opportunities for quantitative evidence integration. Examples of quantitative integration, from simplest to more complex, include (1) combining results for an outcome across sex (within a study); (2) characterizing overall toxicity, as in combining effects that comprise a syndrome, or occur on a continuum (e.g., precursors and eventual overt

toxicity, benign tumors that progress to malignant tumors); and (3) conducting a meta-analysis or meta-regression of all studies addressing a category of important health effects.

Some studies that are used qualitatively for hazard identification may or may not be useful quantitatively for dose-response assessment due to such factors as the lack of quantitative measures of exposure or lack of variability measures for response data. If the needed information cannot be located, semiquantitative analysis may be feasible (e.g., via NOAEL/LOAEL). In this assessment, specific datasets considered for dose-response analysis will be summarized in a tabular format that includes rationales for decisions to proceed (or not) for POD derivation. Table 9-2 presents an example format for how these decisions can be documented, although the specifics in the naphthalene assessment are likely to differ.

In addition, mechanistic evidence that influences the dose-response analyses will be highlighted—for example, evidence related to susceptibility or potential shape of the dose-response curve (i.e., linear, nonlinear, or threshold model). Mode(s) of action summarized as part of hazard identification will be used to highlight information relevant to understanding overall risk. Biological considerations relevant to dose-response for cancer are:

- Is there evidence for direct mutagenicity?
- Does tumor latency decrease with increasing exposure?
- If there are multiple tumor types, which cancers have a longer latency period?
- Is incidence data available (incidence data are preferred to mortality data)?
- Were there different background incidences in different (geographic) populations?
- While benign and malignant tumors of the same cell of origin are generally evaluated together, was there an increase only in malignant tumors?

Table 9-1. Attributes used to evaluate studies for derivation of toxicity values (in addition to the health effect category-specific evidence integration judgment)

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 as well as a reconsideration of the specific limitations identified and their potential impact on dose-response analyses.	
Rationale for choice of species		Human data are preferred over animal data to eliminate interspecies extrapolation uncertainties (e.g., in pharmacodynamics, 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 or PBPK 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 life stage 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*, see section 2.1.1) and facilitate extrapolation to more relevant (generally lower) exposures.	
Subject selection		Studies that provide risk estimates in the most susceptible groups are preferred. Attempts are made to highlight where it might be possible to develop separate risk estimates for a specific population or life stage, or determine whether evidence is available to select a data-derived uncertainty factor (UF).	
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.	

Study attributes	Considerations	
	Human studies	**Animal studies**
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 may 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 (e.g., phthalate syndrome).	
	Among several relevant health outcomes, preference is generally given to those with greater biological significance. When there are multiple endpoints for an organ/system, characterizing the overall impact on this organ/system is considered. For example, if there are multiple histopathological alterations relevant to liver function changes, liver necrosis may be selected as the most representative endpoint to consider for dose-response analysis. For cancer types, consideration is given to the overall risk of multiple types of tumors. Multiple tumor types (if applicable) are discussed, and a rationale given for any grouping.	
Study size and design	Preference is given to studies using designs reasonably expected to have power to detect responses of suitable magnitude.[b] This does not mean that studies with substantial responses but low power would be ignored, but that they should be interpreted in light of a confidence interval or variance for the response. Studies that address changes in the number at risk (through decreased survival, loss to follow-up) are preferred.	

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

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

Table 9-2. Specific example of presenting endpoints considered for dose-response modeling and derivation of points of departure

Endpoint	Study reference/ confidence	Exposure route and duration	Human population or Test species and strain	Lifestage and Sex	POD derivation	Rationale
Endocrine Effects (hazard judgment of evidence indicates [likely])						
Decreased serum total T4	[study 1 author, year, HERO ID]; *high* confidence	Oral Gavage, 90 days	S-D rat	Adult female	Yes ✓	Decreases in total T4 in females were dose-dependent and of a large magnitude (36-53% reduction at ≥3.12 mg/kg-d); effects in males were not prioritized due to body weight loss at the doses causing significant decreases in total T4.
	[study 1 author, year, HERO ID]; *high* confidence	Oral Gavage, 90 days	S-D rat	Adult male	No, X	
Increased thyroid follicular hypertrophy	[study 1 author, year, HERO ID]; *high* confidence	Oral Gavage, 90 days	S-D rat	Adult males and females	Yes ✓	Increases in thyroid follicular hypertrophy incidence were dose-dependent in both sexes at doses that did not affect body weight.
Thyroid weight	[study 2 author, year, HERO ID]; *medium* confidence	Oral Gavage, 90 days	F344 rat	Adult males and females	No, X	Increased thyroid weights were only observed at doses over an order of magnitude higher than those affecting thyroid hormones and histopathology in the other subchronic study (note: this study only tested much higher doses)

1

9.3. CONDUCTING THE DOSE-RESPONSE ASSESSMENT

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

Within the observed dose range, the preferred approach is to use dose-response modeling to incorporate as much of the data set as possible into the analysis for the purpose of deriving a POD; see Section 9.3.1 for more details.

Derivation of cancer risk estimates or reference values nearly always involves extrapolation to exposures lower than the POD and is described in more detail in Sections 9.3.2 and 9.3.3, respectively.

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

For noncancer analyses, IRIS assessments typically derive a candidate value from each suitable data set, whether for human or animal. Evaluating these candidate values grouped within a particular organ/system yields a single organ/system-specific reference value for each organ/system under consideration. Next, evaluation of these organ/system-specific reference values results in the selection of a single overall reference value to cover all health outcomes across all organs/systems. While this overall reference value is the focus of the assessment, the organ/system-specific reference values can be useful for subsequent cumulative risk assessments that consider the combined effect of multiple agents acting at a common organ/system.

For cancer analyses, if there are multiple tumor types in a study population (human or animal), final cancer risk estimates will typically address overall cancer risk (i.e., the risk of developing any combination of modeled tumor types).

9.3.1. Dose-Response Analysis in the Range of Observation

Empirical dose-response modeling is used to fit the data (on the apical outcomes or a key precursor events) in the ranges of observation. For this purpose of empirical dose-response modeling, EPA has developed a standard set of models (https://www.epa.gov/bmds) that can be applied to typical dichotomous and continuous data sets, including those that are nonlinear. In situations where there are alternative models with significant biological support, the users of the assessment can be informed by the presentation of these alternatives along with the models' strengths and uncertainties. EPA has developed guidelines on modeling dose-response data, assessing model fit, selecting suitable models, and reporting modeling results [see the *EPA Benchmark Dose Technical Guidance* (U.S. EPA, 2012b)].

U.S. EPA Benchmark Dose Software (BMDS) is designed to model dose-response datasets in accordance with EPA Benchmark Dose Technical Guidance (U.S. EPA, 2012b). For noncancer (and nonlinear cancer), a benchmark dose lower confidence limit (BMDL) is computed from a model selected from the BMDS suite of models using statistical and graphical criteria. Linear analysis of cancer datasets is generally based on the multistage model, with degree selected following a U.S. EPA Statistical Workgroup technical memo available on the BMDS website (https://cfpub.epa.gov/ncea/bmds/recordisplay.cfm?deid=308382). Modeling of cancer data may in some cases involve additional, specialized methods, particularly for multiple tumors or early removal from observation (due to death or morbidity). Additional judgments or alternative analyses may be used if initial modeling procedures fail to yield results in reasonable agreement with the data. For example, modeling may be restricted to the lower doses, especially if there is competing toxicity at higher doses.

For noncancer (and nonlinear cancer) datasets, EPA recommends (1) application of a preferred set of models that use maximum likelihood estimation (MLE) methods (default models in BMDS) and (2) selection of a POD from a single model based on criteria designed to limit model selection subjectivity (auto-implemented in BMDS version 3 and higher). For the linear analysis of cancer datasets, EPA recommends (1) application of the Multistage MLE model; (2) selection of a single Multistage degree; and (3) in cases where tumors are observed in multiple organ systems, use of a multi-tumor model (i.e., MS-Combo) that appropriately estimates combined tumor risk (both (2) and (3) are available in BMDS).[10]

Version 3.0 and higher of BMDS also provides an alternative modeling approach that uses Bayesian model averaging for dichotomous modeling average (DMA). EPA makes DMA available as alternative approaches but has not yet finalized guidelines for their use.

For each modeled dataset for an outcome, a POD from the observed data should be estimated to mark the beginning of extrapolation to lower doses. The POD is an estimated dose (expressed in human equivalent terms) near the lower end of the observed range without significant extrapolation to lower doses. For linear extrapolation of cancer risk, the POD is used to calculate an OSF or IUR, and for nonlinear extrapolation, the POD is used in calculating an RfD or RfC.

The selection of the response level at which the POD is calculated is guided by the severity of the endpoint. If linear extrapolation is used, selection of a response level corresponding to the point of departure is not highly influential, so standard values near the low end of the observable range are generally used (for example, 10% extra risk for cancer bioassay data, 1% for epidemiologic data, lower for rare cancers). Nonlinear approaches consider both statistical and

[10] The Multistage degree selection process outlined in the memo is auto-implemented in the BMDS multi-tumor model, which can be run on one or more tumor data sets, but only the noncancer model selection process is auto-implemented for individual Multistage model runs in the current version, BMDS 3.3).

biologic considerations. For dichotomous data, a response level of 10% extra risk is generally used for minimally adverse effects, 5% or lower for more severe effects or effects observed in studies with increased statistical sensitivity. Lower BMRs are often supported for developmental toxicity studies. For continuous data, a response level is ideally based on an established definition of biologic significance. In the absence of such definition, one control standard deviation from the control mean is often used for minimally adverse effects, and one-half standard deviation for more severe effects. As with dichotomous endpoints, lower BMRs may also be supported for endpoints observed in studies with greater statistical sensitivity (e.g., developmental toxicity studies). The point of departure is the 95% lower bound on the dose associated with the selected response level.

EPA has developed standard approaches for determining the relevant dose to be used in the dose-response modeling in the absence of appropriate pharmacokinetic modeling. These standard approaches also facilitate comparison across exposure patterns and species:

- Intermittent study exposures are standardized to a daily average over the duration of exposure. For chronic effects, daily exposures are averaged over the lifespan. Exposures during a critical period, however, are not averaged over a longer duration [(U.S. EPA, 2005a), see section 3.1.1; (U.S. EPA, 1991), see section 3.2]. Note that this will typically be done after modeling because the conversion is linear.

- Doses are standardized to equivalent human terms to facilitate comparison of results from different species. Oral doses are scaled allometrically using $mg/kg^{3/4}day$ as the equivalent dose metric across species. Allometric scaling pertains to equivalence across species, not across life stages, and is not used to scale doses from adult humans or mature animals to infants or children [(U.S. EPA, 2011a, 2005a),see Section 3.1.3]. Inhalation exposures are scaled using dosimetry models that apply species-specific physiologic and anatomic factors and consider whether the effect occurs at the site of first contact or after systemic circulation [(U.S. EPA, 2012a, 1994), see Section 3].

- It can be informative to convert doses across exposure routes. If this is done, the assessment describes the underlying data, algorithms, and assumptions [(U.S. EPA, 2005a), see Section 3.1.4].

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

- The preferred approach for dosimetry extrapolation from animals to humans is through PBPK modeling. As explained in Section 9.3.1 and Appendix C.2. 6.4, EPA has selected the naphthalene PBPK model of Kapraun et al. (2020) to compute internal dose metrics relevant to various toxicity studies. The same model will be used to compute human equivalent doses and/or concentrations.

- Briefly, PBPK model simulations will be used to estimate internal dose metrics corresponding to the applied doses for each experimental animal bioassay. By simulating the exposure scenario for each toxicity study (e.g., 6 hours/day, 5 day/week inhalation exposure), the resulting internal dose metric effectively accounts for the difference between

the actual exposure pattern and a nominal 24 hour/day, 7 day/week exposure. The set of internal dose metrics for each toxicity study and endpoint can then be used in dose-response analysis to identify a BMDL or other point-of-departure (POD) for that study. The human version of the PBPK model can then be used to estimate the exposure concentration in air which, given continuous (24 hour/day, 7 day/week) inhalation exposure, would result in a given internal dose POD. Any remaining uncertainty factors, including the factor of 10 for human inter-individual variability (UFH), will then be applied for derivation of the HECs.

- If needed, a similar approach can be applied for oral-to-inhalation route extrapolation for endpoints where toxicity data are available from oral dosimetry studies but not from inhalation.

9.3.2. Dose Metrics

EPA will use the model of Kapraun et al. (2020) to compute internal dose metrics relevant to various toxicity studies. In particular, the five-dose metrics listed in Table 9-3 will be considered. Among the dose metrics described in Table 9-3**Error! Reference source not found.** DM1, DM2, and DM3 should be relevant when the health effect of interest occurs in the DO tissue. DM1 and DM2 reflect an assumption that it is the concentration or delivered dose of naphthalene, itself, that is most predictive of DO toxicity, while DM3 may be more relevant when the health effect occurs in the DO tissue but is correlated more directly with metabolite dose rather than dose of the parent chemical (i.e., naphthalene). DM4 is a general-purpose measure of internal dose and should be relevant when the health effect correlates with systemic, rather than site-specific, dose. Similarly, DM5 is a measure of systemic internal dose, but it should be most relevant when the health effect correlates with metabolite dose rather than dose of the parent chemical. One or more of the five dose metrics described inTable 9-3. Internal dose metrics considered for use in assessing dose-response relationships for naphthalene Appendix C.2. will be used to conduct dose-response analysis for each health effect to obtain a "benchmark dose" or point of departure. Reverse dosimetry (incorporating ¾ body mass scaling for the rate-of-delivery or rate-of-metabolism dose metrics DM2, DM3, and DM5) will then be used to compute a human equivalent external concentration (or oral dose) that corresponds to each benchmark dose.

Table 9-3. Internal dose metrics considered for use in assessing dose-response relationships for naphthalene

Abbreviation	Description
DM 1	Average naphthalene concentration in dorsal olfactory (DO) tissue (µg/mL) (i.e., the total naphthalene mass (µg) in the anterior dorsal olfactory tissue (DO1) and posterior dorsal olfactory tissue (DO2) is computed throughout the simulation and the average concentration is calculated as the area under the curve divided by the total elapsed time and the total volume of DO1 and DO2)
DM 2	Average rate of delivery of naphthalene to DO tissue (µg/cm^2/d) (i.e., the total rate of mass transfer (µg/d) to DO1 and DO2 is computed throughout the simulation and the average rate is calculated as the area under the curve divided by the total elapsed time and the total surface area of DO1 and DO2)
DM 3	Average rate of metabolite production in DO tissue (µg/mL/d) (i.e., the rate at which metabolites are produced in DO1 and DO2 per unit volume are computed throughout the simulation and the average rate is calculated as the area under the curve divided by the total elapsed time)
DM 4	Average naphthalene concentration in blood (µg/mL) (i.e., the total naphthalene mass (µg) in the blood is computed throughout the simulation and the average concentration is calculated as the area under the curve divided by the total elapsed time and the volume of the blood)
DM 5	Average rate of metabolite production in the whole body (µg/kg-d) (i.e., the total rate at which metabolites are produced (µg/d) in olfactory tissue, liver, and other regions of the body are computed throughout the simulation and the average rate is calculated as the area under the curve divided by the total elapsed time divided by the body mass)

For a given toxicological endpoint, the choice of dose metric will be based primarily on biological considerations when possible. In particular, the decision will be based on evidence as to whether the parent (naphthalene) or a metabolite is expected to be the driver of a given toxic effect. Mechanistic data for related toxic effects (e.g., cytotoxicity in hepatocytes vs. respiratory cells) or structurally similar chemicals may also be considered. When dose-response data from multiple studies are available, especially when the dosing regimen or route of administration are varied, a dose metric that explains apparent differences in the response vs. unadjusted dose relationships will be selected. Thus, the extent to which use of a particular dose metric yields consistency in the dose-response relationship will be used to select a dose metric for the purposes of this assessment.

In the event that no mechanistic data are available to inform the choice of dose metric, if only a single dose-response study is available for a given endpoint, or if all existing studies are inherently self-consistent due to similarity of study design, then consistency of the dose metric vs. exposure relationship predicted by the PBPK model for a given dose metric and the observed toxicological response vs. exposure relationship can also be evaluated. For example, metabolic saturation leads to a concave down (negative second derivative) relationship curve for metabolite dose vs. exposure and a concave up (positive second derivative) relationship curve for parent chemical concentration vs. exposure. If the resulting nonlinearity is strong and a similar saturation or concavity occurs in the dose-response curve for a toxic endpoint in the same exposure range, the

consistency between one dose metric option and the dose-response nonlinearity indicates which metric is a better predictor of risk.

However, caution is needed in comparing nonlinearity in the dose vs. exposure relationship with nonlinearity in the response vs. exposure relationship, as nonlinearity in the dose-response relationship can occur due to pharmacodynamic mechanisms that are not related to dosimetry. A modest difference in a statistical correlation coefficient or other measure of goodness of fit is not considered strong evidence for the choice of one dose metric over another. In the absence of compelling mechanistic or exposure-dose-response evidence, the level of uncertainty in the dose metric will also be considered. For example, with respect to CFD-PBPK model predictions, there is less uncertainty in the delivered dose to the olfactory tissue than in the tissue concentration or rate of metabolism in that tissue. The degree to which modeling involving alternate dose metrics yields health protective results (e.g., when a dose metric specific human equivalent dose leads to a lower RfC than does using the nominal dose) will be considered along with the level of uncertainty in each metric.

9.3.3. Dosimetric Modeling Summary

Existing PBPK and inhalation dosimetry models for naphthalene (which are summarized in Appendix D) were identified through a literature search. Of these, the model of Kapraun et al. (2020) was identified as the best for dosimetric applications as it met EPA's quality evaluation criteria, although other dosimetric models have distinct features which are of potential scientific value. Five potentially useful dose metrics were presented in the preceding section and methods for selecting from among them have been proposed. However, as the naphthalene assessment progresses, new information concerning related biology or toxicity mechanisms may be discovered and such information may suggest that alternative model choices or dose metrics should be used or that the proposed methods for estimating human equivalent inhaled concentrations (or oral doses) should be modified. If this is the case, the dosimetry methods proposed for naphthalene in this document may be adjusted.

9.3.4. Extrapolation: Slope Factors and Unit Risk

An OSF or *IUR* facilitates estimation of human cancer risks when low dose linear extrapolation for cancer effects is supported, particularly for chemicals with direct mutagenic activity or those for which the data indicate a linear component below the POD. Low-dose linear extrapolation is also used as a default when the data are insufficient to establish the mode of action (U.S. EPA, 2005a). If data are sufficient to ascertain one or more modes of action consistent with low-dose nonlinearity, or to support their biological plausibility, low-dose extrapolation may use the reference value approach when suitable data are available (U.S. EPA, 2005a); see Section 11.2.3 below.

9.3.5. Extrapolation: Reference Values

Reference value derivation is EPA's most frequently used type of nonlinear extrapolation method. Although it is most commonly used for noncancer effects, this approach is also used for cancer effects if there are sufficient data to ascertain the MOA and conclude that it is not linear at low doses. For these cases, reference values for each relevant route of exposure are developed following EPA's established practices [(U.S. EPA, 2005a), see Section 3.3.4].

For each data set selected for reference value derivation, reference values are estimated by applying relevant adjustments to the PODs to account for the conditions of the reference value definition—for human variation, extrapolation from animals to humans, extrapolation to chronic exposure duration, and extrapolation to a minimal level of risk (if not observed in the data set). The assessment will discuss the scientific bases for estimating these data-based adjustments and UFs:

- *Animal-to-human extrapolation (UF_A)*: If animal results are used to make inferences about humans, the candidate toxicity value incorporates cross-species differences, which may arise from differences in pharmacokinetics or pharmacodynamics. Typically, the pharmacokinetic and pharmacodynamic portions are considered to address an equivalent amount of the total uncertainty factor (i.e., each contributing $10^{0.5}$ or "3" towards the default UF_A of 10). If the POD is standardized to equivalent human terms or is based on pharmacokinetic or dosimetry modeling (U.S. EPA, 2014a, 2011a), a factor of $10^{0.5}$ (rounded to 3) is applied to account for the remaining uncertainty involving pharmacokinetic and pharmacodynamic differences. If a biologically based model adjusts fully for pharmacokinetic and pharmacodynamic differences across species, a factor of 1 is applied. Similarly, although this is not a common scenario, if chemical-specific information is sufficient to reasonably conclude that the experimental animal species is less or equally sensitive as humans, the pharmacodynamic portion of this uncertainty factor (i.e., typically starting at $10^{0.5}$ or "3") can be reduced.

- *Human variation (UF_H)*: This UF accounts for variation in susceptibility across the human population and the possibility that the available data may not be representative of individuals who are most susceptible to the effect. As with the UF_A, this typically considers potential pharmacokinetic and pharmacodynamic differences that might exist across individuals, amongst other considerations (see Table 7-1). If population-based data for the effect or for characterizing the internal dose are available, the potential for data-based adjustments for pharmacodynamics or pharmacokinetics is considered (U.S. EPA, 2014a).[11] Further, "when sufficient data are available, an intraspecies UF either less than or greater than 10× may be justified (U.S. EPA, 2002). However, a reduction from the default (10) is only considered in cases when there is dose-response data for the most susceptible population" (U.S. EPA, 2002). This factor is reduced only if the POD is derived or adjusted specifically for susceptible individuals [not for a general population that includes both

[11]Examples of adjusting the pharmacokinetic portion of interhuman variability include the IRIS boron assessment's use of nonchemical-specific kinetic data [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, 2011c).

susceptible and nonsusceptible individuals [(U.S. EPA, 2002), see Section 4.4.5; (U.S. EPA, 1998a), see Section 4.2; (U.S. EPA, 1996), see Section 4; (U.S. EPA, 1994), see Section 4.3.9.1; (U.S. EPA, 1991), see Section 3.4]. Otherwise, a factor of 10 is generally used to account for this variation. Note that when a PBPK model is available for relating human internal dose to environmental exposure, relevant portions of this UF may be more usefully applied prior to animal-to-human extrapolation, depending on the correspondence of any nonlinearities (e.g., saturation levels) between species (also see **Section 13.2.2**).

- *LOAEL to NOAEL (UF_L)*: If a POD is based on a LOAEL, the assessment must infer an exposure level where such effects are not expected. This can be a matter of great uncertainty if there is no evidence available at lower exposures. The ratio of the doses at the LOAEL and NOAEL are expected to vary considerably across studies and consideration of cross-study information may not be informative. A factor of up to 10 is generally applied to extrapolate to a lower exposure expected to be without appreciable effects. A factor other than 10 may be used depending on the magnitude and nature of the response and the shape of the dose-response curve (U.S. EPA, 2002, 1998a, 1996, 1994, 1991). For example, LOAELs associated with lower response levels or less adverse effects (e.g., a small, minimally biologically significant level of change at the LOAEL) may warrant smaller uncertainty factors, whereas higher response levels likely warrant the default value of 10, or in rare instances, values higher than 10. Regardless, the available data should be carefully evaluated and any decision to apply a non-default value requires adequate discussion in the dose-response section.

- *Subchronic-to-chronic exposure (UF_S)*: Although not always made explicit, the intent of this UF is to address the uncertainty associated with extrapolating from studies with exposure durations shorter than the focus of the toxicity values derived. In IRIS, a lifetime (chronic) reference value is typically the focus and oftentimes PODs are based on subchronic evidence, so the assessment needs to consider whether lifetime exposure could have effects at lower levels of exposure. As a general rule and in the context of subchronic-to-chronic (lifetime) extrapolation, a factor of up to 10 is applied (after adjustment of intermittent exposures to continuous) when using subchronic studies to make inferences about lifetime exposure. A factor other than 10 may be used, depending on the duration and/or timing of the studies and the nature of the response (U.S. EPA, 2002, 1998a, 1994). For example, studies that occur during a sensitive lifestage typically warrant application of a $UF_S = 1$, which would generally be applied regardless of the toxicity value type (e.g., a $UF_S = 1$ for both subchronic and chronic values). A prime example of this is developmental toxicity studies and effects observed in offspring. Typically, developmental toxicity studies use exposure durations either encompassing a specific portion of gestation (e.g., organogenesis) or the entirety of gestation as these are expected to the critical windows of susceptibility for developmental effects. Thus, there is no concern that a longer duration exposure would result in more severe effects and an uncertainty factor would not be applied. This factor may be applied, albeit rarely, for developmental or reproductive effects if exposure covered less than the full critical period. A value different from 10 may be applied if there exists sufficient information from the chemical database. For example, if a chemical database contains subchronic and short-term studies and there is no evidence of an exacerbation of effect when moving from short-term to subchronic exposure durations, an uncertainty factor lower than 10 may be warranted. This UF is not necessarily constrained to a subchronic-to-chronic exposure scenarios: it would also be considered in application to extrapolating from a short-term study to a subchronic toxicity value and might still apply

when extrapolating from a chronic duration study to a lifetime toxicity value if the chronic duration is interpreted as likely to be insensitive. However, no general guidelines exist for the standard values of short-term-to-subchronic, or chronic-to-lifetime extrapolations and chemical-specific data would need to inform the value for these extrapolations assessment to assessment.

