*EPA/635/R-23/337*
**IRIS Assessment Protocol**
www.epa.gov/iris

# Protocol for the Nitrate and Nitrite IRIS Assessment (Oral) (Preliminary Assessment Materials)

[CASRN 14797-55-8 and 147-65-0]
[Sodium nitrate: CASRN 7631-99-4]
[Sodium nitrite: CASRN 7632-00-0]
[Potassium nitrate: CASRN 7757-79-1]
[Potassium nitrite: CASRN 7758-09-0]
[Ammonium nitrate: CASRN 6484-54-2]
[Calcium nitrate: CASRN 10124-37-5 (anhydrous); 13477-34-4 (tetrahydrate)]

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

# CONTENTS

# TABLES

# FIGURES

# ABBREVIATIONS

| | |
|---|---|
| ADME | absorption, distribution, metabolism, or elimination |
| BMD | benchmark dose |
| BMDL | benchmark dose lower confidence limit |
| BW3/4 | body-weight scaling to the 3/4 power |
| BMDS | Benchmark Dose Software |
| CAS | Chemical Abstracts Service |
| CASRN | Chemical Abstracts Service Registry Number |
| CERCLA | Comprehensive Environmental Response, Compensation, and Liability Act |
| CI | confidence interval |
| COI | conflict of interest |
| CPHEA | Center for Public Health and Environmental Assessment |
| EPA | Environmental Protection Agency |
| GLP | good laboratory practices |
| GRADE | Grading of Recommendations Assessment, Development and Evaluation |
| HAWC | Health Assessment Workspace Collaborative |
| HEC | human equivalent concentration |
| HERO | Health and Environmental Research Online |
| IAP | IRIS Assessment Plan |
| IPCS | International Programme on Chemical Safety |
| IRIS | Integrated Risk Information System |
| ITER | International Toxicity Estimates for Risk |
| 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 |
| NOAEL | no-observed-adverse-effect level |
| OCHP | Office of Children's Health Protection |
| OLEM | Office of Land and Emergency Management |
| ORD | Office of Research and Development |
| ORAU | Oak Ridge Associated Universities |
| OSF | oral slope factor |
| OW | Office of Water |
| PBPK | physiologically based pharmacokinetic |
| PECO | populations, exposures, comparators, outcomes |
| PK | pharmacokinetic |
| POD | point of departure |
| RfC | reference concentration |
| RfD | reference dose |
| ROBINS-I | Risk of Bias in Non-Randomized Studies of Interventions |
| SD | standard deviation |
| SE | standard error |
| SEM | systematic evidence map |
| UF | uncertainty factor |

# AUTHORS | CONTRIBUTORS | REVIEWERS

## Assessment Team

| | |
|---|---|
| Krista Christensen, Ph.D. (co-Assessment Manager) | EPA/ORD/CPHEA/CPAD |
| Alexandra Lee, Ph.D. (co-Assessment Manager) | |
| Roman Mezencev, Ph.D. | |
| Amanda Persad, Ph.D. | |
| Martha Powers, Ph.D. | |
| Rachel Shaffer, Ph.D. | |
| Todd Zurlinden, Ph.D. | |

## Contributors

| | |
|---|---|
| Brittany Schulz | Oak Ridge Associated Universities (ORAU) |
| Nicholas Barnett | Contractor |
| Jennifer Lees | |

## Executive Direction

| | |
|---|---|
| Wayne Cascio, M.D. (CPHEA Director) | EPA/ORD/CPHEA |
| Kay Holt, M.S. (CPHEA Deputy Director) | |
| Samantha Jones, Ph.D. (CPHEA Associate Director) | |
| Kristina Thayer, Ph.D. (CPAD Director) | |
| Steve Dutton, Ph.D. (HEEAD Director) | |
| Andrew Kraft, Ph.D. (CPAD Associate Director) | |
| Ravi Subramaniam, Ph.D (CPAD Senior Science Advisor) | |
| Paul White, (CPAD Senior Science Advisor) | |
| Elizabeth Radke, Ph.D. (Branch Chief) | |
| Andrew Hotchkiss, Ph.D. (Branch Chief) | |
| Janice Lee, Ph.D. (Branch Chief) | |
| Kathleen Newhouse, M.S. (Acting Branch Chief) | |
| Glenn Rice, Ph.D. (Branch Chief) | |
| Viktor Morozov, Ph.D. (Branch Chief) | |
| Vicki Soto, B.S. (Branch Chief) | |

## Production Team

| | |
|---|---|
| Maureen Johnson (CPHEA Webmaster) | EPA/ORD/CPHEA |
| Ryan Jones (HERO Director) | |
| Dahnish Shams (Production Team) | |
| Jessica Soto-Hernandez (Production Team) | |
| Samuel Thacker (HERO Team) | |
| Garland Waleko (Production Team) | |

# 1. INTRODUCTION

The Integrated Risk Information System (IRIS) Program is undertaking a reassessment of the health effects of nitrate and nitrite via the oral (ingestion) route of exposure. IRIS assessments provide high quality, publicly available hazard identification and dose-response analyses on chemicals to which the public might be exposed. These assessments are not regulations but provide an important source of toxicity information used by the Environmental Protection Agency (EPA), state and local health agencies, tribes, other federal agencies, and international health organizations.

An IRIS Assessment Plan (IAP) was presented at a public science meeting on September 27–28, 2017 (https://sab.epa.gov/ords/sab/f?p=100:19:3574465722633:::19:P19_ID:904) to seek input on the problem formulation components of the assessment plan. The 2017 IAP specified the EPA need for an assessment of nitrate/nitrite, described the objectives and specific aims of the assessment, provided draft PECO (populations, exposures, comparators, and outcomes) criteria, and described areas of scientific complexity. However, in April 2019 the nitrate/nitrite assessment was suspended due to changes in EPA leadership priorities for the IRIS Program (April 2019 IRIS Program Outlook). During the last nomination cycle, EPA's Office of Water (OW), Office of Children's Health Protection (OCHP), and Region 5 prioritized nitrate and nitrite for assessment by the IRIS Program. In June 2023, the assessment was added to the IRIS Program Outlook to address assessment needs of EPA's Offices and Regions. This assessment may also be used to support actions in other EPA program and regional offices and can inform efforts to address nitrate/nitrite by tribes, states, and international health agencies (see Section 2.2).

The Protocol document includes the IAP content, revised in response to public input and updated EPA scoping needs and presents the methods for conducting the systematic review and dose-response analysis for the assessment. While the IAP described *what* the assessment will cover, this Protocol describes *how* the assessment will be conducted (see Figure 1-1).

The systematic review methods described in this Protocol are based on the Office of Research and Development (ORD) Staff Standard Operating Procedures for Developing Integrated Risk Information System (IRIS) Assessments (Version 2.0, referred to as the "IRIS Handbook") (U.S. EPA, 2022). The methods presented in this Protocol reflect the information provided in the IRIS Handbook which incorporates adjustments made based on a November 2021 National Academy of Sciences, Engineering, and Medicine (NASEM) committee review of that version of the IRIS Handbook (NASEM, 2021; U.S. EPA, 2020a).

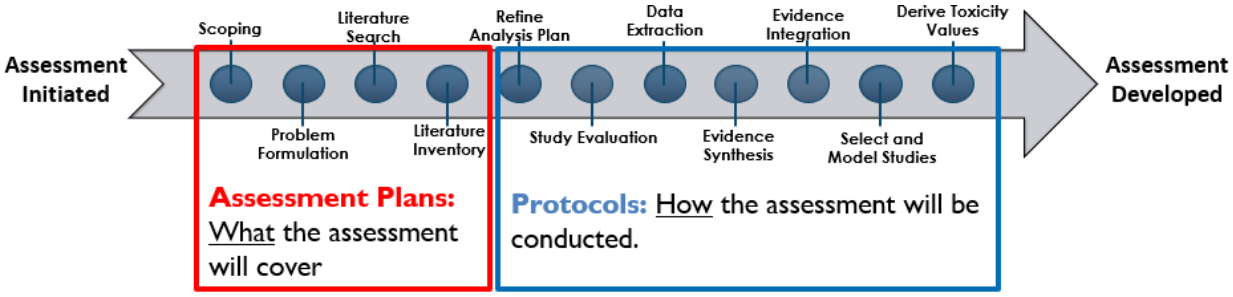



**Figure 1-1. Integrated Risk Information System (IRIS) systematic review problem formulation and method documents.**

# 2. SCOPING AND PROBLEM FORMULATION

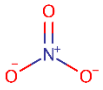
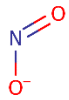
## 2.1. BACKGROUND

Below is a brief overview of aspects of the physiochemical properties, human exposure, and environmental fate characteristics of nitrate and nitrite (Chemical Abstract Services Registry Number [CASRN] 14797-55-8 and 147-65-0). This overview provides a summary of background information for contextual purposes only and is not intended to be comprehensive descriptions of the available information.

### 2.1.1. Physical and Chemical Properties

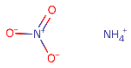
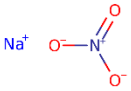
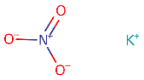
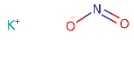
Inorganic nitrate ($NO_3-$) and nitrite ($NO_2-$) are naturally occurring anions formed by fixation of nitrogen and oxygen. Nitrate is a more stable form compared to nitrite although conversion between the two forms can readily occur through biological and chemical processes. Nitrite can also be converted to a class of compounds called N-nitrosamines. There are many organic and inorganic nitrate and nitrite compounds; for the purposes of this assessment the focus is on the following forms: potassium nitrate, potassium nitrite, sodium nitrate, sodium nitrite, ammonium nitrate, and calcium nitrate. Calcium nitrate was not included in the 2017 IAP but has been added to this Protocol based upon recommendation from EPA's OW. This group of inorganic compounds are highly water-soluble and readily dissociate. Selected chemical and physical properties of nitrate and nitrite are listed in Table 2-1 below, while properties of the nitrate and nitrite compounds of interest are listed in Table 2-2.

**Table 2-1. Physicochemical properties of nitrate and nitrite**

| Characteristic or property (unit) | Nitrate | Nitrite |
|---|---|---|
| Chemical structure |  |  |
| CASRN | 14797-55-8 | 14797-65-0 |
| EPA Chemicals Dashboard DTXSID | DTXSID5024217 | DTXSID5024219 |
| Synonyms | Nitrate; Nitric acid, ion(1−) | Nitrite; Nitrite ion; Nitrous acid, ion(1−) |
| Color/form | Varies by specific compound | Varies by specific compound |
| Molecular formula | $NO_{3\,(-)}$ | $NO_{2(-)}$ |
| Molecular weight (g/mol) | 62.005 | 46.006 |
| Log Kow | $4.05 \times 10^{-2}$ | $-5 \times 10^{-3}$ |

[a]U.S. EPA (2021) Chemicals Dashboard: https://comptox.epa.gov/dashboard/chemical/details/DTXSID5024217 and https://comptox.epa.gov/dashboard/chemical/details/DTXSID5024219 (accessed date October 14, 2022). Synonyms are those categorized as "valid" or "good" in the CompTox Chemicals Dashboard excluding foreign language synonyms and United Nation (UN) numbers. Median or average experimental values are used when available; otherwise, median, or average predicted values are used.

**Table 2-2. Physicochemical properties of selected nitrate and nitrite compounds**

| Characteristic or property (unit) | Calcium nitrate | Ammonium nitrate | Sodium nitrate | Sodium nitrite | Potassium nitrate | Potassium nitrite |
|---|---|---|---|---|---|---|
| Chemical structure |  |  |  |  |  |  |
| CASRN | 13477-34-4 (anhydrous) 10124-37-5 (tetrahydrate) | 6484-52-2 | 7631-99-4 | 7632-00-0 | 7757-79-1 | 7758-09-0 |
| EPA Chemicals Dashboard DTXSID | DTXSID1039719 | DTXSID2029688 | DTXSID6020937 | DTXSID0020941 | DTXSID4029692 | DTXSID5042320 |
| Synonyms | Nitric acid, calcium salt | Nitric acid, ammonium salt | Nitric acid, sodium salt | Nitrous acid, sodium salt | Nitric acid, potassium salt | Nitrous acid, potassium salt |

| Characteristic or property (unit) | Calcium nitrate | Ammonium nitrate | Sodium nitrate | Sodium nitrite | Potassium nitrate | Potassium nitrite |
|---|---|---|---|---|---|---|
| | Alternative names: Calcium Dinitrate; Lime nitrate; Norge saltpeter; Norwegian saltpeter; Calcium saltpeter | Alternative Names: Ammonium nitrate; Emulite; EXP 200; German saltpeter; Norway saltpeter; Norge saltpeter; Norwegian saltpeter; Plenco 12203; Varioform I; ZhVK | Alternative Names: Chile saltpeter; Niter; Nitric acid sodium salt; Saltpeter; Soda niter; Nitrate of soda; Cubic niter; Nitratine | Alternative Names: Nitrous acid soda; Nitrous acid sodium salt | Alternative Names: Niter; Nitre; Nitric acid potassium salt; Saltpeter; Saltpetre; Nitrate of potash | Alternative Names: Chile saltpeter; Niter; Nitric acid sodium salt; Salpeter; Soda niter |
| Color/form | White to light gray; Solid | White, colorless, gray, or brown; Solid | White or colorless; Solid | White to pale yellow; Solid | Colorless; Solid | Pale yellow; Solid |
| Molecular formula | $Ca(NO_3)_2$ | $NH_4NO_3$ | $NaNO_3$ | $NaNO_2$ | $KNO_3$ | $KNO_2$ |
| Molecular weight (g/mol) | 164.09 | 80.04 | 84.99 | 68.99 | 101.10 | 85.10 |
| Boiling point (°C) | 142 | 210 | 380 | 320 | 400 | 537 |
| Melting point (°C) | 43 (tetrahydrate) 561 (anhydrous) | 169.7 | 306 | 271 | 334 | 440 |

[a]U.S. EPA (2021) Chemicals Dashboard: https://comptox.epa.gov/dashboard/chemical/details/DTXSID6020937 (sodium nitrate); https://comptox.epa.gov/dashboard/chemical/hazard/DTXSID0020941 (sodium nitrite); https://comptox.epa.gov/dashboard/chemical/details/DTXSID4029692 (potassium nitrate); https://comptox.epa.gov/dashboard/chemical/hazard/DTXSID5042320 (potassium nitrite); https://comptox.epa.gov/dashboard/chemical/hazard/DTXSID2029668 (ammonium nitrate); https://comptox.epa.gov/dashboard/chemical/hazard/DTXSID1039719 (calcium nitrate) (accessed date October 14, 2022).

Synonyms are those categorized as "valid" or "good" in the CompTox Chemicals Dashboard excluding foreign language synonyms and United Nation (UN) numbers. Median or average experimental values are used when available; otherwise, median, or average predicted values are used.

### 2.1.2. Sources, Production, and Use

Nitrate and nitrite play an essential role in Earth's nitrogen cycle. Since 1950, human sources of reactive nitrogen into the environment—released either intentionally (e.g., through fertilizer application) or unintentionally (e.g., as a byproduct of fossil fuel combustion)—have

increased substantially (Fields, 2004). Nitrate salts are mainly used as nitrogen fertilizers and in industrial explosives, fireworks, and glass making; nitrites are largely used as preservatives for meat and fish curing and as color fixatives (IARC, 2010; Pokorny L, 2006). Nonpoint and point sources of nitrate/nitrite include animal waste, urban and agricultural runoff, landfill leachate, storm sewer overflow, vehicle exhaust, septic-system effluent, industrial processes, and industrial or mining wastewater (ATSDR, 2017; Bryan and Loscalzo, 2011; IARC, 2010; Pokorny L, 2006),

### 2.1.3. Environmental Fate and Transport

Nitrates account for most of the available total nitrogen in both ground and surface waters; nitrite levels are generally low in both (Desimone, 2009). According to monitoring data obtained during EPA's third Six-Year Review of National Primary Drinking Water Regulations (U.S. EPA, 2016), nitrate and nitrite were detected in approximately 63.8% and 11.7% of drinking water systems, respectively. The 5th to 95th percentile ranges of detected concentrations for nitrate and nitrite were 84–8,339 µg/L, and 2–1,150 µg/L, respectively (See exhibit 6-1 in the Occurrence Support Document, (U.S. EPA, 2016)). Human activities are responsible for increased levels of nitrate in drinking water sources; (Desimone, 2009) reported that nitrate concentrations greater than 1 mg/L (as N) are levels "considered to result from the effects of human activities in many parts of the United States" and that this level was exceeded in 41.4% of wells surveyed. Populations served by private well water, especially shallow wells in agricultural areas, may be exposed to nitrate at levels several times higher than those served by public water systems (Desimone, 2009; Ward, 2009).

### 2.1.4. Potential for Human Exposure and Populations with Potentially Greater Exposure

The general population is exposed to nitrate in both drinking water and food. Vegetables are the main source of ingested nitrate, with leafy vegetables comprising nearly 80% of nitrate exposure in an average person's diet. Other sources of dietary nitrate include cured meats/fish, cereal grains, dairy products, and beer (ATSDR, 2017; IARC, 2010). In contrast to nitrates, endogenous sources account for approximately 80% of all nitrites in the human body, as 5%–8% of the total nitrate intake is converted into nitrite (WHO, 2016; Mensinga et al., 2003). Almost all exogenous exposure to nitrite comes from food, with relatively higher nitrite concentrations found in cured meats (IARC, 2010). Drinking water is generally a minor source of exposure to nitrite (IARC, 2010).

Populations with potentially greater than average exposures include those living in agricultural areas, users of private well water systems, and those with diets high in concentrations of nitrate/nitrite. Agricultural areas have some of the highest concentrations of nitrates/nitrites in soil, surface, and groundwater in the United States. Populations using private wells tend to be those living in and around these more rural, agricultural areas, where nitrate levels in well water are several times higher than those found in public water systems (Ward, 2009). According to the U.S. Geological Survey (USGS), in 2015 approximately 42.5 million people, or 13% of the U.S. population,

depended on private wells as their main source of drinking water (Hutson, 2004). According to a study of sampled private wells across the United States conducted by the U.S. Geological Survey (USGS) from 1991 to 2004, approximately 4% of all private wells and 25% of private wells in agricultural areas contained levels above the maximum contaminant level (MCL) for nitrates (Desimone, 2009).

## 2.2. SCOPING SUMMARY

During scoping, the IRIS program meets with EPA program and regional offices that have interest in an IRIS assessment for nitrate and nitrite to discuss specific assessment needs. Table 2-3 provides a summary of input from this outreach.

**Table 2-3. EPA program and regional office interest in a reassessment of nitrate and nitrite**

| EPA program or regional office | Oral | Inhalation | Statutes/regulations | Anticipated uses/interest |
|---|---|---|---|---|
| Office of Water | ✓ | | Safe Drinking Water Act (SDWA) – Section 1412 | Six-year review of the National Primary Drinking Water regulations. |
| Region 5[a] | ✓ | | | Evaluation of special provision of the NPDW regulation [40 CFR 141.11(d)] allowing, at the discretion of the state, noncommunity water systems to exceed the nitrate MCL. |
| Office of Children's Health Protection | ✓ | | Executive Order 13045—Protection of Children from Environmental Health Risks and Safety Risks: Policy on Evaluating Health Risks to Children. | |

[a]Region 5 serves Illinois, Indiana, Michigan, Minnesota, Ohio, Wisconsin, and 35 tribes.

The EPA OW regulates nitrates and nitrites under the National Primary Drinking Water Regulations (40 CFR 141, 142); the current MCLs for nitrate and nitrite, promulgated in 1991, are 10 mg/L and 1 mg/L (as nitrogen), respectively (40 CFR 141.62; 56 FR 3594, January 30, 1991). An updated health assessment of nitrate and nitrite is being considered in the ongoing Six-Year Review cycle for National Primary Drinking Water Regulations. A provision of the current regulation [40 CFR 141.11(d)] allows, at the discretion of the state, noncommunity water systems to exceed the nitrate MCL up to 20 mg/L if the supplier can demonstrate that the water will not be available to children under 6 months of age and that no adverse health effects will result. The availability of more recent health effects literature published since 1991 raises questions about whether the

current MCLs for nitrate and nitrite and the provision allowing exceedance of the nitrate MCL up to 20 mg/L provide adequate health protection for the general population (all life stages).

As described above, this assessment will address inorganic forms of nitrate and nitrite and will specifically consider health effect information for the compounds included in Table 2-4. These salts are highly soluble in water and dissociate under environmental conditions; in solution, they exist as ions (ATSDR, 2017). Because the cations are not expected to introduce significant differences in the toxicity of the different salts, toxicity findings from all five compounds are considered relevant to an assessment of nitrate and nitrite toxicity ((EFSA), 2017b). These six compounds listed in Table 2-4 are the most common nitrate and nitrite salts in the environment (ATSDR, 2017). These compounds (except for calcium nitrate) were also the subject of two recent health assessments of nitrate and nitrite (ATSDR, 2017; IARC, 2010)). The decision to develop the assessment of nitrate/nitrite using health effect information for these six compounds was also based on known general population exposure to these six compounds and availability of epidemiological or toxicological information. Specifically, ammonium nitrate is a leading nitrogen fertilizer, and for this reason, has been used in toxicological studies as a component of "California mixture" and "Iowa mixture." These two mixtures are representative of groundwater contamination by fertilizers and pesticides and used for simulations of environmental exposures to pesticides mixtures. Calcium nitrate is similarly used as a fertilizer (Sellars and Nunes, 2021). Sodium nitrate, sodium nitrite, potassium nitrate, and potassium nitrite are used as food additives to cure meats. The National Toxicology Program (NTP) has assessed the toxicities of n-nitroso compounds (NTP, 2021) nitrate and sodium nitrite (NTP, 2001b) in animal toxicology and carcinogenicity studies.

**Table 2-4. Nitrate/nitrite compounds considered for assessment**

| Compound | Chemical formula | CAS Registry Number |
|---|---|---|
| Ammonium nitrate | $NH_4NO_3$ | 6484-52-2 |
| Calcium nitrate | $Ca(NO_3)_2$ | 10124-37-5 (anhydrous);<br>13477-34-4 (tetrahydrate) |
| Sodium nitrate | $NaNO_3$ | 7631-99-4 |
| Sodium nitrite | $NaNO_2$ | 7632-00-0 |
| Potassium nitrate | $KNO_3$ | 7757-79-1 |
| Potassium nitrite | $KNO_2$ | 7758-09-0 |

Assessment of the health effects of nitrate and nitrite following inhalation and dermal routes of exposure will not be included in the scope of this assessment. Inhalation and dermal exposures to nitrate or nitrite in the general population (i.e., populations not exposed occupationally, such as factory and fertilizer workers) are expected to be negligible compared to

oral exposure (ATSDR, 2017). Focusing on the health effects associated with oral exposure to nitrate and nitrite is consistent with the needs of EPA programs and regional offices.