- In addition to the adjustments above, a database UF (UF_D) is applied to address any **database** deficiencies that raise concern that further studies might identify a more sensitive effect (e.g., in an organ system or a lifestage that is not well studied) (U.S. EPA, 2002, 1998a, 1996, 1994, 1991). The size of the factor depends on the nature of the database deficiency. For example, the EPA typically follows the suggestion that a factor of 10 be applied if a prenatal toxicity study and a two-generation reproduction study is both missing, and a factor of $10^{0.5}$ (rounded to 3) if either one or the other is missing [(U.S. EPA, 2002), see Section 4.4.5]. A database UF greater than 1 would still be applied if this type of study were available but considered to be a *low* confidence study based on the evaluation process [described in Chapter 12 of (U.S. EPA, 2022)]**.** However, when deciding what value to apply for this uncertainty factor, assessors need to consider the data missing and available for specific organ systems and/or lifestages, meaning a $UF_D > 1$ can still be applied in scenarios when both developmental and two-generation reproduction studies are available if sufficient evidence is available to raise a concern that effects could occur in other organ systems at lower doses. In addition, a $UF_D > 1$ can still be applied even if the POD being adjusted comes from human data, and information from both human and animal studies should be considered when selecting the value of this factor. Information on structurally-related chemicals could be potentially used to select the value of this factor if it suggests effects in organ systems for which chemical-specific data is missing.

10. PROTOCOL HISTORY

1 Release date:

2 Revisions history:

REFERENCES

Abdo, KM; Grumbein, S; Chou, BJ; Herbert, R. (2001). Toxicity and carcinogenicity study in F344 rats following 2 years of whole-body exposure to naphthalene vapors. Inhal Toxicol 13: 931-950. http://dx.doi.org/10.1080/089583701752378179.

ACGIH (American Conference of Governmental Industrial Hygienists). (1992). Naphthalene. In Documentation of the threshold limit values and biological exposure indices (6th ed.). Cincinnati, OH.

ACGIH (American Conference of Governmental Industrial Hygienists). (2001). Naphthalene. In Documentation of the threshold limit values and biological exposure indices. Cincinnati, OH.

ACGIH (American Conference of Governmental Industrial Hygienists). (2007). 2007 TLVs and BEIs: Based on the documentation of the threshold limit values for chemical substances and physical agents and biological exposure indices [TLV/BEI]. Cincinnati, OH.

AIHA (American Industrial Hygiene Association). (2016). Current ERPG Values (2016). In 2016 ERPG/WEEL Handbook. Fairfax, VA: American Industrial Hygiene Association Guideline Foundation.

Ajao, OG; Adenuga, MO; Lapido, JK. (1988). Colorectal carcinoma in patients under the age of 30 years: A review of 11 cases. Royal College of Surgeons of Edinburgh Journal 33: 277–279.

Anziulewicz, JA; Dick, HJ; Chiarulli, EE. (1959). Transplacental naphthalene poisoning. Am J Obstet Gynecol 78: 519-521. http://dx.doi.org/10.1016/0002-9378(59)90522-8.

ATSDR (Agency for Toxic Substances and Disease Registry). (2005). Toxicological profile for naphthalene, 1-methylnaphthalene, and 2-methylnaphthalene. (PB2006100004). Atlanta, GA. https://ntrl.ntis.gov/NTRL/dashboard/searchResults.xhtml?searchQuery=PB2006100004.

ATSDR (Agency for Toxic Substances and Disease Registry). (2013). Addendum to toxicological profile for jet fuels (JP-5 and JP-9). Atlanta, GA: Division of Toxicology and Human Health Sciences. http://www.atsdr.cdc.gov/toxprofiles/jet_fuels_addendum.pdf.

ATSDR (Agency for Toxic Substances and Disease Registry). (2017). Minimal risk levels (MRLs). Available online at https://web.archive.org/web/20171029000135/https://www.atsdr.cdc.gov/mrls/index.asp (accessed March 26, 2018).

ATSDR (Agency for Toxic Substances and Disease Registry). (2021). Toxic substances portal: Toxicological profiles [Database]. Atlanta, GA. Retrieved from https://www.atsdr.cdc.gov/toxprofiledocs/index.html

Bailey, LA; Rhomberg, LR. (2020). Incorporating ToxCast™ data into naphthalene human health risk assessment. Toxicol In Vitro 67: 104913. http://dx.doi.org/10.1016/j.tiv.2020.104913.

Battelle (Battelle Columbus Laboratories). (1980). Unpublished subchronic toxicity study: Naphthalene (C52904), Fischer 344 rats. (subcontract no. 76-34-106002). Research Triangle Park, NC: National Toxicology Program.

Bogdanffy, MS; Sarangapani, R; Plowchalk, DR; Jarabek, AM; Andersen, ME. (1999). A biologically based risk assessment for vinyl acetate-induced cancer and noncancer inhalation toxicity. Toxicol Sci 51: 19-35. http://dx.doi.org/10.1093/toxsci/51.1.19.

Bois, FY. (2009). GNU MCSim: Bayesian statistical inference for SBML-coded systems biology models. Bioinformatics 25: 1453-1454. http://dx.doi.org/10.1093/bioinformatics/btp162.

Buckpitt, A; Boland, B; Isbell, M; Morin, D; Shultz, M; Baldwin, R; Chan, K; Karlsson, A; Lin, C; Taff, A; West, J; Fanucchi, M; Van Winkle, L; Plopper, C. (2002). Naphthalene-induced respiratory

tract toxicity: Metabolic mechanisms of toxicity [Review]. Drug Metab Rev 34: 791-820. http://dx.doi.org/10.1081/DMR-120015694.

Buckpitt, A; Morin, D; Murphy, S; Edwards, P; Van Winkle, L. (2013). Kinetics of naphthalene metabolism in target and non-target tissues of rodents and in nasal and airway microsomes from the Rhesus monkey. Toxicol Appl Pharmacol 270: 97-105. http://dx.doi.org/10.1016/j.taap.2013.04.006.

Buckpitt, AR; Richieri, P. (1984). Comparative biochemistry and metabolism: Part II: Naphthalene lung toxicity. (AFAMRL-TR-84-058). Wright-Patterson Air Force Base, OH: Air Force Aerospace Medical Research Laboratory.

CalEPA (California Environmental Protection Agency). (2004). Air Toxic hot spots: Adoption of a unit risk value for naphthalene. Sacramento, CA: Office of Environmental Health Hazard Assessment. https://oehha.ca.gov/air/document-hot-spots/air-toxic-hot-spots-adoption-unit-risk-value-naphthalene#downloads.

CalEPA (California Environmental Protection Agency). (2016). OEHHA toxicity criteria database. Sacramento, CA: Office of Environmental Health Hazard Assessment. Retrieved from http://www.oehha.ca.gov/tcdb/index.asp

Campbell, JL; Andersen, ME; Clewell, HJ. (2014). A hybrid CFD-PBPK model for naphthalene in rat and human with IVIVE for nasal tissue metabolism and cross-species dosimetry. Inhal Toxicol 26: 333-344. http://dx.doi.org/10.3109/08958378.2014.896059.

Carlson, GP. (2008). Critical appraisal of the expression of cytochrome P450 enzymes in human lung and evaluation of the possibility that such expression provides evidence of potential styrene tumorigenicity in humans [Review]. Toxicology 254: 1-10. http://dx.doi.org/10.1016/j.tox.2008.09.017.

Cathey, A; Ferguson, KK; McElrath, TF; Cantonwine, DE; Pace, G; Alshawabkeh, A; Cordero, JF; Meeker, JD. (2018). Distribution and predictors of urinary polycyclic aromatic hydrocarbon metabolites in two pregnancy cohort studies. Environ Pollut 232: 556-562. http://dx.doi.org/10.1016/j.envpol.2017.09.087.

CDC (Centers for Disease Control and Prevention). (2022). National Report on Human Exposure to Environmental Chemicals. Atlanta, GA: U.S. Department of Health and Human Services. https://www.cdc.gov/exposurereport/pdf/FourthReport_UpdatedTables_Volume1_Mar2018.pdf.

Celsie, A; Mackay, D; Parnis, JM; Arnot, JA. (2016). A fugacity-based toxicokinetic model for narcotic organic chemicals in fish. Environ Toxicol Chem 35: 1257-1267. http://dx.doi.org/10.1002/etc.3270.

Chuang, JC; Hannan, SW; Wilson, NK. (1987). FIELD COMPARISON OF POLYURETHANE FOAM AND XAD-2 RESIN FOR AIR SAMPLING FOR POLYNUCLEAR AROMATIC HYDROCARBONS. Environ Sci Technol 21: 798-804. http://dx.doi.org/10.1021/es00162a011.

Clewell, HJ; Efremenko, A; Campbell, JL; Dodd, DE; Thomas, RS. (2014). Transcriptional responses in the rat nasal epithelium following subchronic inhalation of naphthalene vapor. Toxicol Appl Pharmacol 280: 78-85. http://dx.doi.org/10.1016/j.taap.2014.06.015.

Cock, TC. (1957). Acute hemolytic anemia in the neonatal period. AMA J Dis Child 94: 77-79. http://dx.doi.org/10.1001/archpedi.1957.04030020079013.

Cok, I; Mazmanci, B; Mazmanci, MA; Turgut, C; Henkelmann, B; Schramm, KW. (2012). Analysis of human milk to assess exposure to PAHs, PCBs and organochlorine pesticides in the vicinity Mediterranean city Mersin, Turkey. Environ Int 40: 63-69. http://dx.doi.org/10.1016/j.envint.2011.11.012.

Coombs, D. (1993). Naphthalene 4-week inhalation study in rats: Lab project number: LDA 1/921559. Unpublished study prepared by Huntingdon Research Centre Ltd. 165 p. (MRID 42934901[ZK3]). Coombs, D.

Corley, RA; Kabilan, S; Kuprat, AP; Carson, JP; Jacob, RE; Minard, KR; Teeguarden, JG; Timchalk, C; Pipavath, S; Glenny, R; Einstein, DR. (2015). Comparative risks of aldehyde constituents in cigarette smoke using transient computational fluid dynamics/physiologically based pharmacokinetic models of the rat and human respiratory tracts. Toxicol Sci 146: 65-88. http://dx.doi.org/10.1093/toxsci/kfv071.

Corley, RA; Kabilan, S; Kuprat, AP; Carson, JP; Minard, KR; Jacob, RE; Timchalk, C; Glenny, R; Pipavath, S; Cox, T; Wallis, C; Larson, RF; Fanucchi, MV; Postlethwait, E; Einstein, DR. (2012). Comparative computational modeling of airflows and vapor dosimetry in the respiratory tracts of a rat, monkey, and human. Toxicol Sci 128: 500-516. http://dx.doi.org/10.1093/toxsci/kfs168.

CT DEEP. Hazardous air pollutants, Sec. 22a-174-29 1-31 (RCSA2015). https://eregulations.ct.gov/eRegsPortal/Browse/getDocument?guid={00D6A654-0300-CC47-9B95-397D2AD21304}

CT DEEP (Connecticut Department of Energy and Environmental Protection). (2018). Technical support document: Recommended numeric criteria for common additional polluting substances and certain alternative criteria. Hartford, CT. https://portal.ct.gov/-/media/DEEP/site_clean_up/remediation_regulations/TechnicalSupportDocumentAPSAltCriteriapdf.pdf.

Dawson, JP; Thayer, WW; Desforges, JF. (1958). Acute hemolytic anemia in the newborn infant due to naphthalene poisoning: report of two cases, with investigations into the mechanism of the disease. Blood 13: 1113-1125.

De Craemer, S; Croes, K; van Larebeke, N; Sioen, I; Schoeters, G; Loots, I; Nawrot, T; Nelen, V; Campo, L; Fustinoni, S; Baeyens, W. (2016). Investigating unmetabolized polycyclic aromatic hydrocarbons in adolescents' urine as biomarkers of environmental exposure. Chemosphere 155: 48-56. http://dx.doi.org/10.1016/j.chemosphere.2016.04.017.

DFG (Deutsche Forschungsgemeinschaft). (2020). List of MAK and BAT values 2020: Permanent senate commission for the investigation of health hazards of chemical compounds in the work area. (Report 56). Bonn, Germany. http://dx.doi.org/10.34865/mbwl_2020_eng.

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

Dobraca, D; Lum, R; Sjödin, A; Calafat, AM; Laurent, CA; Kushi, LH; Windham, GC. (2018). Urinary biomarkers of polycyclic aromatic hydrocarbons in pre- and peri-pubertal girls in Northern California: Predictors of exposure and temporal variability. Environ Res 165: 46-54. http://dx.doi.org/10.1016/j.envres.2017.11.011.

Dodd, DE; Gross, EA; Miller, RA; Wong, BA. (2010). Nasal olfactory epithelial lesions in F344 and SD rats following 1- and 5-day inhalation exposure to naphthalene vapor. Int J Toxicol 29: 175-184. http://dx.doi.org/10.1177/1091581809357955.

DOE (U.S. Department of Energy). (2009). Table 3: Protective Action Criteria (PAC) Rev 25 based on applicable 60-minute AEGLs, ERPGs, or TEELs. The chemicals are listed by CASRN. August 2009. https://edms.energy.gov/pac/docs/Revision_29A_Table3.pdf.

DOE (U.S. Department of Energy). (2018). Table 2: Protective Action Criteria (PAC) Rev. 29a based on applicable 60-minute AEGLs, ERPGs, or TEELs. https://edms3.energy.gov/pac/docs/Revision_29A_Table2.pdf.

Dusseldorp, A; van Bruggen, M; Douwes, J; Janssen, P, .J.C.M; Kelfkens, G. (2011). Health-based guideline values for the indoor environment. (RIVM report 609021044/2007). Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). https://www.rivm.nl/bibliotheek/rapporten/609021044.pdf.

ECB (European Chemicals Bureau). (2003). European Union Risk Assessment Report: Naphthalene CAS No: 91-20-3 EINECS No: 202-049-5. European Commission. http://echa.europa.eu/documents/10162/4c955673-9744-4d1c-a812-2bf97863906a.

Edwards, DA; Andriot, MD; Amoruso, MA; Tummey, AC; Bevan, CJ; Tveit, A; Hayes, LA; Youngren, SH; Nakles, DV. (1997). Total petroleum hydrocarbon criteria working group series: Volume 4: Development of fraction specific reference doses (RFDs) and reference concentrations (RFCs) for total petroleum hydrocarbons (TPH). Amherst, MA: Amherst Scientific Publishers.

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

EIC (ENVIRON International Corporation). (2015). Assessment report on naphthalene for developing ambient air quality objectives 2013 update (pp. 1-87).

Eiguren-Fernandez, A; Miguel, AH; Froines, JR; Thurairatnam, S; Avol, EL. (2004). Seasonal and spatial variation of polycyclic aromatic hydrocarbons in vapor-phase and PM25 in Southern California urban and rural communities. Aerosol Sci Technol 38: 447-455. http://dx.doi.org/10.1080/02786820490449511.

Familusi, JB; Dawodu, AH. (1985). A survey of neonatal jaundice in association with household drugs and chemicals in Nigeria. Ann Trop Paediatr 5: 219-222. http://dx.doi.org/10.1080/02724936.1985.11748397.

Fang, GC; Chang, KF; Lu, CS; Bai, HL. (2004). Estimation of PAHs dry deposition and BaP toxic equivalency factors (TEFs) study at Urban, Industry Park and rural sampling sites in central Taiwan, Taichung. Chemosphere 55: 787-796. http://dx.doi.org/10.1016/j.chemosphere.2003.12.012.

Genter, MB; Marlowe, J; Kevin Kerzee, J; Dragin, N; Puga, A; Dalton, TP; Nebert, DW. (2006). Naphthalene toxicity in mice and aryl hydrocarbon receptor-mediated CYPs. Biochem Biophys Res Commun 348: 120-123. http://dx.doi.org/10.1016/j.bbrc.2006.07.025.

Gerarde, HW. (1960). Toxicology and biochemistry of aromatic hydrocarbons. In Toxicology and Biochemistry of Aromatic Hydrocarbons. Amsterdam, The Netherlands: Elsevier.

Ghanem, A; Shuler, ML. (2000a). Characterization of a perfusion reactor utilizing mammalian cells on microcarrier beads. Biotechnol Prog 16: 471-479. http://dx.doi.org/10.1021/bp000047o.

Ghanem, A; Shuler, ML. (2000b). Combining cell culture analogue reactor designs and PBPK models to probe mechanisms of naphthalene toxicity. Biotechnol Prog 16: 334-345. http://dx.doi.org/10.1021/bp9901522.

Gloede, E; Cichocki, JA; Baldino, JB; Morris, JB. (2011). A validated hybrid computational fluid dynamics-physiologically based pharmacokinetic model for respiratory tract vapor absorption in the human and rat and its application to inhalation dosimetry of diacetyl. Toxicol Sci 123: 231-246. http://dx.doi.org/10.1093/toxsci/kfr165.

Government of Canada. (2021). Publications: Healthy living. Available online at https://www.canada.ca/en/services/health/publications/healthy-living.html (accessed March 30, 2021).

Grigor, WG; Robin, H; Harley, JD. (1966). An Australian variant on "full-moon disease". Med J Aust 2: 1229-1230.

Gustafson, JB; Griffith Tell, J; Orem, D. (1997). Total petroleum hydrocarbon criteria working group series: Volume 3: Selection of representative TPH fractions based on fate and transport considerations. Amherst, MA: Amherst Scientific Publishers. https://netforum.avectra.com/eweb/shopping/shopping.aspx?pager=1&site=aehs&webcode=shopping&prd_key=146b7eaf-c92d-4e4b-af1f-092c5735a10b.

Hanssler, H. (1964). Lebensbedrohliche Naphthalinvergiftung bei einem Saugling durch Vaporin-Dampfe [Life-threatening naphthalene poisoning of an infant by vaporin fumes]. Dtsch Med Wochenschr 89: 1794-1797.

Harrison, RM; Smith D, JT; Luhana, L. (1996). Source apportionment of atmospheric polycyclic aromatic hydrocarbons collected from an urban location in Birmingham, U.K. Environ Sci Technol 30: 825-832. http://dx.doi.org/10.1021/es950252d.

Health Canada. (1996). Health-based tolerable daily intakes/concentrations and tumorigenic doses/concentrations for priority substances. (96-EHD-194). Ottawa, Canada: Environmental Health Directorate, Health Protection Branch. http://publications.gc.ca/collections/Collection/H46-2-96-194E.pdf.

Health Canada. (2013). Residential indoor air quality guideline: Naphthalene. https://www.canada.ca/en/health-canada/services/publications/healthy-living/residential-indoor-air-quality-guideline-naphthalene.html.

Health Canada. (2020). Guidelines for Canadian drinking water quality: Summary table. Ottawa, Ontario: Water and Air Quality Bureau, Healthy Environments and Consumer Safety Branch, Health Canada. https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/ewh-semt/alt_formats/pdf/pubs/water-eau/sum_guide-res_recom/summary-table-EN-2020-02-11.pdf.

Higgins, JPT; Thomas, J; Chandler, J; Cumpston, M; Li, T; Page, MJ; Welch, VA. (2022). Cochrane handbook for systematic reviews of interventions version 6.3. Higgins, JPT; Thomas, J; Chandler, J; Cumpston, M; Li, T; Page, MJ; Welch, VA. http://www.training.cochrane.org/handbook.

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

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

Howard, PH. (1989). Naphthalene. In Handbook of environmental fate and exposure data for organic chemicals: Volume I: Large production and priority pollutants. Chelsea, MI: Lewis Publishers.

HSA (Health and Safety Authority). (2020). Chemical agents and carcinogens: Code of practice 2020. Dublin, Ireland. https://www.hsa.ie/eng/publications_and_forms/publications/codes_of_practice/chemical_agents_cop_2020.pdf.

HSL (Health and Safety Laboratory). (2002). Draft 2nd indicative occupational exposure limit value (IOELV) list: Workplace measurement method summaries. (HSL/2002/23). Bootle, England: Health and Safety Executive. http://www.hse.gov.uk/research/hsl_pdf/2002/hsl02-23.pdf.

IARC (International Agency for Research on Cancer). (2002). Some traditional herbal medicines, some mycotoxins, naphthalene, and styrene [IARC Monograph]. Lyon, France. http://monographs.iarc.fr/ENG/Monographs/vol82/mono82.pdf.

Idaho DEQ. Rules for the control of air pollution in Idaho, 58.01.01 (Department of Environmental Quality2019). https://adminrules.idaho.gov/rules/current/58/580101.pdf

IDEM (Indiana Department of Environmental Management). (2019). Air toxics toxicity information. Available online at https://web.archive.org/web/20200922041251/https://www.in.gov/idem/toxic/2343.htm (accessed July 18, 2019).

IFA (Institut für Arbeitschutz der Deutschen Gesetzlichen Unfallversicherung). (2020). GESTIS international limit values database (Version April 2020) [Database]. Retrieved from https://limitvalue.ifa.dguv.de/

IPCS (International Programme on Chemical Safety). (2010). Characterization and application of physiologically based pharmacokinetic models in risk assessment. (Harmonization Project Document No 9). Geneva, Switzerland: World Health Organization. http://www.inchem.org/documents/harmproj/harmproj/harmproj9.pdf.

Irle, U. (1964). Akute hämolytische Anämie durch Naphthalin-Inhalation bei zwei Frühgeborenen und einem Neugeborenen. Dtsch Med Wochenschr 89: 1798-1800. http://dx.doi.org/10.1055/s-0028-1113200.

Jia, C; Batterman, S. (2010). A critical review of naphthalene sources and exposures relevant to indoor and outdoor air [Review]. Int J Environ Res Public Health 7: 2903-2939. http://dx.doi.org/10.3390/ijerph7072903.

Johns, LE; Cooper, GS; Galizia, A; Meeker, JD. (2015). Exposure assessment issues in epidemiology studies of phthalates [Review]. Environ Int 85: 27-39. http://dx.doi.org/10.1016/j.envint.2015.08.005.

JSOH (Japan Society for Occupational Health). (2017). Recommendation of occupational exposure limits (2017-2018). J Occup Health 59: 436-469. http://dx.doi.org/10.1539/joh.ROEL2017.

Kapraun, DF; Schlosser, PM; Nylander-French, LA; Kim, D; Yost, EE; Druwe, IL. (2020). A physiologically based pharmacokinetic model for naphthalene with inhalation and skin routes of exposure. J Toxicol Sci 177: 377-391. http://dx.doi.org/10.1093/toxsci/kfaa117.

Kim, D; Andersen, ME; Chao, YC; Egeghy, PP; Rappaport, SM; Nylander-French, LA. (2007). PBTK modeling demonstrates contribution of dermal and inhalation exposure components to end-exhaled breath concentrations of naphthalene. Environ Health Perspect 115: 894-901. http://dx.doi.org/10.1289/ehp.9778.

Kim, D; Andersen, ME; Nylander-French, LA. (2006). Dermal absorption and penetration of jet fuel components in humans. Toxicol Lett 165: 11-21. http://dx.doi.org/10.1016/j.toxlet.2006.01.009.

Klotz, K; Schindler, BK; Angerer, J. (2011). 1,2-Dihydroxynaphthalene as biomarker for a naphthalene exposure in humans. Int J Hyg Environ Health 214: 110-114. http://dx.doi.org/10.1016/j.ijheh.2010.11.003.

Lai, CH; Chen, KS; Wang, H. (2009). Influence of rice straw burning on the levels of polycyclic aromatic hydrocarbons in agricultural county of Taiwan. J Environ Sci 21: 1200-1207. http://dx.doi.org/10.1016/S1001-0742(08)62404-3.

Lanza, DL; Code, E; Crespi, CL; Gonzalez, FJ; Yost, GS. (1999). Specific dehydrogenation of 3-methylindole and epoxidation of naphthalene by recombinant human CYP2F1 expressed in lymphoblastoid cells. Drug Metab Dispos 27: 798-803.

Li, L; Carratt, S; Hartog, M; Kovalchuk, N; Jia, K; Wang, Y; Zhang, QY; Edwards, P; Van Winkle, L; Ding, X. (2017). Human CYP2A13 and CYP2F1 mediate naphthalene toxicity in the lung and nasal mucosa of CYP2A13/2F1-humanized mice. Environ Health Perspect 125: UNSP 067004. http://dx.doi.org/10.1289/EHP844.

Li, Z; Romanoff, LC; Lewin, MD; Porter, EN; Trinidad, DA; Needham, LL; Patterson, DG; Sjödin, A. (2010). Variability of urinary concentrations of polycyclic aromatic hydrocarbon metabolite in general population and comparison of spot, first-morning, and 24-h void sampling. J Expo Sci Environ Epidemiol 20: 526-535. http://dx.doi.org/10.1038/jes.2009.41.

Lu, T; Lansing, J; Zhang, W; Bechle, MJ; Hankey, S. (2019). Land Use Regression models for 60 volatile organic compounds: Comparing Google Point of Interest (POI) and city permit data. Sci Total Environ 677: 131-141. http://dx.doi.org/10.1016/j.scitotenv.2019.04.285.

MassDEP (Massachusetts Department of Environmental Protection). (2011). Methodology for updating air guidelines: allowable ambient limits (AALS) and threshold effects exposure limits (TELS). https://www.mass.gov/files/documents/2016/08/tc/aaltel11.pdf.

MassDEP (Massachusetts Department of Environmental Protection). (2019). MassDep Ambient Air Toxics Guidelines. Available online at https://www.mass.gov/service-details/massdep-ambient-air-toxics-guidelines (accessed

MDH (Minnesota Department of Health). (2004). Naphthalene - Guidance for Air: Acute and Chronic Health-Based Values; Inhalation Unit Risk Consultation.

https://www.health.state.mn.us/communities/environment/risk/docs/guidance/air/naphthalene.pdf.

MDH (Minnesota Department of Health). (2019). Air guidance values. Available online at https://www.health.state.mn.us/communities/environment/risk/guidance/air/table.html (accessed

Meeker, JD; Barr, DB; Serdar, B; Rappaport, SM; Hauser, R. (2007). Utility of urinary 1-naphthol and 2-naphthol levels to assess environmental carbaryl and naphthalene exposure in an epidemiology study. J Expo Sci Environ Epidemiol 17: 314-320. http://dx.doi.org/10.1038/sj.jes.7500502.

Michigan DEQ (Michigan Department of Environmental Quality). (2016). Cleanup criteria and screening levels development and application: Remediation and redevelopment division resource materials (draft). Lansing, MI. https://www.michigan.gov/documents/deq/deq-rrd-chem-CleanupCriteriaTSD_527410_7.pdf.

Morris, JB. (2013). Nasal dosimetry of inspired naphthalene vapor in the male and female B6C3F1 mouse. Toxicology 309: 66-72. http://dx.doi.org/10.1016/j.tox.2013.04.009.

Morris, JB; Buckpitt, AR. (2009). Upper respiratory tract uptake of naphthalene. Toxicol Sci 111: 383-391. http://dx.doi.org/10.1093/toxsci/kfp138.

Naiman, JL; Kosoy, MH. (1964). Red cell glucose-6-phosphate dehydrogenase deficiency--a newly recognized cause of neonatal jaundice and kernicterus in Canada. Can Med Assoc J 91: 1243-1249.

NASEM (National Academies of Sciences, Engineering, and Medicine). (2018). Progress toward transforming the Integrated Risk Information System (IRIS) program: A 2018 evaluation. Washington, DC: National Academies Press. http://dx.doi.org/10.17226/25086.

NASEM (National Academies of Sciences, Engineering, and Medicine). (2021). Review of U.S. EPA's ORD staff handbook for developing IRIS assessments: 2020 version. Washington, DC: National Academies Press. http://dx.doi.org/10.17226/26289.

Nazaroff, WW; Singer, BC. (2004). Inhalation of hazardous air pollutants from environmental tobacco smoke in US residences. J Expo Anal Environ Epidemiol 14: S71-S77. http://dx.doi.org/10.1038/sj.jea.7500361.

NC Department of Environmental Quality (North Carolina Department of Environmental Quality). (2014). Toxic air pollutant guidelines. (15A NCAC 02D .1104). https://files.nc.gov/ncdeq/Air%20Quality/rules/rules/D1104.pdf.

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

NIOSH (National Institute for Occupational Safety and Health). (1994). Immediately dangerous to life or health concentrations (IDLH): Naphthalene. Atlanta, GA. https://www.cdc.gov/niosh/idlh/91203.html.

NIOSH (National Institute for Occupational Safety and Health). (2018). NIOSH pocket guide to chemical hazards. Index of chemical abstracts service registry numbers (CAS No.). Atlanta, GA. http://www.cdc.gov/niosh/npg/npgdcas.html.

NJ DEP (New Jersey Department of Environmental Protection). (2020). Toxicity values for inhalation exposure. (NJDEP/DAQ/AQEv). Trenton, NJ. https://www.state.nj.us/dep/aqpp/downloads/risk/ToxAll2020.pdf.

NRC (National Research Council). (2011). Review of the Environmental Protection Agency's draft IRIS assessment of formaldehyde (pp. 1-194). Washington, DC: The National Academies Press. http://dx.doi.org/10.17226/13142.

NRC (National Research Council). (2014). Review of EPA's Integrated Risk Information System (IRIS) process. Washington, DC: The National Academies Press. http://dx.doi.org/10.17226/18764.

NTP (National Toxicology Program). (1991). Final study report and appendix: Developmental toxicity of naphthalene (CAS no. 91-20-3) in Sprague-Dawley (CD) rats on gestational days 6 through 15. (TER91006). Research Triangle Park, NC. https://ntrl.ntis.gov/NTRL/dashboard/searchResults.xhtml?searchQuery=PB92135623.

NTP (National Toxicology Program). (1992). NTP technical report on the toxicology and carcinogenesis studies of naphthalene (CAS no. 91-20-3) in B6C3F1 mice (inhalation studies). (NTP TR 410; NIH Publication No. 92-3141). Research Triangle Park, NC. http://ntp.niehs.nih.gov/ntp/htdocs/LT_rpts/tr410.pdf.

NTP (National Toxicology Program). (2000). NTP technical report on the toxicology and carcinogenesis studies of naphthalene (CAS no. 91-20-3) in F344/N rats (inhalation studies) (pp. 1-173). (NTP TR 500). Research Triangle Park, NC.

NTP (National Toxicology Program). (2011). Naphthalene. In Report on carcinogens: Twelfth edition (12th ed., pp. 276-278). Research Triangle Park, NC. http://www.iaff.org/HS/PDF/12th%20Report%20on%20Carcinogens%20-%202011.pdf.

NTP (National Toxicology Program). (2015). Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration. Research Triangle Park, NC: U.S. Deptartment of Health and Human Services, National Toxicology Program, Office of Health Assessment and Translation. https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf.

NTP (National Toxicology Program). (2019). Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration. Research Triangle, NC: National Institute of Environmental Health Sciences. https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019_508.pdf.

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

O'Brien, PJ. (1991). Molecular mechanisms of quinone cytotoxicity [Review]. Chem Biol Interact 80: 1-41.

OEHHA (California Office of Environmental Health Hazard Assessment). (2000). Chronic toxicity summary - naphthalene (pp. 413-419). Sacramento, CA: California Environmental Protection Agency. http://oehha.ca.gov/media/downloads/crnr/appendixd3final.pdf.

OEHHA (California Office of Environmental Health Hazard Assessment). (2011). Appendix B. Chemical-specific summaries of the information used to derive unit risk and cancer potency values: Naphthalene (pp. 409-431).

Ontario Ministry of Labour. (2020). Current occupational exposure limits for Ontario workplaces required under regulation 833. Ontario, Canada. https://www.labour.gov.on.ca/english/hs/pubs/oel_table.php.

Oregon DEQ (Oregon Department of Environmental Quality). (2018). Ambient benchmark concentrations (ABCs): Based on 2014-2017 ATSAC review. Portland, OR. https://www.oregon.gov/deq/FilterDocs/airtox-abc.pdf.

OSHA (Occupational Safety & Health Administration). (2019). Permissible exposure limits: OSHA annotated table Z-1 [Website]. Washington, DC: United States Department of Labor, Occupational Safety & Health Administration. https://www.osha.gov/dsg/annotated-pels/tablez-1.html.

OSHA (Occupational Safety & Health Administration). (2020a). Air contaminants: Occupational safety and health standards for shipyard employment, subpart Z, toxic and hazardous substances. (OSHA Standard 1915.1000). Washington, DC. https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=10286.

OSHA (Occupational Safety & Health Administration). (2020b). Safety and health regulations for construction: Occupational health and environmental controls: Gases, vapors, fumes, dusts, and mists: Appendix A. Available online at http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=10629 (accessed

Pellizzari, ED; Hartwell, TD; Harris, BS, III; Waddell, RD; Whitaker, DA; Erickson, MD. (1982). Purgeable organic compounds in mother's milk. Bull Environ Contam Toxicol 28: 322-328. http://dx.doi.org/10.1007/BF01608515.

Perrier, F; Giorgis-Allemand, L; Slama, R; Philippat, C. (2016). Within-subject pooling of biological samples to reduce exposure misclassification in biomarker-based studies. Epidemiology 27: 378-388. http://dx.doi.org/10.1097/EDE.0000000000000460.

Piñeiro, R; Jimenez-Relinque, E; Nevshupa, R; Castellote, M. (2021). Primary and Secondary Emissions of VOCs and PAHs in Indoor Air from a Waterproof Coal-Tar Membrane: Diagnosis and Remediation. Int J Environ Res Public Health 18. http://dx.doi.org/10.3390/ijerph182312855.

Powers, M. (2022). Microsoft Teams meeting: Naphthalene biomarker discussion with Antonia Calafat (CDC). Available online at (accessed

Québec. Regulation respecting occupational health and safety, chapter S-2.1, r. 13 (2020). http://legisquebec.gouv.qc.ca/en/showdoc/cr/S-2.1,%20r.%2013?csi_scan_9222d36c6a354dc6=BO9xyrMZ+270UP3j0MGuOD0kZjgFAAAAXrM3HA==&bcsi_scan_filename=S-2.1,%20r.%2013&bcsi_scan_9222d36c6a354dc6=KXzmpPueuN0L1AjnJOB1Zerr85YMAAAAyhrPTg==&bcsi_scan_filename=S-2.1,%20r.%2013

Quick, DJ; Shuler, ML. (1999). Use of in vitro data for construction of a physiologically based pharmacokinetic model for naphthalene in rats and mice to probe species differences. Biotechnol Prog 15: 540-555. http://dx.doi.org/10.1021/bp990057t.

R Core Team (R Development Core Team). (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Radke, EG; Braun, JM; Meeker, JD; Cooper, GS. (2018). Phthalate exposure and male reproductive outcomes: A systematic review of the human epidemiological evidence [Review]. Environ Int 121: 764-793. http://dx.doi.org/10.1016/j.envint.2018.07.029.

Ranjbar, M; Rotondi, MA; Ardern, CI; Kuk, JL. (2015). Urinary biomarkers of polycyclic aromatic hydrocarbons are associated with cardiometabolic health risk. PLoS ONE 10: e0137536. http://dx.doi.org/10.1371/journal.pone.0137536.

Reynolds, S. (1997). Naphthalene: Neurotoxicity study by a single oral gavage administration to CD rats followed by a 14-day observation period: Final report: Lab project number: 96/1140: 96/RCE001/1140: RCE/001. Unpublished study prepared by Huntingdon Life Sciences Ltd. 383 p. (MRID 44282801). Reynolds, S.

RI DEM (Rhode Island Department of Environmental Management). (2008). Rhode island air toxics: Guidelines. http://www.dem.ri.gov/programs/benviron/air/pdf/airtoxgl.pdf.

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

Robbins, MC. (1951). Determination of naphthalene in air. Arch Environ Occup Health 4: 85-87.

Robinson, MS; Anthony, TR; Littau, SR; Herckes, P; Nelson, X; Poplin, GS; Burgess, JL. (2008). Occupational PAH exposures during prescribed pile burns. Ann Occup Hyg 52: 497-508. http://dx.doi.org/10.1093/annhyg/men027.

Safe Work Australia. (2019). Workplace exposure standards for airborne contaminants. Canberra, Australia. https://www.safeworkaustralia.gov.au/system/files/documents/1912/workplace-exposure-standards-airborne-contaminants.pdf.

Schafer, WB. (1951). Acute hemolytic anemia related to naphthalene: Report of a case in a newborn infant. Pediatrics 7: 172-174.

Schünemann, H; Brożek, J; Guyatt, G; Oxman, A. (2013). GRADE handbook. Available online at https://gdt.gradepro.org/app/handbook/handbook.html (accessed April 22, 2022).

Scinicariello, F; Buser, MC. (2014). Urinary Polycyclic Aromatic Hydrocarbons and Childhood Obesity: NHANES (2001-2006). Environ Health Perspect 122: 299-303. http://dx.doi.org/10.1289/ehp.1307234.

Shopp, GM; White, KL, Jr; Holsapple, MP; Barnes, DW; Duke, SS; Anderson, AC; Condie, LW, Jr; Hayes, JR; Borzelleca, JF. (1984). Naphthalene toxicity in CD-1 mice: general toxicology and immunotoxicology. Toxicol Sci 4: 406-419. http://dx.doi.org/10.1093/toxsci/4.3part1.406.

Shuler, ML; Ghanem, A; Quick, D; Wong, MC; Miller, P. (1996). A self-regulating cell culture analog device to mimic animal and human toxicological responses. Biotechnol Bioeng 52: 45-60. http://dx.doi.org/10.1002/(SICI)1097-0290(19961005)52:1<45::AID-BIT5>3.0.CO;2-Z.

Shultz, MA; Choudary, PV; Buckpitt, AR. (1999). Role of murine cytochrome P-450 2F2 in metabolic activation of naphthalene and metabolism of other xenobiotics. J Pharmacol Exp Ther 290: 281-288.

Smith, MT; Guyton, KZ; Gibbons, CF; Fritz, JM; Portier, CJ; Rusyn, I; DeMarini, DM; Caldwell, JC; Kavlock, RJ; Lambert, PF; Hecht, SS; Bucher, JR; Stewart, BW; Baan, RA; Cogliano, VJ; Straif, K. (2016). Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis [Review]. Environ Health Perspect 124: 713-721. 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. BMJ 355: i4919. http://dx.doi.org/10.1136/bmj.i4919.

Strandberg, B; Julander, A; Sjöström, M; Lewné, M; Koca Akdeva, H; Bigert, C. (2018). Evaluation of polyurethane foam passive air sampler (PUF) as a tool for occupational PAH measurements. Chemosphere 190: 35-42. http://dx.doi.org/10.1016/j.chemosphere.2017.09.106.

Su, T; Bao, ZP; Zhang, QY; Smith, TJ; Hong, JY; Ding, XX. (2000). Human cytochrome p450 CYP2A13: Predominant expression in the respiratory tract and its high efficiency metabolic activation of a tobacco-specific carcinogen, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. Cancer Res 60: 5074-5079.

SWCAA (Southwest Clean Air Agency). (2019). SWCAA pollutant details: Naphthalene. Available online at http://www.swcleanair.org/epages/pollutantdetail.asp?id=2723 (accessed July 23, 2019).

SWCAA (Southwest Clean Air Agency). (2021). Pollutant search [Database]. Retrieved from http://www.swcleanair.gov/epages/polsrch.asp

Sweeney, LM; Shuler, ML; Babish, JG; Ghanem, A. (1995). A cell culture analog of rodent physiology: Application to naphthalene toxicology. Toxicol In Vitro 9: 307-316. http://dx.doi.org/10.1016/0887-2333(95)00007-U.

Sweeney, LM; Shuler, ML; Quick, DJ; Babish, JG. (1996). A preliminary physiologically based pharmacokinetic model for naphthalene and naphthalene oxide in mice and rats. Ann Biomed Eng 24: 305-320. http://dx.doi.org/10.1007/BF02667357.

TCEQ (Texas Commission on Environmental Quality). (2018). TRRP protective concentration levels: April 2018 PCL and supporting tables [Database]. Retrieved from https://www.tceq.texas.gov/remediation/trrp/trrppcls.html

TCEQ (Texas Commission on Environmental Quality). (2021). Final development support documents (DSDs). Available online at https://www.tceq.texas.gov/toxicology/dsd/final (accessed

Teeguarden, JG; Bogdanffy, MS; Covington, TR; Tan, C; Jarabek, AM. (2008). A PBPK model for evaluating the impact of aldehyde dehydrogenase polymorphisms on comparative rat and human nasal tissue acetaldehyde dosimetry. Inhal Toxicol 20: 375-390. http://dx.doi.org/10.1080/08958370801903750.

TERA (Toxicology Excellence for Risk Assessment). (2021). ITER database [Database]. Cincinnati, OH. Retrieved from https://iter.tera.org/

Thayer, KA; Shaffer, RM; Angrish, M; Arzuaga, X; Carlson, LM; Davis, A; Dishaw, L; Druwe, I; Gibbons, C; Glenn, B; Jones, R; Kaiser, JP; Keshava, C; Keshava, N; Kraft, A; Lizarraga, L; Markey, K; Persad, A; Radke, EG; ... Yost, E. (2022). Use of systematic evidence maps within the US environmental protection agency (EPA) integrated risk information system (IRIS) program: Advancements to date and looking ahead [Comment]. Environ Int 169: 107363. http://dx.doi.org/10.1016/j.envint.2022.107363.

Tiesjema, B; Baars, AJ. (2009). Re-evaluation of some human-toxicological Maximum Permissible Risk levels earlier evaluated in the period 1991-2001. (RIVM Report 711701092). Bilthoven, the Netherlands: National Institute for Public Health and the Environment (Netherlands). http://www.rivm.nl/bibliotheek/rapporten/711701092.pdf.

ToxNet Hazardous Substances Data Bank. (2017). HSDB: Naphthalene. Bethesda, MD: National Institute of Health, U.S. National Library of Medicine. Retrieved from https://toxnet.nlm.nih.gov/newtoxnet/hsdb.htm

Tsang, HL; Wu, S; Leung, CK; Tao, S; Wong, MH. (2011). Body burden of POPs of Hong Kong residents, based on human milk, maternal and cord serum. Environ Int 37: 142-151. http://dx.doi.org/10.1016/j.envint.2010.08.010.

U.S. APHC (U.S. Army Public Health Command). (2013). Environmental health risk assessment and chemical exposure guidelines for deployed military personnel. (Technical guide 230, 2013 revision). Aberdeen Proving Ground, MD. https://phc.amedd.army.mil/PHC%20Resource%20Library/TG230-DeploymentEHRA-and-MEGs-2013-Revision.pdf.

U.S. EPA (U.S. Environmental Protection Agency). (1986). Guidelines for carcinogen risk assessment [EPA Report] (pp. 33993-34003). (EPA/630/R-00/004). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. https://www.epa.gov/iris/basic-information-about-integrated-risk-information-system#risk.

U.S. EPA (U.S. Environmental Protection Agency). (1988). Recommendations for and documentation of biological values for use in risk assessment [EPA Report]. (EPA600687008). Cincinnati, OH. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=34855.

U.S. EPA (U.S. Environmental Protection Agency). (1991). Guidelines for developmental toxicity risk assessment. Fed Reg 56: 63798-63826.

U.S. EPA (U.S. Environmental Protection Agency). (1993). NATICH: Data base report on state, local and EPA air toxics activities. (EPA-453/R-93-041). Research Triangle Park, NC. http://nepis.epa.gov/exe/ZyPURL.cgi?Dockey=2000NS7S.txt.

U.S. EPA (U.S. Environmental Protection Agency). (1994). Methods for derivation of inhalation reference concentrations and application of inhalation dosimetry [EPA Report]. (EPA600890066F). Research Triangle Park, NC.

https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=71993&CFID=51174829&CFTOKEN=25006317.

U.S. EPA (U.S. Environmental Protection Agency). (1996). Guidelines for reproductive toxicity risk assessment [EPA Report]. (EPA/630/R-96/009). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=30004YQB.txt.

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

U.S. EPA (U.S. Environmental Protection Agency). (1998b). Toxicological review of naphthalene (CAS no. 91-20-3) in support of summary information on the Integrated Risk Information System (IRIS) [EPA Report]. Washington, DC: Integrated Risk Information System. https://iris.epa.gov/static/pdfs/0436tr.pdf.

U.S. EPA (U.S. Environmental Protection Agency). (1999). Compendium of methods for the determination of toxic organic compounds in ambient air. second edition. (EPA/625/R-96/010b). https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/PB99172355.xhtml.

U.S. EPA (U.S. Environmental Protection Agency). (2002). A review of the reference dose and reference concentration processes. (EPA630P02002F). Washington, DC. https://www.epa.gov/sites/production/files/2014-12/documents/rfd-final.pdf.

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

U.S. EPA (U.S. Environmental Protection Agency). (2005a). Guidelines for carcinogen risk assessment [EPA Report]. (EPA630P03001F). Washington, DC. https://www.epa.gov/sites/production/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf.

U.S. EPA (U.S. Environmental Protection Agency). (2005b). Supplemental guidance for assessing susceptibility from early-life exposure to carcinogens [EPA Report]. (EPA/630/R-03/003F). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. https://www.epa.gov/risk/supplemental-guidance-assessing-susceptibility-early-life-exposure-carcinogens.

U.S. EPA (U.S. Environmental Protection Agency). (2011a). Recommended use of body weight 3/4 as the default method in derivation of the oral reference dose. (EPA100R110001). Washington, DC. https://www.epa.gov/sites/production/files/2013-09/documents/recommended-use-of-bw34.pdf.

U.S. EPA (U.S. Environmental Protection Agency). (2011b). Toxicological review of dichloromethane (methylene chloride) (CASRN 75-09-2): In support of summary information on the Integrated Risk Information System (IRIS) [EPA Report]. (EPA/635/R-10/003F). Washington, DC. https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0070tr.pdf.

U.S. EPA (U.S. Environmental Protection Agency). (2011c). Toxicological review of trichloroethylene (CAS No. 79-01-6) in support of summary Information on the Integrated Risk Information System (IRIS) [EPA Report]. (EPA635R09011F). Washington, DC. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100CB6V.txt.

U.S. EPA (U.S. Environmental Protection Agency). (2012a). Advances in inhalation gas dosimetry for derivation of a reference concentration (RfC) and use in risk assessment (pp. 1-140). (EPA/600/R-12/044). Washington, DC.

https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=244650&CFID=50524762&CFTOKEN=17139189.

U.S. EPA (U.S. Environmental Protection Agency). (2012b). Benchmark dose technical guidance [EPA Report]. (EPA100R12001). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. https://www.epa.gov/risk/benchmark-dose-technical-guidance.

U.S. EPA (U.S. Environmental Protection Agency). (2012c). Inert use information [Database]. Washington, D.C.: U.S. Environmental Protection Agency. Office of Pesticide Program. Retrieved from https://iaspub.epa.gov/apex/pesticides/f?p=INERTFINDER:1:0::NO:1::

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

U.S. EPA (U.S. Environmental Protection Agency). (2014b). Summary report: State-of-the-science workshop on chemically-induced mouse lung tumors: Applications to human health assessments. (600R14002). http://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=291094&CFID=40888511&CFTOKEN=98632970.

U.S. EPA (U.S. Environmental Protection Agency). (2015a). Naphthalene acetates; pesticide tolerances. Fed Reg 80: 77255-77260.

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

U.S. EPA (U.S. Environmental Protection Agency). (2016). Public database 2016 chemical data reporting (May 2017 release). Washington, DC: US Environmental Protection Agency, Office of Pollution Prevention and Toxics. https://www.epa.gov/chemical-data-reporting.

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

U.S. EPA (U.S. Environmental Protection Agency). (2018a). 2018 Edition of the drinking water standards and health advisories tables [EPA Report]. (EPA 822-F-18-001). Washington, DC: U.S. Environmental Protection Agency, Office of Water. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100U7U8.txt.

U.S. EPA (U.S. Environmental Protection Agency). (2018b). Access acute exposure guideline levels (AEGLs) values database [Database]. Washington, DC: National Research Council, National Academy of Sciences. Retrieved from https://www.epa.gov/aegl/access-acute-exposure-guideline-levels-aegls-values#chemicals

U.S. EPA (U.S. Environmental Protection Agency). (2018c). Naphthalene: Human health risk assessment in support of registration review. (EPA-HQ-OPP-2016-0113-0018). https://www.regulations.gov/document?D=EPA-HQ-OPP-2016-0113-0018.