Given input received during scoping, the IRIS assessment will include evaluation of noncancer and cancer human health hazards associated with ingested nitrate and nitrite. Although all health effects will be considered for hazard identification, the assessment will take a different approach for hematological outcomes. A hematological hazard has already been established through the known association between methemoglobinemia and nitrate/nitrite (Ward et al., 2005; Walton, 1951). Therefore, EPA will not re-consider the hematological domain during hazard identification. Instead, any new studies identified for methemoglobinemia and supporting hematological endpoints will be examined for information on the quantitative relationship with nitrate/nitrite and the potential to support dose-response analysis. For cancer, EPA will develop a qualitative assessment of the carcinogenic potential of nitrate and nitrite and will explore the feasibility of developing a quantitative assessment (for details, see Sections 8 and 9). EPA anticipates that a quantitative cancer assessment will be particularly challenging, given the co-occurrence of nitrosatable compounds and antioxidants in dietary sources, conflicting results across studies, and design limitations of epidemiological studies investigating the association between cancer and nitrate/nitrite exposure at different sites.

## 2.3. PROBLEM FORMULATION

The IRIS program currently does not include cancer risk values for nitrate or nitrite. The International Agency for Research on Cancer (IARC) has determined that there is "inadequate" evidence of carcinogenicity of nitrate in food or drinking water, "limited" evidence for the carcinogenicity of nitrite in food, and "sufficient" evidence for the carcinogenicity of nitrite in combination with amines or amides. IARC concludes that "ingested nitrate and nitrite under conditions that result in endogenous nitrosation is probably carcinogenic to humans (Group 2A)" (IARC, 2010).

The IRIS program lists reference dose (RfD) values of 1.6 mg/kg-day for nitrate and 0.1 mg/kd-g-day for nitrite, based on a critical effect of methemoglobinemia. Agency for Toxic Substances and Disease Registry (ATSDR) has determined minimal risk levels of 4 mg/kg-day for nitrate and 0.1 mg/kg-day for nitrite (applicable for acute, intermediate, and chronic durations of oral exposure) based upon the same health endpoint (ATSDR, 2017). The Joint FAO/WHO Expert Committee on Food Additives (JECFA) has also determined acceptable daily intake values of 3.7 mg/kg-day for nitrate and 0.07 mg/kg-day for nitrite (based on heart and lung effects in rats) (WHO, 2003; JECFA, 1995).

EPA's MCLs for nitrate and nitrite are 10 mg/L (or ppm) and 1 mg/L (or ppm), respectively. These are equivalent to ~44 mg nitrate/L as nitrate-nitrogen and ~3.3 mg nitrite/L as nitrite-nitrogen. California's Office of Environmental Health Hazard Assessment lists public health goals (PHGs) of 45 mg/L and 3 mg/L for nitrate and nitrite, respectively (the joint nitrate/nitrite PHG is

10 mg/L) (CalEPA, 2018). The FDA uses these same values for allowable levels in bottled water (FDA, 2021), and these are also the same values that Health Canada has determined for maximum allowable concentration values (Water and Air Quality Bureau, 2013).

Federal agencies (OSHA, NIOSH, ATSDR, EPA) have not set legal or recommended limits for nitrate or nitrite in air, largely due to lack of adequate data. A May 2023 summary of existing human health reference values for oral exposure to nitrates/nitrites is provided in Figure 2-1. See Appendix A (Table A-1) for a tabular summary, including derivation details of the displayed values.

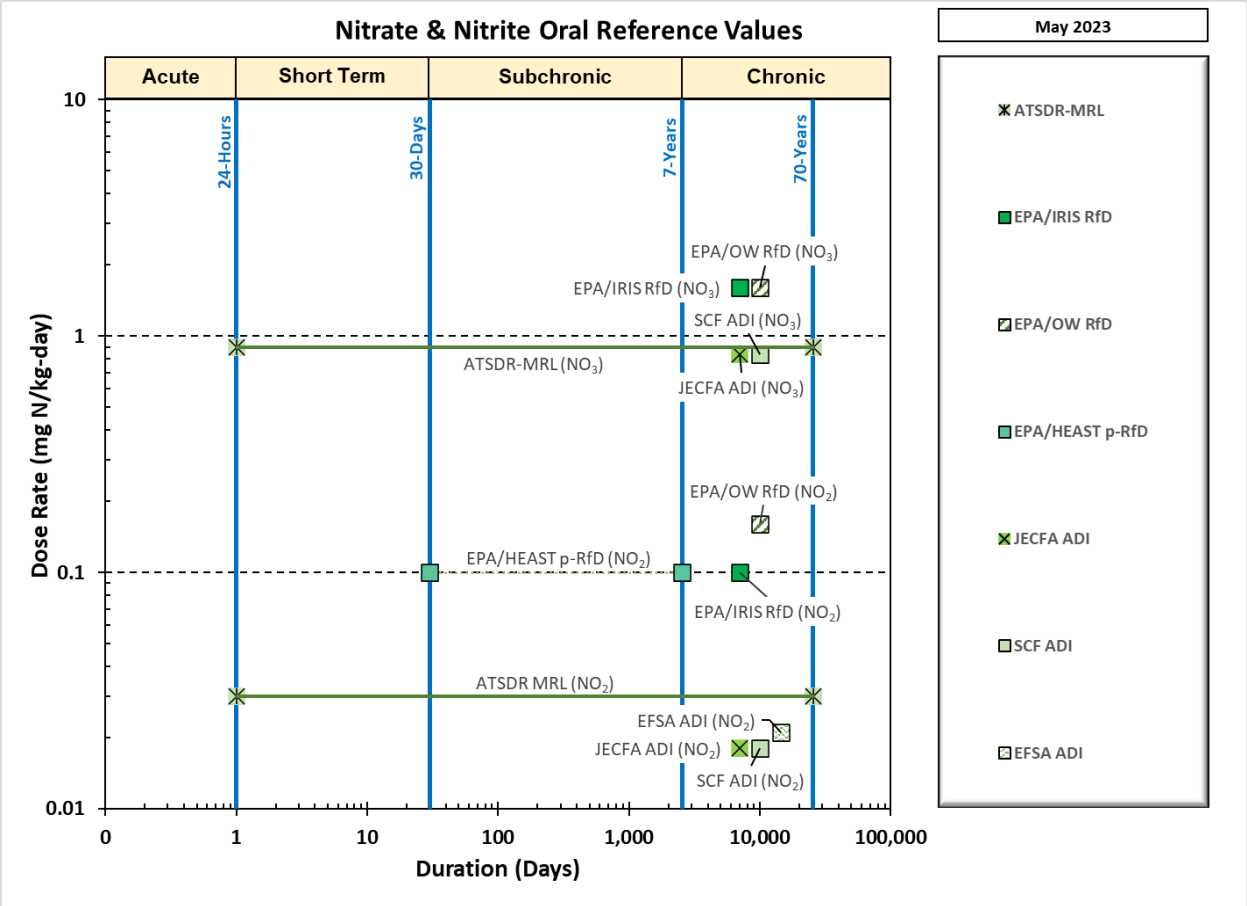



**Figure 2-1. Available health effect reference values for oral exposure to nitrate and nitrite.**

To identify noncancer and cancer health outcomes for which possible association with exposure to nitrate/nitrite has been investigated, a preliminary literature survey was performed using health assessments produced by other federal, state, and international health agencies (CalEPA, 2018; ATSDR, 2017; WHO, 2016; Water and Air Quality Bureau, 2013; IARC, 2010; IPCS, 2005). In particular, EPA relied on the ATSDR *Toxicological Profile for Nitrate and Nitrite* (ATSDR, 2017), as the most recent authoritative health agency assessment, to identify the pertinent health effect literature through 2016. ATSDR (ATSDR, 2017) updated the comprehensive review of the

cancer epidemiological literature provided in IARC (IARC, 2010) (i.e., literature published up to approximately 2007), and the IARC monograph also was used to identify the cancer literature. To identify studies published since the end of the period covered by the ATSDR Toxicological Profile (i.e., from 2016 to 2022), a literature search update was performed by EPA. The search strategy and literature screening are described further in Sections 3, 4 and 5 and Appendices B through D. The details of the preliminary literature survey, also referred to as a systematic evidence map (SEM), are described in Appendix A.

The SEM revealed many randomized, controlled trial human studies reporting potential association between controlled nitrate/nitrite exposure and beneficial cardiovascular outcomes. Because IRIS assessments focus on the adverse effects associated with exposure to environmental chemicals, a systematic review of the potential beneficial outcomes to the cardiovascular system associated with the intake of nitrate or nitrite will not be included in this assessment but will be identified as potentially relevant supplementary material.

## 2.4. KEY SCIENCE ISSUES

The SEM identified the following key scientific issues and potential mode-of-action hypotheses as warranting evaluation in this assessment.

- Nitrate and nitrite are generated endogenously as part of the nitrate-nitrite-nitric oxide cycle that controls the availability of nitric oxide, which is a ubiquitous signaling molecule involved in the regulation of numerous physiological and pathological processes, including vasodilation, platelet activation, metabolic regulation, neurotransmission, and host defense (inflammation). The roles of endogenous versus exogenous nitrate and nitrite in toxicity, particularly methemoglobinemia in infants, have been debated in the scientific literature.

- Several susceptible populations and life stages have been identified for methemoglobinemia. These include infants under 6 months of age; individuals with higher-than-normal gastric pH; individuals with glucose-6-phosphate dehydrogenase or NADH (nicotinamide adenine dinucleotide (NAD) + hydrogen)-dependent methemoglobin reductase deficiency; individuals with diseases such as anemia, cardiovascular disease, lung disease, and sepsis; individuals with abnormal hemoglobin species including carboxyhemoglobin, sulfhemoglobin, and sickle hemoglobin.

- A physiologically based pharmacokinetic (PBPK) model structure for simulating the kinetics of methemoglobinemia formation after oral exposure to nitrate in adults is available (Zeilmaker et al., 2010; Zeilmaker et al., 1996). An updated parameterization of this model using recent human data (Lin et al., 2020) needs to be evaluated against the original model fit (Zeilmaker et al., 2010) for its potential to inform human variability in the dose-response assessment.

- Previously published assessments by Health Canada (Water and Air Quality Bureau, 2013), ATSDR (ATSDR, 2017), IARC (IARC, 2010), the California EPA (CalEPA, 2018) and the WHO (WHO, 2016) and newer animal and epidemiological studies published after 2014 raise the following issues related to cancer risk:

- Risk associated with intake of nitrates, nitrites, or both from cured meats, vegetables, and drinking water could differ because of co-occurrence with antioxidants (e.g., vitamin C, vitamin E) in vegetables, amines in fish and meats, and calcium in drinking water. Consequently, risks associated with dietary intake, intake through drinking water, and total intake may need to be assessed separately.

- There may be susceptible populations with increased cancer risk associated with intake of nitrate/nitrite due to increased exposure or intrinsic factors.

- Populations vary significantly in the ability to reduce salivary nitrate by oral bacteria (e.g., actinomyces and veilonella) (Bryan and Petrosino, 2017). For example, patients with migraines were shown to have higher abundance of nitrate, nitrite, and nitric oxide reductase genes in their oral bacterial metagenome (Gonzalez et al., 2016). In contrast, the use of antiseptic mouthwashes appears to deplete nitrate-reducing oral bacteria and affect some nitrite-mediated biological processes (Kapil et al., 2013). Individuals from some subgroups may be able to convert more nitrate to nitrite and consequently produce more carcinogenic n-nitroso derivatives.

# 3. OVERALL OBJECTIVES AND SPECIFIC AIMS

The overall objective of this assessment is to identify adverse health effects of nitrate and nitrite ingestion exposure and characterize exposure-response relationships for these effects to support development of toxicity values. This assessment will use systematic review methods to evaluate the epidemiological and toxicological literature, including consideration of relevant mechanistic evidence, for the specified forms of nitrate/nitrite. The assessment methods described in this Protocol utilize EPA guidelines.[1]

## 3.1. SPECIFIC AIMS

- To aid problem formulation, develop a SEM to identify epidemiological (i.e., human), toxicological (i.e., experimental animal), and supplemental literature pertinent to characterizing the health effects of ingestion exposure to nitrate and nitrite.
    - Epidemiological studies, toxicological studies, and PBPK models are identified for inclusion based on predefined PECO criteria. The problem formulation PECO used to develop the SEM is intended to identify the amount and type of evidence available to address a particular topic and is a useful scoping tool for health effects assessments (NASEM, 2021; Wolffe et al., 2019).
    - Supplemental material content includes mechanistic studies, including in vivo, in vitro, ex vivo, or in silico models; nonmammalian model systems; pharmacokinetic and absorption, distribution, metabolism, and excretion (ADME) studies; human exposure characteristics (no health outcome); human biomarker studies with a health outcome; mixture studies; non-ingestion routes of exposure; case studies or case series; records with no original data; and conference abstracts.
- Use the results of the SEM to (1) develop refined PECO criteria for the assessment (referred to as "assessment PECO"); (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 assessment PECO criteria.
- 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.

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

- Conduct data extraction (summarizing study methods and results) from epidemiological and animal toxicological studies that meet the 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 may 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 (strength of evidence) and evidence integration (weight of evidence) conclusions in an evidence profile table.

- As supported by the currently available evidence, derive chronic and subchronic oral reference doses (RfDs) and organ- or system-specific RfDs, and cancer oral slope factors (OSFs). Apply pharmacokinetic and dosimetry modeling (possibly including PBPK modeling) to account for interspecies differences, as appropriate. Characterize confidence in any toxicity values that are derived.

- Characterize uncertainties and identify key data gaps and research needs, such as limitations of the evidence base, 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 LITERATURE INVENTORY

The literature search and screening processes described in this Section were used to develop an SEM using the problem formulation PECO (see Section 4.1) and supplemental screening criteria (see Section 4.2) to guide the inclusion of studies. The resulting inventory of studies identified in the SEM was used to develop the assessment PECO criteria and identify priority analyses of supplemental material (described in Section 5). The initial literature search as well as all subsequent literature search updates use the same literature search and screening process, and therefore the literature inventory is continually updated with new studies as the assessment progresses.

## 4.1. POPULATIONS, COMPARATORS, EXPOSURES, OUTCOMES CRITERIA FOR THE SYSTEMATIC EVIDENCE MAP

PECO criteria are used to focus the assessment question(s), search terms, and inclusion criteria. To meet the PECO criteria a study must meet all PECO elements. The problem formulation PECO criteria used to develop the SEM were intentionally broad (see Table 4-1) to identify all the available evidence in humans and animal models.

During problem formulation, exposure to nitrates/nitrites from routes other than ingestion, were determined to be out of scope for this assessment. Studies of beneficial health effects were identified but not included in the study evaluation process since the focus of the assessment is on hazard identification and dose-response analysis for adverse health effects.

**Table 4-1. Problem formulation populations, exposures, comparators, outcomes (PECO) criteria for the nitrate and nitrite assessment**

| PECO element | Evidence |
|---|---|
| **Populations** | **Human:** Any population and lifestage (occupational or general population, including children and other sensitive populations).<br><br>**Animal**: Nonhuman mammalian animal species (whole organism) of any lifestage (including fetal, early postnatal, adolescents and adults) that are informative for human health risk assessment. |
| **Exposures** | **Human**: Any exposure to the nitrate/nitrite forms below via the oral route for any duration. Studies will also be included if biomarkers of exposure are evaluated (e.g., measured chemical or metabolite levels in tissues or bodily fluids) AND there is additional information to allow estimation/attribution of nitrate/nitrite ingestion (e.g., measures of nitrate/nitrite in environmental media). If there is no additional information, but the exposure route is unclear or likely from multiple routes, the study will be tagged as "potentially relevant supplemental material." Other exposure routes, such as those that are clearly inhalation or dermal, will be tracked during title and abstract screening and tagged as "potentially relevant supplemental material."<br><br>**Animal**: Any exposure to the nitrate/nitrite forms below. Studies involving exposures to mixtures will be included only if they include an experimental arm with exposure to the nitrate/nitrite forms below, alone. Other exposure routes, including inhalation or dermal, will be tracked during title and abstract as "potentially relevant supplemental material."<br><br>*Relevant forms of nitrate/nitrite*: Calcium nitrate, Ammonium nitrate, Potassium nitrate, Potassium nitrite, Sodium nitrate, Sodium nitrite. |
| **Comparators** | **Human**: A comparison or referent population with exposure to lower levels, no exposure, or exposure below detection limits; exposure for shorter periods of time; or cases versus controls; or a repeated measures design. Worker surveillance studies are considered to meet PECO criteria even if no statistical analyses using a referent group is presented. Case reports or case series of >3 people will be considered to meet PECO criteria, while case reports describing findings in 1–3 people will be tracked as "potentially relevant supplemental material."<br><br>**Animal**: A concurrent control group exposed to vehicle-only treatment and/or untreated control. The control could be a baseline measurement (e.g., acute toxicity studies of mortality) or a repeated measure design. |
| **Outcomes** | All health outcomes are considered relevant (i.e., both cancer and noncancer). In general, endpoints related to clinical diagnostic criteria, disease outcomes, biochemical, histopathological examination, or other apical/phenotypic outcomes are considered to meet PECO criteria. We continue to include relevant studies of methemoglobinemia even though, for this outcome, the hazard is established. However, the focus is on studies that inform quantitative dose-response relationships. |

## 4.2. SUPPLEMENTAL CONTENT SCREENING CRITERIA

During the literature screening process, studies containing information that may be 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 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 | Evidence |
|---|---|
| **Classical pharmacokinetic (PK) or physiologically based pharmacokinetic (PBPK) model studies** | **Classical pharmacokinetic or dosimetry model studies**: Classical PK or dosimetry modeling usually divides the body into just one or two compartments, which are not specified by physiology, wherein movement of a chemical into, between, and out of the compartments is quantified empirically by fitting model parameters to absorption, distribution, metabolism, and excretion (ADME) data. This category is for papers that provide detailed descriptions of PK models that are not physiologically based PK (PBPK) models.<br>• The data are typically the concentration time course in blood or plasma after oral and or intravenous exposure, but other exposure routes can be described.<br>• A classical PK model might be elaborated from the basic structure applied in standard PK software, for example to include dermal or inhalation exposure, or growth of body mass over time, but otherwise does not use specific tissue volumes or blood flow rates as model parameters.<br>• Such models can be used for extrapolation similar to PBPK models, although such use might be more limited.<br>Note: ADME studies often report classical PK parameters, such as bioavailability (fraction of an oral dose absorbed), volume of distribution, clearance rate, and/or half-life or half-lives. If a paper provides such results only in tables with minimal description of the underlying model or software (i.e., uses standard PK software without elaboration), including "noncompartmental analysis," it should only be listed as a supplemental material ADME study.<br><br>**Physiologically based pharmacokinetic or mechanistic dosimetry model studies**: PBPK models represent the body as various compartments (e.g., liver, lung, slowly perfused tissue, richly perfused tissue) to quantify the movement of chemicals or particles into and out of the body (compartments) by defined routes of exposure, metabolism, and elimination, and thereby estimate concentrations in blood or target tissues.<br>• Usually specific to humans or defined animal species; often a single model structure is calibrated for multiple species.<br>• Some mechanistic dosimetry models might not be compartmental PBPK models but predict dose to the body or specific regions or tissues based on mechanistic data, such as ventilation rate and airway geometry.<br>• A defining characteristic is that key parameters are determined from a substance's physicochemical parameters (e.g., particle size and distribution, octanol-water partition coefficient) and physiological parameters (e.g., ventilation rate, tissue volumes); that is, data that are independent of in vivo ADME data that are otherwise used to estimate model parameters. |

| Category | Evidence |
|---|---|
| | • Chemical-specific information on metabolism (e.g., $V_{max}$, $K_m$) or other molecular processes (e.g., protein binding) might be obtained by fitting the model to in vivo ADME data or determined from in vitro experiments and extrapolated to in vivo predictions.<br><br>Allow extrapolation between species, routes of exposure, or exposure durations and levels; that is, they do not just quantify ADME for specific experiments to which they have been fitted. |
| **Pharmacokinetic (ADME)** | Pharmacokinetic (ADME) studies are primarily controlled experiments in which 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.<br><br>• These data are used to estimate the amount absorbed (A), distributed to different organs (D), metabolized (M), and/or excreted (E) through urine, breath, or feces.<br><br>• The most informative studies involve measurements over time such that the initial increase and subsequent concentration decline is observed, preferably at multiple exposure levels.<br><br>• Data collected from multiple tissues or excreta at a single time point also inform distribution.<br><br>• 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).<br><br>• 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.<br><br>Note: Studies describing environmental fate and transport or metabolism in bacteria or model systems not applicable to humans or animals should not be tagged. |
| **Mechanistic** | Studies reporting measurements related to a health outcome that inform the biological or chemical events associated with phenotypic effects, in both mammalian and nonmammalian model systems, including in vitro, in vivo (by various routes of exposure), ex vivo, and in silico studies. Studies in which 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). |
| **Non-PECO animal model (i.e., nonmammalian systems)** | Studies reporting outcomes in animal models that meet the outcome criteria but do not meet the "P" in the PECO criteria. Depending on the endpoints measured in these studies, they can also provide mechanistic information (in these cases studies should also be tagged "mechanistic endpoints"). |
| **Non-PECO route of exposure** | Studies using routes of exposure that fall outside the PECO scope. |

| Category | Evidence |
|---|---|
| **Human exposure and biomonitoring (no health outcome)** | Exposure characteristic studies include data that are unrelated to toxicological endpoints, but which provide information on exposure sources or measurement properties of the environmental agent (e.g., demonstrate a biomarker of exposure). |
| **Biomarker studies for which exposure route is unknown and cannot be inferred** | Studies evaluate health effects in relation to biomarkers of nitrate and/or nitrite exposure (e.g., urinary, or salivary levels) without additional information to inform exposure via ingestion. |
| **Mixture studies** | Mixture studies that are not considered PECO relevant because they do not contain an exposure or treatment group assessing only the chemical of interest. This categorization generally does not apply to epidemiological studies in which the exposure source might be unclear. |
| **Case reports or case series** | Case reports describing health outcomes after exposure are tracked as potentially relevant supplemental information when the number of subjects is ≤3. |
| **Records with no original data** | Records that do not contain original data, such as other agency assessments, informative scientific literature reviews, editorials, or commentaries. |
| **Posters, conference abstracts, abstract-only** | Records that do not contain sufficient documentation to support study evaluation and data extraction. |

## 4.3. LITERATURE SEARCH STRATEGIES

### 4.3.1. Database Search Term Development

The database search terms focused only on the chemical names and CASRNs, limited to publication years 2016–2022 with the exception that no year limit was placed on the search for calcium nitrate as it was not considered in the earlier (2017) IAP.

### 4.3.2. Database Searches

The literature search focused on studies published after the period covered by the ATSDR Toxicological Profile (ATSDR, 2017), namely 1/1/2016 onward. This literature search was initially conducted in August 2022 and regular updates performed with the most recent update occurring in August 2022. The databases listed below are searched by an EPA information specialist and stored in the Health and Environmental Research Online (HERO)[2] database.