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

U.S. EPA (U.S. Environmental Protection Agency). (2020a). Dose-response assessment for assessing health risks associated with exposure to hazardous air pollutants. Available online at https://www.epa.gov/fera/dose-response-assessment-assessing-health-risks-associated-exposure-hazardous-air-pollutants (accessed March 30, 2021).

U.S. EPA (U.S. Environmental Protection Agency). (2020b). ORD staff handbook for developing IRIS assessments (public comment draft) [EPA Report]. (EPA/600/R-20/137). Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development, Center for Public Health and Environmental Assessment. https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086.

U.S. EPA (U.S. Environmental Protection Agency). (2020c). Umbrella quality assurance project plan (QAPP) for dosimetry and mechanism-based models. (EPA QAPP ID Number: L-CPAD-0032188-QP-1-2). Research Triangle Park, NC.

U.S. EPA (U.S. Environmental Protection Agency). (2021a). Integrated risk information system (IRIS) database [Database]. Washington, DC. Retrieved from http://www.epa.gov/iris/

U.S. EPA (U.S. Environmental Protection Agency). (2021b). Pesticide chemical search [Database]. Retrieved from https://iaspub.epa.gov/apex/pesticides/f?p=chemicalsearch:1

U.S. EPA (U.S. Environmental Protection Agency). (2022). ORD staff handbook for developing IRIS assessments [EPA Report]. (EPA 600/R-22/268). Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development, Center for Public Health and Environmental Assessment. https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=356370.

U.S. EPA (U.S. Environmental Protection Agency). (2023). The Superfund Enterprise Management System (SEMS). Available online at https://www.epa.gov/enviro/sems-overview (accessed March 9, 2023).

Valaes, T; Doxiadis, SA; Fessas, P. (1963). Acute hemolysis due to naphthalene inhalation. J Paediatr Child Health 63: 904-915. http://dx.doi.org/10.1016/S0022-3476(63)80221-8.

Viravaidya, K; Shuler, ML. (2004). Incorporation of 3T3-L1 cells to mimic bioaccumulation in a microscale cell culture analog device for toxicity studies. Biotechnol Prog 20: 590-597. http://dx.doi.org/10.1021/bp034238d.

Viravaidya, K; Sin, A; Shuler, ML. (2004). Development of a microscale cell culture analog to probe naphthalene toxicity. Biotechnol Prog 20: 316-323. http://dx.doi.org/10.1021/bp0341996.

VT ANR. Air pollution control regulations, Subchapter III (State of Vermont agency of natural resources2018). https://dec.vermont.gov/sites/dec/files/aqc/laws-regs/documents/AQCD%20Regulations%20ADOPTED_Dec132018.pdf#page=127

Washington State Legislature. Table of ASIL, SQER and de minimis emission values, Sec. 173-460-150 (WSL2009). https://apps.leg.wa.gov/WAC/default.aspx?cite=173-460-150

Wheeler, AJ; Dobbin, NA; Héroux, ME; Fisher, M; Sun, L; Khoury, CF; Hauser, R; Walker, M; Ramsay, T; Bienvenu, JF; Leblanc, A; Daigle, E; Gaudreau, E; Belanger, P; Feeley, M; Ayotte, P; Arbuckle, TE. (2014). Urinary and breast milk biomarkers to assess exposure to naphthalene in pregnant women: an investigation of personal and indoor air sources. Environ Health 13: 30. http://dx.doi.org/10.1186/1476-069X-13-30.

WHO (World Health Organization). (2010). Guidelines for indoor air quality: Selected pollutants. Geneva. http://www.euro.who.int/__data/assets/pdf_file/0009/128169/e94535.pdf.

WHO (World Health Organization). (2017). Guidelines for drinking-water quality, 4th edition, incorporating the 1st addendum (4th ed.). Geneva, Switzerland. https://www.who.int/publications/i/item/9789241549950.

WHO (World Health Organization). (2021). Online catalog for the Environmental Health Criteria (EHC) monographs [Database]. Geneva, Switzerland: World Health Organization (WHO). Retrieved from http://www.who.int/ipcs/publications/ehc/en/

Willems, B; Melnick, R; Kohn, M; Portier, C. (2001). A physiologically based pharmacokinetic model for inhalation and intravenous administration of naphthalene in rats and mice. Toxicol Appl Pharmacol 176: 81-91. http://dx.doi.org/10.1006/taap.2001.9269.

Wolf, O. (1976). Krebserkrankungen bei Chemiearbeitern einer ehemaligen Naphthalinreinigung [Cancer diseases in chemical workers in a former naphthalene cleaning plant]. Dtsch Gesundheitsw 31: 996-999.

Wolf, O. (1978). [Cancer of the larynx in naphthalene cleaners]. Z Gesamte Hyg 24: 737-739.

Wolffe, TAM; Whaley, P; Halsall, C; Rooney, AA; Walker, VR. (2019). Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management. Environ Int 130: 104871. http://dx.doi.org/10.1016/j.envint.2019.05.065.

Worksafe (Worksafe New Zealand). (2018). Workplace exposure standards and biological exposure indices (10th ed.). Wellington, New Zealand. https://www.worksafe.govt.nz/topic-and-industry/work-related-health/monitoring/exposure-standards-and-biological-exposure-indices/.

Worksafe (Worksafe New Zealand). (2022). Workplace exposure standards and biological exposure indices (13th ed.). Wellington, New Zealand. https://www.worksafe.govt.nz/dmsdocument/20238-workplace-exposure-standards-and-biological-exposure-indices-edition-13/latest.

Yang, P; Sun, H; Gong, YJ; Wang, YX; Liu, C; Chen, YJ; Sun, L; Huang, LL; Ai, SH; Lu, WQ; Zeng, Q. (2017a). Repeated measures of urinary polycyclic aromatic hydrocarbon metabolites in relation to altered reproductive hormones: A cross-sectional study in China. Int J Hyg Environ Health 220: 1340-1346. http://dx.doi.org/10.1016/j.ijheh.2017.09.004.

Yang, P; Wang, YX; Sun, L; Chen, YJ; Liu, C; Huang, LL; Lu, WQ; Zeng, Q. (2017b). Urinary metabolites of polycyclic aromatic hydrocarbons, sperm DNA damage and spermatozoa apoptosis. J Hazard Mater 329: 241-248. http://dx.doi.org/10.1016/j.jhazmat.2017.01.053.

Yin, S; Tang, M; Chen, F; Li, T; Liu, W. (2017). Environmental exposure to polycyclic aromatic hydrocarbons (PAHs): The correlation with and impact on reproductive hormones in umbilical cord serum. Environ Pollut 220: 1429-1437. http://dx.doi.org/10.1016/j.envpol.2016.10.090.

Yost, EE; Galizia, A; Kapraun, DF; Persad, AS; Vulimiri, SV; Angrish, M; Lee, JS; Druwe, IL. (2021). Health effects of naphthalene exposure: A systematic evidence map and analysis of potential considerations for dose-response evaluation. Environ Health Perspect 129: 76002. http://dx.doi.org/10.1289/EHP7381.

Zhang, Z; Kleinstreuer, C. (2011). Deposition of naphthalene and tetradecane vapors in models of the human respiratory system. Inhal Toxicol 23: 44-57. http://dx.doi.org/10.3109/08958378.2010.540261.

Zhu, H; Martinez-Moral, MP; Kannan, K. (2021). Variability in urinary biomarkers of human exposure to polycyclic aromatic hydrocarbons and its association with oxidative stress. Environ Int 156: 106720. http://dx.doi.org/10.1016/j.envint.2021.106720.

Zinkham, WH; Childs, B. (1957). Effect of vitamin K and naphthalene metabolites on glutathione metabolism of erythrocytes from normal newborns and patients with naphthalene hemolytic anemia. Arch Pediatr Adolesc Med 94: 420-423.

Zinkham, WH; Childs, B. (1958). A defect of glutathione metabolism in erythrocytes from patients with a naphthalene-induced hemolytic anemia. Pediatrics 22: 461-471.

APPENDICES

APPENDIX A. SURVEY OF EXISTING REFERENCE VALUES FOR NAPHTHALENE

Table A-1 lists websites which were searched for relevant human health reference values for naphthalene, along with indications of the results of the search. In addition to these sources, the ToxValDB on EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard/chemical_lists/TOXVAL_V5) was also searched for additional reference values that were not captured by other sources. When values were identified for naphthalene, they are shown in Figures 1-2 and described in Tables A-2 and A-3 if details were provided on how the values were derived. When values were identified from sources that did not provide derivation details, they are described in Table A-4 but not shown in Figures 1-2. The values in these tables are current as of August 2022.

Table A-1. Sources searched for naphthalene heath effect reference values

Source	Search Results	Reference
American Conference of Governmental Industrial Hygienists (ACGIH)	See Appendix Table A2.	ACGIH (2007)
American Industrial Hygiene Association (AIHA)	No search results found.	AIHA (2016)
Agency for Toxic Substances and Disease Registry (ATSDR)	See Appendix Tables A2 and A3.	ATSDR (2021)
		ATSDR (2017)
California Environmental Protection Agency (CalEPA)	See Appendix Table A2.	CalEPA (2016)
Connecticut Department of Energy & Environmental Protection (CT DEEP)	See Appendix Tables A2 and A3.	CT DEEP (2015)
		CT DEEP (2018)
Deutsche Forschungsgemeinschaft, German Research Foundation (DFG)	No search results found.	DFG (2020)
Drinking Water Standards and Health Advisories (DWSHA)	See Appendix Table A3.	U.S. EPA (2018a)
Acute Exposure Level Guidelines from the U.S. Environmental Protection Agency and National Research Council) (EPA/NRC AEGL)	No search results found.	U.S. EPA (2018b)
Health Canada	See Appendix Table A2.	Government of Canada (2021)
	No values found.	Health Canada (2020)
	No values found.	Health Canada (1996)
Health and Safety Authority (HSA)	See Appendix Table A2.	HSA (2020)
Health and Safety Laboratory (HSL)	No values found.	HSL (2002)

Source	Search Results	Reference
Indiana Department of Environmental Management (IDEM)	See Appendix Table A2.	IDEM (2019)
Idaho Department of Environmental Quality (ID DEQ)	See Appendix Table A4.	Idaho DEQ (2019)
Institut für Arbeitsschutz, The Institute for Occupational Safety and Health (IFA)	See Appendix Table A4.	IFA (2020)
Integrated Risk Information System (IRIS)	See Appendix Tables A2 and A3.	U.S. EPA (2021a)
International Toxicity Estimates for Risk (ITER)	No unique search results found.	TERA (2021)
Japan Society for Occupational Health (JSOH)	No values found.	JSOH (2017)
Massachusetts Department of Environmental Protection (MassDEP)	See Appendix Table A4.	MassDEP (2019)
Minnesota Department of Health (MDH)	See Appendix Table A2.	MDH (2019)
Michigan Department of Environment, Great Lakes & Energy (MI EGLE)	See Appendix Tables A2 and A3.	Michigan DEQ (2016)
National Air Toxics Information Clearinghouse (NATICH)	See Appendix Tables A2 and A4.	U.S. EPA (1993)
North Carolina Department of Environmental Quality (NC DEQ)	No values found.	NC Department of Environmental Quality (2014)
Nevada Division of Environmental Protection (NDEP)	See Appendix Table A2.	NDEP (2017)
National Institute for Occupational Safety and Health (NIOSH)	See Appendix Table A2.	NIOSH (2018)
New Jersey Department of Environmental Protection (NJ DEP)	See Appendix Table A2.	NJ DEP (2020)
New York State Department of Environmental Conservation (NY DEC)	See Appendix Tables A2 and A3.	NYSDEC (2006)
Office of Air Quality Planning and Standards (OAQPS)	No unique search results found.	U.S. EPA (2020a)
Ontario Ministry of Labour	See Appendix Table A2.	Ontario Ministry of Labour (2020)
Office of Pesticide Programs (OPP)	See Appendix Table A3.	U.S. EPA (2021b)
Oregon Department of Environmental Quality (OR DEQ)	See Appendix Table A2.	Oregon DEQ (2018)
Occupational Safety and Health Administration (OSHA)	See Appendix Table A2.	OSHA (2019)
		OSHA (2020a)
		OSHA (2020b)
Protective Action Criteria (PAC) Database	See Appendix Table A2.	DOE (2018)
Publications Quebec	See Appendix Table A2.	Québec (2020)
Rhode Island Department of Environmental Management (RI DEM)	See Appendix Table A2.	RI DEM (2008)
	No values found.	Tiesjema and Baars (2009)

Source	Search Results	Reference
Rijksinstituut voor Volksgezondheid en Milieu (RIVM), The Netherlands Institute for Public Health and the Environment	See Appendix Table A2.	Dusseldorp et al. (2011)
	No values found.	RIVM (2001)
Safe Work Australia	See Appendix Table A2.	Safe Work Australia (2019)
Southwest Clean Air Association (SWCAA)	See Appendix Table A4.	SWCAA (2021)
Texas Commission on Environmental Quality (TCEQ)	No values found.	TCEQ (2021)
	See Appendix Tables A2 and A3.	TCEQ (2018)
United States Army Public Health Center (USAPHC)	See Appendix Table A4.	U.S. APHC (2013)
Vermont Department of Environmental Conservation (VT DEC)	See Appendix Table A4.	VT ANR (2018)
Washington State Dept. of Ecology	See Appendix Table A4.	Washington State Legislature (2009)
Worksafe	See Appendix Table A4.	Worksafe (2018)
World Health Organization (WHO)	No values found.	WHO (2017)
		WHO (2021)

Table A-2. Details on derivation of the available health effect reference values for inhalation exposure to naphthalene (from Figure 2-1 of the main text)

	Reference Value Name	Duration	Reference Value		Health Effect	Point of Departure	Qualifier	Source	Uncertainty Factors[a]	Notes on Derivation	Review Status
			(mg/m^3)	(ppm)							
Emergency Response	PAC-3	1 hr	2,600	500	Adopted previous IDLH	--	--	(NIOSH, 1994)	--	Adopted previous IDLH	Final (DOE, 2018)
	PAC-2	1 hr	430	83	Based on PAC-3	--	--	--	--	Based on PAC-3[b]	
	PAC-1	1 hr	79	15	Adopted NIOSH REL-STEL	--	--	--	--	Adopted NIOSH REL-STEL	

	Reference Value Name	Duration	Reference Value		Health Effect	Point of Departure	Qualifier	Source	Uncertainty Factors[a]	Notes on Derivation	Review Status
			(mg/m^3)	(ppm)							
Occupational	**NIOSH REL (TWA)**	10-hr TWA	50	10	NR	NR	NR		NR		Final (NIOSH, 1994)
	NIOSH REL-STEL	15 min	75	15	NR	NR	NR		NR		
	NIOSH IDLH	30 min	1,300	250	Acute oral toxicity	NR	NR	(Gerarde, 1960)	NR	Route-to-route extrapolation applied	
	ACGIH TLV-TWA [Skin][c]	8-hr TWA	52	10	Eye irritation at 15 ppm, acute hemolysis, and hepatoxicity in humans	NR	NR	(Robbins, 1951); (Hanssler, 1964); (Grigor et al., 1966), (Irle, 1964); (Naiman and Kosov, 1964); (Valaes et al., 1963); (Dawson et al., 1958); (Cock, 1957); (Schafer, 1951)	NR		Final (ACGIH, 2001)
	ACGIH TLV-STEL [Skin][d]	15 min	79	15							
	OSHA PEL (TWA)e	8-hr TWA	50	10	NR	NR	NR		NR		Final (OSHA, 2019)
	Cal-OSHA PEL (TWA)	8-hr TWA	0.5	0.1	NR	NR	NR		NR		

	Reference Value Name	Duration	Reference Value		Health Effect	Point of Departure	Qualifier	Source	Uncertainty Factors[a]	Notes on Derivation	Review Status
			(mg/m^3)	(ppm)							
General Public	**U.S. EPA Chronic RfC (IRIS)[f]**	Chronic	0.003	0.0006	Hyperplasia in the respiratory epithelium and metaplasia in the olfactory epithelium of adult male and female mice	10 ppm 9.3 mg/m^3 9.3 mg/m^3	LOAEL LOAELADJ LOAELHEC	(NTP, 1992)	Total UF = 3,000 UF_A = 10 UF_H = 10 UF_L = 10 UF_{DB} = 3	Duration adjusted: (6-hr/24-hr) × (5-d/7-d) HEC Adjusted[g]	Final (U.S. EPA, 1998b)
	ATSDR MRL	Chronic (>1 yr)	0.0036	0.0007	Nonneoplastic lesions in nasal olfactory epithelium and respiratory epithelium of adult male and female rats and mice	10 ppm 1.8 ppm 0.2 ppm	LOAEL LOAELADJ LOAELHEC	(Abdo et al., 2001); (NTP, 2000); (NTP, 1992)	Total UF = 300 UF_A = 3 UF_H = 10 UF_L = 10	Duration adjusted: (6-hr/24-hr) × (5-d/7-d) HEC Adjusted[h]	Final (ATSDR, 2005)
	OEHHA RELi	Chronic	0.009	0.002	Nasal inflammation, olfactory epithelial metaplasia, and respiratory epithelial hyperplasia in adult male and female mice	10 ppm 1.8 ppm	LOAEL LOAELADJ	(NTP, 1992)	Total UF = 1,000 UF_A = 10 UF_H = 10 UF_L = 10 UF_S = 1	Duration adjusted: (6-hr/24-hr) × (5-d/7-d)	Final (OEHHA, 2000)

	Reference Value Name	Duration	Reference Value		Health Effect	Point of Departure	Qualifier	Source	Uncertainty Factors[a]	Notes on Derivation	Review Status
			(mg/m^3)	(ppm)							
	MDH HBV	Acute (1 hr)	0.2	0.038	Respiratory cell swelling and sloughing in rats and nausea, vomiting, abdominal pain, and hemolytic anemia in humans	204 mg/m^3	NOAEL	(Buckpitt and Richieri, 1984)	Total UF = 1,000 UF_A = 10 UF_H = 10 UF_{DB} = 10		Final (MDH, 2004)
		Chronic (1 yr)	0.009	0.002	Nasal effects in adult rats and mice	10 ppm 9.3 mg/m^3	LOAEL LOAELADJ	(NTP, 2000); (NTP, 1992)	Total UF = 1,000 UF_A = 10 UF_H = 10 UF_L = 10	Duration adjusted: (6-hr/24-hr) × (5-d/7-d)	
	RIVM TCA	Chronic	0.025	0.0048	Local toxic effect on the nasal mucous membrane in adult rats exposed for 28 d	5 mg/m^3	LOAEL	(Coombs, 1993)	Total UF = 200 UF_A = 10 UF_H = 10 UF_L = 2	No time extrapolation Based on EU Risk Assessment: (ECB, 2003)	Final (Dusseldorp et al., 2011)
	Health Canada Residential Indoor RfC	Chronic	0.01	0.0019	Nasal epithelial cytotoxicity in adult rats	52 mg/m^3 9.3 mg/m^3	LOAEL LOAELADJ	(NTP, 2000)	Total UF = 1,000 UF_A = 10 UF_H = 10 UF_{DB} = 10	Duration adjusted: (6-hr/24-hr) × (5-d/7-d)	Final (Health Canada, 2013)

General Public (Other State Values)	**RI DEM AAL**	24 hr	0.003	0.0006	Adopted IRIS RfC as 24-hr. AAL	--	--	--	--	Adopted IRIS RfC as 24-hr. AAL	Final (RI DEM, 2008)
		1 yr	0.00003	0.0000056	Cancer	0.000034 $(\mu g/m^3)^{-1}$	OEHHA Cancer URF	(OEHHA, 2011)	NA	Calculated[j]	
	OR DEQ ABC	1 yr	0.00003	0.0000056	Cancer	0.000034 $(\mu g/m^3)^{-1}$	OEHHA Cancer URF	(OEHHA, 2011)	NA	Calculated[k]	Final (Oregon DEQ, 2018)
	CT DEEP HLV	30 min	5	1	NR	NR	NR		NR	NA	Final (CT DEEP, 2015)
		8 hr	1	0.2	NR	52 mg/m^3	ACGIH TLV-TWA	(ACGIH, 1992)	Total UF = 50	Details reported to NATICH	
	NDEP BCL	Chronic (Cancer)	0.0000826	0.000016	Cancer	0.000034 $(\mu g/m^3)^{-1}$	OEHHA Cancer URF	(OEHHA, 2011)	NA	Calculated[l]	Final (NDEP, 2017)

AAL = Acceptable Ambient Level; ABC = Ambient Benchmark Concentration; ACGIH = American Conference of Governmental Industrial Hygienists; ADJ = adjusted; ATSDR = Agency for Toxic Substances and Disease Registry; BCL = Basic Comparison Level; Cal-OSHA = California Division of Occupational Safety and Health; CT DEEP = Connecticut Department of Energy and Environmental Protection; DOE = Department of Energy; ECB = European Chemicals Bureau; EU = European Union; HBV = Health-Based Value; HEC = human equivalent concentration; HLV = Hazard Limiting Value; IDLH = Immediately Dangerous to Life and Health; IRIS = Integrated Risk Information System; LOAEL = lowest-observed-adverse-effect level; MDH = Minnesota Department of Health; MRL = Minimal Risk Level; NA = Not applicable; NATICH = National Air Toxics Information Clearinghouse; NDEP = Nevada Division of Environmental Protection; NIOSH = National Institute for Occupational Safety and Health; NOAEL = no-observed-adverse-effect level; NR = Not reported; NTP = National Toxicology Program; OEHHA = California Environmental Protection Agency Office of Environmental Health Hazard Assessment; OR DEQ = Oregon Department of Environmental Quality; OSHA = Occupational Safety and Health Administration; PAC = Protective Action Criteria; PEL = Permissible Exposure Limit; REL = Recommended Exposure Limit (NIOSH) or Reference Exposure Level (OEHHA); RfC = Reference Concentration; RI DEM = Rhode Island Department of Environmental Management; RIVM = *Rijksinstituut voor Volksgezondheid en Milieu,* The Netherlands Institute for Public Health and the Environment; STEL = Short-term Exposure Limit; TCA = Tolerable Concentration; TLV = Threshold Limit Value; TWA = Time-weighted average; UF = uncertainty factor; UF_H = inter-human variability; UF_A = animal to human variability; UF_L = LOAEL to NOAEL adjustment; UF_S = subchronic to chronic adjustment; UF_{DB} = database uncertainty; URF = unit risk factor; U.S. EPA = United States Environmental Protection Agency

[a] "Uncertainty factors" refer to modifying factors and other adjustment factors used by some organizations or in older EPA assessments.

[b] PAC-2 = PAC-3 / 6 = 500 ppm / 6 = 83 ppm

[c] Support documentation states: "systemic poisoning following dermal contact and absorption of naphthalene warrants a Skin notation." Agencies of Ontario, Quebec, Ireland, Australia, Austria, Belgium, Spain, and Singapore report identical values.

[d] Agencies of Quebec, Australia, Belgium, China, Singapore, South Korea, Spain, Sweden, and the Netherlands report identical values.
[e] Agencies of Denmark, France, Hungary, Italy, Latvia, China, Romania, South Korea, Sweden, Switzerland, the Netherlands, and Turkey report identical values.
[f] The EPA IRIS RfC has been adopted as a state value by the Texas Commission on Environmental Quality, Indiana Department of Environmental Management, Pennsylvania Department of Environmental Protection, Alaska Department of Environmental Conservation, New Jersey Department of Environmental Protection, and Michigan Department of Environment, Great Lakes & Energy.
[g] $LOAEL_{HEC} = LOAEL_{ADJ} \times RGDR = 9.3\ mg/m^3 \times 1 = 9.3\ mg/m^3$
[h] $LOAEL_{HEC} = LOAEL_{ADJ} \times RGDR = 1.8\ ppm \times 0.132 = 0.2\ ppm$
[i] The OEHHA REL value has been adopted by New York DEC
[j] $AAL = 1 / URF / 10^6 = 1 / 0.000034\ (\mu g/m^3)^{-1} / 10^6 = 0.03\ \mu g/m^3$
[k] $ABC = 1 / URF / 10^6 = 1 / 0.000034\ (\mu g/m^3)^{-1} / 10^6 = 0.03\ \mu g/m^3$
[l] $BCL = TR \times AT / (ET \times EF \times ED \times URF) = (10^{-6} \times 70\ yr \times 365\ d/yr \times 24\ hrs/d) / [24\ hrs/d \times 350\ d/yr \times 26\ yrs \times 0.000034\ (\mu g/m^3)^{-1}] = 0.0826\ \mu g/m^3$

Table A-3. Details on derivation of the available health effect reference values for oral exposure to naphthalene (from Figure 2-2 of the main text)

	Reference Value Name	Duration	Reference Value (mg/kg-d)	Health Effect	Point of Departure	Qualifier	Source	Uncertainty Factors[a]	Notes on Derivation	Review Status
General Public	**U.S. EPA RfD (IRIS)**[b]	Chronic	0.02	Decreased body wt. in adult in male rats exposed 13 wks.	100 mg/kg-d 71 mg/kg-d	NOAEL $NOAEL_{ADJ}$	(Battelle, 1980)	Total UF = 3,000 UF_A = 10 UF_H = 10 UF_S = 10 UF_{DB} = 3	Duration adjusted: 5-d/7-d	Final (U.S. EPA, 1998b)
	U.S. EPA RfD (OPP)[c]	Acute	0.4	Neurotoxicity in adult male and female rats, such as head shaking and reduced motor activity.	400 mg/kg-d	LOAEL	(Reynolds, 1997)	Total UF = 1,000 UF_A = 10 UF_H = 10 UF_L = 10		Final (U.S. EPA, 2018c)
		Chronic	0.1	Renal toxicity in adult male rats and decreased body weight in males and females exposed 13 wks.	100 mg/kg-d	NOAEL	(Battelle, 1980)	Total UF = 1,000 UF_A = 10 UF_H = 10 UF_S = 10		
	ATSDR MRL	Acute (1–14 d)	0.6	Transient clinical toxicity in pregnant rats exposed on GD 6–15.	50 mg/kg-d	LOAEL	(NTP, 1991)	Total UF = 90 UF_A = 10 UF_H = 3 UF_L = 3		Final (ATSDR, 2005)
		Intermediate (15–365 d)	0.6							
	RIVM TDI[d]	Chronic	0.04	Decreased body wt. and increased kidney and liver wt. in laboratory animals (further details not provided).	NR	NR	(Edwards et al., 1997); (Gustafson et al., 1997)	NR	Based on TPHCWG approach	Final (RIVM, 2001)

1

ADJ = adjusted; ATSDR = Agency for Toxic Substances and Disease Registry; GD = Gestation day; IRIS = Integrated Risk Information System; LOAEL = lowest-observed-adverse-effect level; MRL = Minimal Risk Level; NOAEL = no-observed-adverse-effect level; NR = Not reported; OPP = Office of Pesticide Programs; RfD = Reference Dose; RIVM = *Rijksinstituut voor Volksgezondheid en Milieu*; TDI = Tolerable Daily Intake; TPHCWG = Total Petroleum Hydrocarbon Criteria Working Group; UF = uncertainty factor; UF_H = inter-human variability; UF_A = animal to human variability; UF_L = LOAEL to NOAEL adjustment; UF_S = subchronic to chronic adjustment; UF_{DB} = database uncertainty; U.S. EPA = U.S. Environmental Protection Agency

[a] "Uncertainty factors" refer to modifying factors and other adjustment factors used by some organizations or in older EPA assessments.

[b] The U.S. EPA IRIS RfD has been adopted by the Office of Water, Health Canada, Alaska Department of Environmental Conservation, Pennsylvania Department of Environmental Protection, Connecticut Department of Energy & Environmental Protection, Nevada Division of Environmental Protection, New York State Department of Environmental Conservation, and Texas Commission on Environmental Quality.

[c] The U.S. EPA OPP chronic RfD has been adopted as a state value by Michigan Department of Environment, Great Lakes & Energy.

[d] The RIVM TDI value applies individually to non-carcinogenic polycyclic aromatic hydrocarbons "with equivalent carbon numbers of >9–16 (i.e., anthracene, fluorene and naphthalene)."

Table A-4. Details on additional inhalation values based on another agency's values or lacking derivation descriptions

	Reference Value Name	Duration	Reference Value		Health Effect	Point of Departure	Qualifier	Source	Uncertainty Factors[a]	Notes on Derivation	Review Status
			(mg/m^3)	(ppm)							
Special Use	**USAPHC MEG – Critical (MEG-C)**	1 hr	1,300	250	Adopted 2009 PAC-3	--	--	(DOE, 2009)	--	Adopted 2009 PAC-3	Final (U.S. APHC, 2013)
	USAPHC MEG – Marginal (MEG-M)	1 hr	75	15	Adopted 2009 PAC-2	--	--		--	Adopted 2009 PAC-2	
	USAPHC MEG – Negligible (MEG-N)	1 hr	75	15	Adopted 2009 PAC-1	--	--		--	Adopted 2009 PAC-1	
		8 hr	52	10	Adopted ACGIH TLV-TWA	--	--	--	--	Adopted ACGIH TLV-TWA	
		14 d	18	3.5	Based on ACGIH TLV-TWA	--	--	--	--	Based on ACGIH TLV-TWA[b]	
		1 yr	0.0021	0.0004	Based on IRIS RfC	--	--	--	--	Based on IRIS RfC[c]	
Occupational (International)	**Finland Limit Value**	15 min	10	2	NR	NR	NR		NR		Final (IFA, 2020)
		8-hr TWA	5	1							
	Denmark Limit Value	Short-term	100	20	NR	NR	NR		NR		
	Interdepartmental Commission MAC (Poland)	15 min	50	10	NR	NR	NR		NR		
		8-hr TWA	20	3.8							
	Worksafe WES (New Zealand) [Skin]	15 min	10	2	NR	NR	NR		NR		Final (Worksafe, 2022)
		8-hr TWA	2.6	0.5							

1

	Reference Value Name[a]	Duration	Reference Value		Health Effect[b]	Point of Departure[b]	Qualifier[b]	Source	Uncertainty Factors[b]	Notes on Derivation	Review Status
			(mg/m^3)	(ppm)							
General Public (Limited Details)	**ID DEQ AAC**	24 hr	2.5	0.48	NR	NR	NR		NR		Final (Idaho DEQ, 2019)
	VT DEC HAAS	1 yr	0.0003	0.000056	NR	NR	NR		NR		Final (VT ANR, 2018)
	Washington State Dept. of Ecology ASIL	1 yr	0.0000294	0.0000056	NR	NR	NR		NR		Final (Washington State Legislature, 2009)
	SWCAA ASIL	24 hr	0.17	0.033	NR	NR	NR		NR	Adopted 1998 Washington State ASIL	Final (SWCAA, 2019)
	MassDEP TEL[d]	24 hr	0.01425	0.00272	NR	NR	NR		NR	Values derived in accordance with this protocol: (MassDEP, 2011)	Final (MassDEP, 2019)
	MassDEP AAL[d]	1 yr	0.01425	0.00272	NR	NR	NR		NR		
	ADEQ AQG	1 hr	0.63	0.12	Based on ACGIH TLV-STEL	--	--	--	--	Based on ACGIH TLV-STEL[e]	Final (U.S. EPA, 1993)[g]
		24 hr	0.4	0.077	Based on ACGIH TLV-TWA	--	--	--	--	Based on ACGIH TLV-TWA[f]	
	Broward County ONRP AAC[h]	8 hr	0.5	0.096	NR	52 mg/m^3	ACGIH TLV-TWA	(ACGIH, 1992)	Total UF[i] = 100		

	Reference Value Name[a]	Duration	Reference Value		Health Effect[b]	Point of Departure[b]	Qualifier[b]	Source	Uncertainty Factors[b]	Notes on Derivation	Review Status
			(mg/m^3)	(ppm)							
	Pinellas County Air Pollution Control Board AAC	24 hr	0.12	0.023	NR	NR	NR		NR		
	ME DEP AAL	15 min	7.9	1.52	NR	NR	NR		NR		
		24 hr	0.87	0.17							
		1 yr	0.014	0.0027							
	ND Dept. of Health ACG	1 hr	0.79	0.15	NR	79 mg/m^3	ACGIH TLV-STEL	(ACGIH, 1992)	Total UF = 100		
		8 hr	0.52	0.1	NR	52 mg/m^3	ACGIH TLV-TWA				
	NDEP AAC	8 hr	1.19	0.23	NR	52 mg/m^3	ACGIH TLV-TWA	(ACGIH, 1992)	Total UF = 42		
	NY DEC AAL	1 yr	0.167	0.032	NR	52 mg/m^3	ACGIH TLV-TWA	(ACGIH, 1992)	Total UF = 300		
	OK Dept. of Health AAC	24 hr	50	10	NR	NR	NR		Total UF[j] = 50	Based on occupational values	
	SC DHEC AAL	24 hr	1.25	0.24	NR	52 mg/m^3	ACGIH TLV-TWA	(ACGIH, 1992)	Total UF = 40		
	TX Air Control Board AAC	30 min	0.44	0.085	NR	NR	NR		NR		
		1 yr	0.05	0.01							
	VA Air Pollution Control AAC	24 hr	0.87	0.17	NR	52 mg/m^3	ACGIH TLV-TWA	(ACGIH, 1992)	Total UF[k] = 60		
	WI DNR Bureau of Air Management AQG	24 hr	1.2	0.23	Based on ACGIH TLV-TWA	--	--		--	Based on ACGIH TLV-TWA[l]	

1

AAC = Acceptable Ambient Concentration; AAL = Allowable Ambient Limit; ACG = Ambient Concentration Guideline; ACGIH = American Conference of Governmental Industrial Hygienists; ADEQ = Arizona Department of Environmental Quality; AQG = Air Quality Guideline; ASIL = Acceptable Source Impact Level; HAAS = Hazardous Ambient Air Standard; ID DEQ = Idaho Department of Environmental Quality; IRIS = Integrated Risk Information System; MAC =

Maximum Admissible Concentration; MassDEP = Massachusetts Department of Environmental Protection; ME DEP = Maine Department of Environmental Protection; MEG = Military Exposure Guidelines; ND = North Dakota; NDEP = Nevada Division of Environmental Protection; NR = Not reported; NY DEC = New York Department of Environmental Conservation; OK = Oklahoma; ONRP = Office of Natural Resource Protection; PAC = Protective Action Criteria; RfC = Reference Concentration ; SC DHEC = South Carolina Department of Health and Environmental Control; STEL = Short-term Exposure Limit; SWCAA = Southwest Clean Air Agency; TEL = Threshold Effects Exposure Limit; TLV = Threshold Limit Value; TWA = Time-weighted average; TX = Texas; UF = uncertainty factor; USAPHC = United States Army Public Health Center; VA = Virginia; VT DEC = Vermont Department of Environmental Conservation; WES = workplace exposure standard; WI DNR = Wisconsin Department of Natural Resources

[a] "Uncertainty factors" refer to modifying factors and other adjustment factors used by some organizations or in older EPA assessments.
[b] MEG = TLV × ($IR_{Occupational}$ / $IR_{Military}$) = 52 × (10 m^3/d / 29.2 m^3/d) = 18 mg/m^3
[c] MEG = RfC × ($IR_{General\ pop.}$ / $IR_{Military}$) = 0.003 mg/m^3 × (20 m^3/d / 29.2 m^3/d) = 0.0021 mg/m^3
[d] MassDEP TEL and AAL values apply to the sum of naphthalene and 2-methylnaphthalene.
[e] 1-hr. AQG = TLV / 120 = 79 mg/m^3 / 120 = 0.63 mg/m^3
[f] 24-hr. AQG = TLV / 126 = 52 mg/m^3 / 126 = 0.4 mg/m^3
[g] This document was compiled by the U.S. Environmental Protection Agency in 1993. Values from this document may have since been archived or updated by the state agencies which reported them.
[h] The Hillsborough Co. Environmental Protection Commission and Pinellas County Air Control Board report the same value.
[i] A factor of 100 is applied "for category A substances."
[j] A factor of 50 is applied for category B substances.
[k] A factor of 60 is applied for non-carcinogens.
[l] 24-hr. AQG = TLV × 0.024 = 52 mg/m^3 × 0.024 = 1.2 mg/m^3

APPENDIX B. ELECTRONIC DATABASE SEARCH STRATEGIES

Table B-1. Core database search strategy

Database Search Date	Query String
PubMed	
1/11/2022	("naphthalene"[nm] AND 2021/01/01:2022/01/11[mhda]) OR (("naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR "mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND "Naphthalenes"[mh:noexp] AND 2021/01/01:2022/01/11[mhda]) OR ((("naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR "mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND (2021/01/01:2022/01/11[edat] OR 2021/01/01:2022/01/11[crdt])) NOT medline[sb])
1/28/2021	("naphthalene"[nm] AND 2018/12/01 : 2021/01/31[mhda]) OR (("naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR "mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND "Naphthalenes"[mh:noexp] AND 2018/12/01 : 2021/01/31 [mhda]) OR ((("naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR "mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND (2018/12/01 : 2021/01/31[edat] OR 2018/12/01 2021/01/31[crdt])) NOT medline[sb])
2/8/2019	("naphthalene"[nm] AND 2017/10/01 : 2019/01/01[mhda]) OR (("naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR "mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND "Naphthalenes"[mh:noexp] AND 2017/10/01 : 2019/01/01[mhda]) OR ((("naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR "mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND (2017/10/01 : 2019/01/01[edat] OR 2017/10/01 : 2019/01/01[crdt])) NOT medline[sb])
9/29/2017	("naphthalene"[nm] AND 2017/02/01 : 3000[mhda]) OR (("naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR "mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND "Naphthalenes"[mh:noexp] AND 2017/02/01 : 3000[mhda]) OR ((("naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR

Database Search Date	Query String
	"mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND (2014/10/01 : 3000[edat] OR 2017/02/01 : 3000[crdt])) NOT medline[sb])
01/04/2017	((524-42-5[rn] OR 130-15-4[rn] OR 7234-04-0[rn] OR 277-50-9[rn]) OR (("1,2-Dihydro-1,2-diketo-naphthalene"[tw] OR "1,2-Naphthalenedione"[tw] OR "1,2-Naphthaquinone"[tw] OR "beta-Naphthoquinone"[tw] OR "o-Naphthoquinone"[tw] OR "1,4-Dihydro-1,4-diketonaphthalene"[tw] OR "1,4-Naphthalenedione"[tw] OR "1,4-Naphthoquinone"[tw] OR "1,4-Naphthylquinone"[tw] OR "alpha-Naphthoquinone"[tw] OR "p-Naphthoquinone"[tw] OR "1,2-Dihydronaphthalene-1,2-diol"[tw] OR "1,2-Dihydroxy-1,2-dihydronaphthalene"[tw] OR "1,2-dihydro-1,2-Naphthalenediol"[tw] OR "Naphthalene-1,2-dihydrodiol"[tw] OR "trans-1,2-Dihydroxy-1,2-dihydronaphthalene"[tw] OR "Naphthalene 1,2-oxide"[tw] OR "Naphthalene oxide"[tw] OR "Naphth(1,2-b)oxirene"[tw]) NOT medline[sb])) OR (("naphthalene"[nm] AND 2015/10/01 : 3000[mhda]) OR (("naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR "mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND "Naphthalenes"[mh:noexp] AND 2015/10/01 : 3000[mhda]) OR ((("naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR "mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND (2015/10/01 : 3000[edat] OR 2015/10/01 : 3000[crdt])) NOT medline[sb]))
11/06/2015	("naphthalene"[nm] AND 2014/10/01 : 3000[mhda]) OR (("naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR "mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND "Naphthalenes"[mh:noexp] AND 2014/10/01 : 3000[mhda]) OR ((("naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR "mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND (2014/10/01 : 3000[edat] OR 2014/10/01 : 3000[crdt])) NOT medline[sb])
12/16/2014	("naphthalene"[nm] AND 2012/12/01 : 3000[mhda]) OR ("Naphthalenes"[mh:noexp] AND ("91-20-3"[tw] OR "naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR "mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND 2012/12/01 : 3000[mhda]) OR ((("91-20-3"[tw] OR "naphthalene"[tw] OR "albocarbon"[tw] OR "naphthalin"[tw] OR "naphthaline"[tw] OR "naphthene"[tw] OR "naphtalene"[tw] OR "camphor tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR "mothballs"[tw] OR "Naphtalinum"[tw] OR "Naphthalinum"[tw] OR "Dezodorator"[tw] OR "Mighty 150"[tw] OR "Mighty RD1"[tw]) AND (2012/12/01 : 3000[crdat] OR 2012/12/01 : 3000[edat])) NOT medline[sb])
02/17/2013	(((91-20-3[rn]) OR (("91-20-3"[tw] OR naphthalene[tw] OR albocarbon[tw] OR naphthalin[tw] OR naphthaline[tw] OR naphthene[tw] OR naphtalene[tw] OR "camph[tw] OR tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR mothballs[tw]) AND ("naphthalenes"[mh:noexp]))) AND (("naphthalenes/toxicity"[MeSH Terms] OR

Database Search Date	**Query String**
	"naphthalenes/adverse effects"[MeSH Terms] OR "naphthalenes/poisoning"[MeSH Terms] OR "naphthalenes/pharmacokinetics"[MeSH Terms]) OR ("naphthalenes/blood"[MeSH Terms] OR "naphthalenes/cerebrospinal fluid"[MeSH Terms] OR "naphthalenes/urine"[MeSH Terms]) OR ("naphthalenes/metabolism"[MeSH Terms] AND ("humans"[MeSH Terms] OR "animals"[MeSH Terms])) OR ("naphthalenes/antagonists and inhibitors"[MeSH Terms]) OR ("chemically induced"[MeSH Subheading] OR "environmental exposure"[MeSH Terms]) OR ("endocrine system"[mh] OR "hormones, hormone substitutes, and hormone antagonists"[mh] OR "endocrine disruptors"[mh]) OR (cancer[sb]) OR ("Computational biology"[mh] OR "Medical Informatics"[mh] OR Genomics[mh] OR Genome[mh] OR Proteomics[mh] OR Proteome[mh] OR Metabolomics[mh] OR Metabolome[mh] OR Genes[mh] OR "Gene expression"[mh] OR Phenotype[mh] OR genetics[mh] OR genotype[mh] OR Transcriptome[mh] OR ("Systems Biology"[mh] AND ("Environmental Exposure"[mh] OR "Epidemiological Monitoring"[mh] OR analysis[sh])) OR "Transcription, Genetic "[mh] OR "Reverse transcription"[mh] OR "Transcriptional activation"[mh] OR "Transcription factors"[mh] OR ("biosynthesis"[sh] AND (RNA[mh] OR DNA[mh])) OR "RNA, Messenger "[mh] OR "RNA, Transfer"[mh] OR "peptide biosynthesis"[mh] OR "protein biosynthesis"[mh] OR "Reverse Transcriptase Polymerase Chain Reaction"[mh] OR "Base Sequence"[mh] OR "Trans-activators"[mh] OR "Gene Expression Profiling"[mh]) OR (rat[tw] OR rats[tw] OR mouse[tw] OR mice[tw] OR muridae[tw] OR rabbit[tw] OR rabbits[tw] OR hamster[tw] OR hamsters[tw] OR ferret[tw] OR ferrets[tw] OR gerbil[tw] OR gerbils[tw] OR rodent[tw] OR rodents[tw] OR rodentia[tw] OR dog[tw] OR dogs[tw] OR beagle[tw] OR beagles[tw] OR canine[tw] OR cats[tw] OR feline[tw] OR pig[tw] OR pigs[tw] OR swine[tw] OR porcine[tw] OR monkey[tw] OR monkeys[tw] OR macaque[tw] OR macaques[tw] OR baboon[tw] OR baboons[tw] OR marmoset[tw] OR marmosets[tw] OR "animals, laboratory"[mh]) OR (((pharmacokinetics[mh] OR metabolism[mh]) AND (humans[mh] OR animals[mh])) OR "dose-response relationship, drug"[mh] OR risk[mh]))) OR (("91-20-3"[tw] OR naphthalene[tw] OR albocarbon[tw] OR naphthalin[tw] OR naphthaline[tw] OR naphthene[tw] OR naphtalene[tw] OR "camph[tw] OR tar"[tw] OR "tar camphor"[tw] OR "white tar"[tw] OR "moth balls"[tw] OR "moth flakes"[tw] OR mothballs[tw]) NOT medline[sb])
Web of Science	
1/11/2022	(TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS="mothballs" OR TS="Naphtalinum" OR TS="Naphthalinum" OR TS="Dezodorator" OR TS="Mighty 150" OR TS="Mighty RD1") AND ((WC=("Toxicology" OR "Endocrinology & Metabolism" OR "Gastroenterology & Hepatology" OR "Gastroenterology & Hepatology" OR "Hematology" OR "Neurosciences" OR "Obstetrics & Gynecology" OR "Pharmacology & Pharmacy" OR "Physiology" OR "Respiratory System" OR "Urology & Nephrology" OR "Anatomy & Morphology" OR "Andrology" OR "Pathology" OR "Veterinary Sciences" OR "Otorhinolaryngology" OR "Ophthalmology" OR "Pediatrics" OR "Oncology" OR "Reproductive Biology" OR "Developmental Biology" OR "Biology" OR "Dermatology" OR "Allergy" OR "Public, Environmental & Occupational Health") OR SU=("Anatomy & Morphology" OR "Cardiovascular System & Cardiology" OR "Developmental Biology" OR "Endocrinology & Metabolism" OR "Gastroenterology & Hepatology" OR "Hematology" OR "Immunology" OR "Neurosciences & Neurology" OR "Obstetrics & Gynecology" OR "Oncology" OR "Ophthalmology" OR "Pathology" OR "Pediatrics" OR "Pharmacology & Pharmacy" OR "Physiology" OR "Public, Environmental & Occupational Health" OR "Respiratory System" OR "Toxicology" OR "Urology & Nephrology" OR "Reproductive Biology" OR "Dermatology" OR "Allergy")) AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR

Database Search Date	**Query String**
	TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset* OR TS=toxic*) AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset*) OR (TS="child" OR TS="children" OR TS=adolescen* OR TS=infant* OR TS="WORKER" OR TS="WORKERS" OR TS="HUMAN" OR TS=patient* OR TS="mother" OR TS="fetal" OR TS="fetus" OR TS="citizens" OR TS="milk" OR TS="formula")) AND PY=(2021-2022)
1/28/2021	(TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS="mothballs" OR TS="Naphtalinum" OR TS="Naphthalinum" OR TS="Dezodorator" OR TS="Mighty 150" OR TS="Mighty RD1") AND ((WC=("Toxicology" OR "Endocrinology & Metabolism" OR "Gastroenterology & Hepatology" OR "Gastroenterology & Hepatology" OR "Hematology" OR "Neurosciences" OR "Obstetrics & Gynecology" OR "Pharmacology & Pharmacy" OR "Physiology" OR "Respiratory System" OR "Urology & Nephrology" OR "Anatomy & Morphology" OR "Andrology" OR "Pathology" OR "Veterinary Sciences" OR "Otorhinolaryngology" OR "Ophthalmology" OR "Pediatrics" OR "Oncology" OR "Reproductive Biology" OR "Developmental Biology" OR "Biology" OR "Dermatology" OR "Allergy" OR "Public, Environmental & Occupational Health") OR SU=("Anatomy & Morphology" OR "Cardiovascular System & Cardiology" OR "Developmental Biology" OR "Endocrinology & Metabolism" OR "Gastroenterology & Hepatology" OR "Hematology" OR "Immunology" OR "Neurosciences & Neurology" OR "Obstetrics & Gynecology" OR "Oncology" OR "Ophthalmology" OR "Pathology" OR "Pediatrics" OR "Pharmacology & Pharmacy" OR "Physiology" OR "Public, Environmental & Occupational Health" OR "Respiratory System" OR "Toxicology" OR "Urology & Nephrology" OR "Reproductive Biology" OR "Dermatology" OR "Allergy")) AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset* OR TS=toxic*) AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset*) OR (TS="child" OR TS="children" OR TS=adolescen* OR TS=infant* OR TS="WORKER" OR TS="WORKERS" OR TS="HUMAN" OR TS=patient* OR TS="mother" OR TS="fetal" OR TS="fetus" OR TS="citizens" OR TS="milk" OR TS="formula")) AND PY=(2019-2021)
2/8/2019	(TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS="mothballs" OR TS="Naphtalinum" OR TS="Naphthalinum" OR TS="Dezodorator" OR TS="Mighty 150" OR TS="Mighty RD1") AND ((WC="Toxicology" OR WC="Endocrinology & Metabolism" OR WC="Gastroenterology & Hepatology" OR WC="Gastroenterology & Hepatology" OR WC="Hematology" OR WC="Neurosciences" OR WC="Obstetrics & Gynecology" OR WC="Pharmacology & Pharmacy" OR WC="Physiology" OR WC="Respiratory System" OR WC="Urology & Nephrology" OR WC="Anatomy & Morphology" OR WC="Andrology" OR WC="Pathology" OR