- PubMed (National Library of Medicine)
- Web of Science (WoS; Thomson Reuters); given the number of records identified from an initial WoS search, a more targeted WoS search strategy was used to identify the records most likely to be applicable to human health (see Appendix A)

After deduplication in HERO, records are imported into SWIFT Review software (Howard et al., 2016) to identify those references most likely to be applicable to a human health assessment. In brief, SWIFT Review has preset literature search strategies ("filters") developed and applied by information specialists to identify studies more likely to be useful for identifying human health content from those that likely are not (e.g., analytical methods). The filters function like a typical search strategy in which studies are tagged as belonging to a certain filter if the terms in the filter literature search strategy appear in title, abstract, keyword or medical subject headings (*MeSH)* fields content. The applied SWIFT Review filters focused on lines of evidence: human, animal models for human health, and in vitro studies. The details of the search strategies that underlie the filters are available online. Studies not retrieved using these filters are not considered further. Studies that included one or more of the search terms in the title, abstract, keyword, or *MeSH* fields are exported as a RIS (Research Information System) file for further screening as described below. The impact of application of the SWIFT evidence stream filters on the number of studies for title and abstract screening is presented in Appendix A.

The literature searches are updated throughout the assessment's development and review process to identify newly published literature. During this period the literature search terms do not change from that used in the initial search and studies are screened according to both the problem formulation and assessment PECO criteria. Thus, the literature inventory is updated during the

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

process of developing the draft assessment. The last full literature search update is conducted several months prior to the planned release of the draft document for public comment. Studies identified after peer review begins are only considered for inclusion if they are directly relevant to the assessment PECO criteria and are expected to fundamentally alter the draft assessment conclusions.

### 4.3.3. Searching Other Sources

For this assessment, the starting point is the 2017 ATSDR Toxicological Profile, thus the literature search aimed to identify studies published in 01/2016 or later. The literature search will be expanded in subsequent stages of assessment development, to identify any potentially missed studies from previous years (or published after the end of the current literature search timeframe) for the health effect categories selected for hazard characterization.

The literature search strategy described above was designed to be broad, but like any search strategy, studies can be missed [e.g., cases in which 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 database searches, the sources below are used to identify studies that could have been missed based on the database search. Searching of these resources occurs during preparation of the initial literature inventory when assembling the SEM. After preparation of the initial literature inventory, references can be identified during public comment periods, by technical consultants, and during peer review. Records that appear to meet the initial PECO criteria are uploaded into DistillerSR, annotated with respect to source of the record, and screened using the methods described in Section 4.4. Appendix D.1 describes the specific methods and results for searching the sources below. Searching of these sources is summarized to include the source type or name, the search string (when applicable), number of results present within the resource, and the URL (uniform resource locator, when available and applicable). The list of other sources consulted includes:

- Manual review (at the title level) of the reference list from other publicly available final or draft assessments from other non-EPA Agencies including studies published after 2015 (e.g., ATSDR Toxicological Profile) or published journal review specifically focused on human health. Reviews can be identified from the database search or from the resources listed in Appendix D.

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

- EPA ChemView database (U.S. EPA, 2019a) to identify unpublished studies, information submitted to EPA under Toxic Substances Control Act Section 4 (chemical testing results), Section 8(d) (health and safety studies), Section 8(e) (substantial risk of injury to health or the environment notices), and FYI (for your information, voluntary documents). Other

databases accessible via ChemView include EPA's High Production Volume Challenge database and the Toxic Release Inventory database.

- The NTP database of study results and research projects (https://ntp.niehs.nih.gov/data).
- The Organization for Economic Cooperation and Development Screening Information DataSet (SIDS) High Production Volume Chemicals (https://www.echemportal.org/echemportal/).
- The EPA CompTox (Computational Toxicology Program) Chemical Dashboard (U.S. EPA, 2019b) to retrieve a summary of any ToxCast or Tox21 high throughput screening information. This data will be evaluated and, if amenable, used to generate mechanistic insight, predict adverse outcome, and potentially inform dose-response modeling. Their importance for outcome prediction and dose-response modeling depends on the context, size and quality, and information value of retrieved results and the lack of availability of other data typically used for these purposes.
- The National Institute of Health Gene Expression Omnibus (GEO) (http://ncbi.nlm.nih.gov/geo/) and the European Bioinformatics Institute (EMBL-EBI) Array Express (http://ebi.ac.uk/biostudies/arrayexpress) repositories to retrieve functional genomics data from appropriate in vitro and in vivo studies. If available, this data will be evaluated and potentially used to generate mechanistic insight, predict adverse outcomes, and inform dose-response assessment.
- Review of the list of references in the ECOTOX database for the chemical(s) of interest.
- Comparative Toxicogenomics Database (CTDB), available at http://ctdbase.org/.
- References identified during public comment periods, by technical consultants, and during peer review.

#### 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 may be identified during assessment development. In these instances, the 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 when 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, 2015). This independent, contractor driven, peer review would include an evaluation of the study similar to that for peer review of a journal publication. The contractor would identify and 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 EPA, and the names of the peer reviewers would be identified. In certain cases, IRIS will assess the utility of a data 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 if the information is made publicly available. If such ancillary information is acquired, it is documented in the Health Assessment Workspace Collaborative (HAWC) or HERO project page (depending on the nature of the information received).

## 4.4. LITERATURE SCREENING

Records identified from the literature searches are housed in the HERO system and imported into SWIFT-Active Screener (https://www.sciome.com/swift-activescreener/) for an initial title abstract screen using machine learning followed by import into DistillerSR (Evidence Partners; https://distillercer.com/products/distillersr-systematic-review-software/) for full-text screening. Both title-and-abstract (TIAB) and full-text screening are conducted by two independent reviewers.

### 4.4.1. Title-and-Abstract Screening

The studies identified from the searches described above are imported into SWIFT-Active Screener (https://www.sciome.com/swift-activescreener/) for TIAB screening. SWIFT-Active Screener is a web-based collaborative software application that utilizes active machine learning approaches to reduce the screening effort (Howard et al., 2020). Following a pilot phase to calibrate screening guidance, two screeners independently perform a TIAB screen using a structured form. Studies considered "relevant" or "unclear" based on meeting all problem formulation PECO criteria at the TIAB level are considered for inclusion and advanced to full-text screening. TIAB screening is conducted by two independent reviewers and any screening conflicts are resolved by discussion between the primary screeners with consultation by a third reviewer, if needed. For citations with no abstract, articles are initially screened based on the following: title relevance (title should indicate clear relevance), and page length (articles two pages in length or less are assumed to be conference reports, editorials, or letters). Eligibility status of non-English studies is assessed using the same approach with online translation tools or engagement with a native speaker.

The machine learning screening process is designed to prioritize references that appear to meet the problem formulation PECO criteria or supplemental material content for manual review (i.e., both types of references are screened as "include" for machine learning purposes). Screening continues until SWIFT-Active Screener indicates that it was likely at least 95% of the relevant studies are identified, a percent identification often used to evaluate the performance of machine learning applications and considered comparable to human error rates (Bannach-Brown et al.,

2018; Howard et al., 2016; Cohen et al., 2006). Any studies with "partially screened" status at the time of reaching the 95% threshold are then fully screened. Studies identified as meeting the problem formulation PECO criteria, unclear, or supplemental material during TIAB screening are then imported into DistillerSR software (https://www.evidencepartners.com/products/distillersr-systematic-review-software/) either for conflict resolution or for an additional round of more specific TIAB tagging (i.e., to separate studies meeting PECO criteria versus supplemental content and to tag the evidence stream or specific type of supplemental content). In DistillerSR, TIAB screening is conducted by two independent reviewers and any screening conflicts resolved by discussion between the primary screeners with consultation by a third reviewer, if needed. Conflicts between screeners in applying the supplemental tags, which primarily occur at the TIAB level, are resolved similarly, erring on the side of over-tagging based on TIAB content. Note that more granular sub-tagging of supplemental material occurs during preparation of the literature inventory as described in Section 4.5.2.

### 4.4.2. Full-Text Screening

Full-text references are sought through the EPA's HERO database for studies screened as meeting the problem formulation PECO criteria or "unclear" based on the TIAB screening. Full-text screening occurs in Distiller SR. Full-text copies of these records are retrieved, stored in the HERO database, and independently assessed by two screeners using a structured form in DistillerSR to confirm eligibility. 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). Rationales for excluding studies are documented, e.g., study did not meet PECO, full-text not available. Approaches for language translation include online translation tools or engagement of a native speaker. 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 are multiple publications using the same or overlapping data, all publications are included, with one selected for use as the primary study; the others are considered as secondary publications with annotation in HAWC and HERO 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 typically the one with the longest duration of exposure, the largest sample size, or with the outcome(s) most informative to the initial PECO. For both epidemiology and animal studies, the assessments include relevant data from all publications of the study, although if the same data are reported in more than one study, the data are extracted only once (see Section 7). For corrections, retractions, and other companion documents to the included publications, a similar approach to annotation is taken and the most recently published data are incorporated into the assessments.

### 4.4.4. Literature Flow Diagrams

The results of the screening process are posted on the project page for the assessment in the HERO database (https://heronet.epa.gov/heronet/index.cfm/project/page/project_id/2367). Results are also summarized in a literature flow diagram and interactive HAWC literature trees (where additional sub-tagging beyond what is presented in HERO is documented and visualized, e.g., more details on the nature of mechanistic or ADME studies).

## 4.5. LITERATURE INVENTORY

During TIAB or full-text level screening, studies that meet the problem formulation PECO criteria are categorized by evidence type (human or animal) or category of supplemental information (e.g., mechanistic, PB the ADME properties are dynamic). 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 is 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.

### 4.5.1. Studies That Meet Problem Formulation PECO Criteria

Human and animal studies that met the problem formulation PECO criteria after TIAB and full-text review are briefly summarized using data extraction forms in HAWC (hawc.epa.gov; see Figure 4-1). The literature inventories are used to inform the assessment PECO criteria and assessment approach. More detail on the process of summarizing studies is presented in Section 7 (Data Extraction of Study Methods and Results).

### 4.5.2. Organizational Approach for Supplemental Material

The results of the supplemental material tagging conducted in DistillerSR are imported into the literature review module in HAWC, where more granular sub-tagging within a type of supplemental material content category can be conducted. A single study can have multiple tags. The degree of sub-tagging depends on the extent of content for a given type of supplemental material and needs of the assessment with respect to developing human health hazard conclusions and derivation of toxicity values. Tagging judgments in HAWC are made by one assessment member and confirmed during preparation of draft assessment by another member of the assessment team. The overall tagging structure for supplemental material content is presented in Figure 4-1, with details on sub-tagging presented in the following sections under the specific type of supplemental content (i.e., mechanistic, ADME and PK/PBPK).

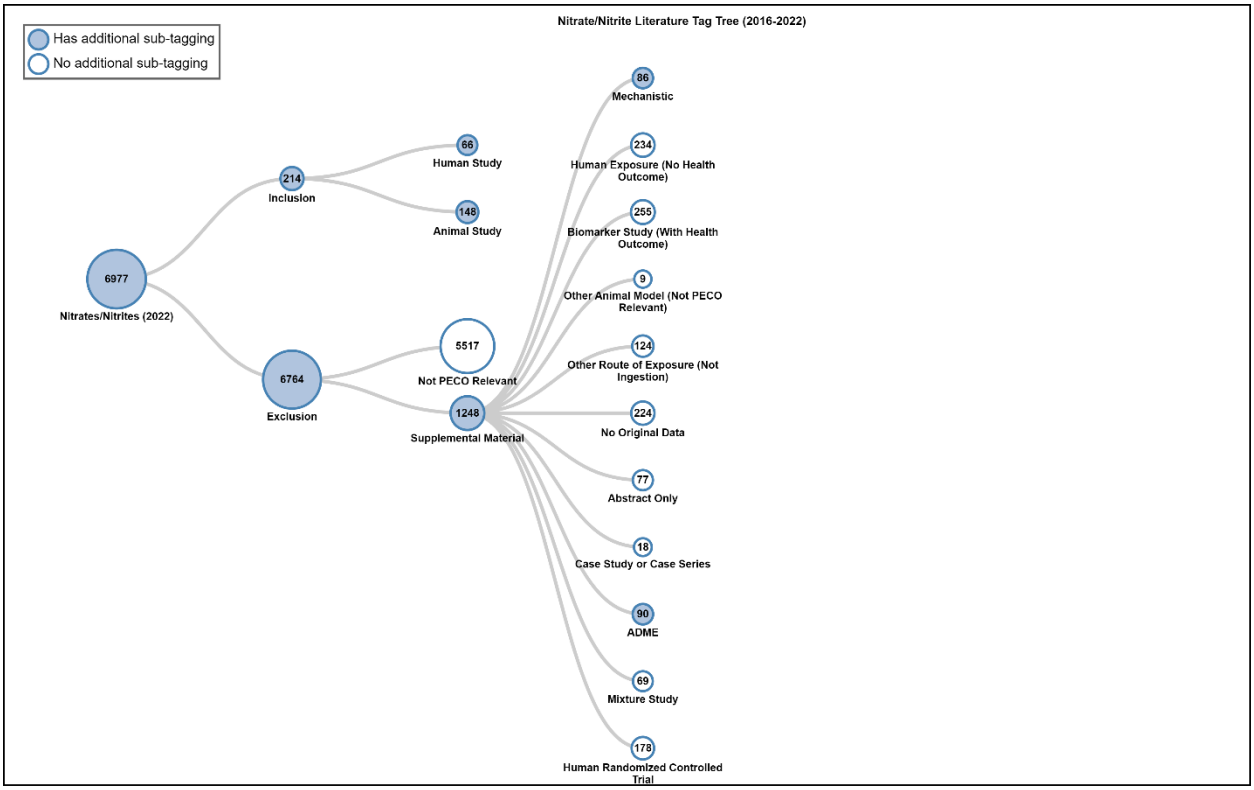



**Figure 4-1. Visual summary of approach for tagging major categories of supplemental material.**

### *Organization of Mechanistic Information*

The literature inventory of mechanistic information is used to develop the assessment approach (see Section 5), in particular to help assess whether any units of analysis should be defined to include mechanistic information and to identify prioritized analyses. The sub-tagging structure applied to mechanistic evidence is based on 10 specific mechanism pathways or events, listed below:

1) Mitochondrial function.
2) Inflammation
3) Oxidative and nitrosative stress
4) Genotoxicity
5) Nitrosation of amines/production of nitrosamines
6) S-Nitrosation
7) Generation of methemoglobin
8) Endothelial function

9) NO-mediated cell signaling

10) Modulation of enzyme activity

Figure 4-2 illustrates how the categories are represented in the overall tagging approach for mechanistic information.

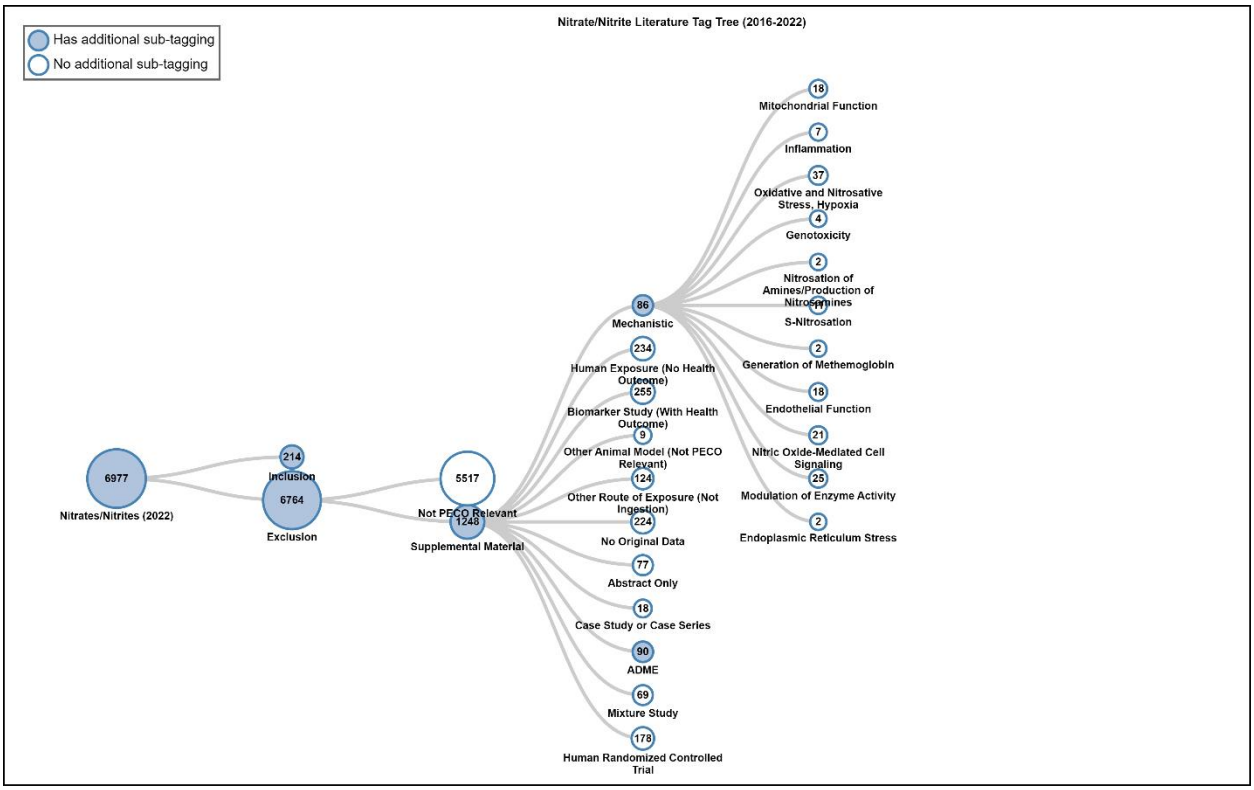



**Figure 4-2. Visual summary of overall tagging structure for mechanistic studies.**

### *Organization of ADME and PK/PBPK Model Information*

ADME and PK/PBPK model evidence are tagged as supplemental material in DistillerSR as outlined in Table 4-2. Tagged ADME studies and PK/PBPK models were imported into the HAWC Literature Review module and underwent more detailed tagging by disciplinary experts. Primary data ADME studies are tagged as absorption, distribution, metabolism, or elimination (using a tag all that apply approach). PK/PBPK models are tagged according to species applicability, i.e., animal, human, or multiple species (to include human). See Figure 4-3 for organizational structure.

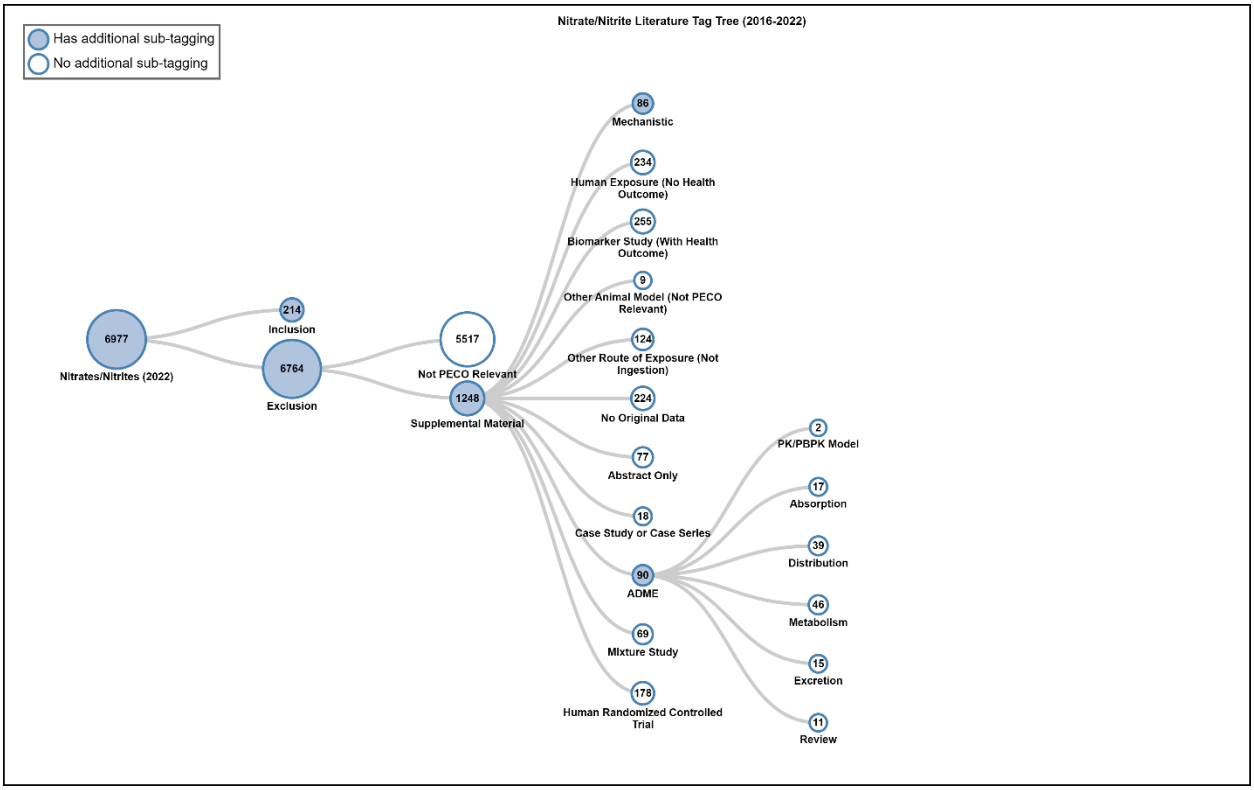



**Figure 4-3. Visual summary of tagging structure for ADME and PK/PBPK studies.**

# 5. REFINED PROBLEM FORMULATION AND ASSESSMENT APPROACH

## 5.1. ASSESSMENT PECO CRITERIA

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 or 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). In some assessments, the units of analysis may include predefined categories of mechanistic evidence (e.g., biomarkers or precursors relating to other outcomes within the unit of analysis, evidence that provides support for grouping together biologically linked endpoints into a unit of analysis).

Based on the results of the initial literature inventories for the SEM, the problem formulation PECO criteria were refined to the "assessment" PECO criteria (see Table 5-1). The assessment PECO criteria 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 literature search identified 178 human randomized control trials; since these studies concerned protective health effects of nitrate/nitrite exposure, they will be considered as supplemental material. Note that there were no studies identified during the primary literature search that evaluated hazards of oral exposure to calcium nitrate.

The systematic review will focus on the health outcome categories that appear to have sufficient information available to support hazard identification, based upon the availability of animal and human studies as cited in ATSDR (ATSDR, 2017) and IARC (IARC, 2010), and the updated literature search conducted by EPA. EPA anticipates conducting a systematic review for the following health effect categories, for which the available epidemiology and experimental animal studies are likely to be sufficient for drawing conclusions about human hazard:

**Cancer**

ATSDR concluded that "*In general, outcomes of cohort and case-control studies have found no or weak associations between nitrate intakes and cancer in humans, with stronger associations for exposures to nitrite or intake of high-nitrite foods such as cured meat*" and that "*Associations between intake of nitrite and a variety of cancer types has been studied; however, the strongest and most consistent evidence for carcinogenicity of nitrite derives from studies of gastrointestinal cancers and,*

*in particular, gastric cancer (Buiatti et al. 1990; Engel et al. 2003; La Vecchia et al. 1994, 1997; Mayne et al. 2001; Palli et al. 2001; Risch et al. 1985; Rogers et al. 1995; Ward et al. 2007, 2008). In general, these studies have found significant positive trends for cancer risk (risk increases with increasing intake), and three studies found elevated cancer risk (Engel et al. 2003; Kim et al. 2007; Risch et al. 1985)*." Since the conclusion of the ATSDR literature search period, 28 human epidemiology studies have been published that evaluate associations between nitrate/nitrite in water and diet with cancer at various sites (including eight studies of colorectal cancer, with smaller numbers of studies evaluating other cancer types). The human cancer studies are supported by two animal studies evaluating neoplasms in rodent models (one study concerns colorectal cancer, the other multiple cancers).

**Cardiovascular effects**

ATSDR found few studies evaluating risk of cardiovascular effects, reflected by the conclusion in the 2017 IAP. However, EPA's updated literature search identified 13 total new human epidemiology studies evaluating endpoints including cardiovascular disease mortality (5 studies) as well as cardiovascular and cerebrovascular disease (6 studies) and blood pressure (6 studies). In addition, a large number (48 studies) of toxicology studies were found to have evaluated cardiovascular endpoints.

**Developmental effects**

ATSDR identified several studies of developmental effects following nitrate/nitrite exposure in early life (including in utero), with many of the human studies focusing on risk of congenital malformation. However, they note that "*Several population-based, case-control studies evaluated possible associations between developmental end points and exposure to nitrate from drinking water sources. The results are not adequate for quantitative risk assessment because estimations of nitrate intakes were typically based on measurements of nitrate levels in drinking water sources at selected time points and self-reported estimates of water consumption, possible confounding by other potential toxicants was not evaluated, and most studies did not account for dietary nitrate or nitrite intake which is typically the major source of ingested nitrate and nitrite. Statistically significant associations between nitrate in the drinking water and selected developmental end points (e.g., birth defects, spontaneous abortions) were reported by some investigators, but were not observed by others."* EPA found that since the conclusion of the ATSDR search, six new human studies have been published including two studies evaluating birth defects, three evaluating measures of early life size and growth and one evaluating offspring mortality. The new body of human studies is complemented by three new toxicology studies.

**Endocrine effects**

Nitrate is a competitive inhibitor of the sodium iodide symporter, therefore endocrine effects due to nitrate/nitrite exposure are of concern. ATSDR noted that "*Available human data provide suggestive evidence that elevated levels of nitrate in drinking water and/or nitrate-rich diets may be associated with signs of thyroid dysfunction. However, limitations of these studies include lack of individual dose-response data, quantification of iodine intake, and control for other potential substances that may affect the thyroid; one study relied on self-reported thyroid status and self-reported dietary nitrate intake."* Since then, a small number of new human and animal studies (four human and two animal) have been published evaluating hypothyroidism and thyroid abnormalities in humans, and thyroid hormone levels and function in animals.

**Hematological effects**

All existing toxicity values have been based upon methemoglobinemia. EPA identified one new human study with this endpoint, which will be evaluated for its potential to support dose-response characterization. Additionally, 11 new animal studies have been published that evaluate both methemoglobin levels and other hematologic endpoints. While the hazard for hematological endpoints is considered well-established and will not be revisited, new studies evaluating methemoglobinemia and related endpoints will be considered for their potential to support dose-response evaluation.

**Hepatic effects**

ATSDR did not identify any human studies of hepatic effects and noted that the five animal studies identified did not show associations with nitrate/nitrite exposure. However, since that time, 22 new animal studies have been published evaluating a variety of endpoints including liver function biomarkers (such as alanine aminotransferase and aspartate aminotransferase) and liver histopathology. In addition, there has been one new human study evaluating mortality due to chronic liver disease.

**Metabolic effects**

ATSDR identified a number of human studies (but few animal studies) evaluating metabolic effects, noting that "*Possible associations between nitrate and/or nitrite in drinking water and/or food sources and risk of type 1 diabetes have been investigated in a number of epidemiological studies (Casu et al. 2000; Dahlquist et al. 1990; Kostraba et al. 1992; Moltchanova et al. 2004; Parslow et al. 1997; van Maanen et al. 2000; Zhao et al. 2001). Statistically significant associations between estimated nitrate and/or nitrite intake were reported by some investigators but were not observed by others. Limitations of studies include the lack of quantitative dose-response data and the likelihood of confounding by other potential toxicants. Therefore, there is considerable uncertainty regarding*

*nitrate or nitrite intake and risk of type 1 childhood diabetes*." EPA identified one new study in humans evaluating type 1 diabetes; two other human studies evaluating type 2 diabetes; and metabolic dysfunction and one study evaluating mortality due to diabetes mellitus. However, a large number (50) of new animal studies have measured a variety of endpoints related to lipid levels, insulin, and glucose homeostasis that may inform human health risk for endocrine outcomes.

**Nervous system effects**

ATSDR identified few studies of nervous system effects (two human studies reporting headache, and three animal studies). However, since that time three new human studies have been published that evaluated nervous system effects in both adolescents (depressive symptoms) as well as in middle-aged and older adults (cognitive function, mortality due to Alzheimer's disease. In addition, seven new animal studies have evaluated endpoints, including tremor, sensory endpoints, learning, and memory.

**Reproductive effects**

Much of the evidence identified in the ATSDR Toxicological Profile is described under developmental effects. However, EPA identified new human studies that evaluated time to pregnancy (one study) or gestation duration (five studies), and 12 new animal studies that evaluated reproductive endpoints in both male and female animals including reproductive hormone levels, reproductive organ histopathology and fertility.

**Urinary effects**

The ATSDR Toxicological Profile did not find any human studies, and only one animal study evaluating urinary system effects. However, EPA identified one new human study that evaluated risk of chronic kidney disease and one new human study evaluating mortality due to kidney disease. A larger number (14) of new animal studies have evaluated urinary system effects, mainly kidney function and histopathology.

**Other health effect categories (not considered further)**

The health effect categories listed in this section are those for which the ATSDR Toxicological Profile found limited or no epidemiological or toxicological evidence. Further, EPA's updated literature search identified no new substantial evidence. Primarily, these studies investigated health protective effects, which is outside the scope of this assessment. Several also only administered nitrate/nitrite with the purpose of inducing toxicity, using doses high enough to be of limited use to generalizing dose-response analysis to target populations. Therefore, none of the following categories will be carried forward for hazard evaluation based on the literature available in 2022, although new evidence may be identified with future literature search updates:

*Dermal effects*: ATSDR only identified one case study; there were no new human studies and only one new animal study. This study only administered nitrate/nitrite with the purpose of inducing toxicity and had limited results reporting.

*Gastrointestinal effects*: ATSDR identified one human study of acid reflux and a few animal studies of forestomach epithelial hyperplasia. There was one new human study evaluating risk of diarrheal disease, and 11 new animal studies. The human study was ecological in design, correlating diarrheal disease case counts with nitrate levels in water samples (no associations found (Kulinkina et al., 2016)). Most of the animal studies examined protective effects of nitrate/nitrite and one only administered nitrate/nitrite with the purpose of inducing toxicity.

*Immune effects*: ATSDR did not identify any animal or human studies. There was one new human study evaluating risk of diarrheal disease (discussed above (Kulinkina et al., 2016)), one new human study evaluating type 1 diabetes and islet autoimmunity ((Mattila et al., 2020) included under metabolic effects), and one human study evaluating mortality due to infection. The bulk of the new animal studies (n = 25) evaluated cytokine levels, markers of inflammation and oxidative stress, or white blood cell counts (classified as hematological endpoints). One new animal study examined metabolic islet autoimmunity (included under metabolic effects) and one animal study evaluated beneficial effect of nitrate in an animal model of colitis (included under gastrointestinal effects).

*Musculoskeletal effects*: ATSDR did not identify any animal or human studies. There were two new human studies evaluating muscle function (EPA inventoried these as 'whole body' effects), and 10 new animal studies—however, these focused on identifying beneficial effects of exposure to nitrate/nitrite rather than hazard. Both human studies found protective effects. Three of the animal studies found non-conclusive or adverse effects, but the rest provided evidence of therapeutic effects of nitrate/nitrite.

*Ocular effects*: ATSDR did not identify any animal or human studies. There were three new human studies evaluating risk of glaucoma (two studies) and retinal microvasculature (one study), and one animal study evaluating features of macular degeneration. Each human study investigated and found protective effects of nitrate/nitrite against ocular disease. The animal study looked for any impact of nitrate/nitrite and found evidence of adverse effects, though at relatively high doses of exposure.

*Respiratory effects*: ATSDR did not identify any animal or human studies. One new human study evaluated mortality from respiratory disease, and three new animal studies evaluated pulmonary function or histopathology. Of the animal studies, two investigated and found health protective effects. The third only administered nitrate/nitrite at a very high dose with the purpose of inducing toxicity.

As noted in Section 4, the literature inventory initially developed using the problem formulation. PECO is continually updated as the assessment progresses in part to ensure that emerging areas of potential health concern are monitored. The only adjustment made to the approaches used to tag supplemental material presented in Table 4-2 was the addition of human randomized controlled trials as a supplemental material category.

**Table 5-1. Assessment PECO criteria for the nitrate/nitrite (oral) assessment**

| PECO element | Evidence |
|---|---|
| **Populations** | **Human:** Any population and lifestage (occupational or general population, including children and other sensitive populations).<br><br>**Animal:** Nonhuman mammalian animal species (whole organism) of any lifestage (including fetal, early postnatal, adolescents and adults) that are informative for human health risk assessment.<br><br>*Examples:*<br>• PECO-relevant: humans and laboratory animals, such as mice, rats, guinea pigs, monkeys, hamsters, dogs, etc.<br>• Supplemental: zebrafish in developmental studies, hens in neurotoxicology studies, frog embryos for teratogenicity; in vitro assays will be tagged as "mechanistic."<br>• Not PECO-relevant: birds, trout, salmon, algae, seedlings, hens in feather growth; farm animals (especially multi-stomach animals) like cattle, sheep, pigs, etc. |
| **Exposures** | **Human**: Any exposure to the nitrate/nitrite forms below via the oral route for any duration. Studies will also be included if biomarkers of exposure are evaluated (e.g., measured chemical or metabolite levels in tissues or bodily fluids) AND there is additional information to allow estimation/attribution of nitrate/nitrite ingestion. If there is no additional information, but the exposure route is unclear or likely from multiple routes, the study will be tagged as "potentially relevant supplemental material." Other exposure routes, such as those that are clearly inhalation or dermal, will be tracked during title and abstract screening and tagged as "potentially relevant supplemental material."<br><br>**Animal**: Any exposure to the nitrate/nitrite forms below. Studies involving exposures to mixtures will be included only if they include an experimental arm with exposure to the nitrate/nitrite forms below, alone. Other exposure routes, including inhalation or dermal, will be tracked during title and abstract as "potentially relevant supplemental material."<br><br>*Relevant forms of nitrate/nitrite*: Calcium nitrate, Ammonium nitrate, Potassium nitrate, Potassium nitrite, Sodium nitrate, Sodium nitrite. |

| PECO element | Evidence |
|---|---|
| **Comparators** | **Human**: A comparison or referent population with exposure to lower levels, no exposure, or exposure below detection limits; exposure for shorter periods of time; or cases versus controls; or a repeated measures design. Worker surveillance studies are considered to meet PECO criteria even if no statistical analyses using a referent group is presented. Case reports or case series of >3 people will be considered to meet PECO criteria, while case reports describing findings in 1–3 people will be tracked as "potentially relevant supplemental material."<br><br>**Animal**: A concurrent control group exposed to vehicle-only treatment and/or untreated control. The control could be a baseline measurement (e.g., acute toxicity studies of mortality) or a repeated measure design. |
| **Outcomes** | All health endpoints for the following health effect categories are considered relevant: cancer; cardiovascular; developmental; endocrine; hematopoietic; hepatic; metabolic; nervous; reproductive; urinary. In general, endpoints related to clinical diagnostic criteria, disease outcomes, biochemical, histopathological examination, or other apical/phenotypic outcomes are considered to meet PECO criteria. We continue to include relevant studies of methemoglobinemia even though, for this outcome, the hazard is established. However, the focus is on studies that inform quantitative dose-response relationships. Human randomized controlled trials examining the protective effects of nitrate/nitrite exposure will be considered "potentially relevant supplemental material". |

Underlined text shows changes made to the assessment PECO criteria compared to the initial PECO criteria.

### 5.1.1. Other Exclusions Based on Full-Text Content

In addition to failure to meet PECO criteria (described above), epidemiological and toxicological studies may be excluded at the full-text level due to critical reporting limitations. Reporting limitations can be identified during full-text screening but are more commonly identified during subsequent phases of the assessment (e.g., literature inventory, study evaluation). Regardless of when the limitation is identified, exclusions based on full-text content are documented at the level of full-text exclusions in literature flow diagrams with a rationale of "critical reporting limitation."

A similar approach is taken for in vitro studies that are prioritized for focused analysis during assessment development (i.e., the critical reporting deficiency may preclude them from consideration). Critical reporting information for different study types are summarized below. For each piece of information, if the information can be inferred (when not directly stated) for an exposure/endpoint combination, the study should be included.

Epidemiology studies

- Sample size
- Exposure characterization and/or measurement method
- Outcome ascertainment method

- Study design

Animal studies

- Species
- Test article name
- Levels and duration of exposure
- Route of exposure
- Quantitative or qualitative (e.g., photomicrographs; author-reported lack of an effect on the outcome) results for at least one endpoint of interest

In vitro studies prioritized for focused analysis

- Cell/tissue type(s) or test system
- Test article name
- Concentration and duration of treatment
- Quantitative or qualitative results for at least one endpoint of interest

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

The planned units of analysis based on outcomes identified in the assessment PECO are summarized in Table 5-2. General considerations for defining the units of analysis are presented in the IRIS Handbook. Each unit of analysis is initially synthesized and judged separately within an evidence stream (see Section 8.1). Depending on the specific health endpoint or outcome, PK data, mechanistic information, and other supporting evidence (e.g., from studies of non-PECO routes of exposure) may be included in a unit of analysis.

The units of analysis can also include or be framed to focus on precursor events (e.g., biomarkers). 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).

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

| 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** |
| **Cardiovascular** | • Cardiovascular disease and mortality; cerebrovascular disease<br>• Blood pressure | • Blood pressure and other measures of vascular function<br>• Heart and vessel morphology<br>• Heart function |
| **Developmental** | • Fetal viability/pregnancy outcomes (spontaneous abortion)<br>• Congenital malformations<br>• Size and weight in early life | • Fetal viability/survival or other birth parameters (e.g., resorptions, number of pups per litter)<br>• Fetal growth (e.g., weight or length)<br><br>(Note: 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).) |
| **Endocrine** | • Thyroid hormones and antibodies; goiter | • Thyroid hormones<br>• Thyroid morphology/histopathology |
| **Hematopoietic** (focus on studies to support dose-response) | • Methemoglobin | • Methemoglobin |
| **Hepatic** | • (None identified) | • Liver function biomarkers (including liver enzymes)<br>• Liver histopathology |
| **Metabolic** | • Metabolic dysfunction, including diabetes | • Serum lipid measures (e.g., triglycerides; cholesterol) |

| 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** |
| | | • Indicators of insulin production and glucose homeostasis<br>• Adiposity |
| **Nervous** | • Cognitive function in adulthood<br>• Depressive symptoms<br>• Neurodegenerative disease | • Learning/memory<br>• Brain morphology/histopathology<br>• Neurodegenerative disease<br>• Sensory processing |
| **Reproductive** | • Gestational length (e.g., preterm birth) | • Reproductive hormone levels<br>• Sperm parameters<br>• Reproductive organ morphology/histopathology<br>• Fertility |
| **Urinary** | • Kidney disease | • Kidney function biomarkers<br>• Kidney morphology/histopathology |

| 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** |
| **Carcinogenicity** | • Colorectal cancer<br>• Breast cancer<br>• Gastrointestinal tract cancer<br>• Bladder cancer<br>• Kidney cancer<br>• Central nervous system cancer<br>• Thyroid cancer<br>• Liver cancer<br>• Cancer of reproductive organs<br>• Reticuloendothelial cancer<br>• Cancer mortality | • Colorectal cancer precursors<br>• All other cancer endpoints observed as part of general toxicity assays |

## 5.3. CONSIDERATION OF SUPPLEMENTAL MATERIAL

### 5.3.1. Noncancer MOA Mechanistic Information

The non-carcinogenesis mechanistic studies were screened and tagged according to the relevant target organ/health system as described in Section 4.5.2. Findings from newly identified studies will be briefly summarized in tabular format. Nonmammalian model systems were included in this analysis. These summary conclusions regarding mechanisms of toxicity for nitrate and nitrite will be used to support evidence integration conclusions for specific health system hazard analyses as well as describe general features of mechanisms of toxicity.

### 5.3.2. ADME and PK/PBPK Model Information

Studies containing ADME and PK/PBPK content were screened and tagged as described in Section 4.5.2. Oral pharmacokinetics of nitrates and nitrites are the primary focus since the current assessment focuses on the derivation of oral toxicity values. However, pharmacokinetic studies from alternate routes of exposure can still inform various aspects of ADME and are also considered.

For supplemental material studies categorized as PK/PBPK models, only three such models were identified: (Zeilmaker et al., 2010) (an application of the Zeilmaker, 1996, 3859914 model), (Lin et al., 2020) (an updated parametrization of the Zeilmaker, 1996, 3859914 model), and (Coggan and Thies, 2020). With the limited number of studies, an initial scoping process is not needed, and all three models will be evaluated for their suitability for deriving toxicity values for the nitrates/nitrites assessment (for more detail, see Section 6.6, Pharmacokinetic Model Evaluation). Model determination will include the evaluation of underlying pharmacokinetic data for training, model assumptions relative to known ADME, and the ability to predict internal dose metrics of interest.

### 5.3.3. Other Supplemental Material Content

Structured approaches to organize evidence like those presented for genotoxicity mechanistic studies, noncancer MOA, and ADME/PK/PBPK were not developed for other types of supplemental material. Instead, the tagged material was reviewed during preparation of the draft to see if studies were available to address specific uncertainties of the health study evidence base, inform susceptibility conclusions, and ensure completeness of identifying primary data papers most pertinent to the assessment.

- Titles of studies tagged as exposure-only are reviewed to see if they provided information pertinent to establish study evaluation considerations for the exposure domain.

- Titles of review articles are reviewed to identify those that are directly pertinent to the scope of the assessment. The reference lists of such reviews are scanned to identify primary data studies that might have been missed from database search queries. The reviews may

also be used to provide perspective on interpretation of foundational science cited in the assessment.

- Other types of supplemental material did not undergo additional analysis because the information was not considered likely to impact toxicity value development (including application of uncertainty factors). The specific categories are case reports, mixtures, or conference abstracts.

# 6. STUDY EVALUATION (RISK OF BIAS AND SENSITIVITY)

The general approach for evaluating primary health effect studies that meet PECO is described in Section 6.1 Instructional and informational materials for study evaluations are available at https://hawc.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 6.4, 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, 2018), which is summarized below (see Section 6.6).

## 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<br>• Outcome ascertainment<br>• Participant selection<br>• Confounding<br>• Analysis<br>• Selective reporting<br>• Sensitivity | • Allocation<br>• Observational bias/blinding<br>• Confounding<br>• Attrition<br>• Chemical administration and characterization<br>• Endpoint measurement<br>• Results presentation<br>• Selective reporting<br>• Sensitivity | • Observational bias/blinding<br>• Variable control<br>• Selective reporting<br>• Chemical administration and characterization<br>• Endpoint measurement<br>• Results presentation<br>• Sensitivity |

**(b) Domain level judgements 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 the EPA's version of HAWC (https://hawc.epa.gov/assessment/100500308/) 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, 1998, 1996, 1991a).

Study authors may be queried for information, especially if manuscripts are missing key information on study design or relevant results. Queries may also be made to inquire about additional analyses that could address major 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 of the attempt to contact. Only information or data that can be made publicly available (e.g., within HAWC or HERO) will be considered.

When evaluating studies 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 predefined 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 usable 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 will be conducted for the following domains: exposure measurement, outcome ascertainment, participant selection, potential confounding, analysis, study sensitivity, and selective reporting. Bias can result in false positives and negatives (i.e., Types I and II errors), while study sensitivity is typically concerned with identifying the latter.

The principles and framework used for evaluating epidemiology studies are based on the Cochrane Risk of Bias in Nonrandomized Studies of Interventions [ROBINS-I; (Sterne et al., 2016)] but modified to address environmental and occupational exposures. Core and prompting questions, shown in Table 6-1, are used to collect information to guide evaluation of each domain. Core questions represent key concepts, while the prompting questions help the reviewer focus on relevant details under each key domain. Table 6-1 also includes criteria that apply to all exposures and outcomes.