Database Search Date	**Query String**
	WC="Otorhinolaryngology" OR WC="Ophthalmology" OR WC="Pediatrics" OR WC="Oncology" OR WC="Reproductive Biology" OR WC="Developmental Biology" OR WC="Biology" OR WC="Dermatology" OR WC="Allergy" OR WC="Public, Environmental & Occupational Health" OR SU="Anatomy & Morphology" OR SU="Cardiovascular System & Cardiology" OR SU="Developmental Biology" OR SU="Endocrinology & Metabolism" OR SU="Gastroenterology & Hepatology" OR SU="Hematology" OR SU="Immunology" OR SU="Neurosciences & Neurology" OR SU="Obstetrics & Gynecology" OR SU="Oncology" OR SU="Ophthalmology" OR SU="Pathology" OR SU="Pediatrics" OR SU="Pharmacology & Pharmacy" OR SU="Physiology" OR SU="Public, Environmental & Occupational Health" OR SU="Respiratory System" OR SU="Toxicology" OR SU="Urology & Nephrology" OR SU="Reproductive Biology" OR SU="Dermatology" OR SU="Allergy") OR (WC="veterinary sciences" AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset*)) OR (TS=toxic* AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset*) OR (TS="child" OR TS="children" OR TS=adolescen* OR TS=infant* OR TS="WORKER" OR TS="WORKERS" OR TS="HUMAN" OR TS=patient* OR TS=mother OR TS=fetal OR TS=fetus OR TS=citizens OR TS=milk OR TS=formula)) OR TI=toxic*) AND PY=(2017-2019)
9/29/2017	(TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS="mothballs" OR TS="Naphtalinum" OR TS="Naphthalinum" OR TS="Dezodorator" OR TS="Mighty 150" OR TS="Mighty RD1") AND ((WC=("Toxicology" OR "Endocrinology & Metabolism" OR "Gastroenterology & Hepatology" OR "Gastroenterology & Hepatology" OR "Hematology" OR "Neurosciences" OR "Obstetrics & Gynecology" OR "Pharmacology & Pharmacy" OR "Physiology" OR "Respiratory System" OR "Urology & Nephrology" OR "Anatomy & Morphology" OR "Andrology" OR "Pathology" OR "Otorhinolaryngology" OR "Ophthalmology" OR "Pediatrics" OR "Oncology" OR "Reproductive Biology" OR "Developmental Biology" OR "Biology" OR "Dermatology" OR "Allergy" OR "Public, Environmental & Occupational Health") OR SU=("Anatomy & Morphology" OR "Cardiovascular System & Cardiology" OR "Developmental Biology" OR "Endocrinology & Metabolism" OR "Gastroenterology & Hepatology" OR "Hematology" OR "Immunology" OR "Neurosciences & Neurology" OR "Obstetrics & Gynecology" OR "Oncology" OR "Ophthalmology" OR "Pathology" OR "Pediatrics" OR "Pharmacology & Pharmacy" OR "Physiology" OR "Public, Environmental & Occupational Health" OR "Respiratory System" OR "Toxicology" OR "Urology & Nephrology" OR "Reproductive Biology" OR "Dermatology" OR "Allergy")) OR (WC="veterinary sciences" AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset*)) OR (TS=toxic* AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs"

Database Search Date	Query String
	OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset*) OR (TS="child" OR TS="children" OR TS=adolescen* OR TS=infant* OR TS="WORKER" OR TS="WORKERS" OR TS="HUMAN" OR TS=patient* OR TS=mother OR TS=fetal OR TS=fetus OR TS=citizens OR TS=milk OR TS=formula)) OR TI=toxic*) AND PY=(2017-2017)
01/04/2017	(TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS="mothballs" OR TS="Naphtalinum" OR TS="Naphthalinum" OR TS="Dezodorator" OR TS="Mighty 150" OR TS="Mighty RD1") AND ((WC=("Toxicology" OR "Endocrinology & Metabolism" OR "Gastroenterology & Hepatology" OR "Gastroenterology & Hepatology" OR "Hematology" OR "Neurosciences" OR "Obstetrics & Gynecology" OR "Pharmacology & Pharmacy" OR "Physiology" OR "Respiratory System" OR "Urology & Nephrology" OR "Anatomy & Morphology" OR "Andrology" OR "Pathology" OR "Otorhinolaryngology" OR "Ophthalmology" OR "Pediatrics" OR "Oncology" OR "Reproductive Biology" OR "Developmental Biology" OR "Biology" OR "Dermatology" OR "Allergy" OR "Public, Environmental & Occupational Health") OR SU=("Anatomy & Morphology" OR "Cardiovascular System & Cardiology" OR "Developmental Biology" OR "Endocrinology & Metabolism" OR "Gastroenterology & Hepatology" OR "Hematology" OR "Immunology" OR "Neurosciences & Neurology" OR "Obstetrics & Gynecology" OR "Oncology" OR "Ophthalmology" OR "Pathology" OR "Pediatrics" OR "Pharmacology & Pharmacy" OR "Physiology" OR "Public, Environmental & Occupational Health" OR "Respiratory System" OR "Toxicology" OR "Urology & Nephrology" OR "Reproductive Biology" OR "Dermatology" OR "Allergy")) OR (WC="veterinary sciences" AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset*)) OR (TS=toxic* AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset*) OR (TS="child" OR TS="children" OR TS=adolescen* OR TS=infant* OR TS="WORKER" OR TS="WORKERS" OR TS="HUMAN" OR TS=patient* OR TS=mother OR TS=fetal OR TS=fetus OR TS=citizens OR TS=milk OR TS=formula)) OR TI=toxic*) AND PY=(2015-2017)
11/04/2015	(TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS="mothballs" OR TS="Naphtalinum" OR TS="Naphthalinum" OR TS="Dezodorator" OR TS="Mighty 150" OR TS="Mighty RD1") AND ((WC=("Toxicology" OR "Endocrinology & Metabolism" OR "Gastroenterology & Hepatology" OR "Gastroenterology & Hepatology" OR "Hematology" OR "Neurosciences" OR "Obstetrics & Gynecology" OR "Pharmacology & Pharmacy" OR "Physiology" OR "Respiratory System" OR "Urology & Nephrology" OR "Anatomy & Morphology" OR "Andrology" OR "Pathology" OR "Otorhinolaryngology" OR "Ophthalmology" OR "Pediatrics" OR "Oncology" OR "Reproductive Biology" OR "Developmental Biology" OR "Biology" OR "Dermatology" OR "Allergy" OR "Public, Environmental & Occupational Health") OR SU=("Anatomy & Morphology" OR "Cardiovascular System & Cardiology" OR "Developmental Biology" OR "Endocrinology & Metabolism" OR "Gastroenterology & Hepatology" OR "Hematology" OR "Immunology" OR "Neurosciences & Neurology" OR "Obstetrics & Gynecology" OR "Oncology" OR "Ophthalmology" OR "Pathology" OR "Pediatrics" OR "Pharmacology & Pharmacy" OR "Physiology" OR "Public, Environmental &

Database Search Date	**Query String**
	Occupational Health" OR "Respiratory System" OR "Toxicology" OR "Urology & Nephrology" OR "Reproductive Biology" OR "Dermatology" OR "Allergy")) OR (WC="veterinary sciences" AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset*)) OR (TS=toxic* AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset*) OR (TS="child" OR TS="children" OR TS=adolescen* OR TS=infant* OR TS="WORKER" OR TS="WORKERS" OR TS="HUMAN" OR TS=patient* OR TS=mother OR TS=fetal OR TS=fetus OR TS=citizens OR TS=milk OR TS=formula)) OR TI=toxic*) AND PY=(2014-2016)
12/16/2014	((TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS="mothballs" OR TS="Naphtalinum" OR TS="Naphthalinum" OR TS="Dezodorator" OR TS="Mighty 150" OR TS="Mighty RD1") AND ((WC=("Toxicology" OR "Endocrinology & Metabolism" OR "Gastroenterology & Hepatology" OR "Gastroenterology & Hepatology" OR "Hematology" OR "Neurosciences" OR "Obstetrics & Gynecology" OR "Pharmacology & Pharmacy" OR "Physiology" OR "Respiratory System" OR "Urology & Nephrology" OR "Anatomy & Morphology" OR "Andrology" OR "Pathology" OR "Otorhinolaryngology" OR "Ophthalmology" OR "Pediatrics" OR "Oncology" OR "Reproductive Biology" OR "Developmental Biology" OR "Biology" OR "Dermatology" OR "Allergy" OR "Public, Environmental & Occupational Health") OR SU=("Anatomy & Morphology" OR "Cardiovascular System & Cardiology" OR "Developmental Biology" OR "Endocrinology & Metabolism" OR "Gastroenterology & Hepatology" OR "Hematology" OR "Immunology" OR "Neurosciences & Neurology" OR "Obstetrics & Gynecology" OR "Oncology" OR "Ophthalmology" OR "Pathology" OR "Pediatrics" OR "Pharmacology & Pharmacy" OR "Physiology" OR "Public, Environmental & Occupational Health" OR "Respiratory System" OR "Toxicology" OR "Urology & Nephrology" OR "Reproductive Biology" OR "Dermatology" OR "Allergy")) OR (WC="veterinary sciences" AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset*)) OR (TS=toxic* AND (TS="rat" OR TS="rats" OR TS="mouse" OR TS="murine" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset*) OR (TS="child" OR TS="children" OR TS=adolescen* OR TS=infant* OR TS="WORKER" OR TS="WORKERS" OR TS="HUMAN" OR TS=patient* OR TS=mother OR TS=fetal OR TS=fetus OR TS=citizens OR TS=milk OR TS=formula)) OR TI=toxic*)) AND PY=2012-2015
02/21/2013	((TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS=mothballs) NOT TS="naphthalene acetic acid") AND (TS="chronic" OR TS=immun* OR TS=lymph* OR TS=neurotox* OR TS=toxicokin* OR

Database Search Date	Query String
	TS=pharmacokin* OR TS=biomarker* OR TS=neurolog* OR TS="subchronic" OR TS="pbpk" OR TS=epidemiolog* OR TS="acute" OR TS="subacute" OR TS="ld50")
	((TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS=mothballs) NOT TS="naphthalene acetic acid") AND (TS="lc50" OR TS=inhal* OR TS=pulmon* OR TS="nasal" OR TS=lung* OR TS=respir* OR TS=occupation* OR TS="workplace" OR TS=worker* OR TS="oral" OR TS="orally" OR TS=ingest* OR TS="gavage" OR TS="diet" OR TS="diets" OR TS="dietary" OR TS="drinking" OR TS=gastr* OR TS=intestin* OR TS=liver* OR TS=hepat* OR TS=kidney* OR TS=nephr*)
	((TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS=mothballs) NOT TS="naphthalene acetic acid") AND (TS="gut" OR TS=sensitiz* OR TS=abort* OR TS=abnormalit* OR TS=embryo* OR TS=cleft* OR TS=fetus* OR TS=foetus* OR TS=fetal* OR TS=foetal* OR TS=fertilit* OR TS=infertil* OR TS="fertilization" OR TS="fertilisation" OR TS=malform* OR TS="ovum" OR TS="ova" OR TS="ovary" OR TS="ovaries" OR TS="ovarian" OR TS=placenta* OR TS=pregnan*)
	((TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS=mothballs) NOT TS="naphthalene acetic acid") AND (TS=dermal* OR TS="dermis" OR TS="skin" OR TS=epiderm* OR TS="cutaneous" OR TS=carcinog* OR TS=cocarcinog* OR TS="cancer" OR TS="precancer" OR TS=neoplas* OR TS=tumor* OR TS=tumour* OR TS=oncogen* OR TS=lymphoma* OR TS=carcinom* OR TS=genetox* OR TS=genotox* OR TS=mutagen* OR TS=nephrotox* OR TS=hepatotox* OR TS=endocrin* OR TS=estrogen* OR TS=androgen*)
	((TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS=mothballs) NOT TS="naphthalene acetic acid") AND (TS=hormon* OR TS="blood" OR TS="serum" OR TS="urine" OR TS="bone" OR TS="bones" OR TS=skelet* OR TS="rat" OR TS="rats" OR TS="mouse")
	((TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS=mothballs) NOT TS="naphthalene acetic acid") AND (TS="mice" OR TS="guinea" OR TS="muridae" OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS="dog" OR TS="dogs" OR TS=beagle* OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset* OR TS=toxic* OR TS="adverse" OR TS="poisoning")
	((TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS=mothballs) NOT TS="naphthalene acetic acid") AND (TS="prenatal" OR TS="perinatal" OR TS="postnatal" OR TS="reproduce" OR TS=reproduct* OR TS=steril* OR TS=teratogen* OR TS=sperm* OR TS=neonat* OR TS=newborn* OR TS=development* OR TS=zygote* OR TS="child" OR TS="children" OR TS=adolescen* OR TS=infant* OR TS=wean* OR TS="offspring" OR TS="age factor" OR TS="age factors")
	((TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS=mothballs) NOT TS="naphthalene acetic

Database Search Date	Query String
	acid") AND (TS="Genomics" OR TS="Proteomics" OR TS="Metabolic Profile" OR TS="Metabolome" OR TS="Metabolomics" OR TS="Microarray" OR TS="Nanoarray")
	((TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS=mothballs) NOT TS="naphthalene acetic acid") AND (TS="Gene expression" OR TS="Transcript expression" OR TS="transcriptomes" OR TS="transcriptome" OR TS="Phenotype" OR TS="Transcription" OR TS="Trans-act*" OR TS="transact*" OR TS="trans act*" OR TS=genetic OR TS="genetics" OR TS="genotype")
	((TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS=mothballs) NOT TS="naphthalene acetic acid") AND (TS="Informatics" OR (TS="Information Science" AND TS=Medical OR TS="Systems biology" OR (TS="Biological systems" AND (TS=monit* OR TS=data OR TS=analysis))))
	((TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS=mothballs) NOT TS="naphthalene acetic acid") AND (TS="Genetic transcription" OR TS="Gene transcription" OR TS="Gene Activation" OR TS="Genetic induction" OR TS="Reverse transcription" OR TS="Transcriptional activation" OR TS="Transcription factors" OR (TS="Biosynthesis" AND (TS=RNA OR TS=DNA)) OR TS="mRNA")
	((TS="naphthalene" OR TS="albocarbon" OR TS="naphthalin" OR TS="naphthaline" OR TS="naphthene" OR TS="naphtalene" OR TS="camphor tar" OR TS="tar camphor" OR TS="white tar" OR TS="moth balls" OR TS="moth flakes" OR TS=mothballs) NOT TS="naphthalene acetic acid") AND (TS="messenger RNA" OR TS="transfer RNA" OR TS="peptide biosynthesis" OR TS="protein biosynthesis" OR TS="protein synthesis" OR TS="RT-PCR" OR TS="RTPCR" OR TS="Reverse Transcriptase Polymerase Chain Reaction" OR TS="DNA sequence")
ToxLine	
2/8/2019	@syn0+@AND+@OR+(naphthalene+albocarbon+naphthalin+naphthaline+ naphthene+naphtalene+""camphor+tar"+"tar+camphor"+"white+tar"+"moth+balls" +"moth+flakes"+mothballs+Naphtalinum+Naphthalinum+Dezodorator+ "Mighty+150"+"Mighty+RD1"+@term+@rn+91+20+3) +@and+@range+yr+2017+2019+@not+@org+pubmed
9/29/2017	@syn0+@AND+@OR+(naphthalene+albocarbon+naphthalin+naphthaline+naphthene+naphtalene +"camphor+tar"+"tar+camphor"+"white+tar"+"moth+balls"+"moth+flakes"+mothballs+Naphtalinum+Naphthalinum+Dezodorator+"Mighty+150"+"Mighty+RD1"+@term+@rn+91+20+3)+@and+@range+yr+2017+@not+@org+pubmed
01/04/2017	@syn0+@OR+(piscesqcorrection+naphthalene+albocarbon+naphthalin+naphthaline+naphthene+ naphtalene+"camphor tar"+"tar camphor"+"white tar"+"moth balls"+"moth flakes"+mothballs+Naphtalinum+Naphthalinum+Dezodorator+"Mighty 150"+"Mighty RD1"+@term+@rn+91-20-3)+@and+@range+yr+2015+2017+@not+@org+pubmed+pubdart+"nih+reporter"+tscats
11/09/2015	@syn0+@OR+(piscesqcorrection+naphthalene+albocarbon+naphthalin+naphthaline+naphthene+ naphtalene+"camphor tar"+"tar camphor"+"white tar"+"moth balls"+"moth flakes"+mothballs+Naphtalinum+Naphthalinum+Dezodorator+"Mighty 150"+"Mighty RD1"+@term+@rn+91-20-3)+@and+@range+yr+2014+2016+@not+@org+pubmed+pubdart+"nih+reporter"+tscats

Database Search Date	Query String
12/16/2014	@OR+(naphthalene+albocarbon+naphthalin+naphthaline+naphthene+naphtalene+mothballs+@term+@rn+91-20-3)+@AND+@range+yr+2012+2015+@NOT+@org+pubmed+pubdart+"nih+reporter"+tscats
	@OR+("camphor+tar"+"tar+camphor"+"white+tar"+"moth+balls"+"moth+flakes")+@AND+@range+yr+2012+2015+@NOT+@org+pubmed+pubdart+"nih+reporter"+tscats
02/18/2013	@OR+(naphthalene+albocarbon+naphthalin+naphthaline+naphthene+naphtalene+mothballs+@term+@rn+91-20-3)+@NOT+@org+pubmed+pubdart+crisp+tscats
	@OR+("camphor+tar"+"tar+camphor"+"white+tar"+"moth+balls"+"moth+flakes")+@NOT+@org+pubmed+pubdart+crisp+tscats

Table B-2. Targeted database search for PBPK models for naphthalene

Database Search Date	Query String
PubMed	
8/17/2022	(pbpk[tiab] OR "pb-pk"[tiab] OR pbtk[tiab] OR "pb-tk"[tiab] OR pbk[tiab] OR httk[tiab] OR pk-model*[tiab] OR tk-model*[tiab] OR (("physiologically based"[tiab] OR "biologically based"[tiab]) AND (pharmacokinetic*[tiab] OR toxicokinetic*[tiab] OR kinetic[tiab] OR model*[tiab] OR pharmacokinetics[mh] OR toxicokinetics[mh:noexp] OR pharmacokinetics[sh]))) AND naphthalene

Table B-3. Toxic Substances Control Act Test Submissions (TSCATS) search strategy

Database Search Date	Query String
TSCATS via CDAT[a]	
01/04/2017	91-20-3 Mail Received Date Range 10/01/2015 to 01/04/2017
11/04/2015	91-20-3 Mail Received Date Range 01/01/2014 to 11/04/2015
TSCATS 2[b]	
01/04/2017	91-20-3 EPA receipt date 10/01/2015 to date of search
12/16/2014	91-20-3 EPA receipt date 02/01/2013 to date of search
05/01/2013	91-20-3 date limited, 2000 to date of search
TSCATS 1[c]	
02/18/2013	@term+@rn+91-20-3+@AND+@org+tscats
TSCA section 8e/FYI recent submissions[d]	
01/04/2017	Google: 91-20-3 (8e or fyi) tsca

Database Search Date	Query String
12/16/2014	Google: 91-20-3 (8e or fyi) tsca
05/01/2013	Google: 91-20-3 (8e or fyi) tsca

[a] CDAT (Chemical Data Access Tool); formerly available at http://java.epa.gov/oppt_chemical_search/. Information from CDAT has since been incorporated into EPA's ChemView database at https://chemview.epa.gov/chemview.

[b] TSCATS 2 was searched via the following database URL: https://catalog.data.gov/dataset/toxic-substances-control-act-test-submissions-2-0-tscats-2-01

[c] TSCATS 1 was searched via Toxline

[d] TSCA section 8e/FYI recent submissions were searched via Google

Table B-4. Processes used to augment the search of core databases for naphthalene

System Used	Selected Reference(s) or Sources	Date	Additional References Identified
Toxic Substances Control Act Test Submissions (TSCATS)	CDAT (Chemical Data Access Tool) 91-20-3 Mail Received Date Range 10/01/2015 to 01/04/2017 91-20-3 Mail Received Date Range 01/01/2014 to 11/04/2015	01/2017	
Manual search of citations from published reviews	Bailey et al. (2015). "Hypothesis-based weight-of-evidence evaluation and risk assessment for naphthalene carcinogenesis." Critical Reviews in Toxicology: 1-42	12/2015	12 citations added
	Lewis (2012). "Naphthalene animal carcinogenicity and human relevancy: overview of industries with naphthalene-containing streams." Regulatory Toxicology and Pharmacology 62(1): 131-137	12/2015	1 citations added
	Piccirillo et al. (2012). "Preliminary evaluation of the human relevance of respiratory tumors observed in rodents exposed to naphthalene." Regulatory Toxicology and Pharmacology 62(3): 433-440.	12/2015	0 citations added
	Magee et al. (2010). "Screening-level population risk assessment of nasal tumors in the US due to naphthalene exposure." Regulatory Toxicology and Pharmacology 57(2-3): 168-180.	12/2015	0 citations added
	Rhomberg et al. (2010). "Hypothesis-based weight of evidence: a tool for evaluating and communicating uncertainties and inconsistencies in the large body of evidence in proposing a carcinogenic mode of action--naphthalene as an example." Critical Reviews in Toxicology 40(8): 671-696.	12/2015	0 citations added
Manual search of citations from national and international	NTP (2021). Naphthalene. In Report on Carcinogens, 15th Edition. National Toxicology Program.	8/2022	2 citations added
	NTP (2016). Naphthalene (14th ed.). Research Triangle Park, NC: National Toxicology Program. https://ntp.niehs.nih.gov/ntp/roc/content/profiles/naphthalene.pdf	1/2017	0 citations added

System Used	Selected Reference(s) or Sources	Date	Additional References Identified
health agency documents	ACGIH (2001). Naphthalene. Documentation of the threshold limit values and biological exposure indices. Cincinnati, OH: American Conference of Industrial Hygienists.	5/2013	4 citations added
	ATSDR (2005). Toxicological Profile for Naphthalene, 1-Methylnaphthalene, and 2-Methylnaphthalene. Atlanta, GA: Agency for Toxic Substances and Disease Registry.	5/2013	7 citations added
	IARC (2002). IARC Monographs on the evaluation of the carcinogenic risk of chemicals to humans: Some traditional herbal medicines, some mycotoxins, naphthalene, and styrene [IARC Monograph]. Lyon, France. http://monographs.iarc.fr/ENG/Monographs/vol82/mono82.pdf	5/2013	3 citations added
	NTP (2011). Naphthalene. In Report on Carcinogens, 12th Edition. National Toxicology Program.	5/2013	0 citations added
	WHO (1998). Selected non-heterocyclic polycyclic aromatic hydrocarbons. Environmental Health Criteria, 202. Geneva, Switzerland, World Health Organization.	5/2013	2 citations added
Web of Science, "forward" searcha	Abdo et al. (2001). Toxicity and carcinogenicity study in F344 rats following 2 years of whole-body exposure to naphthalene vapors. Inhalation Toxicology 13:931-950.	1/2017	0 citations added
		5/2013	0 citations added
	Dodd et al. (2012). Nasal epithelial lesions in F344 rats following a 90-day inhalation exposure to naphthalene. Inhalation Toxicology 24:70-79.	1/2017	0 citations added
		5/2013	0 citations added
	Shopp et al. (1984). Naphthalene toxicity in CD-1 mice: general toxicology and immunotoxicology. Toxicological Sciences 4:406-419.	1/2017	0 citations added
		5/2013	0 citations added
Web of Science, "backward" searchb	Abdo et al. (2001). Toxicity and carcinogenicity study in F344 rats following 2 years of whole-body exposure to naphthalene vapors. Inhalation Toxicology 13:931-950.	5/2013	2 citations added
	Dodd et al. (2012). Nasal epithelial lesions in F344 rats following a 90-day inhalation exposure to naphthalene. Inhalation Toxicology 24:70-79.	5/2013	0 citations added
	Shopp et al. (1984). Naphthalene toxicity in CD-1 mice: general toxicology and immunotoxicology. Toxicological Sciences 4:406-419.	5/2013	5 citations added
References obtained during the assessment process	References that had been previously added to the HERO project page for the naphthalene assessment during the development of earlier draft materials.	3/2017	2 citations added
		1/2017	9 citations added
		12/2015	22 citations added