**Table 6-1. Domains, questions, and general considerations to guide the evaluation of epidemiology studies**

| Domain and core question | Prompting questions | Follow-up questions | Criteria that apply to most exposures and outcomes |
|---|---|---|---|
| Exposure measurement<br>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:<br>• Does the exposure measure capture the variability in exposure among the participants, considering intensity, frequency, and duration of exposure?<br>• 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?<br>• Was the exposure measurement likely to be affected by knowledge of the outcome?<br>• Was the exposure measurement likely to be affected by the presence of the outcome (i.e., reverse causality)?<br>For case-control studies of occupational exposures:<br>• Is exposure based on a comprehensive job history describing tasks, setting, period, and use of specific materials? | Is the degree of exposure misclassification likely to vary by exposure level?<br><br>If the correlation between exposure measurements is moderate, is there an adequate statistical approach to ameliorate variability in measurements?<br><br>If potential for bias is a concern, is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)? | **Good**<br>• Valid exposure assessment methods used, which represent the etiologically relevant period of interest.<br>• Exposure misclassification is expected to be minimal.<br>**Adequate**<br>• Valid exposure assessment methods used, which represent the etiologically relevant period of interest.<br>• Exposure misclassification could exist but is not expected to greatly change the effect estimate.<br>**Deficient**<br>• Valid exposure assessment methods used, which represent the etiologically relevant time period of interest. Specific knowledge about the exposure and outcome raises concerns about reverse causality, but whether it is influencing the effect estimate is uncertain.<br>• Exposed groups are expected to contain a notable proportion of unexposed or minimally exposed individuals, the method did not capture important temporal or spatial variation, or other evidence of exposure misclassification would be expected to notably change the effect estimate.<br>**Critically deficient**<br>• Exposure measurement does not characterize the etiologically relevant period of exposure or is not valid.<br>• Evidence exists that reverse causality is very likely to account for the observed association. |

| Domain and core question | Prompting questions | Follow-up questions | Criteria that apply to most exposures and outcomes |
|---|---|---|---|
| | For biomarkers of exposure, general population:<br>• 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?<br>• What exposure period is reflected by the biomarker? If the half-life is short, what is the correlation between serial measurements of exposure? | | • Exposure measurement was not independent of outcome status. |
| <u>Outcome ascertainment</u><br>Does the outcome measure reliably distinguish the presence or absence (or degree of severity) of the outcome? | For all:<br>• Is outcome ascertainment likely affected by knowledge, or presence, of exposure (e.g., consider access to health care, if based on self-reported history of diagnosis)?<br>For case-control studies:<br>• 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?<br>For mortality measures: | Is there a concern that any outcome misclassification is nondifferential, differential, or both?<br><br>What is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)? | **Good**<br>• High certainty in the outcome definition (i.e., specificity and sensitivity), minimal concerns with respect to misclassification.<br>• Assessment instrument was validated in a population comparable to the one from which the study group was selected.<br>**Adequate**<br>• Moderate confidence that outcome definition was specific and sensitive, some uncertainty with respect to misclassification but not expected to greatly change the effect estimate.<br>• Assessment instrument was validated but not necessarily in a population comparable to the study group. |

| Domain and core question | Prompting questions | Follow-up questions | Criteria that apply to most exposures and outcomes |
|---|---|---|---|
| | • 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?<br>For diagnosis of disease measures:<br>• 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?<br>For laboratory-based measures (e.g., hormone levels):<br>• 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? | | **Deficient**<br>• Outcome definition was not specific or sensitive.<br>• Uncertainty regarding validity of assessment instrument.<br>**Critically deficient**<br>• Invalid/insensitive marker of outcome.<br>• Outcome ascertainment is very likely to be affected by knowledge of, or presence of, exposure.<br>Note: Lack of blinding should not be automatically construed to be *critically deficient*. |
| <u>Participant selection</u><br>Is there evidence that selection into or out of the study (or analysis sample) was jointly related to exposure and to outcome? | For longitudinal cohort:<br>• Did participants volunteer for the cohort on the basis of knowledge of exposure or preclinical disease symptoms? Was entry into, or continuation in, the cohort related to exposure and outcome?<br>For occupational cohort:<br>• Did entry into the cohort begin with the start of the exposure? | Were differences in participant enrollment and follow-up evaluated to assess bias?<br>If potential for bias is a concern, what is the predicted direction or distortion of the bias on the effect | **Good**<br>• Minimal concern for selection bias based on description of recruitment process and follow-up (e.g., selection of comparison population, population-based random sample selection, recruitment from sampling frame including current and previous employees).<br>• Exclusion and inclusion criteria specified and would not induce bias.<br>• Participation rate is reported at all steps of study (e.g., initial enrollment, follow-up, selection into analysis sample). If rate is not high, appropriate rationale is given for why it is |

| Domain and core question | Prompting questions | Follow-up questions | Criteria that apply to most exposures and outcomes |
|---|---|---|---|
| | • Was follow-up or outcome assessment incomplete, and if so, was follow-up related to both exposure and outcome status?<br>• Could exposure produce symptoms that would result in a change in work assignment/work status ("healthy worker survivor effect")?<br>For case-control study:<br>• Were controls representative of population and periods from which cases were drawn?<br>• Are hospital controls selected from a group whose reason for admission is independent of exposure?<br>• Could recruitment strategies, eligibility criteria, or participation rates result in differential participation relating to both disease and exposure?<br>For population-based survey:<br>• Was recruitment based on advertisement to people with knowledge of exposure, outcome, and hypothesis? | estimate (if there is enough information)?<br><br>Were appropriate analyses performed to address changing exposures over time relative to symptoms?<br><br>Is there a comparison of participants and nonparticipants to address whether differential selection or study retention/continuation is likely? | unlikely to be related to exposure (e.g., comparison between participants and nonparticipants or other available information indicates differential selection is not likely).<br>**Adequate**<br>• Enough of a description of the recruitment process to be comfortable that there is no serious risk of bias.<br>• Inclusion and exclusion criteria specified and would not induce bias.<br>• Participation rate is incompletely reported but available information indicates participation is unlikely to be related to exposure.<br>**Deficient**<br>• Little information on recruitment process, selection strategy, sampling framework, and participation OR aspects of these processes raises the potential for bias (e.g., healthy worker effect, survivor bias).<br>**Critically deficient**<br>• Aspects of the processes for recruitment, selection strategy, sampling framework, or participation result in concern that selection bias is likely to have had a large impact on effect estimates (e.g., convenience sample with no information about recruitment and selection, cases and controls are recruited from different sources with different likelihood of exposure, recruitment materials stated outcome of interest and potential participants are aware of or are concerned about specific exposures). |
| <u>Confounding</u> | Is confounding adequately addressed by considerations in: | If potential for bias is a concern, what is the predicted | **Good**<br>• Conveys strategy for identifying key confounders, including co-exposures. This may include a priori biological |

| Domain and core question | Prompting questions | Follow-up questions | Criteria that apply to most exposures and outcomes |
|---|---|---|---|
| Is confounding of the effect of the exposure likely? | • Participant selection (matching or restriction)?<br>• Accurate information on potential confounders and statistical adjustment procedures?<br>• Lack of association between confounder and outcome, or confounder and exposure in the study?<br>• Information from other sources?<br>• 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)? | direction or distortion of the bias on the effect estimate (if there is enough information)? | consideration, published literature, causal diagrams, or statistical analyses, with the recognition that not all "risk factors" are confounders.<br>• Inclusion of potential confounders in statistical models not based solely on statistical significance criteria (e.g., $p < 0.05$ from stepwise regression).<br>• Does not include variables in the models likely to be influential colliders or intermediates on the causal pathway.<br>• Key confounders are evaluated appropriately and considered unlikely sources of substantial confounding. This often will include:<br>○ Presenting the distribution of potential confounders by levels of the exposure of interest or the outcomes of interest (with amount of missing data noted);<br>○ Consideration that potential confounders were rare among the study population, or were expected to be poorly correlated with exposure of interest;<br>○ Consideration of the most relevant functional forms of potential confounders;<br>○ Examination of the potential impact of measurement error or missing data on confounder adjustment; or<br>○ Presenting a progression of model results with adjustments for different potential confounders, if warranted.<br>**Adequate**<br>• Similar to *good* but might not have included all key confounders, or less detail might be available on the evaluation of confounders (e.g., sub-bullets in *good*). That |

| Domain and core question | Prompting questions | Follow-up questions | Criteria that apply to most exposures and outcomes |
|---|---|---|---|
| | | | residual confounding could explain part of the observed effect is possible, but concern is minimal.<br>**Deficient**<br>• Does not include variables in the models shown to be influential colliders or intermediates on the causal pathway.<br>• And any of the following:<br>o The potential for bias to explain some results is high based on an inability to rule out residual confounding, such as a lack of demonstration that key confounders of the exposure-outcome relationships were considered;<br>o Descriptive information on key confounders (e.g., their relationship relative to the outcomes and exposure levels) are not presented; or<br>o Strategy of evaluating confounding is unclear or is not recommended (e.g., only based on statistical significance criteria or stepwise regression [forward or backward elimination]).<br>**Critically deficient**<br>• Includes variables in the models that are colliders or intermediates in the causal pathway, indicating that substantial bias is likely from this adjustment; or<br>• Confounding is likely present and not accounted for, indicating that all results were most likely due to bias. |
| <u>Analysis</u><br>Does the analysis strategy and presentation convey the necessary | • Are missing outcome, exposure, and covariate data recognized, and if necessary, accounted for in the analysis? | If potential for bias is a concern, what is the predicted direction or distortion of the bias on the effect | **Good**<br>• Use of an optimal characterization of the outcome variable, including presentation of subgroup- or lifestage-specific comparisons (as appropriate for the outcome). |

| Domain and core question | Prompting questions | Follow-up questions | Criteria that apply to most exposures and outcomes |
|---|---|---|---|
| familiarity with the data and assumptions? | • Does the analysis appropriately consider variable distributions and modeling assumptions?<br>• Does the analysis appropriately consider subgroups or lifestages of interest (e.g., based on variability in exposure level or duration or susceptibility)?<br>• Is an appropriate analysis used for the study design?<br>• Is effect modification considered, based on considerations developed a priori?<br>• Does the study include additional analyses addressing potential biases or limitations (i.e., sensitivity analyses)? | estimate (if there is enough information)? | • Quantitative results presented (effect estimates and confidence limits or variability in estimates) (i.e., not presented only as a *p*-value or "significant"/"not significant").<br>• Descriptive information about outcome and exposure provided (where applicable).<br>• Amount of missing data noted and addressed appropriately (discussion of selection issues—missing at random vs. differential).<br>• Where applicable, for exposure, includes Limit of detection LOD (and percentage below the LOD), and decision to use log transformation.<br>• 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.<br>• No deficiencies in analysis evident. Discussion of some details might be absent (e.g., examination of outliers).<br>**Adequate**<br>• Same as 'Good,' except:<br>• Descriptive information about exposure provided (where applicable) but might be incomplete; might not have discussed missing data, cut-points, or shape of distribution(s).<br>• 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 | Criteria that apply to most exposures and outcomes |
|---|---|---|---|
| | | | **Deficient**<br>• Does not conduct analysis using optimal characterization of the outcome variable.<br>• Descriptive information about exposure levels not provided (where applicable).<br>• Effect estimate and p-value presented, without standard error or confidence interval.<br>• Results presented as statistically "significant"/"not significant."<br>**Critically deficient**<br>• Analysis methods are not appropriate for design or data of the study. |
| Selective reporting<br>Is there reason to be concerned about selective reporting? | • Were results provided for all the primary analyses described in the methods section?<br>• Is appropriate justification given for restricting the amount and type of results shown?<br>• Are only statistically significant results presented? | If potential for bias is a concern, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)? | **Good**<br>• The results reported by study authors are consistent with the primary and secondary analyses described in a registered protocol or methods paper.<br>**Adequate**<br>• The authors described their primary (and secondary) analyses in the methods section and results were reported for all primary analyses.<br>**Deficient**<br>• Concerns were raised based on previous publications, a methods paper, or a registered protocol indicating that analyses were planned or conducted that were not reported, or that hypotheses originally considered to be secondary were represented as primary in the reviewed paper. |

| Domain and core question | Prompting questions | Follow-up questions | Criteria that apply to most exposures and outcomes |
|---|---|---|---|
| | | | • Only subgroup analyses were reported, suggesting that results for the entire group were omitted.<br>• Only statistically significant results were reported. |
| Sensitivity<br>Is there a concern that sensitivity of the study is not adequate to detect an effect? | • Is the exposure contrast adequate to detect associations and exposure-response relationships?<br>• Was the appropriate population or lifestage included?<br>• Was the length of follow-up adequate? Is the time/age of outcome ascertainment optimal given the interval of exposure and the health outcome?<br>• Do other aspects related to risk of bias or otherwise raise concerns about sensitivity? | | **Good**<br>• There is sufficient variability/contrast in exposure to evaluate primary hypotheses.<br>• The study population was sensitive to the development of the outcomes of interest (e.g., ages, lifestage, sex).<br>• The timing of outcome ascertainment was appropriate given expected latency for outcome development (i.e., adequate follow-up interval).<br>• The study was adequately powered to observe an effect.<br>• No other concerns raised regarding study sensitivity.<br>**Adequate**<br>Same considerations as *Good*, except:<br>• There may be issues identified that could reduce sensitivity, but they are considered unlikely to substantially impact the overall findings of the study.<br>**Deficient**<br>• Concerns were raised about the considerations described for *Good* that are expected to notably decrease the sensitivity of the study to detect associations for the outcome.<br>**Critically deficient**<br>• Severe concerns were raised about the sensitivity of the study such that any observed associations are likely to be explained by bias. |

## 6.3. EXPERIMENTAL ANIMAL STUDY EVALUATION

Using the principles described in Section 6.1, the animal studies of health effects are evaluated for the following domains to assess risk of bias and sensitivity: allocation, observational bias/blinding, confounding, selective reporting, attrition, chemical administration and characterization, endpoint measurement and validity, results presentation and comparisons, and sensitivity (see Table 6-2).

The rationale for judgments is documented at the outcome level. The evaluation documentation in HAWC includes the identified limitations and their expected impact on the overall confidence level. To the extent possible, the rationale will reflect an interpretation of the potential influence on the outcome-specific results, including the direction or magnitude of influence (or both).

**Table 6-2. Domains, questions, and general considerations to guide the evaluation of animal toxicology studies**

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| **Allocation**<br>Were animals assigned to experimental groups using a method that minimizes selection bias? | For each study:<br>Did each animal or litter have an equal chance of being assigned to any experimental group (i.e., random allocation)?[a]<br>Is the allocation method described?<br>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.<br>A judgment and rationale for this domain should be given for each cohort or experiment in the study.<br>**Good**: Experimental groups were randomized, and any specific randomization procedure was described or inferable (e.g., computer-generated scheme. Note that normalization is not the same as randomization [see response for *adequate*]).<br>**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).<br>**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.<br>**Critically deficient**: Bias in the animal allocations was reported or inferable. |
| **Observational bias/blinding**<br>Did the study implement measures to reduce observational bias? | For each endpoint/outcome or grouping of endpoints/outcomes in a study:<br>Does the study report blinding or other procedures for reducing observational bias?<br>If not, did the study use a design or approach for which such procedures can be inferred?<br>What is the expected impact of failure to implement (or report implementation) of these procedures on results? | These considerations typically do not need to be refined by the assessment teams. (Note that it can be useful for teams to identify highly subjective measures of endpoints/outcomes when observational bias may strongly influence results prior to performing evaluations.)<br>A judgment and rationale for this domain should be given for each endpoint/outcome or group of endpoints/outcomes investigated in the study.<br>**Good**: Measures to reduce observational bias were described (e.g., blinding to conceal treatment groups during endpoint evaluation; consensus-based evaluations of histopathology-lesions).[b]<br>**Adequate**: Methods for reducing observational bias (e.g., blinding) can be inferred or were reported but described incompletely.<br>**Not reported**: Measures to reduce observational bias were not described.<br>**(Interpreted as *adequate*)** The potential concern for bias was mitigated based on use of automated/computer driven systems, standard laboratory kits, relatively simple, objective measures (e.g., body or tissue weight), or screening-level evaluations of histopathology.<br>**(Interpreted as *deficient*)** The potential impact on the results is major (e.g., outcome measures are highly subjective). |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| | | **Critically deficient**: Strong evidence for observational bias that impacted the results. |
| **Confounding**<br>Are variables with the potential to confound or modify results controlled for and consistent across experimental groups?<br><br>*Note:*<br>*Consideration of overt toxicity (possibly masking more specific effects) is addressed under endpoint measurement reliability.* | For each study:<br>Are there difference across the treatment groups, considering both differences related to the exposure (e.g., co-exposures, vehicle, diet, palatability) and other aspects of the study design or animal groups (e.g., animal source, husbandry, or health status), that could bias the results?<br>If differences are identified, to what extent are they expected, based on a specific scientific understanding, 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.<br>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.<br>**Good**: Outside of the exposure of interest, variables that are likely to confound or modify results appear to be controlled for and consistent across experimental groups.<br>**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.<br>**Deficient**: Notable concern that potentially confounding variables were uncontrolled or inconsistent across groups and are expected based on to substantially impact the results.<br>**Critically deficient**: Confounding variables were presumed to be uncontrolled or inconsistent across groups and are expected to be a primary driver of the results. |
| **Attrition**<br>Did the study report results for all tested animals? | For each study:<br>Are all animals accounted for in the results?<br>If there is attrition, do authors provide an explanation (e.g., death or unscheduled sacrifice during the study)?<br>If unexplained attrition of animals for outcome assessment is identified, what is the expected impact on the interpretation of the results? | These considerations typically do not need to be refined by assessment teams.<br>A judgment and rationale for this domain should be given for each cohort or experiment in the study.<br>**Good**: Results were reported for all animals. If animal attrition is identified, the authors provide an explanation, and these are not expected to impact the interpretation of the results.<br>**Adequate**: Results are reported for most animals. Attrition is not explained but this is not expected to significantly impact the interpretation of the results.<br>**Deficient**: *Moderate*-to-*high* level of animal attrition that is not explained and may significantly impact the interpretation of the results.<br>**Critically deficient**: Extensive animal attrition that prevents comparisons of results across treatment groups. |
| **Chemical administration and characterization**<br>Did the study adequately characterize exposure to the chemical of interest and | For each study:<br>Are there concerns [specific to this chemical] regarding the source and purity and/or composition (e.g., identity | It is essential that these considerations 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).<br>A judgment and rationale for this domain should be given for each cohort or experiment in the study. |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| the exposure administration methods?<br>*Note:*<br>*Consideration of the appropriateness of the route of exposure (not the administration method) is not a risk of bias consideration. Relevance and utility of the routes of exposure are considered in the PECO criteria for study inclusion and during evidence synthesis. Relatedly, consideration of exposure level selection (e.g., were levels sufficiently high to elicit effects) is addressed during evidence synthesis and is not a risk of bias consideration.* | and percent distribution of different isomers) of the chemical?<br>Was independent analytical verification of the test article (e.g., composition, homogeneity, and purity) performed?<br>Were nominal exposure levels verified analytically? Are there concerns about the methods used to administer the chemical (e.g., inhalation chamber type, gavage volume)? | **Good**: Chemical administration and characterization is complete (i.e., source and purity are provided or can be obtained from the supplier and test article is analytically verified). There are no notable concerns about the composition, stability, or purity of the administered chemical, or the specific methods of administration. Exposure levels are verified using reliable analytical methods.<br>**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., purity of the test article is suboptimal but interpreted as unlikely to have a significant impact; analytical verification of exposure levels is not reported or verified with non-preferred methods).<br>**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, and composition is not independently verified; impurities are substantial or concerning; administration methods are considered likely to introduce confounders, such as use of static inhalation chambers or a gavage volume considered too large for the species or lifestage at exposure).<br>**Critically deficient**: Uncertainties in the exposure characterization are identified and there is reasonable certainty that the study 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**<br>Are the selected procedures, protocols and animal models adequately described and appropriate for the endpoint(s)/outcome(s) of interest?<br>*Notes:*<br>*Considerations related to the sensitivity of the animal model and timing of endpoint measurement are* | For each endpoint/outcome or grouping of endpoints/outcomes in a study:<br>Are the evaluation methods and animal model adequately described and appropriate?<br>Are there concerns regarding the methodology selected for endpoint evaluation?<br>Are there concerns about the specificity of the experimental design?<br>Are there serious concerns regarding the sample size or how endpoints were sampled? | Considerations for this domain are highly variable depending on the endpoint(s)/outcome(s) of interest and typically must be refined by assessment teams.<br>A judgment and rationale for this domain should be given for each endpoint/outcome or group of endpoints/outcomes investigated in the study.<br>Some considerations include the following:<br>**Good**:<br>• Adequate description of methods and animal models.<br>• Use of generally accepted and reliable endpoint methods. |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| *evaluated under Sensitivity Considerations related to adjustments/corrections to endpoint measurements (e.g., organ weight corrected for body weight) are addressed under results presentation.* | Are appropriate control groups for the study/assay type included? | • 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 (e.g., sampling procedures for histological analysis).<br>• Includes appropriate control groups and any use of nonconcurrent or historical control data (e.g., for evaluation of rare tumors) is justified (e.g., authors or evaluators considered the similarity between current experimental animals and laboratory conditions to historical controls).<br>Ratings of **Adequate, Deficient, and Critically Deficient** are generally defined as follows:<br>**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.<br>**Deficient**: Concerns are raised that are expected to notably affect endpoint measurement and reduce the reliability of the study findings<br>**Critically deficient**: Severe concerns are raised about endpoint measurement and any findings are likely to be largely explained by these limitations<br>• 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:<br>• 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 strain, sex, or lifestage of the animals).<br>• 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.[a]<br>• Overt toxicity (e.g., mortality, extreme weight loss) is observed or expected based on findings from similarly designed studies and may mask interpretation of outcome(s) of interest. |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| | | • 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).[b]<br>• Control groups are not included, considered inappropriate, or comparisons to non-concurrent or historical controls are not adequately justified. |
| **Results presentation**<br>Are the results presented and compared in a way that is appropriate and transparent? | For each endpoint/outcome or grouping of endpoints/outcomes in a study:<br>Does the level of detail allow for an informed interpretation of the results?<br>Are the data 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 typically must be refined by assessment teams.<br>A judgment and rationale for this domain should be given for each endpoint/outcome or group of endpoints/outcomes investigated in the study.<br>Some considerations include the following:<br>**Good**:<br>• No concerns with how the data are presented.<br>• 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).<br>• No concerns with completeness of the results reporting.[c]<br>• Ratings of **Adequate**, **Deficient**, and **Critically Deficient** are generally defined as follows:<br>**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.<br>**Deficient:** Concerns with results presentation are identified and expected to substantially impact results interpretation and reduce the reliability of the study findings.<br>**Critically deficient:** Severe concerns about results presentation were identified and study findings are likely to be largely explained by these limitations.<br>• The following specific examples of relevant concerns are typically associated with a **Deficient** rating but **Adequate** or **Critically Deficient** might be applied |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| | | depending on expected impact of limitations on the reliability and interpretation of the results:<br>• Nonpreferred presentation of data (e.g., developmental toxicity data averaged across pups in a treatment group, when litter responses are more appropriate; presentation of only absolute organ weight data when relative weights are more appropriate).<br>• Pooling data when responses are known or expected to differ substantially (e.g., across sexes or ages).<br>• Incomplete presentation of the data[c] (e.g., presentation of mean without variance data; concurrent control data are not presented; dichotomizing or truncating continuous data). |
| **Selective reporting**<br>Did the study report results for all prespecified outcomes?<br>*Note:*<br>*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:<br>Are results presented for all endpoints/outcomes described in the methods (see note)?<br>If unexplained results omissions are identified, what is the expected impact on the interpretation of the results? | These considerations typically do not need to be refined by assessment teams.<br>A judgment and rationale for this domain should be given for each cohort or experiment in the study.<br>**Good**: Quantitative or qualitative results were reported for all prespecified outcomes (explicitly stated or inferred), exposure groups and evaluation time points. Data not reported in the primary article 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.<br>**Adequate**: Quantitative or qualitative results are reported for most prespecified outcomes (explicitly stated or inferred) and evaluation time points. Omissions and are not explained but are not expected to significantly impact the interpretation of the results.<br>**Deficient**: Quantitative or qualitative results are missing for many prespecified outcomes (explicitly stated or inferred), omissions are not explained and may significantly impact the interpretation of the results.<br>**Critically deficient**: Extensive results omission is identified and prevents comparisons of results across treatment groups. |
| **Sensitivity**<br>Are there concerns that sensitivity in the study is not | Was the exposure period, timing (e.g., lifestage), frequency, and duration sensitive for the outcome(s) of interest? | These considerations may require customization to the specific exposure and outcomes. 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, |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| adequate to detect an effect?<br>*Note:*<br>*Consideration of exposure level selection (e.g., were levels sufficiently high to elicit effects) is addressed during evidence synthesis and is not a study sensitivity consideration.* | Based on knowledge of the health hazard of concern, did the selection of species, strain, and/or sex of the animal model reduce study sensitivity?<br>Are there concerns regarding the timing (e.g., lifestage) of the outcome evaluation?<br>Are there aspects related to risk of bias domains that raise concerns about insensitivity (e.g., selection of protocols that are known to be insensitive or nonspecific for the outcome(s) of interest) | along with any features that have not been addressed elsewhere. Some considerations include:<br>**Good**<br>• The experimental design (considering exposure period, timing, frequency, and duration) is appropriate and sensitive for evaluating the outcome(s) of interest.<br>• The selected animal model (considering species, strain, sex, and/or lifestage) is known or assumed to be appropriate and sensitive for evaluating the outcome(s) of interest.<br>• No significant concerns with the ability of the experimental design to detect the specific outcome(s) of interest. (e.g., outcomes evaluated at the appropriate lifestage; study designed to address known endpoint variability that is unrelated to treatment, such as estrous cyclicity or time of day).<br>• Timing of endpoint measurement in relation to the chemical exposure is appropriate and sensitive (e.g., behavioral testing is not performed during a transient period of test chemical-induced depressant or irritant effects; endpoint testing does not occur only after a prolonged period, such as weeks or months, of nonexposure).<br>• Potential sources of bias toward the null are not a substantial concern.<br>**Adequate**<br>Same considerations as *Good*, except:<br>• The duration and frequency of the exposure was appropriate, and the exposure covered most of the critical window (if known) for the outcome(s) of interest.<br>• Potential issues are identified that could reduce sensitivity, but they are unlikely to impact the overall findings of the study. |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| | | **Deficient**<br>• Concerns were raised about the considerations described for *Good* or Adequate that are expected to notably decrease the sensitivity of the study to detect a response in the exposed group(s).<br>**Critically deficient**<br>• 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**<br>Considering the identified strengths and limitations, what is the overall confidence rating for the endpoint(s)/outcome(s) of interest? | For each endpoint/outcome or grouping of endpoints/outcomes in a study:<br>Were concerns (i.e., limitations or uncertainties) related to the risk of bias or sensitivity identified?<br>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. Reviewers should mark studies that are rated lower than high confidence only due to low sensitivity (i.e., bias toward the null) for additional consideration during evidence synthesis. If the study is otherwise well conducted and an effect is observed, it may increase the strength of evidence judgment.<br>A confidence rating and rationale should be given for each endpoint/outcome or group of endpoints/outcomes investigated in the study. Confidence ratings are described above (see Section 6.1). |

[a]These limitations typically also raise a concern for insensitivity.

[b]Sample size alone is not a reason to conclude an individual study is critically deficient.

[c]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) is addressed under Selective Reporting.

## 6.4. CONTROLLED HUMAN EXPOSURE STUDY EVALUATION

Controlled human health studies are seldom available for IRIS assessments. In the case of nitrate/nitrite there is a substantial body of literature evaluating potential *beneficial* health effects (namely, cardiovascular benefits) of controlled exposure to nitrate/nitrite, but no controlled human exposure studies evaluated risk of *adverse* health effects. However, if any such studies are identified during literature search updates, evaluation criteria will be developed incorporating aspects of the approaches used for epidemiology studies and experimental animal studies, as well as the Cochrane risk of bias tools for randomized trials (ROB2) (Sterne et al., 2019) and the ROBINS-I tool discussed in Section 6.2 (Sterne et al., 2016). Controlled human exposure studies will be evaluated for important attributes of experimental studies, including randomization of exposure assignments, blinding of subjects and investigators, exposure generation, inclusion of a clean air control exposure (if applicable), outcome ascertainment, missing data, deviations from the intended intervention, study sensitivity, and other aspects of the exposure protocol. Evaluation will also include confirmation that the study protocol was approved by an institutional review board.

## 6.5. IN VITRO AND OTHER MECHANISTIC STUDY EVALUATION

Mechanistic studies will be evaluated using the considerations presented in Table 6-3 for the following domains: risk of bias (observational bias/blinding, variable control, specificity, selective reporting, chemical administration and characterization, endpoint measurement validity, and results presentation and comparisons) and study sensitivity. Mechanistically relevant endpoints reported in human and in vivo animal studies are evaluated using the domains for epidemiology and experimental animal studies presented in the previous sections. Assay-specific considerations are applied when evaluating the sensitivity domains and will be recorded in their evaluations in the HAWC database.

**Table 6-3. Domains, questions, and general considerations to guide the evaluation of in vitro studies**

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| **Observational bias/blinding**<br>Did the study implement measures, where possible, to reduce observational bias?<br>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:<br>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)?<br>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?<br>Were the assays evaluated using automated approaches (e.g., microplate readers) that reduce concern for observational bias?<br>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.<br>A judgment and rationale for this domain should be given for each assay or endpoint or group of endpoints investigated in the study.<br>**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.<br>**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.<br>**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).<br>**Critically deficient**: Strong evidence for observational bias that could have impacted the results. |
| **Variable control**<br>Are all introduced variables with the potential to affect the results of interest controlled for and consistent across experimental groups? | For each study:<br>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?<br>Did the study address features inherent to the physico-chemical properties of the test substance(s) that have the potential to bias the | 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.<br>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.<br>**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.<br>**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 |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| | 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.)*<br>Are there known variations in cellular signaling unique to the model system that could influence the possibility of detecting the effect(s) of interest?<br>Are there concerns regarding the negative (untreated and/or vehicle) controls used? Were negative controls run concurrently? | uncontrolled or inconsistent across groups but are expected to have a minimal impact on the results.<br>**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.<br>**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**<br>Did the study present results, quantitatively or qualitatively, for all prespecified assays or endpoints and replicates described in the methods?<br>*Note: The appropriateness of the analysis or results presentation is considered under results presentation.* | For each study:<br>Are results presented for all endpoints/outcomes described in the methods?<br>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)?<br>If unexplained results omissions are identified, what is the expected impact on the interpretation of the results? | These considerations typically do not need to be refined by assessment teams.<br>A judgment and rationale for this domain should be given for each assay or endpoint in the study.<br>**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.<br>**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.<br>**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. |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| | | **Critically Deficient**: Extensive results omissions are identified, preventing comparisons of results across treatment groups. |
| **Chemical administration and characterization**<br>Did the study adequately characterize exposure to the chemical of interest and the exposure administration methods? | For each study:<br>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)?<br>Was independent analytical verification of the test article purity and composition performed? If not, is this a significant concern for this substance?<br>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)?<br>Are there concerns about the preparation or storage conditions of the test substance?<br>Are there concerns about the methods used to administer the chemical? | 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).<br>A judgment and rationale for this domain should be given for each experiment in the study.<br>**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.<br>**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).<br>**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).<br>**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). |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| **Endpoint measurement**<br>Are the selected protocols, procedures, and test systems adequately described and appropriate for evaluating the endpoint(s) of interest?<br>*Notes:*<br>*Considerations related to adjustments or corrections to endpoint measurements are addressed under results presentation.*<br>*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:<br>Are the evaluation methods and test systems adequately described and appropriate?<br>Are there concerns regarding the methodology selected (e.g., accepted guidelines, established criteria) for endpoint evaluation?<br>Are there concerns about the specificity of the experimental design? Did the study address features inherent to the test system or experiment that have the potential to lead to bias away from the null?<br>Are there serious concerns about the number of replicates or sample size in the study?<br>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?<br>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.<br>A judgment and rationale for this domain should be given for each assay or endpoint or group of endpoints investigated in the study.<br>Some considerations include the following:<br>**Good**:<br>• Adequate description of methods and test system.<br>• 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.<br>• 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.<br>• 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).<br>Ratings of Adequate, Deficient, and Critically Deficient are generally defined as follows:<br>**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.<br>**Deficient**: Concerns are raised that are expected to notably affect endpoint measurement and reduce the reliability of the study findings<br>**Critically deficient**: Severe concerns are raised about endpoint measurement and any findings are likely to be largely explained by these limitations.<br>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: |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| | | • 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).<br>• 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.[a]<br>• Cytotoxicity is observed or expected based on findings from similarly designed studies and may mask interpretation of outcome(s) of interest.<br>• 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).[b]<br>• Controls are not included or considered inappropriate. |
| **Results presentation**<br>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:<br>Does the level of detail allow for an informed interpretation of the results?<br>If applicable, was the assay signal normalized to account for nonbiological differences across replicates and exposure groups?<br>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.<br>A judgment and rationale for this domain should be given for each assay or endpoint or group of endpoints investigated in the study.<br>Some considerations include the following:<br>**Good**:<br>• No concerns with how the data are presented.<br>• 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).<br>• No concerns with completeness of the results reporting.[c] |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| | | • Ratings of **Adequate**, **Deficient**, and **Critically Deficient** are generally defined as follows:<br>**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.<br>**Deficient:** Concerns with results presentation are identified and expected to substantially impact results interpretation and reduce the reliability of the study findings.<br>**Critically deficient:** Severe concerns about results presentation were identified and study findings are likely to be largely explained by these limitations.<br>• 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:<br>• Nonpreferred presentation of data (e.g., averaging technical replicates rather than independent replicates).<br>• Failure to present quantitative results.<br>• Pooling data when responses are known or expected to differ substantially (e.g., across cell types or passage number).<br>• Incomplete presentation of the data[c] (e.g., presentation of mean without variance data; concurrent control data are not presented; failure to report or address overt cytotoxicity). |
| **Sensitivity**<br>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, | 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?<br>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 |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| | particularly in the absence of a positive control?<br>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)?<br>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? | 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.<br>Some considerations include:<br>**Good**<br>• The experimental design (considering exposure period, timing, frequency, and duration) is appropriate and sensitive for evaluating the outcome(s) of interest.<br>• 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).<br>• 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).<br>• Timing of endpoint measurement in relation to the chemical exposure is appropriate and sensitive (e.g., cultures adequately express mature cell markers).<br>• Potential sources of bias toward the null is not a substantial concern.<br>**Adequate**<br>• 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.<br>**Deficient**<br>• 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). |

| Domain and core question | Prompting questions | General considerations |
|---|---|---|
| | | **Critically deficient**<br>• 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**<br>Considering the identified strengths and limitations, what is the overall confidence rating for the assay(s) or endpoint(s) of interest?<br>*Note:*<br>*Reviewers should mark studies for additional consideration during evidence synthesis if, due to low sensitivity only (i.e., bias toward the null), these studies are rated as lower than high confidence. If the study is otherwise well conducted and an effect is observed, the confidence may be increased.* | For each assay or endpoint or grouping of endpoints in a study:<br>• Were concerns (i.e., limitations or uncertainties) related to the risk of bias or sensitivity identified?<br>• 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.<br>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). |

[a]These limitations typically also raise a concern for insensitivity.

[b]Sample size alone is not a reason to conclude an individual study is critically deficient.

[c]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.

## 6.6. PHARMACOKINETIC MODEL EVALUATION

PBPK (or classical pharmacokinetic [PK]) models should be used in an assessment when a validated and 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 may come from studies across various species and may be in vitro or in vivo in design. Specific details for this evaluation are provided below and in the Umbrella quality assurance project plan (QAPP) for dosimetry and mechanism-based models (U.S. EPA, 2020b).

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

PBPK modeling is the preferred approach for calculating a human equivalent dose according to the hierarchy of approaches outlined in EPA guidelines (U.S. EPA, 2011a). As PK/PBPK studies had been evaluated in the 2001 EPA oral assessment, a literature search was conducted for PK/PBPK studies published since 2000. As described in Section 4.2, PK/PBPK studies identified in our search were tagged as supplemental material.

Following literature searches, a stepwise approach is taken that includes conducting an initial scoping of the supplemental material studies categorized as PK/PBPK models. Then, an in-depth full model evaluation is implemented to identify PBPK models that are potentially suitable for deriving toxicity values for the nitrate/nitrite assessment.

The initial scoping process is distinct from the full model evaluation. The scoping process provides a rapid assessment and communication of the availability, structure, and potential uses of PBPK/PK models, but is not a full evaluation. Full model evaluation—the complete and thorough assessment of the quality and utility of a particular model—is conducted if the initial scoping identifies one or more models that are available and considered appropriate for one or more applications in the assessment. The model evaluation is then conducted for the selected application(s). As shown below in Table 6-4 for example, key information from identified PBPK models during the scoping process is summarized in tabular format for further in-depth model evaluation following the evaluation approaches summarized in Section 6.6.2.

**Table 6-4. Example descriptive summary for a physiologically based pharmacokinetic (PBPK) model**

| Study detail | Description/notes | | | | |
|---|---|---|---|---|---|
| Author | Smith et al. (2003) | | | | |
| Contact email | xxxxx@email.com | | | | |
| Contact phone | xxx-xxx-xxxx | | | | |
| Sponsor | N/A | | | | |
| Model summary | | | | | |
| Species | Rat | | | | |
| Strain | F433 | | | | |
| Sex | Male and female | | | | |
| Life stage | Adult | | | | |
| Exposure routes | Inhalation | Oral | I.V. | Skin | |
| Tissue dosimetry | Blood | Liver | Kidney | Urine | Lung |
| Model evaluation | | | | | |
| Language | ACSL 11.8 | | | | |
| Code available | YES | Effort to recreate model | | COMPLETE | |
| Code received | YES | Effort to migrate to open software | | SIGNIFICANT | |
| Structure evaluated | YES | | | | |
| Math evaluated | YES | | | | |
| Code evaluated | YES. Issue (minor): Incorrect units listed in comments for liver metabolism (line 233). Issue (major): Mass balance error in stomach compartment. | | | | |
| Available PK data | Urine (cumulative amount excreted) and blood (concentration) time course data for oral (gavage) and inhalation (6 hr/day for 4 days) exposure. In vitro skin permeation. | | | | |

### 6.6.2. Pharmacokinetic (PK)/Physiologically Based Pharmacokinetic (PBPK) Model Evaluation

Once available PBPK models are summarized, the assessment team undertakes model evaluation in accordance with criteria outlined by U.S. EPA (2020b). 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 the 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, the use of scientifically supported and biologically consistent parameters in the model, and the reproducibility of model results reported in journal publications and other documents. This approach stresses (1) clarity in the documentation of model purpose, structure, and biological characterization; (2) validation of mathematical descriptions, parameter values, and computer implementation; and (3) the ability of the model to predict each plausible dose metric such as nitrate and nitrite concentrations in the blood and the production of relevant metabolites. The in-depth analysis is used to evaluate the potential value and cost of developing a new model or substantially revising an existing one. PBPK models adapted, modified, or developed by EPA during the assessment will undergo peer review, either as a component of the draft assessment or by publication in a journal article.

In brief, a major strength of a PBPK model is its capacity to provide quantitative descriptions of ADME of chemicals by accounting for the dynamic but complex relationships among physiological, biochemical, and metabolic determinants. When describing a published PBPK model, two components must be evaluated: 1) the underlying biological assumptions and resulting mathematical equations giving rise to the model structure and 2) the parameterization of these mathematical equations using experimental pharmacokinetic data (such as concentration vs. time data). Taken together, these two components of model structure and model parameters constitute a unique PBPK model. To this end, three PBPK models exist for nitrates/nitrites: (Zeilmaker et al., 2010), (Lin et al., 2020), and (Coggan et al., 2021). Of these models, (Zeilmaker et al., 2010) and (Lin et al., 2020) share the same underlying model structure originally introduced in (Zeilmaker et al., 1996) with different in vivo datasets used to parameterize the model structure.

Biotransformation of nitrate to nitrite through gut and salivary bacteria is thought to be a major source of dietary nitrate toxicity. Therefore, the PBPK model(s) selected for the nitrate/nitrite assessment should reflect the underlying mechanisms and anatomical location for this biotransformation and any additional mechanisms of action for specific toxicological endpoints when estimating relevant dose metrics (U.S. EPA, 2018). For example, nitrite is known to react with hemoglobin in the blood to form methemoglobin. Inclusion of this mechanism will be important for linking exposure-response information for effect of nitrite on risk of methemoglobinemia, to exogenous nitrate exposure.

The available PBPK models aim to describe the pharmacokinetics of nitrate and nitrite following nitrate absorption in the stomach and biotransformation to nitrite throughout the body. Briefly, the (Zeilmaker et al., 1996) model structure assumes exposure only to nitrate. In this model structure, nitrate is absorbed into a central compartment and secreted into a salivary compartment where it undergoes conversion to nitrite. Following this conversion, nitrite is absorbed through the stomach into the central compartment where it reacts with hemoglobin to create methemoglobin.

The (Zeilmaker et al., 2010) model then parameterizes this model structure using data from human volunteers to characterize nitrate and nitrite levels in blood and saliva following a known exposure to nitrate. Comparatively, the (Coggan et al., 2021) model structure assumes exogenous exposure to both nitrate and nitrite in which nitrate is transformed to nitrite in the central compartment through first order kinetics. Using a cohort of elderly volunteers, this model structure is parameterized using plasma nitrate and nitrite concentrations. Finally, (Lin et al., 2020) uses the same model structure as (Zeilmaker et al., 2010) and updates the parameters using an additional human nitrate dataset. Further evaluation of these models will be conducted according to EPA's Umbrella QAPP for Dosimetry and Mechanism-Based Models (U.S. EPA, 2020b). It may be that none of the existing PBPK models adequately fulfills all assessment applications. In this case, a hybrid model could be created that merges elements from the existing models to achieve this objective if needed and feasible under the time constraints for the assessment.

**Table 6-5. Criteria for evaluation of physiologically based pharmacokinetic (PBPK) models**

| Criteria | Example information |
|---|---|
| **Scientific** | Biological basis for the model is accurate.<br>• Consistent with mechanisms that significantly impact dosimetry.<br>• Predicts dose metrics expected to be relevant.<br>• 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.<br>• Can the model describe critical behavior, such as nonlinear kinetics in a relevant dose range, better than the default (i.e., $BW^{3/4}$ scaling)?<br>• Is the available metric a better predictor of risk than the default? (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 (e.g., 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 metric). |
| | Principle of parsimony<br>• 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. |
| | Well-documented model code is readily available to the EPA and the public. |

| Criteria | Example information |
|---|---|
| **Initial technical** | A set of published parameters is clearly identified, including origin/derivation. |
| | Parameters do not vary unpredictably with dose (e.g., any dose dependence in absorption constants is predictable across the dose ranges relevant for animal and human modeling). |
| | Sensitivity and uncertainty analysis have been conducted for relevant exposure levels (local sensitivity analysis is sufficient, but a global analysis provides more information).<br>• If a sensitivity analysis was not conducted, EPA may decide to independently conduct this additional work before using the model in the assessment.<br>• A sound explanation should be provided when sensitivity of the dose metric to model parameters differs from what is reasonably expected based on experience. |

### 6.6.3. Selection of the Appropriate Dose Metric

The level of confidence in using a pharmacokinetic (PK) or PBPK model depends on its ability to provide a reliable estimation of dose metrics based on biological plausibility and MOA considerations. Thus, one needs to take into consideration mechanism(s) relevant to the endpoint(s) of interest, data availability and uncertainties in estimating that dose metric. For nitrates and nitrites, hemoglobin is an established target of toxicity, although other toxicities might exist. Existing noncancer reference values for nitrate are derived from its transformation to nitrite and resulting risk for methemoglobinemia. An existing model for nitrate exposure (Zeilmaker et al., 2010) includes the nitrite-dependent transformation of hemoglobin to methemoglobin mechanism of action. Therefore, the production of methemoglobin from nitrite will serve as the dose metric for the methemoglobinemia endpoint.

Compared to methemoglobin production, it remains less understood what the appropriate dose metric for other toxicities should be. N-nitrosamines, formed via N-nitrosation, are considered strong carcinogens. Absent a model predicting the formation of N-nitrosamines from parent compounds, surrogate dose metrics such as nitrate/nitrite (average) daily area under the curve will be evaluated for this toxicity. If required, the addition of an N-nitrosamine pathway could be included in existing models if the appropriate pharmacokinetic data exists.

# 7. DATA EXTRACTION OF STUDY METHODS AND RESULTS

The process of summarizing study methods and results is referred to as data extraction. Studies that met initial PECO criteria after full-text review are briefly summarized in data extraction forms available in HAWC. These study summaries are used to create interactive literature inventory visualizations to display the extent and nature of the available evidence in HAWC.

For experimental animal studies, which are typically studies in rodents, the following information is captured: chemical form, study type (acute [<24 hours], short term [<7 days], short term [7–27 days], subchronic [28–90 days], chronic [>90 days[3]] and developmental, which includes multigeneration studies), duration of treatment, route, species, strain, sex, dose or concentration levels tested, dose units, health system and specific endpoints assessed.

For human studies, the following information is summarized in HAWC data extraction forms: chemical form, population type (e.g., general population-adult, occupational, pregnant women, infants, and children), study type (e.g., cross-sectional, cohort, case-control), sex, major route of exposure (if known), description of how exposure was assessed, health system studied, and specific endpoints assessed.

For epidemiology and animal studies that met the assessment PECO criteria, HAWC is used for study evaluation and for full extraction of study methods and results. Compared with the literature inventory, full data extraction in HAWC includes summarizing more details of study design and gathering effect size information. Instructions on how to conduct data extraction in HAWC are available at https://hawcproject.org/resources/. An additional resource used to implement use of a consistent vocabulary to summarize endpoints assessed in animal studies is available in HAWC (the Environmental Health Vocabulary; https://hawc.epa.gov/vocab/ehv/).

In some cases, EPA may conduct their 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.

Data extraction for in vivo and in vitro studies prioritized to assess mechanisms of nitrate/nitrite is conducted in Microsoft Word and presented in tabular format.

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

[3]EPA considers chronic exposure to be more than approximately 10% of the life span in humans. For typical laboratory rodent species, this can lead to consideration of exposure durations of approximately 90 days to 2 years. However, studies in duration of 1–2 years are typical of what is considered representative of chronic exposure rather than durations just over 90 days.

qualitative). For quality control, studies were summarized by one member of the evaluation team and independently verified by at least one other member. Discrepancies were 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. The literature inventories are presented in the HAWC Visualization module, with options to link to the native Tableau application where the underlying information is available for download. Download of full data extraction for animal studies is done directly in HAWC.

For non-English studies online translation tools (e.g., Google translator) or engagement with a native speaker can be used to summarize 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 hazard identification or dose-response to facilitate study evaluation and 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.

# 8. EVIDENCE SYNTHESIS AND INTEGRATION

Within-stream evidence synthesis is conducted separately for human, animal, and mechanistic evidence to directly inform the integration across the streams of evidence and draw overall conclusions for each of the assessed human health effects. 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). A structured framework approach is used to guide both evidence synthesis and integration. This structured framework includes consideration of mechanistic information during both evidence synthesis and integration, although the focus of the analysis differs. Similarly other types of supplemental information (e.g., ADME, non-PECO route of exposure) can also inform evidence synthesis and integration analyses.

- Evidence synthesis: Judgment(s) regarding the strength of the evidence for hazard for each unit of analysis from the available human and animal studies are made in parallel, but separately. These judgments can incorporate PK, mechanistic, and other supplemental evidence when the unit of analysis is defined as such (see Section 5.2). The units of analysis can also include or be framed to focus on precursor events (e.g., biomarkers). In addition, this includes 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 streams, potential susceptibility, and other critical inferences (e.g., biological plausibility) 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, a synthesis (strength of 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, evidence synthesis judgments are used to inform evidence integration (weight of evidence) 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[4] typically focus on the unit of analysis with the strongest evidence synthesis judgments, although exceptions may occur. 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").