System Used	Selected Reference(s) or Sources	Date	Additional References Identified
		5/2013	36 citations added
Search of Online Chemical Assessment-Related Websites	Searched a combination of CASRNs and synonyms on the following databases: American Conference of Governmental Industrial Hygienists (ACGIH): https://www.acgih.org/ American Industrial Hygiene Association (AIHA): Workplace Environmental Exposure Levels (WEELs) (https://www.tera.org/OARS/PDF_documents/OARS_WEEL_Table.pdf) Emergency Response Planning Guidelines (ERPGs) (https://www.aiha.org/get-involved/AIHAGuidelineFoundation/EmergencyResponsePlanningGuidelines/Pages/default.aspx) Agency for Toxic Substances and Disease Registry (ATSDR): https://wwwn.cdc.gov/TSP/index.aspx CalEPA Office of Environmental Health Hazard Assessment (OEHHA): http://www.oehha.ca.gov/risk.html OEHHA Toxicity Criteria Database (http://www.oehha.ca.gov/tcdb/index.asp) Biomonitoring California-Priority Chemicals (https://biomonitoring.ca.gov/chemicals/priority-chemicals) Biomonitoring California-Designated Chemicals (https://biomonitoring.ca.gov/chemicals/designated-chemicals) Cal/Ecotox Database (https://ecotox.oehha.ca.gov/) OEHHA Fact Sheets (http://www.oehha.ca.gov/public_info/facts/index.html) Non-cancer health effects [reference exposure levels (RELs)] (http://www.oehha.ca.gov/air/allrels.html) Cancer Potency Factors (Appendix A and B) (http://www.oehha.ca.gov/air/hot_spots/tsd052909.html) Consumer Product Safety Commission (CPSC): http://www.cpsc.gov Centre for Chemical Safety Assessment (ECETOC): http://www.ecetoc.org/publications European Chemicals Agency (ECHA): General site (http://echa.europa.eu/information-on-chemicals) Registered Substances (https://echa.europa.eu/information-on-chemicals/registered-substances) Existing Substances Regulation (ESR) (http://echa.europa.eu/information-on-chemicals/information-from-existing-substances-regulation) Environment Canada: Toxic Substances Managed Under Canadian Environmental Protection Act (http://www.ec.gc.ca/toxiques-toxics/Default.asp?lang=En&n=98E80CC6-1)	8/2022	23 citations added
		1/2017	1 citation added
		12/2015	13 citations added
		4/2012	19 citations added

System Used	Selected Reference(s) or Sources	Date	Additional References Identified
	Final Assessments (http://www.ec.gc.ca/lcpe-cepa/default.asp?lang=En&xml=09F567A7-B1EE-1FEE-73DB-8AE6C1EB7658) Draft Assessments (http://www.ec.gc.ca/lcpe-cepa/default.asp?lang=En&xml=6892C255-5597-C162-95FC-4B905320F8C9) Federal Docket: www.regulations.gov Health Canada: Health Canada Drinking Water Documents (http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/index-eng.php#tech_doc) Health Canada First Priority List Assessments (http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/psl1-lsp1/index-eng.php) Health Canada Second Priority List Assessments (http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/psl2-lsp2/index-eng.php) International Agency for Research on Cancer (IARC): http://monographs.iarc.fr/ENG/Monographs/vol101/mono101-B02-B03.pdf International Toxicity Estimates for Risk (ITER): https://iter.tera.org/ Japan Existing Chemical Data Base: http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPageENG.jsp National Academies of Sciences, Engineering, and Medicine (NASEM): http://www.nap.edu/ National Cancer Institute (NCI): http://www.cancer.gov National Industrial Chemicals Notification and Assessment Scheme (NICNAS) (Australia): Australian Inventory of Chemical Substances (AICS) (http://www.cirs-reach.com/Inventory/Australian_Inventory_of_Chemical_Substances_AICS.html) National Institute of Environmental Health Sciences (NIEHS): http://www.niehs.nih.gov/ National Institute of Occupational Safety and Health (NIOSH): All Workplace Safety & Health Topics (http://www.cdc.gov/niosh/topics/) NIOSHTIC 2 Publications Search: http://www2a.cdc.gov/nioshtic-2/ Registry of Toxic Effects of Chemical Substances (https://www.cdc.gov/niosh/rtecs/default.html) National Institute of Technology and Evaluation Chemical Risk Information Platform (NITE-CHIRP) (Japan): http://www.safe.nite.go.jp/english/db.html National Toxicology Program (NTP): Report on Carcinogens (RoC) (https://ntp.niehs.nih.gov/whatwestudy/assessments/cancer/roc/index.html) NTP Site Search (https://ntpsearch.niehs.nih.gov/)		

System Used	Selected Reference(s) or Sources	Date	Additional References Identified
	Occupational Safety and Health Administration (OSHA): http://www.osha.gov/dts/chemicalsampling/toc/toc_chemsamp.html Organisation for Economic Cooperation and Development (OECD)c: eChemPortal (https://www.echemportal.org/echemportal/substance-search) OECD Existing Chemicals Database (https://hpvchemicals.oecd.org/ui/Search.aspx) U.S. Environmental Protection Agency (EPA): Acute Exposure Guideline Levels (https://www.epa.gov/aegl/access-acute-exposure-guideline-levels-aegls-values#chemicals) Integrated Risk Information System (IRIS) (http://www.epa.gov/iris/) National Service Center for Environmental Publications (NSCEP) (https://www.epa.gov/nscep) RfD/RfC and Carcinogen Risk Assessment Verification Endeavor (CRAVE) meeting notes Science Inventory (http://cfpub.epa.gov/si/) High Production Volume Information System (HPVIS) (https://ofmpub.epa.gov/oppthpv/metadata.html) Chemical Data Access Tool (formerly available at http://java.epa.gov/oppt_chemical_search/; information from CDAT has been incorporated into EPA's ChemView database at https://chemview.epa.gov/chemview) Office of Pesticide Programs (http://iaspub.epa.gov/apex/pesticides/f?p=chemicalsearch:1) U.S. Food and Drug Administration (FDA): http://www.fda.gov/ National Center for Toxicological Research (NCTR) (http://www.fda.gov/AboutFDA/CentersOffices/OC/OfficeofScientificandMedicalPrograms/NCTR/default.htm)		

[a] "Forward" search for records that cite included studies

[b] "Backward" search for records cited by included studies

[c] Searched for OECD High Production Volume (HPV) chemicals, Screening Information Dataset (SIDS) International Uniform Chemicals Information Database (IUCLID), and SIDS United Nations Environment Programme (UNEP).

B.1. ELECTRONIC SCREENING

For literature searches conducted through November 2015, all identified records were first electronically screened with a set of terms intended to prioritize "on-topic" references for title and abstract review. The electronic screening process creates two broad categories: one comprising all records that contain (in title, abstract, or keywords) at least one inclusion/exclusion term (listed in Table A-3) related to health outcomes, epidemiological or toxicological study design, toxicokinetics, or mechanistic information, and one that does not contain any of the terms. Some of the electronic inclusion/exclusion terms are generic (i.e., not chemical specific) and are intended to capture health effect studies of any type. Other terms are specific to naphthalene and are based on previous knowledge of health effects and possible mechanisms of toxicity. Records that contained at least one inclusion/exclusion term were moved forward for title and abstract screening.

Citations that did not contain at least one inclusion/exclusion term in Table A-3 were subjected to a quality control check to verify that relevant references are not missed. Specifically, a random sample (~10%) of the electronically excluded citations were subjected to title/abstract review by a scientist (toxicologist or epidemiologist) to confirm that the electronic screening process produced acceptable results (i.e., no relevant citations were inadvertently missed). If the random sample contained at least one potentially relevant citation, the list of electronic screening terms was revised to add terms pertaining to the missing citation, and the electronic screening process was repeated. This quality control and revision process was repeated as many times as necessary to ensure that relevant studies are retained for title/abstract screening. Citations that did not contain at least one term inclusion/exclusion term were excluded from further review.

Table B-5. Electronic screening inclusion terms for naphthalene (listed alphabetically within each organ/system category)

Category	Terms			
Organ/System Specific Terms				
Cardiovascular	angio aort arrhythm artery, arteri blood AND pressure	blood AND vessel capillar cardiac, cardio, cardium circulat coronary	endotheli heart hypertens infarct myocardi	thrombus valve vascular, vaso vein, venous ventricle
Dermal/ Integumentary system	blister bulla, bullous cutaneous dermal, dermis	epiderm, epidermal erythema hair keratin, kerato	nail pruritus sebaceous skin	sweat, perspiration tooth, teeth
Developmental	abnormalit abort cleft congenital defect development embryo	fetal, fetus, foetal, foetus gestation implantation malform neonat newborn neural AND tube	parturition perinatal postnatal puberty pregnan prenatal resorption	terato uterus, uterine viable, viabil visceral wean zygote
Endocrine	adipokine adipocyt adrenal hormone	hypothalamus insulin pancreas, pancreat pineal	pituitary triiodo tetraiodo thymus, thymic	thyro
Gastrointestinal	abdomen anus, anal bucca bowel cecum, cecal colon	constipation diarrhea digestive duoden esophagus gastric	gastrointestinal ileum, ileal, ileus intestin jejunum, jejunal mouth oral AND cavity	peptic rectum, rectal salivary stomach tongue

Category	Terms			
Organ/System Specific Terms				
Hematologic	albumin anemia, anemic, anaemia, anaemic blood cholesterol clot coagulat	cytopenia erythro hemoly, haemoly hemat hemocoagulat hemoglobin	histamine hypoxemi granulocyt plasma platelet polycythemia	RBC (red blood cell) reticulocyt serum thrombo
Hepatic	alkaline AND phosphatase aminotransferase bile, biliary bilirubin centrilobular	cholesta cholangio cirrho gall AND bladder glycogen	glutamyltransferase hepat hydropic Ito Kuppfer	liver peroxisome portal, periportal steatosis stellate
Immune	adenopath allerg anaphyla antibod antigen asthma basophil, basopenia B-cell cytokine chemokine	complement dendrocyt, dendritic eosinophil, eosinopenia epitope globulin granuloma hapten humoral hypersensit immun	inflamm interferon leukocyt lymph macrophag major histocompatibility complex, MHC marrow mast AND cell macroglobulin	monocyt natural AND killer neutrophil, neutropenia phagocyt polymorphonuclear sensitize, sensitis sensitivity spleen, splenous WBC (white blood cell) T-cell
Musculoskeletal	articular bone bursa calcitonin	cartilage collagen connective ligament	muscle, muscul osteo pyridinoline skelet	tendon vertebra
Nervous	autonomic axon behavior, behaviour brain CNS (central nervous system) Cognitive dendrite	efferent electrophysiol encephalo fatigue FOB (functional observational battery) ganglia, ganglio	memory myelin AND sheath locomotor nerve nervous AND system neuro parasympathetic	PNS (peripheral nervous system) Ranvier Schwann sensory, sensori spinal AND cord sympathetic synap

Category	Terms			
Organ/System Specific Terms				
Ocular	cataract cornea eye	harderian lachrymal, lacrimal lens, lenticular	ocular ophthalm retina	
Reproductive	androgen breast cervical, cervix coagulating AND gland corpora lutea, corpus luteum endometrium epididym estrogen, estradiol estrus, estrous fallopian	fertilit follicle FSH gamete gonad infertility lacto, lacta LH (luteinizing hormone) lordosis mammar	ova, ovum penis placenta primordial progesterone prolactin prostate reproduct scrotum seminal AND vesicle	seminiferous sexual sperm sterility testes, testic, testis testosterone urogenital vagina vulva
Respiratory	airway alveolar BAL (bronchoalveolar lavage) bleb bronch chest	cough crackle diffusing AND capacity dyspnea FEV, forced AND expiratory FVC, forced AND vital	intratrach laryn lung nasal nose olfactory	pharyn pneumon pulmonary rale respir trach
Urinary	alpha 2u globulin anion AND gap BUN bladder Bowman's	creatinine dilation, dilatation genitourinary glomerul Henle	kidney nephro proximal AND tubule, distal AND tubule renal	urethra uria urinalysis urinary urine
Nonspecific Terms				
Epidemiology	case-control, case AND control case AND report, case AND series	cohort epidemiol	occupation	survey
Animal	animals baboon beagle cat, cats, feline chimp	dog, dogs, canine ferret gerbil guinea hamster	macaque marmoset monkey mouse, mice, murine pig, pigs, porcine, swine	primate rabbit rat, rats rodent

Cell lines, single-celled organisms, and other in vitro and ex vivo terms	Bacillus Drosophila E. Coli	Escher Explant photobacterium	Saccharomyces Salmonella V79	
Survival and general toxicity	anorexi body AND weight	weight AND loss death, mortality, survival	poison	
General cancer terms	adenoma hemangioma biops	cancer carcino CDC2	malignan metasta neoplas	oncogen sarcoma tumor, tumour
General gross and microscopic pathology terms	apoptosis, apoptotic amyloid atrophy atypic, atypia biometr congest cyst degenerat dysplas dystroph	edema endoplasmic epitheli fibros, fibrotic hemorrhag histiocytic histometr histolog, histopatholog hyaline	hyperplas hypertroph hypoxi infiltrat lesion medulla metaplas microdissected mitochondria mucosa	necrosis, necrotic nodul parenchyma phenotyp radiographic tubul vacuol vesicul
Nonspecific clinical chemistry	calcium	clinical AND chemistry	glucos	
Inflammation/oxidative stress	buthionine AND sulfoximine, BSO diethyl AND maleate, DEM glutathione, GSH	lipid AND peroxidation oxidative AND stress reactive AND oxygen AND species,	ROS thiobarbituric, TBARS TNF	
Genotox/mutagenicity	aber ames assay ames test aneuploid anisokaryo, anisonucleo binuclea	chromati, chromosom clastogen cytogen DNA dominant AND lethal gene, genes, genetic	genom genotox hyperploid karyo micronucle mitotic	mutagen mutat polyploid recessive AND lethal sister AND chromatid, SCE, SCEs
ADME/TK	absorb, absorp cytochrome, CYP deposit distribut	excret metabol microsom PBPK	PBTK pharmacokinetic protein AND binding stereo	tolerance toxicokinetic
Naphthalene-specific Terms				

Toxicity terms	Clara	club AND cell	clubbing (of the nail)	
ADME/TK	aldoketo dihydrodiol	epoxide naphthoquinone	oxide	
Mechanistic terms	CC10 (Clara cell 10-kDa protein) CC16 (Clara cell 16-kDa protein) CCSP (Clara cell 10 kDa secretory protein) CGRP (calcitonin gene-related peptide)	cyclin dependent kinase 1, CDK1 EGF (epidermal growth factor) metalloproteinase, MMP NEB, NEBs (neuroepithelial body) nerve growth factor, NGF Neurotrophic tyrosine kinase	PNEC (pulmonary neuroendocrine cell) signal transducer and activator of transcription 3, STAT3 SCGB1A1 (Secretoglobin 1A1) sulfhydryl	TFF, trefoil (trefoil factor) trk1 (Neurotrophic tyrosine kinase receptor 1) TrkA (tropomyosin receptor kinase A)

1

APPENDIX C. INITIAL LITERATURE INVENTORY FOR NAPHTHALENE (SYSTEMATIC EVIDENCE MAP)

An SEM for naphthalene was conducted according to the methods for literature search, screening, and inventory described in Section 4 and was used to develop a literature inventory of human and animal health effect studies and PBPK models meeting the problem formulation PECO criteria described in Section 4.1. A literature flow diagram summarizing the literature search and screening results is shown in Figure C-1. Literature search and screening results can also be viewed on the HERO project page for this assessment (https://hero.epa.gov/hero/index.cfm/project/page/project_id/367).

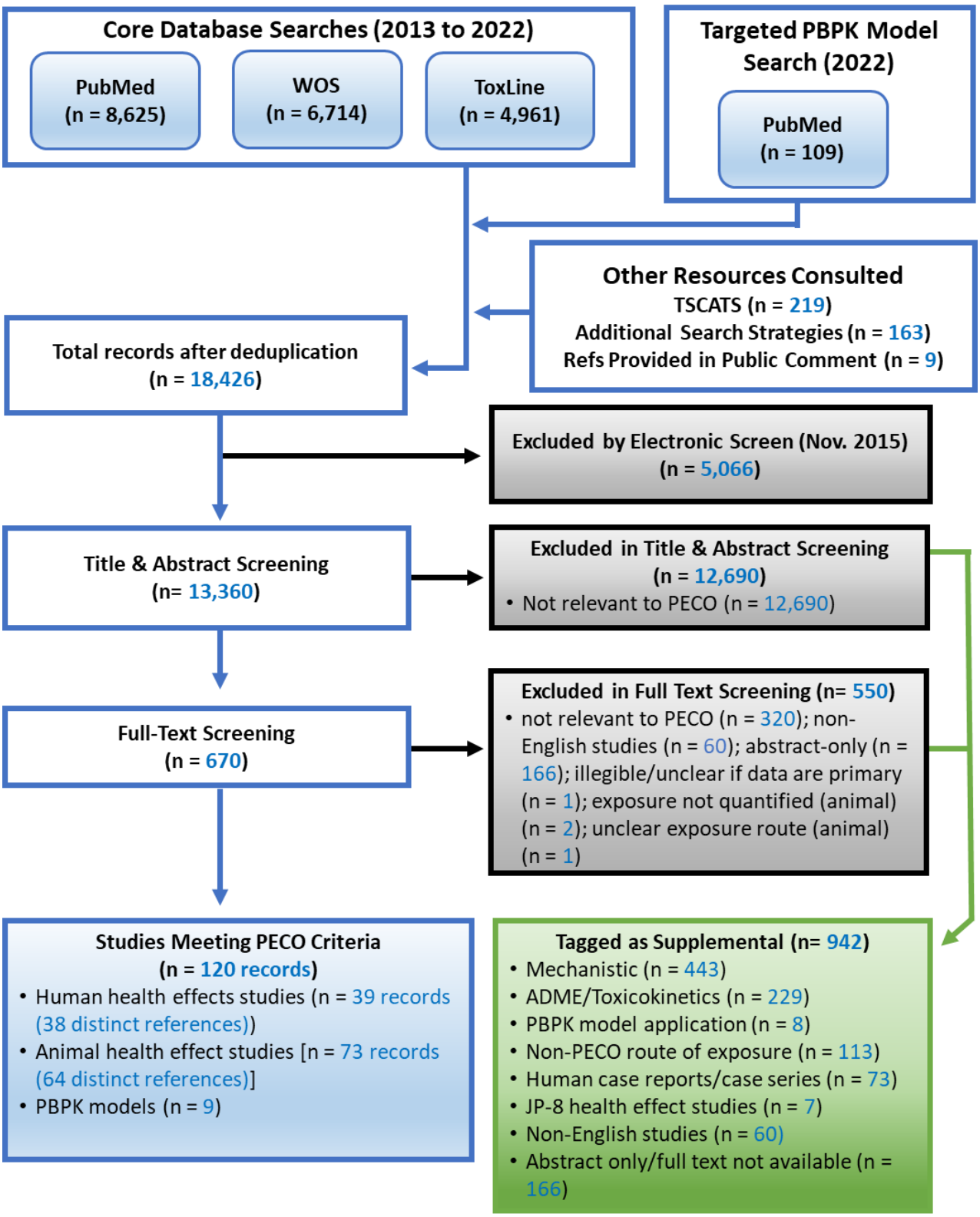


Figure C-1. Literature flow diagram for naphthalene.

C.1. HUMAN AND ANIMAL HEALTH EFFECT STUDIES

A survey of study designs and health systems assessed in the human studies that met the problem formulation PECO criteria is provided in Figure C-2. A total of 38 epidemiological studies were identified that evaluated effects in several population types (occupational, general population, pregnant women/infants, and children). Studies classified as "inhalation" exposure quantified naphthalene levels in air, whereas studies classified as "nonspecific" exposure used biomonitoring to assess naphthalene or naphthalene metabolites in blood or urine. The epidemiological studies that evaluated pulmonary, nasal, hematological, immune, reproductive, or developmental effects meet the assessment PECO criteria (see Section 5.1) and therefore will be included in the assessment-specific approach as described in Section 5 (29 studies total).

A survey of study designs and health systems evaluated in the 64 animal studies that met the problem formulation PECO criteria is provided in Figure C-3. Studies with inhalation and oral routes of exposure were identified. Durations of exposure ranged from acute to chronic, and there were several oral exposure studies that exposed animals during gestation. Inhalation exposure studies were conducted in rats and mice, and oral exposure studies were conducted in rats, mice, and rabbits. Seventeen of these studies met assessment PECO criteria based on the considerations described in Section 5.1 and will be included in the assessment-specific approach.

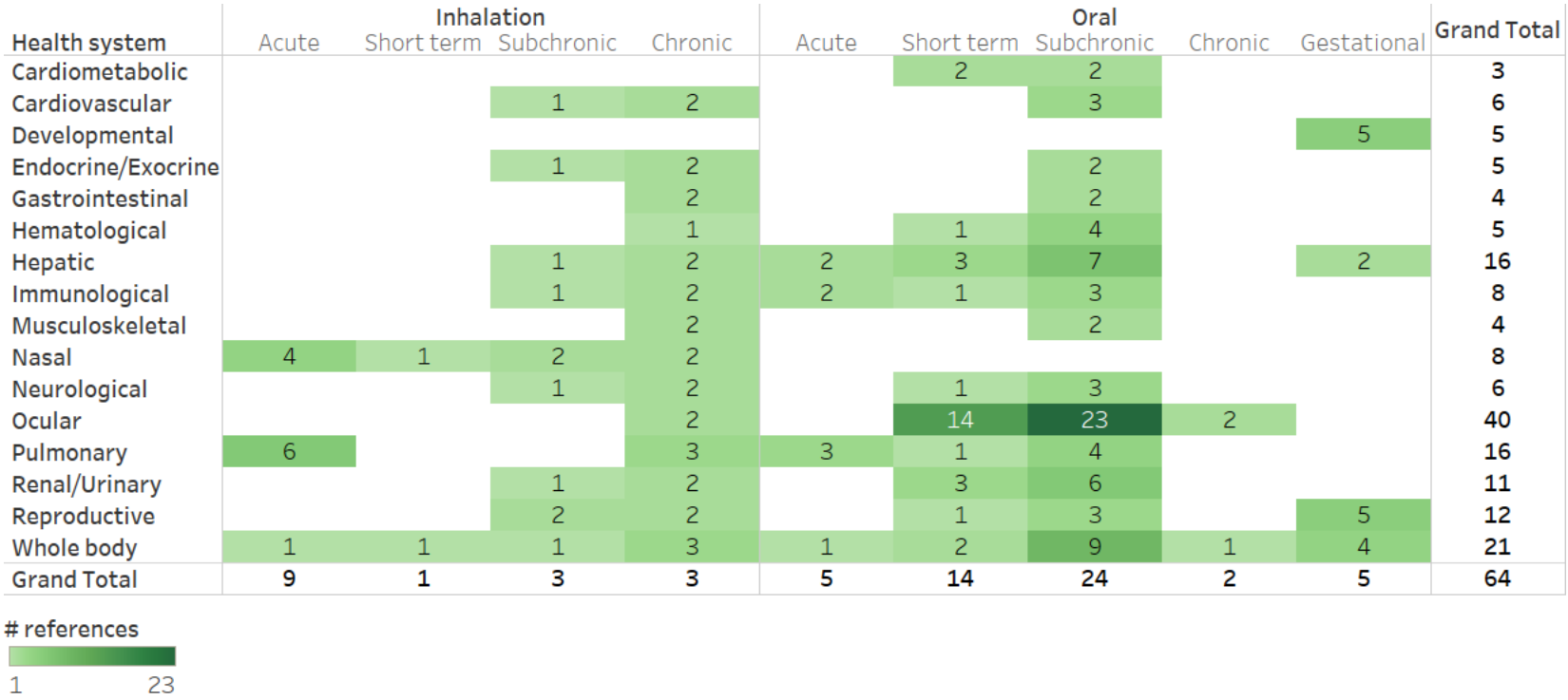
Interactive versions of these literature inventory figures that include a more detailed description of study designs and results are available on a Tableau Public dashboard, which also allows users to filter for the subset of studies that are included under the assessment PECO criteria (see Section 5).