[4]In some cases, as discussed in Section 8.2, it will 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 predefined manifestations of potential toxicity. For example, within the health effect category of immune effects, separate and different evidence integration judgments might be appropriate for immunosuppression, 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 (1991a)). 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 (strength of evidence) 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):<br><br>⊕⊕⊕ *Evidence demonstrates*<br>⊕⊕⊙ *Evidence indicates (likely)*<br>⊕⊙⊙ *Evidence suggests*<br>⊙⊙⊙ *Evidence inadequate*<br>*– – – Strong evidence supports no effect*<br><br>Highlight the primary supporting evidence for each integration judgment[a]<br><br>Present inferences and conclusions on:<br><br>• Human relevance of findings in animals[a]<br>• Cross-stream coherence[a]<br>• Potential susceptibility[a]<br>• Understanding of biological plausibility and MOA[a]<br>• Other critical inferences[a] |
| **Evidence from human studies** | | | | | |
| • Unit of analysis #1<br>• Studies considered and study confidence | • Description of the primary results | • All/Mostly *medium* or *high* confidence studies<br>• Consistency<br>• Dose-response gradient<br>• Large or concerning magnitude of effect<br>• Coherence[a] | • All/Mostly *low* confidence studies<br>• Unexplained inconsistency<br>• Imprecision<br>• Concerns about biological significance[a]<br>• Indirect outcome measures[a]<br>• Lack of expected coherence[a] | • **Judgment reached for each unit of analysis[a]**<br>⊕⊕⊕ *Robust*<br>⊕⊕⊙ *Moderate*<br>⊕⊙⊙ *Slight*<br>⊙⊙⊙ *Indeterminate*<br>*– – – Compelling evidence of no effect* | |
| • Unit of analysis #2<br>• Studies considered and study confidence | • Description of the primary results | | | | |
| **Evidence from animal studies** | | | | | |
| • Unit of analysis #1<br>• Studies considered and study confidence | • Description of the primary results | • All/Mostly *medium* or *high* confidence studies<br>• Consistency<br>• Dose-response gradient<br>• Large or concerning magnitude of effect<br>• Coherence[a] | • All/Mostly l*ow* confidence studies<br>• Unexplained inconsistency<br>• Imprecision<br>• Concerns about biological significance[a]<br>• Indirect outcome measures[a]<br>• Lack of expected coherence[a] | • **Judgment reached for each unit of analysis**<br>⊕⊕⊕ *Robust*<br>⊕⊕⊙ *Moderate*<br>⊕⊙⊙ *Slight*<br>⊙⊙⊙ *Indeterminate*<br>*– – – Compelling evidence of no effect* | |
| • Unit of analysis #2<br>• Studies considered and study confidence | • Description of the primary results | | | | |

[a]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 can be presented separately, e.g., by exposure route or key uncertainty addressed.<br><br>Each analysis can include multiple rows separated by biological events or other feature of the approach used for the analysis.<br>• Generally, will cite mechanistic synthesis (e.g., for references, for detailed analysis)<br>• Does not have to be chemical-specific (e.g., read-across) | Can include separate summaries, for example by study type (e.g., new approach methods vs. in vivo biomarkers), dose, or design.<br><br>*Interpretation*: Summary of expert interpretation for the body of evidence and supporting rationale<br><br>*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., adverse outcome pathway).<br>• Includes the primary evidence supporting the interpretation(s)<br>• Describes and informs the extent to which the evidence influences inferences across evidence streams<br>• Characterizes the limitations of the evaluation and highlights existing data gaps<br>• May have overlap with factors summarized for other streams |

1

## 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 as evidence for hazard (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 [RoB] 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 6-3; see additional discussion in (U.S. EPA, 2022, 2005a, 1994)]. These factors are similar to the domains considered in the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) Quality of Evidence framework (Schünemann et al., 2013). Each of the considered factors and the certainty of evidence judgments requires 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).

As previously described, the units of analysis may include predefined categories of mechanistic evidence or other supplemental information (e.g., from studies of non-PECO routes of exposure). This may include consideration of biomarkers or precursor events. Biological understanding (e.g., knowledge of how an effect is manifest or progresses) or mechanistic inference (e.g., dependency on a conserved key event across outcomes) can also be used to define which related outcomes are considered as a unit of analysis. These considerations also inform the evaluation of coherence and adversity within a unit of analysis and coherence with other units of analyses. Mechanistic analyses outside the context of defining and evaluating the units of analysis during evidence synthesis are considered as part of across stream evidence integration (see Section 8.2).

Typically, human and animal evidence synthesis sections are structured similarly across different units of analysis, health effects, and assessments. In contrast, the presentation, and analyses of mechanistic and other types of supplemental information often differs within and across assessments. This is due to the diversity of supplemental data that may be available and the complexity of conducting supplemental analyses. For example, these data may inform unit of analysis considerations, evidence integration judgments, or both. Each of the key analyses informing the synthesis judgments are described in the narrative and summarized in an evidence profile table.

Five levels of certainty in the evidence for (or against) 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 Table 8-4

and Table 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. Observations that increase certainty are having consistency across *high* or *medium* confidence studies or experiments, the presence of *medium* or *high* confidence studies with a strong dose-response gradient or observing a large or concerning magnitude of effect, and coherent findings across *medium* or *high* confidence studies for closely related endpoints (can include mechanistic endpoints) within the unit of analysis within an evidence stream. Evidence from *low* confidence studies can further strengthen observations from *medium* or *high* confidence studies but do not increase certainty on their own. Observations that decrease certainty are having an evidence base of mostly *low* confidence studies, unexplained inconsistency, lack of expected coherence, imprecision, unclear biological significance, null findings with concerns for insensitivity (which decreases certainty in the lack of an effect), or indirect measures of outcomes. Table 8-3 provides additional detail on how these factors are considered when evaluating units of analysis.

**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 and 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.<br>• This factor should not be used if no other factors would increase or decrease the confidence for a given unit of analysis.<br>• 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 strength. 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 in which an association is identified because the expected impact of study insensitivity is toward the null.<br>• An evidence base of mostly null findings in which insensitivity is a serious concern decreases certainty that the evidence is sufficient to support a lack of health effect or association.<br>• 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 | • 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 might be hard to assess consistency. An evidence base of a single or a few studies in which consistency cannot be accurately assessed does not, alone, increase or decrease evidence certainty. Similarly, a reasonable explanation for inconsistency does not necessarily result in an increase in evidence certainty. |

| 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]) |
|---|---|---|
| | given the most attention during evidence synthesis. | |
| **Effect magnitude and imprecision** | • Evidence of a large or concerning magnitude of effect can increase strength (generally only when observed in *medium* or *high* confidence studies).<br>• Judgments on effect magnitude and imprecision consider the rarity and severity of the effect. | • Certainty could be decreased if the findings are considered not likely to be biologically significant. Effects that are small in magnitude might not be considered biologically significant (adverse[b]) based on information such as historical responses and variability. However, effects that appear to be of small magnitude could be meaningful at the population level e.g., IQ shifts); I such cases, certainty would not be decreased.<br>• Certainty might also be decreased for imprecision, particularly if there are only a few studies available to evaluate consistency in effect magnitude across studies. |
| **Dose-response** | • Evidence of dose-response or exposure-response in *high* or *medium* confidence studies increases certainty. Dose-response can be demonstrated across studies or within studies and it can be dose- or duration-dependent. It could also not be a monotonic dose-response (monotonicity should not necessarily be expected as different outcomes might be expected at low vs. high doses 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.<br>• Decreases in a response (e.g., symptoms of current asthma) after a documented cessation of exposure also might 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 on the basis of biological understanding can decrease certainty in the evidence. If the data are not adequate to evaluate a dose-response pattern, however, certainty is neither increased nor decreased.<br>• 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 (1998)], endpoint severity, judgments regarding the potential for delayed or secondary effects, the underlying mechanism(s) involved, and the exposure context focus of the assessment (e.g., addressing intermittent or short-term exposures). |

| 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]) |
|---|---|---|
| **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 for measures that have an unclear linkage to an apical or clinical (adverse[b]) outcome. Scenarios in which 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 of outcome measures.<br>• Related to indirectness, certainty in the evidence can 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). |
| **Coherence** | • Biologically related findings within or across studies, within an organ system or across populations (e.g., sex), increase certainty (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.<br>• Coherence across findings within a unit of analysis (e.g., consistent changes in disease markers and biological precursors in exposed | • An observed lack of expected coherent changes (e.g., in well-established biological relationships) within or across biologically related units of analysis will typically decrease evidence certainty. This includes mechanistic changes when included in the unit of analysis. However, as described for decisions to increase certainty, confidence in the understanding of 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 certainty 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. |

| 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]) |
|---|---|---|
| | humans) can increase certainty in the evidence for an effect.<br>• Coherence within or across biologically related units of analysis can also increase certainty 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.<br>• 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. | |
| **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]Although 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 could 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 8-4 and 8-5 (for human and animal evidence, respectively) provide the 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 a conclusion 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 increased certainty in the evidence for hazard. 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 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**

| Evidence synthesis judgment | Description |
|---|---|
| *Robust* (⊕⊕⊕) …evidence in human studies<br><br>*(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) could raise the certainty in the 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<br><br>*(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 certainty in the evidence. For multiple studies, there is primarily consistent evidence of an association with reasonable support for adversity, but there might be some uncertainty due to potential chance, bias, or confounding or because of the indirectness of some measures. When only a single study is available in the unit of analysis, 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. Supplemental evidence included in the unit of analysis might address the above factors and raise certainty in the evidence to *moderate* for a set of studies that otherwise would be described as *slight* or, in exceptional cases, could support raising to *moderate* evidence that would otherwise be described as *indeterminate*. Mechanistic 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. |
| *Slight* (⊕⊙⊙) …evidence in human studies<br><br>*(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 within the unit of analysis 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 when 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*. |

| Evidence synthesis judgment | Description |
|---|---|
| *Indeterminate* (⊙⊙⊙) …evidence in human studies<br><br>*(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 might include situations in which 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<br><br>*(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 (e.g., 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. Supplemental evidence can help to address the above considerations or, when included in the unit of analysis, provide additional support for this judgment. |

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

| Evidence synthesis judgment | Description |
|---|---|
| *Robust* (⊕⊕⊕) …evidence in animal studies<br><br>*(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 certainty in the evidence: coherent effects across multiple related endpoints (within or across biologically related units of analysis); 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. Supplemental evidence included in the unit of analysis (e.g., mechanistic studies in exposed animals or animal cells) might raise the certainty 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 animal evidence, the directness of the outcome measures, and the biological significance of the findings. |

| Evidence synthesis judgment | Description |
|---|---|
| *Moderate* (⊕⊕⊙) …evidence in animal studies<br><br>*(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 and additional information increasing certainty in the evidence. For multiple studies or a single study, the evidence is primarily consistent or coherent with 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 certainty in the evidence, such as consistent effects across laboratories or species; coherent effects across multiple related endpoints (can include mechanistic endpoints within the unit of analysis); an unusual magnitude of effect, rarity, age at onset, or severity; a strong dose-response relationship; or consistent observations across exposure scenarios (e.g., route, timing, duration), sexes, or animal strains. Supplemental evidence included in the unit of analysis could address the above factors and raise certainty in the evidence to *moderate* for a set of studies that otherwise would be described as *slight* or, in exceptional cases, might support raising to *moderate* evidence that would otherwise be described as *indeterminate*. Mechanistic evidence not included in the unit of analysis can also inform evaluations of the coherence of the animal evidence, the directness of the outcome measures, and the biological significance of the findings. |
| *Slight* (⊕⊙⊙) …evidence in animal studies<br>*(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 within the unit of analysis without additional information increasing certainty in the evidence (e.g., coherent findings within the same study or from other studies). Biological evidence from mechanistic studies could also be independently interpreted as s*light*. This category serves primarily to encourage additional study for which 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<br>*(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 might include situations in which 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 c*ompelling evidence of no effect*. |
| *Compelling evidence of no effect* (- - -) …in animal studies<br>*(strong signal for lack of an effect with little uncertainty)* | 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. Supplemental evidence can help to address the above considerations or, when included in the unit of analysis, 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 or MOA, and potentially 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 during problem formulation). To the extent it can be characterized prior to conducting dose-response analyses, exposure context is also provided.

With respect to susceptibility, the assessment describes the evidence (i.e., human, animal, mechanistic) on populations and lifestages most likely to be susceptible to the hazards identified and, to the extent possible, the factors that increase their risk for the hazards. In addition to assessment-specific health effects evidence, background information about biological mechanisms or ADME, as well as biochemical and physiological differences among lifestages and sexes, may be used. At a minimum, particular consideration is given to infants and children, pregnant women, and women of childbearing age. Many of the foundational analyses for summarizing susceptibility in the evidence integration narrative are undertaken during evidence synthesis as patterns across studies are evaluated with respect to consistency, coherence, and the magnitude and direction of effect measures. Relevant factors for exploring patterns may include intrinsic factors (e.g., age, sex, genetics, health status, behaviors) and certain extrinsic factors (e.g., socioeconomic status, access to health care), although information on the latter is rarely available in human health studies of environmental chemicals.

**Table 8-6. Considerations that inform evidence integration judgments**

| Judgment | Description |
|---|---|
| **Human relevance of findings** | Used to describe and justify the interpreted relevance of the data from experimental animals (or other model systems) to humans. In the absence of chemical-specific evidence informing human relevance, the evidence integration narrative will briefly describe the interpreted underlying biological similarity across species. 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. Thus, in the absence of specific evidence or cross-species understanding of the underlying biology, it is appropriate to use a statement such as, "without evidence to the contrary, [health effect] responses in animals are presumed relevant to humans." |
| **Cross-stream coherence** | Used to address the concordance of biologically related findings across human, animal, and mechanistic studies, considering features of the available evidence such as exposure timing and cancer), it is not necessary or expected that effects manifest in humans are identical to those observed in animals (e.g., tumors in animals can be predictive of carcinogenic potential |

| Judgment | Description |
|---|---|
| | in humans, but not necessarily at the same site), although this typically provides stronger evidence. Biological understanding of the manner in which the outcomes are manifest in different species can inform cross-stream coherence. Evidence supporting a biologically plausible mechanistic pathway across species adds coherence (see below). |
| **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 on the influence of intrinsic biological factors but can also include consideration of mechanistic and ADME evidence. |
| **Biological plausibility and MOA considerations** | Used to summarize the interpreted biological plausibility of an association between exposure and the health effect, based primarily on the extent to which the available evidence comports with the known development and characteristics of the health effect (and thus dependent on sufficient information being available to draw such an interpretation). Importantly, because this interpretation is dependent on canonical scientific knowledge about the health effect, the lack of such understanding does not provide a rationale to decrease certainty in the evidence for an effect (NTP, 2015; NRC, 2014). These analyses can be detailed (e.g., when attempting to establish MOA understanding) and, if so, are typically conducted separately (e.g., as part of the mechanistic evidence synthesis) and then referenced in the evidence integration narrative. |
| **Other critical inferences (optional)** | Can be used to describe the consideration of other evidence or non-chemical-specific information that informs evidence integration judgments (e.g., use of read-across analyses or ADME understanding used to inform the other considerations described below; judgments on other health effects expected to be linked to the health effect under evaluation). |

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

Using a structured framework approach, one of five phrases is used to summarize the evidence integration judgment based on the integration of the evidence synthesis judgments, taking into account the additional considerations assessed across evidence streams: *evidence demonstrates*, *evidence indicates (likely)*, *evidence suggests*, *evidence is inadequate*, or *strong evidence supports no effect* (see Table 8-7). The five evidence integration judgment levels reflect the differences in the amount and quality of the data that inform the evaluation of whether exposure is interpreted as capable of causing the health effect(s). As it is assumed that any identified health hazards will only be 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 (see Section 8). More than one evidence integration judgment level 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).

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 judgments (i.e., "***evidence demonstrates***" aligns with "carcinogenic to humans") but not in all cases. For example, the evidence integration judgments 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), including application of the MOA framework (incorporating an evaluation of evidence relevant to potential mutagenicity).

Similar to the description for summarizing noncancer judgments above, the cancer descriptor and evidence integration narrative for carcinogenicity also consider the conditions of carcinogenicity, including exposure (e.g., route; level) and susceptibility (e.g., genetics; lifestage), as the data allow (Farland, 2005; U.S. EPA, 2005a, b). As with noncancer effects, the specific exposure conditions necessary for carcinogenicity are further defined during dose-response analysis.

**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.<br>• This conclusion level is used if there is *robust* human evidence supporting an effect.<br>• 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.<br>• 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.<br>• 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 given sufficient exposure conditions. This conclusion is based on studies of [humans or animals] that assessed [exposure or dose] levels | **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.<br>• This conclusion level is used if there is *slight* human evidence and *indeterminate*-to-*slight* animal evidence. |

| Summary evidence integration judgment[a] in narrative | Evidence integration judgment level | Explanation and example scenarios[b] |
|---|---|---|
| of [range of concentrations or specific cutoff level concentration]. | | • This conclusion level is also used with *slight* animal evidence and *indeterminate*-to-*slight* human evidence.<br>• 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.<br>• 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]<br>• This conclusion level is used if there is *indeterminate* human and animal evidence.<br>• This conclusion level is also used with *slight* animal evidence and *compelling evidence of no effect* human evidence.<br>• 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.<br>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. |

| Summary evidence integration judgment[a] in narrative | Evidence integration judgment level | Explanation and example scenarios[b] |
|---|---|---|
| **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.<br>• 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.<br>• 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.<br>• 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," are 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 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.

# 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, 2012), 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[5] to the chemical of interest, if supported by existing data. For noncancer hazards, an oral reference dose (RfD) will be derived. (Inhalation toxicity values will not be derived in this assessment of nitrate/nitrite.) An 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). In addition to an RfD, this assessment will attempt to derive organ- or system-specific RfDs (osRfDs) when the data are sufficiently strong (i.e., noncancer conclusions of *evidence demonstrate* or *evidence indicates [likely]*). An RfD may also be derived for cancer effects in cases in which a nonlinear MOA is concluded that indicates a key precursor event necessary for carcinogenicity does not occur below a specific exposure level ((U.S. EPA, 2005a), §3.3.4). In addition to an RfD, 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.

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

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 OSF 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 OSF is a plausible upper-bound lifetime cancer risk from chronic ingestion of a chemical (expressed as mg/kg-day). In contrast with reference doses (RfDs), an OSF can be used in conjunction with exposure information to estimate cancer risk at a given dose.

The derivation of toxicity values also 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 in which suggestive evidence might be used to develop cancer risk estimates or noncancer toxicity value include when the evidence base includes a well-conducted study (overall *medium* or *high* confidence for the outcome), 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 life stages, (3) how dose response modeling will be informed by pharmacokinetic information, and (4) the identification of biologically based BMR levels. 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 the draft and final assessments, specific endpoints considered for dose-response are summarized in a tabular format that includes rationales for decisions to proceed (or not) for POD derivation.

In addition, mechanistic evidence that influences the dose-response analyses is 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 is summarized including any interactions between them relevant to understanding overall risk. For cancer dose-response, 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**

| Study attributes | | Considerations | |
|---|---|---|---|
| | | **Human studies** | **Animal studies** |
| Study confidence | | *High* or *medium* confidence studies are highly preferred over *low* confidence studies. The selection of low confidence studies should include an additional explanatory justification (e.g., only low confidence studies had adequate data for toxicity value derivation). The available *high* and *medium* confidence studies are further differentiated on the basis of 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, dose-response pattern in relevant dose range, relevance of specific health outcomes to humans). | Animal studies provide supporting evidence when adequate human studies are available, and they are considered the studies of primary interest 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*, §2.1.1) and facilitate extrapolation to more relevant (generally lower) exposures. | |
| Subject selection | | Studies that provide risk estimates in the most susceptible groups are preferred. | |
| Controls for possible confounding[a] | | Studies with a design (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. |
| 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. For example, individual data allow you to characterize experimental variability more realistically and 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 outcomes with less concern for indirectness or with greater biological significance. | |
| Study size and design | Preference is given to studies using designs reasonably expected to have power to detect responses of suitable magnitude.[b] This does not mean that studies with substantial responses but low power would be ignored, but that they should be interpreted in light of a confidence interval or variance for the response. Studies that address changes in the number at risk (through decreased survival, loss to follow-up) are preferred. | |

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

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

## 9.3. CONDUCTING DOSE-RESPONSE ASSESSMENTS

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

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

2) 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.

### 9.3.1. Dose-Response Analysis in the Range of Observation

For conducting a dose response assessment, pharmacodynamic ("biologically based") modeling can be used when there are sufficient data to ascertain the mode of action and quantitatively support model parameters that represent rates and other quantities associated with the key precursor events of the modes of action. If there is not an applicable pharmacodynamic model available to assess health effects associated with ingestion exposure to nitrate/nitrite, 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 (http://www.epa.gov/ncea/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. The 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, 2012)].

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, 2012). 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. Modeling may also need to accommodate cases of nonlinear dose-response data.

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 for which 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).[6]

Version 3.2 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. DMA may be applied to nitrate/nitrite as a supplemental analysis; see the section on Supplemental Dose-Response Analyses below for details.

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, and for nonlinear extrapolation, the POD is used in calculating an RfD.

[6]The Multistage degree selection process outlined in the memo is auto-implemented in the BMDS multitumor model, which can be run on one or more tumor data sets, but only the noncancer model selection process is auto-implemented for individual Multistage model runs in the current version, BMDS 3.2).

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 POD 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 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, 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 POD 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), §3.1.1; (U.S. EPA, 1991a), §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, 2011, 2005a), §3.1.3).

- 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), §3.1.4).

- In the absence of study specific data on, for example, intake rates or body weight, the 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. Elements of more than one published model can be combined if the effort involved is minimal and no one model has all the features desired.

- Briefly, PBPK model simulations are 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, the resulting internal metric

effectively accounts for the difference between the pattern and a nominal daily 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 POD for individual animal toxicity studies. In this assessment, the internal dose metric is either the tissue-specific rate of oxidative metabolism or a daily average blood concentration. The human version of the PBPK model can then be used to estimate the exposure dose that would result in internal dose at the 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.

### 9.3.2. Extrapolation: Slope Factors and Unit Risk

An OSF 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).

An inhalation unit risk (IUR) was not included in the scope for this assessment.

### 9.3.3. 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). In general, it has been the IRIS program's preference to base cancer reference values on key precursor events in the MOA that are necessary for tumor formation rather than on the incidence of tumors themselves. For example, see the ethylene glycol monobutyl ether assessment in which the cancer RfD was based on hemosiderin deposition in the liver vs. liver tumor incidence (HEROID: 4442193).

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). Increasingly, data-based adjustments (U.S. EPA, 2014) and Bayesian methods for characterizing population variability (NRC, 2014) are feasible and may be distinguished from the uncertainty factor (UF) considerations outlined below. The assessment will discuss the scientific bases for estimating these data-based adjustments and UFs:

- *Animal-to-human extrapolation*: If animal results are used to make inferences about humans, the reference value derivation incorporates the potential for cross-species

differences, which may arise from differences in pharmacokinetics or pharmacodynamics. If available, a biologically based model that adjusts fully for pharmacokinetic and pharmacodynamic differences across species may be used. Otherwise, the POD is standardized to equivalent human terms or is based on pharmacokinetic or dosimetry modeling, which may range from detailed chemical-specific to default approaches (U.S. EPA, 2014, 2011), and a factor of $10^{1/2}$ (rounded to 3) is applied to account for the remaining uncertainty involving pharmacokinetic and pharmacodynamic differences.