Health system	Inhalation			Nonspecific				Grand Total
	Occupational	General population	Children	Occupational	General population	Pregnant women/infants	Children	
Cardiometabolic					2		2	4
Developmental						4	1	5
Endocrine/Exocrine					3			3
Gastrointestinal	1							1
Hematological				1	1		1	3
Hepatic						2		2
Immunological	1		2		1		5	9
Nasal	1							1
Neurological	1							1
Pulmonary	1	1			1			3
Reproductive					6	2		8
Grand Total	**4**	**1**	**2**	**1**	**13**	**8**	**9**	**38**

references

1 6

Figure C-2. Survey of human studies that met PECO criteria by study design and health systems assessed. The numbers indicate the number of studies that investigated a particular health system, not the number of studies that observed an association with naphthalene exposure. If a study evaluated multiple health outcomes, it is shown here multiple times. An interactive version of this figure that includes a more detailed description of study designs and results is available at the following URL: https://public.tableau.com/app/profile/literature.inventory/viz/NaphthaleneEvidenceMapUSEPAIRISSystematicReviewProtocol2022/ReadMe?publish=yes



Health system	Inhalation				Oral					Grand Total
	Acute	Short term	Subchronic	Chronic	Acute	Short term	Subchronic	Chronic	Gestational	
Cardiometabolic						2	2			3
Cardiovascular			1	2			3			6
Developmental									5	5
Endocrine/Exocrine			1	2			2			5
Gastrointestinal				2			2			4
Hematological				1		1	4			5
Hepatic			1	2	2	3	7		2	16
Immunological			1	2	2	1	3			8
Musculoskeletal				2			2			4
Nasal	4	1	2	2						8
Neurological			1	2		1	3			6
Ocular				2		14	23	2		40
Pulmonary	6			3	3	1	4			16
Renal/Urinary			1	2		3	6			11
Reproductive			2	2		1	3		5	12
Whole body	1	1	1	3	1	2	9	1	4	21
Grand Total	9	1	3	3	5	14	24	2	5	64

Figure C-3. Survey of animal studies that met PECO criteria by exposure duration, species, and health systems assessed. The numbers indicate the number of studies that investigated a particular health system, not the number of studies that observed an association with naphthalene exposure. If a study evaluated multiple species, study designs, or health outcomes, it is shown here multiple times. An interactive version of this figure that includes a more detailed description of study designs and results is available at the following URL: https://public.tableau.com/app/profile/literature.inventory/viz/NapthaleneEvidenceMapUSEPAIRISSystematicReviewProtocol2022/ReadMe?publish=yeshttps://public.tableau.com/app/profile/literature.inventory/viz/NapthaleneEvidenceMapUSEPAIRISSystematicReviewProtocol2022/ReadMe?publish=yes

C.2. PHARMACOKINETIC (PK)/PHYSIOLOGICALLY BASED PHARMACOKINETIC (PBPK) MODELS

The literature search identified nine peer-reviewed publications that describe novel, whole-organism PBPK models for naphthalene (Kapraun et al., 2020; Celsie et al., 2016; Campbell et al., 2014; Morris, 2013; Kim et al., 2007; Willems et al., 2001; NTP, 2000; Quick and Shuler, 1999; Sweeney et al., 1996) and eight additional peer-reviewed publications that describe applications of PBPK models for naphthalene (Bailey and Rhomberg, 2020; Clewell et al., 2014; Viravaidya et al., 2004; Viravaidya and Shuler, 2004; Ghanem and Shuler, 2000a, b; Shuler et al., 1996; Sweeney et al., 1995). Of the publications describing the application of PBPK models, six describe cell culture analogs (CCAs) of PBPK models (Viravaidya et al., 2004; Viravaidya and Shuler, 2004; Ghanem and Shuler, 2000a, b; Shuler et al., 1996; Sweeney et al., 1995). CCA models are constructed as in vitro cell culture systems rather than in silico mathematical descriptions of whole organisms; thus, CCA models cannot be efficiently utilized for risk assessment dosimetry calculations. The two remaining publications involving applications of PBPK models describe studies that made use of existing PBPK models.

The paragraphs that follow provide details of the nine publications that describe novel, whole-organism PBPK models for naphthalene, as well as two publications (Corley et al., 2012; Zhang and Kleinstreuer, 2011) that describe computational fluid dynamics (CFD) models that inform naphthalene dosimetry. Table C-1 provides summary information for these eleven models.

Table C-1. Summary of Novel PBPK and Airway Dosimetry Models for Naphthalene

Citation	Species	Exposure routes	Metabolism[a]	Respiratory tract details
(Sweeney et al., 1996)	• Mouse • Rat	• Oral • Intraperitoneal	• Liver • Lung • Naphthalene oxidation • Naphthalene oxide: ○ Hydrolysis ○ GSH conjugation ○ Rearrangement ○ Covalent binding	None: A "lung" compartment is included in the model as a site of metabolism, but the model does not describe inhalation exposure.
(Quick and Shuler, 1999)	• Mouse • Rat	• Oral • Intraperitoneal • Intravenous • Inhalation	• Liver • Lung • Naphthalene oxidation • Naphthalene oxide: ○ Hydrolysis ○ GSH conjugation ○ Rearrangement ○ Covalent binding	Pulmonary gas exchange
(NTP, 2000)	• Mouse • Rat	• Inhalation	• Liver: ○ Michaelis-Menten ○ Hill • Lung: ○ Michaelis-Menten • Naphthalene oxidation	Pulmonary gas exchange
(Willems et al., 2001)	• Mouse • Rat	• Inhalation • Intravenous	• Liver • Lung • Naphthalene oxidation • Naphthalene oxide: ○ Hydrolysis ○ GSH conjugation	Pulmonary gas exchange

Citation	Species	Exposure routes	Metabolism[a]	Respiratory tract details
(Kim et al., 2007)	• Human	• Inhalation • Dermal	• Liver • Naphthalene oxidation	Pulmonary gas exchange
(Morris, 2013)	• Mouse	• Inhalation	• Nasal • Naphthalene oxidation	Nasal airway compartments with air-tissue mass transfer based on computational fluid dynamics (CFD)
(Zhang and Kleinstreuer, 2011)	• Human	• Inhalation	• None[b]	Full CFD model of airways through the upper tracheobronchial region
(Corley et al., 2012)[c]	• Rat • Monkey • Human	• Inhalation	• Nasal • Conducting airways • Secondary bronchi • Bronchioles	Full CFD model of airways through the secondary bronchi and bronchioles
(Celsie et al., 2016)	• Fish	• Gills	• Liver: ○ First order[d] • Naphthalene oxidation	None: Exchange of naphthalene exchange between aqueous environment and blood in gills similar to pulmonary gas exchange in mammals.
(Campbell et al., 2014)	• Rat • Human	• Inhalation	• Liver • Lung • Nasal • Naphthalene oxidation	• Nasal airway compartments with air-tissue mass transfer based on CFD • Pulmonary gas exchange

Citation	Species	Exposure routes	Metabolism[a]	Respiratory tract details
(Kapraun et al., 2020)	• Rat • Human	• Inhalation • Dermal • Intravenous	• Liver • Lung • Nasal • Naphthalene oxidation	• Nasal airway compartments with air-tissue mass transfer based on CFD • Pulmonary gas exchange

[a]Unless otherwise indicated, metabolism is described using Michaelis-Menten rate equations.

[b]Zhang and Kleinstreuer (2011) only described the concentration distribution in the airways but assumed that uptake by airway tissues is proportional to air concentration. Thus, there is an implicit assumption of ongoing first-order removal of naphthalene from the airway lining and that metabolism may contribute to that removal.

[c]The model of Corley et al. (2012) was not parameterized for naphthalene, but it is included in this summary because it is the most advanced air-phase vapor deposition model for the rat, monkey, and human respiratory tracts, and as such, it could potentially inform naphthalene inhalation dosimetry.

[d]Celsie et al. (2016) included a term for first-order elimination of naphthalene in a liver compartment but no value for the parameter was given and a later statement indicates that it was set to zero for the analysis of short duration exposures.

The first naphthalene PBPK model published in the peer reviewed literature (Sweeney et al., 1996) describes the kinetics of naphthalene and naphthalene oxide in mice and rats. This model was subsequently revised and extended by Quick and Shuler (1999). The original model of Sweeney et al. (1996) contained five compartments (lung, fat, kidney, liver, and combined "other tissues"), with saturable metabolism of naphthalene to the enantiomers of naphthalene oxide, as well as subsequent hydrolysis to the 1,2-dihyrodiol, conjugation to GSH, non-enzymatic rearrangement to 1-naphthol, and covalent binding to intracellular protein occurring in lung and liver compartments. Kinetic parameters for these processes were selected based on optimal fit to published in vitro reaction data. The model facilitated predictions of internal doses following oral and intraperitoneal (IP) exposures; however, rates of oral uptake were estimated in the absence of sufficient data. The model was used to simulate available pharmacokinetic data for naphthalene, including GSH conjugation and re-synthesis, covalent binding in lung and liver, and GSH concentration in lung and liver; however, the simulation results were not evaluated against pharmacokinetic data for naphthalene or its metabolites in blood or tissues.

The updated model published by Quick and Shuler (1999) has a structure similar to the Sweeney et al. (1996) model, but it includes explicit arterial and venous blood compartments with added intravenous (IV) and inhalation exposure routes. Kinetic parameters for metabolism of naphthalene as well as metabolism, protein binding, and non-enzymatic rearrangement of naphthalene oxide in mouse were updated using a separate whole cell model describing Club (formerly Clara) cells and hepatocytes. Kinetic parameters in rat were fit to microsomal data and, in the case of liver kinetics, adjusted based on data from the mouse whole cell model. The Quick and Shuler (1999) model has several notable shortcomings. Though equations are given, the inhalation route of exposure is not described in the methods, and a blood-to-air partition coefficient is not stated in the text or in tables. Also, while the motivation for using whole cell rather than cellular fraction (e.g., microsomal) kinetic data in the PBPK model is conceptually sound, particularly given the heterogeneity of lung tissue and its potential role in the site- and species-specificity of naphthalene toxicity, the description of how this was done is not sufficiently clear. Following IV exposure, model simulation of naphthalene in blood by the mouse model over-predicted alpha phase elimination and under-predicted beta phase elimination; predictions generated using the rat model were more comparable to observed data during the beta phase, but still over-predicted alpha phase elimination. Predictions of GSH and protein binding are reasonably accurate when compared to available data, and improve upon the Sweeney et al. (1996) model simulations, while the revised model did not accurately predict GSH depletion and re-synthesis. Ultimately, though the authors' approach to describing naphthalene metabolism has conceptual merit, the model is not robust or accurate enough for use.

A novel PBPK model for naphthalene is described in the National Toxicology Program Report on the Toxicology and Carcinogenesis Studies of Naphthalene in F344/N Rats (Inhalation Studies) (NTP, 2000). The authors of the NTP (2000) report claimed, "this model was the best

fitting product [for the data analyzed in the report] after testing several alternative models." The "alternative models" included the models of Sweeney et al. (1996) and Quick and Shuler (1999). The NTP (2000) model included a second metabolic rate term "in the form of a Hill equation" into the equation describing the amount of naphthalene in the liver. This second Hill term for metabolism was not included in any of the other identified PBPK models for naphthalene. Notably, the NTP (2000) model was constructed based on an assumption of diffusion-limited, rather than perfusion-limited kinetics. That is, for each of the five tissues represented in the model (lung, liver, kidney, fat, and "other"), the model includes one state variable for amount in the tissue and another for amount in the capillary blood of that tissue. In the model, diffusion between capillary blood and tissue depends on the difference in concentrations in those two compartments as well as a parameter describing capillary permeability.

The rat and mouse model of Willems et al. (2001) uses parallel sub-models for naphthalene (parent) and naphthalene oxide (metabolite) as described by Sweeney et al. (1996) and Quick and Shuler (1999), but incorporates diffusion-limited compartments in the parent sub-model as was done in the NTP (2000) model. As in the model of Quick and Shuler (1999), each sub-model includes compartments for lung, fat, kidney, liver, and "other" tissues, as well as explicit arterial and venous blood compartments. Saturable metabolism of naphthalene was included in lung and liver compartments. Metabolic rate and tissue permeability constants were optimized from blood time-course data from inhalation exposures. Performance of the rat model was evaluated against naphthalene (but not naphthalene oxide) blood time course concentration data following IV exposure; the mouse model was not evaluated against independent pharmacokinetic data. The predictions of IV rat data are reasonably accurate, though the data suggest naphthalene may be eliminated more slowly from blood than model predictions indicate. The authors state that the model as written does not explain the apparent interspecies differences in naphthalene toxicity in the lung, nor does it address nasal toxicity in either species.

The human model of Kim et al. (2007) describes the PK behavior of naphthalene as a surrogate for jet propulsion fuel 8 (JP-8). The model contains five compartments — two representing layers of skin (the exposed portion of the stratum corneum, and viable epidermis directly beneath this) and three representing the rest of the body (blood, fat, and combined other tissues) — and simulates dermal and inhalation exposures. First order metabolism of naphthalene to naphthalene-oxide by the liver is included in the blood compartment. Notably, the authors report measurement of a human blood-to-air partition coefficient of 10.3, which is considerably lower than the rodent value of 571 reported by Willems et al. (2001). Rate constants describing uptake and diffusion in the skin compartments and partition coefficients for fat-to-blood and other-tissues-to-blood were optimized to fit time course blood concentration data for each of 10 subjects included in a controlled dermal exposure study (Kim et al., 2006). Average parameter values were then used to define an "optimal" overall parameter set. The optimized model was used to predict concentrations of naphthalene in exhaled breath of 53 U.S. Air Force personnel exposed to

naphthalene via inhalation (without dermal contact), as well as 3 U.S. Air Force personnel exposed via inhalation and dermal contact. These predictions consistently overestimated observed data by 1 to 2 orders of magnitude unless inhalation concentrations were adjusted; while some rationale for this adjustment was provided, details of the adjustment were not described. Inadequate model validation and a limited treatment of respiratory tissues relevant to naphthalene toxicity reduce the utility of the Kim et al. (2007) model for the purposes of this assessment.

A hybrid computational fluid dynamic (CFD) and PBPK model (i.e., a "hybrid CFD-PBPK model") for nasal dosimetry of naphthalene in the upper respiratory tract (URT) of mice was described by Morris (2013). (Note that while the terms "CFD-PBPK model" and "hybrid CFD-PBPK model" are commonly used to describe PBPK models that have been informed by CFD models of airways [e.g., to determine parameters that describe rates and proportions of deposition for PBPK model compartments representing parts of the respiratory tract], these "hybrid" models do not actually incorporate CFD partial differential equations.) The model structure was based on that of the Gloede et al. (2011) CFD-PBPK model for diacetyl: stacks of compartments corresponding to the airspace, mucus, epithelium, and submucosa are described for relevant portions of the URT (including dorsal and ventral respiratory regions and a dorsal olfactory region). Other body tissues are not explicitly described, only the nasal epithelium and sub-mucosa. The model assumes saturable rates of metabolism in the epithelial and submucosal sub-compartments, with maximal rates specified for each region of the respiratory tract. Model prediction of uptake efficiency by the entire URT (i.e., all compartments representing components of the URT) was accurate when compared to observed data on vapor uptake in isolated URTs of mice; however, dosimetry predictions for the described individual portions of the URT could not be evaluated since PK data specific to the URT sub-regions is not available, and therefore the validity of the model's complex nasal structure cannot be confirmed.

Zhang and Kleinstreuer (2011) developed a full CFD model that predicts deposition of naphthalene in the human respiratory tract. Note that the Zhang and Kleinstreuer (2011) model is not a PBPK model, but a type of dosimetry model. The model uses a geometrically accurate model of the airways through the upper tracheobronchial region, with a level of resolution that is lost in the development of hybrid CFD-PBPK models. However, the model of Zhang and Kleinstreuer (2011) does not have airway tissue compartments and assumes a rate of uptake by the airway lining that is simply proportional to the concentration of naphthalene in the adjacent air, i.e., it does not account for metabolic saturation but implicitly assumes ongoing elimination of naphthalene such that it does not accumulate in the airway lining. Results from this model might still be valid at low exposure levels, below saturation, but could only be used in extrapolation of naphthalene deposition or tissue flux predicted by a rodent CFD-PBPK model. Further, tabulated results reported by the authors only give uptake by major anatomical region; the nasal cavity is not sub-divided into olfactory and respiratory tissues. Thus, the model is limited in utility and does not incorporate the human vs. rodent differences in metabolic rate observed in vitro.

More recently, Corley and colleagues developed (Corley et al., 2012) and applied (Corley et al., 2015) a full CFD model for rats, monkeys, and humans, which includes two tissue layers (mucus + epithelium and sub-epithelium) and which allows for removal by a first-order pathway and a saturable metabolic pathway in each layer, plus blood perfusion in the sub-epithelium. The model defines separate areas for respiratory and olfactory epithelia in the nose. While the models of Corley et al. (2012) might not include compartments for the rest of the body, they otherwise represent the most anatomically accurate representation of airway geometry and vapor disposition in rats, monkeys, and humans, with a good level of detail for the airway tissues. The primary barrier to further consideration of these CFD models is that they have not been parameterized for naphthalene, which would require setting the metabolic parameters in each airway region appropriately. The Corley et al. (2012) model is not a whole-body PBPK model but includes compartments for respiratory tissues with parameters set based on anatomical and physiological data, like the model of Morris (2013). While it was not parameterized for naphthalene, it is described here because it is the most advanced model of air-phase vapor deposition for the rat, monkey, and human airways, with high anatomical accuracy and the capacity to incorporate first-order and saturable metabolism in the tissues.

Celsie et al. (2016) developed a PBPK model for narcotic organic chemicals in fish and parameterized the model for describing naphthalene concentrations in fathead minnows. This model includes compartments for gills, blood, liver, rapidly perfused tissue, and slowly perfused tissue, as well as a compartment for "membrane," which is the assumed target site of toxicity. The Celsie et al. (2016) model allows for simulations of aquatic exposures via the gills, which are analogous to but anatomically and physiologically different from mammalian lungs. Furthermore, the Celsie et al. (2016) model equations are constructed in the "fugacity format," making them quite different from PBPK model equations typically used for mammalian species. The state variables of the model are time-varying fugacities (Pascals), and these can be used along with constant "fugacity capacities" (moles per cubic meter per Pascal) to calculate concentrations in the various model compartments. While the Celsie et al. (2016) model could potentially be adapted to create a PBPK model for mammalian dosimetry, the resulting model would need to be evaluated using naphthalene PK data in the species of interest. Also, the Celsie et al. (2016) model lacks URT compartments which allow for tissue- and site-specific dosimetry in the URT. Thus, this model is not ideal for the current human health assessment application.

Campbell et al. (2014) published a CFD-PBPK model for naphthalene in rats and humans. Unlike the model of Morris (2013), this model includes compartments representative of the entire body rather than just the URT. The URT components were based on published inhalation-route models for vinyl acetate (Bogdanffy et al., 1999) and acetaldehyde (Teeguarden et al., 2008) and are organized into two parallel airways: (1) the dorsal path, comprising sequential compartments for respiratory epithelium and one or two olfactory compartments; and (2) the ventral path, comprising two respiratory epithelium compartments. One dorsal olfactory compartment was used

for the human model, whereas two dorsal olfactory compartments were used for the rat model. Similar to the Morris (2013) CFD-PBPK model for mice, the Campbell et al. (2014) model represents each of the URT compartments with multiple layers. In the case of the Campbell et al. (2014) model, each URT compartment consists of lumen, epithelial cell layer, and submucosal tissue sub-compartments. In addition to the URT compartments, the model includes compartments for lung, fat, liver, slowly perfused, and rapidly perfused tissues. Time-course data for naphthalene concentrations in rat blood after single IV doses (Quick and Shuler, 1999) and 6-hour inhalation exposures (NTP, 2000), as well as rat upper respiratory tract extraction data at fixed inspiratory flow rates (Morris and Buckpitt, 2009), were used to evaluate the accuracy of rat model predictions. As was the case for the Morris and Buckpitt (2009) model, dosimetry predictions for distinct sub-regions of the URT could not be evaluated since PK data specific to the represented URT sub-regions is not available. Also, while Campbell et al. (2014) showed that their rat model simulations generally reproduced observed rat data to within a factor of 2 (and in the worst case, to within a factor of 3), time-course data for humans exposed to naphthalene via the inhalation route were not available to evaluate the human model predictions.

Kapraun et al. (2020) revised the PBPK model of Campbell et al. (2014) by adding compartments that allow one to simulate skin exposure. (See Table C-2 for descriptive summary.) This enhancement allowed Kapraun et al. (2020) to evaluate their PBPK model using data from a controlled skin exposure study in human subjects (Kim et al., 2006) and demonstrate that model predictions of time-course blood concentrations of naphthalene generally agree with observed human in vivo data to within a factor of two. Such agreement supports the general practice that PBPK model dosimetry, rather than allometric scaling or other default approaches, are preferred for dosimetry calculations (U.S. EPA, 2020c; IPCS, 2010). Kapraun et al. (2020) implemented the model using R version 3.6.1 (R Core Team, 2019) and MCSim (Bois, 2009) and applied the quality assurance guidelines of U.S. EPA (2018d) to verify parameter values and various other aspects of the software implementation of the model. A complete set of model implementation files for the Kapraun et al. (2020) PBPK model are available through the U.S. EPA Environmental Dataset Gateway (https://doi.org/10.23719/1519044). When the skin compartments of the Kapraun et al. (2020) model are turned "off" (by setting the volumes and blood flow rates for those compartments to zero), that PBPK model is functionally equivalent to the PBPK model of Campbell et al. (2014). The Kapraun et al. (2020) model will be used for this assessment. Further details of this model can be found in Table C-2.

As discussed in the preceding paragraphs, the validity of the Morris (2013), Campbell et al. (2014) and Kapraun et al. (2020) models' complex nasal structures cannot be confirmed. The lack of validation data for URT sub-regions is an issue common to most CFD-PBPK models since measurement of regional tissue samples is technically challenging, and ongoing metabolism or volatilization of an inhaled gas from the tissue during collection and initial processing of tissue would confound any attempt to make such measurements.

Whenever model predictions of total URT uptake have been validated (as is the case for the Campbell et al. (2014) and Morris (2013) models), the primary remaining question is whether or not the model correctly predicts the fraction of uptake (and removal) assigned to each sub-region. As long as the regional model structures and parameters are consistent, or varied according to anatomical, biochemical, and physiological data, one can have reasonable confidence in the model predictions. If the model under-predicts uptake in one URT sub-region, it must over-predict uptake in another region in order to achieve the overall mass balance. It should be noted, however, that if the predicted differences in uptake between sub-regions are based on conservation of mass, anatomical features, and CFD predictions based on the anatomy, it is unlikely that predictions for two different regions would have significant errors in opposite directions. Thus, whenever total URT uptake has been validated using experimental data, CFD-PBPK model predictions for sub-regions of the URT can be assumed to be reasonably accurate. In some cases, Monte Carlo simulations have been performed with PBPK models to assess uncertainty and variability in dose metrics [e.g., in the IRIS Toxicological Review of Dichloromethane (U.S. EPA, 2011b)]. However, performing a Monte Carlo (MC) analysis with the Campbell et al. (2014) and Kapraun et al. (2020) PBPK models for naphthalene would be because the values used for parameters that describe the respiratory tract have only been defined for humans and rats of specific sizes (i.e., body masses) — the way these parameters vary for animals and humans with different body sizes has not been characterized.

Table C-2. Descriptive summary for the Kapraun et al. (2020) CFD-PBPK model

Study detail	Description/notes				
Author	Kapraun et al. (2020)				
Contact email	mkapraun.dustin@epa.gov				
Contact phone	919-541-4045				
Sponsor	U.S. EPA				
Model summary					
Species	Rat	Human			
Strain	F433	N/A			
Sex	Male and female				
Life stage	Adult				
Exposure routes	Inhalation	IV	Skin		
Tissue dosimetry	Blood	URT tissues			
Model evaluation					
Language	R and MCSim				
Code available	YES	Effort to recreate model			COMPLETE
Code received	YES	Effort to migrate to open software (R/MCSim)			COMPLETE
Structure evaluated	YES				
Math evaluated	YES				
Code evaluated	YES				
Available PK data	Time-course data for naphthalene concentrations in rat blood after single intravenous doses (Quick and Shuler, 1999); time-course data for naphthalene concentrations in rat blood following 6-hour inhalation exposures (NTP, 2000); rat upper respiratory tract extraction data at fixed inspiratory flow rates (Morris and Buckpitt, 2009); and time-course dermal penetration (tape strip) and blood concentration data following controlled dermal exposure in humans (Kim et al., 2006).				

1