- *Human variation*: The assessment accounts for variation in susceptibility across the human population and the possibility that the available data may not represent individuals who are most susceptible to the effect, by using a data-based adjustment or UF or a combination of the two. Where appropriate data or models for the effect or for characterizing the internal dose are available, the potential for data-based adjustments for pharmacodynamics or pharmacokinetics is considered (U.S. EPA, 2014, 2002).[7] [8] When sufficient data are available, an intraspecies UF either less than or greater than 10-fold may be justified (U.S. EPA, 2002). This factor may be reduced if the POD is derived from or adjusted specifically for susceptible individuals [not for a general population that includes both susceptible and non-susceptible individuals; ((U.S. EPA, 2002), §4.4.5;(U.S. EPA, 1998), §4.2;(U.S. EPA, 1996),§4;(U.S. EPA, 1994), §4.3.9.1;(U.S. EPA, 1991a),§3.4)]. When the use of such data or modeling is not supported, a UF with a default value of 10 is considered.

- *LOAEL to NOAEL*: If a POD is based on a LOAEL, the assessment includes an adjustment to 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. A factor of 3 or 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, 1998, 1996, 1994, 1991a).

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

- *Database deficiencies*: In addition to the adjustments above, if database deficiencies raise concern that further studies might identify a more sensitive effect, organ system, or life stage, the assessment may apply a database UF (U.S. EPA, 2002, 1998, 1996, 1994, 1991a). The size of the factor depends on the nature of the database deficiency. For

[7]Examples of adjusting the pharmacokinetic portion of interhuman variability include the IRIS boron assessment's use of nonchemical-specific kinetic data [e.g., glomerular filtration rate in pregnant humans as a surrogate for boron clearance (U.S. EPA, 2004)] and the IRIS trichloroethylene assessment's use of population variability in trichloroethylene metabolism, via a PBPK model, to estimate the lower 1st percentile of the dose metric distribution for each POD (Mina et al., 2021).

[8]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.

example, the EPA typically follows the recommendation that a factor of 10 be applied if both a prenatal toxicity study and a two-generation reproduction study are missing and a factor of $10^{1/2}$ (i.e., 3) if either one or the other is missing ((U.S. EPA, 2002), §4.4.5).

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

# APPENDIX A. SYSTEMATIC EVIDENCE MAP FOR HEALTH EFFECTS OF NITRATES AND NITRITES

## A.1. INTRODUCTION

This systematic evidence map (SEM) was developed based on the IRIS Assessment Plan (IAP) developed for nitrate and nitrite. Nitrate and nitrite are considered together, as both are chemically related and metabolically linked, and their biological effects are determined by conversion of nitrate to nitrite and vice versa. Review of the health effect literature for both chemicals in a single health assessment also follows the approach taken by other health agencies (CalEPA, 2018; ATSDR, 2017; WHO, 2016; Water and Air Quality Bureau, 2013; IARC, 2010; IPCS, 2005). More specifically, this SEM includes information for the six inorganic forms of nitrate and nitrite listed in Table 4 of the Protocol, comprising: ammonium nitrate, sodium nitrate, sodium nitrite, potassium nitrate, potassium nitrite, and calcium nitrate. These nitrate and nitrite salts are the most common in the environment (ATSDR, 2017). These salts are highly soluble in water and dissociate under environmental conditions; in solution, they exist as ions (ATSDR, 2017). Because the cations are not expected to introduce significant differences in the toxicity of the different salts, toxicity findings from all six compounds are considered relevant to an assessment of nitrate and nitrite toxicity.

## A.2. METHODS

The systematic review methods used to conduct the evidence map are described in the Protocol document and follow the Office of Research and Development (ORD) Staff Standard Operating Procedures for Developing Integrated Risk Information System (IRIS) Assessments (Version 2.0, referred to as the "IRIS Handbook") (U.S. EPA, 2022).

### A.2.1. Specific Aims

The specific aims for the SEM are presented below:

- Identify epidemiological (i.e., human) and toxicological (i.e., experimental animal) literature reporting health effects of exposure to nitrates and nitrites as outlined in the problem formulation populations, exposures, comparators, and outcomes (PECO) criteria (shown in Table 4-2 of the Protocol).

- Identify supplemental material as outlined in Table 4-2 of the Protocol. Supplemental material content includes mechanistic studies; non-PECO-relevant species/model systems; toxicokinetic and *absorption, distribution, metabolism, and excretion* (ADME)

studies; pharmacokinetic (PK) or physiologically based pharmacokinetic (PBPK) model studies; exposure characteristics (no health outcome); human exposure biomarker studies with health outcome; mixture studies; routes of exposure not pertinent to the PECO; case studies; records with no original data; and conference abstracts.

- Create a literature inventory of PECO-relevant studies. The literature inventory summarizes basic features of study design, and health system(s) assessed.

- Provide an overview of the evidence base, including the degree to which it supports conducting a formal assessment for the effects of nitrates and nitrites on the specified health effect categories.

### A.2.2. Literature Search and Screening Strategies

***Survey of Existing Regulatory Toxicity Values***

Toxicity value is a broad term that encompasses reference values, probabilistic risk estimates (i.e., slope factors and unit risk estimates), and assessment-based points of departure (PODs). The term reference value applies to values designed to provide a "benchmark" or exposure limit from which some level of protection to human life and health can be inferred. Reference values are the most common final output from the dose-response assessment component of the risk assessment paradigm set forth by the National Research Council (NRC, 2009) and are based on an observed or estimated threshold for an effect, usually noncancer.

Health-based reference values for noncancer effects are presented either in units of concentration (e.g., mg/L) or in terms of dose (e.g., milligrams per kilogram of body weight per day, mg/kg-day). Reference values generally are derived by applying uncertainty and adjustment factors to the exposure/dose level that elicits an effect observed in studies with human subjects or in controlled animal experiments, the POD. The derivation methods and factors used in moving from a POD to a final reference value vary according to the organization developing the values, often with consideration of how the resulting values will be applied. Oral reference values often are used as the basis for deriving standards for drinking water or acceptable levels in food.

Probabilistic risk estimates are most often developed for cancer effects when the default assumption is that there is no level of exposure without some effect (i.e., non-threshold effects); however, probabilistic approaches to estimate ranges for noncancer effect levels have also been developed (Blessinger et al., 2020). Probabilistic risk estimates are used to determine exposure levels associated with an acceptable risk range (e.g., less than one-in-a-million probability for risks above background for an adverse health effect). Assessment-based PODs are identified using the same process as used in the derivation of reference values and are used in evaluations of risk when specific conditions of use are part of a decision process to determine exposure or consumption levels associated with acceptable level of risk.

A visual representation was developed to illustrate the available toxicity values for oral exposure to nitrate/nitrite (see Figure 2-1 of the Protocol). The information displayed on this graphical array of toxicity values was collected from searches of a number of authoritative sources;

these sources, cited in Appendix B, were manually searched for health risk assessments for the oral route of exposure. In addition to these sources, the ToxVal database on the EPA Chemicals Dashboard (https://comptox.epa.gov/dashboard/chemical_lists/TOXVAL_V5) was searched for reference values, risk estimate values, and PODs as described in Appendix C.

### A.2.3. Literature Inventory

The literature search and screening methods are described in Section 4 of the Protocol document. Human and animal studies that met problem formulation PECO criteria after full-text review were briefly summarized using data extraction forms in the Health Assessment Workspace Collaborative (HAWC; hawc.epa.gov). These study summaries are referred to as literature inventories and are used to create interactive visualizations.

For animal studies, the following information was captured: chemical assessed, study type (acute [<24 hours], short term [1–30 days], subchronic [30–90 days], chronic [>90 days, multigenerational, peripubertal, developmental]), duration of treatment, route, species, strain, sex, dose, or concentration levels tested, dose units, health system and specific endpoints assessed. For epidemiological studies, the following information was summarized: chemical assessed, population type (e.g., general population-adult, occupational, pregnant women, infants, and children), study type (e.g., cross-sectional, cohort, case-control), sex, major route of exposure (if known), health system and specific outcomes assessed. Summaries were extracted into HAWC by one team member and the extracted data were quality checked by at least one other team member.

## A.3. RESULTS

### A.3.1. Available Health Values

The available health values are shown in Table A-1 and Figure 2-1 of the Protocol. The IRIS program currently does not include cancer risk values for nitrate or nitrite. The International Agency for Research on Cancer (IARC) has determined that there is "inadequate" evidence of carcinogenicity of nitrate in food or drinking water, "limited" evidence for the carcinogenicity of nitrite in food, and "sufficient" evidence for the carcinogenicity of nitrite in combination with amines or amides. IARC concludes that "ingested nitrate and nitrite under conditions that result in endogenous nitrosation is probably carcinogenic to humans (Group 2A)" (IARC, 2010).

The IRIS program lists reference dose (RfD) values of 1.6 mg/kg-day for nitrate and 0.1 mg/kd-g-day for nitrite, based on a critical effect of methemoglobinemia. ATSDR has determined minimal risk levels (MRLs) of 4 mg/kg-day for nitrate and 0.1 mg/kg-day for nitrite (applicable for acute, intermediate, and chronic durations of oral exposure) based upon the same health endpoint (ATSDR, 2017). The Joint FAO/WHO Expert Committee on Food Additives (JECFA) has also determined acceptable daily intake (ADI) values of 3.7 mg/kg-day for nitrate and 0.07 mg/kg-day for nitrite (based on heart and lung effects in rats) (WHO, 2003; JECFA, 1995).

The EPA's maximum contaminant levels for nitrate and nitrite are 10 mg/L (or ppm) and 1 mg/L (or ppm), respectively. These are equivalent to ~44 mg nitrate/L as nitrate-nitrogen and ~3.3 mg nitrite/L as nitrite-nitrogen. California's Office of Environmental Health Hazard Assessment lists public health goals (PHGs) of 45 mg/L and 3 mg/L for nitrate and nitrite, respectively (the joint nitrate/nitrite PHG is 10 mg/L) (CalEPA, 2018). The FDA uses these same values for allowable levels in bottled water (FDA, 2021), and these are also the same values that Health Canada has determined for maximum allowable concentration values (Water and Air Quality Bureau, 2013).

**Table A-1. Details on derivation of the available health effect reference values for oral exposure to nitrate and nitrite**

| Reference value name | Chemical form | Duration | Reference value | Health effect | Point of departure | Qualifier | Source | Uncertainty/ modifying factors | Notes on derivation | Review status |
|---|---|---|---|---|---|---|---|---|---|---|
| **EPA RfD (IRIS)**[a] | Nitrate | Chronic | 1.6 mg N/kg-d | Early clinical signs of methemoglobinemia in infants | 10 mg nitrate-nitrogen/L | NOAEL | Bosch et al. (1950) and Walton (1951) | Total UF = 1 | Dose calculated[b] | Final U.S. EPA (1991b) |
| | Nitrite | | 0.1 mg N/kg-d | | 10 mg N/L | NOEL | Walton (1951) | Total UF = 1 $MF^{c}$ = 10 | Dose calculated[d] | Final U.S. EPA (1987) |
| **EPA p-RfD (HEAST)** | Nitrite | Subchronic | 0.1 mg N/kg-d | Adopted IRIS RfD | – | – | – | – | Adopted chronic IRIS RfD for subchronic duration | Provisional U.S. EPA (1997) |
| **EPA RfD (OW)** | Nitrate | Chronic | 1.6 mg N/kg-d | Methemoglobin concentration in infants >10% | 1.6 mg nitrate-nitrogen/kg-d | NOAEL | Bosch et al. (1950) and Walton (1951) | Total $UF^{e}$ = 1 | WOE approach | Final U.S. EPA (1990) |
| | Nitrite | | 0.16 mg N/kg-d | Based on nitrate RfD | – | – | – | – | RfD adjusted[f] | |
| **ATSDR MRL** | Nitrate | Acute (1–14 d) | 4 mg $NO_3$/kg-d | Methemoglobinemia in infants due to nitrate-contaminated water | 44 mg/L | NOAEL | Walton (1951) | Total UF = 1 $UF_H$ = 1 | Dose calculated[g] | Final ATSDR (2017) |
| | | Intermediate (15–365 d) | 4 mg $NO_3$/kg-d | | | | | | | |

| Reference value name | Chemical form | Duration | Reference value | Health effect | Point of departure | Qualifier | Source | Uncertainty/ modifying factors | Notes on derivation | Review status |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Chronic (>1 y) | 4 mg $NO_3$/kg-d | | | | | | No duration adjustment[h] | |
| | Nitrite | Acute (1–14 d) | 0.1 mg $NO_2$/kg-d | | 0.2 mg/kg-d | NOAEL | | Total UF = 1<br>$UF_H$ = 1<br>MF[i] = 2 | Dose calculated[j] | |
| | | Intermediate (15–365 d) | 0.1 mg $NO_2$/kg-d | | | | | | | |
| | | Chronic (>1 y) | 0.1 mg $NO_2$/kg-d | | | | | | | |
| **JECFA ADI** | Nitrate | Chronic | 3.7 mg $NO_3$/kg-d | No effects noted in rats | 370 mg/kg-d | NOEL | Speijers et al. (1989) | Total UF = 100 | Derived values not protective of infants below the age of 3 mo | Final JECFA (1995) and WHO (2003) |
| | Nitrite | | 0.06 mg $NO_2$/kg-d | Hypertrophy of the adrenal zona glomerulosa in rats exposed for 90 d | 5.4 mg/kg-d | NOEL | Til et al. (1988) and Kuper F (1995) | | | |
| | | | | Methemoglobin formation, dilated bronchi and arteries, lymphocyte infiltration, and alveolar hyperinflation in rats | 6.7 mg/kg-d | NOEL | Speijers et al. (1989) | | | |
| **SCF ADI** | Nitrate[k] | Chronic | 3.7 mg $NO_3$/kg-d | No toxicity in rats | 2,500 mg $NaNO_3$/kg-d | NOEL | Maekawa et al. (1982) | Total UF = 500 | MW adjustment[l] | Final CEC (1992) and SCF (1997) |
| | Nitrite | | 0.06 mg $NO_2$/kg-d | Hypertrophy of the adrenal zona glomerulosa in the rat | 5.4 mg/kg-d | NOEL | Til et al. (1988) and Kuper F (1995) | Total UF = 100 | NA | |

| Reference value name | Chemical form | Duration | Reference value | Health effect | Point of departure | Qualifier | Source | Uncertainty/ modifying factors | Notes on derivation | Review status |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Histological changes in the lung and heart of rats | 6.7 mg/kg-d | NOEL | Speijers et al. (1989) | | | |
| **EFSA ADI** | Nitrite | Chronic | 0.07 mg $NO_2$/kg-d | Increased methemoglobin levels | 9.63 mg $NaNO_2$/kg-d | BMDL | NTP (2001a) | Total UF = 100 | MW adjustment[m] | Final(EFSA) (2017b) |

[a]The IRIS RfDs have been adopted by NDEP, TCEQ, and MDEQ (TCEQ, 2023; NDEP, 2020; Michigan DEQ, 2015).
[b]Dose = NOAEL × water intake ÷ BW = 10 mg/L × 0.64 L/day ÷ 4 kg = 1.6 mg/kg-d.
[c]IRIS documentation states: "A modifying factor of 10 was applied because of the direct toxicity of nitrite."
[d]Dose = NOEL × water intake ÷ BW = 10 mg/L × 1 L/day ÷ 10 kg = 1.0 mg/kg-d.
[e]No uncertainty factor is required since the POD is a NOAEL based on a sensitive subpopulation.
[f]$NO_2$ RfD = $NO_3$ RfD × conversion factor = 1.6 mg nitrate-nitrogen/kg-d × 0.1 mg nitrite-nitrogen/ mg nitrate-nitrogen = 0.16 mg nitrite-nitrogen/kg-d.
[g]Dose = NOAEL × water intake ÷ BW = 44 mg/L × 0.525 L/day ÷ 5.33 kg = 4.33 mg/kg-d.
[h]The toxicological profile states: "Repeated ingestion for intermediate- or chronic-duration time periods would be expected to result in changes in methemoglobin levels similar to those elicited from a single exposure."
[i]A modifying factor is applied due to the increased susceptibility of infants to methemoglobinemia.
[j]$NO_2$ dose = $NO_3$ dose × 0.05 = 4 mg/kg-d × 0.05 = 0.2 mg/kg-d. "The ingestion of 0.2 mg nitrite/kg/day by an adult would be expected to result in a nitrite blood level similar to that achieved following ingestion of 4 mg nitrate/kg/day" (ATSDR, 2017).
[k]EFSA concurs with the nitrate ADI established by the Scientific Committee for Food (EFSA) (2017a).
[l]ADI = NOEL ÷ UF × $NO_3$ MW ÷ $NaNO_3$ MW = 2,500 mg/kg-d ÷ 500 × 62 g/mol ÷ 85 g/mol = 3.7 mg/kg-d.
[m]ADI = BMDL ÷ UF × $NO_2$ MW ÷ $NaNO_2$ MW = 9.63 mg/kg-d ÷ 100 × 46 g/mol ÷ 69 g/mol = 0.07 mg/kg-d.
ADI = acceptable daily intake; ATSDR = Agency for Toxic Substances and Disease Registry; BMDL = benchmark dose level; BW = body weight; CEC = Commission of the European Communities; EFSA = European Food Safety Authority; EPA = U.S. Environmental Protection Agency; HEAST = Health Effects Assessment Summary Table; IRIS = Integrated Risk Information System; JECFA = Joint FAO/WHO Expert Committee on Food Additives; MDEQ = Michigan Department of Environmental Quality; MF = modifying factor; MRL = minimal risk level; MW = molecular weight; $NaNO_2$ = sodium nitrate; $NaNO_3$ = sodium nitrate; NDEP = Nevada Division of Environmental Protection; $NO_2$ = nitrite; $NO_3$ = nitrate; NOAEL = no-observed-adverse-effect level; NOEL = no-observed-effect level; NTP = National Toxicology Program; OW = Office of Water; RfD = reference dose; SCF = Scientific Committee for Food; TCEQ = Texas Commission on Environmental Quality; UF = uncertainty factor; $UF_H$ = inter-human variability; WHO = World Health Organization; WOE = weight of evidence.

### A.3.2. Literature Screening Results

The flow of studies for nitrate/nitrite during the screening process is summarized in Figure A-1 and available in an interactive format in a HAWC literature tree. The database searches yielded 73,395 unique records. Application of the SWIFT Review filters (human, animal/human health models, and in vitro) reduced the number of studies for TIAB screening to 18,495. After TIAB screening, 5,549 studies were excluded as not PECO relevant and another 1,080 were tagged as supplemental material, leaving 557 studies that advanced to full-text screening. The remaining 11,374 studies were identified by the SWIFT-AS machine learning algorithm as not relevant. The supplemental literature search yielded an additional 56 studies from other sources for a total of 613 studies that were considered for full-text screening.

The studies identified for full-text screening were processed in DistillerSR. Of these, 65 were excluded as not meeting PECO criteria, text was unable to be obtained for 4, and 166 were tagged as supplemental material. A total of 391 studies were considered PECO relevant, of which 244 were human studies (178 human randomized controlled trials and 66 human observational studies) and 148 were animal studies (one study evaluated health endpoints in both animals and humans).

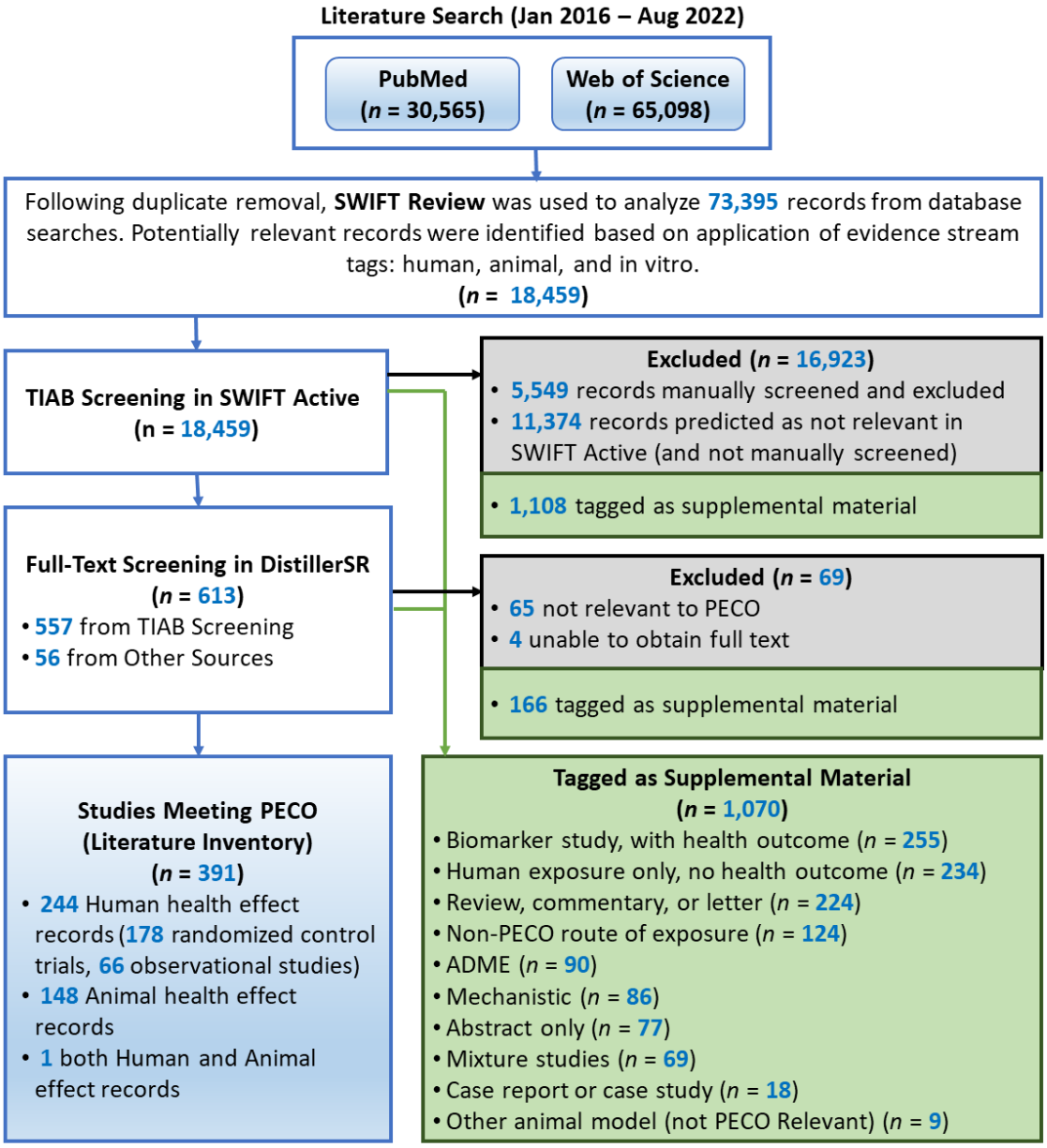



**Figure A-1. Nitrate/ nitrite literature flow diagram.**

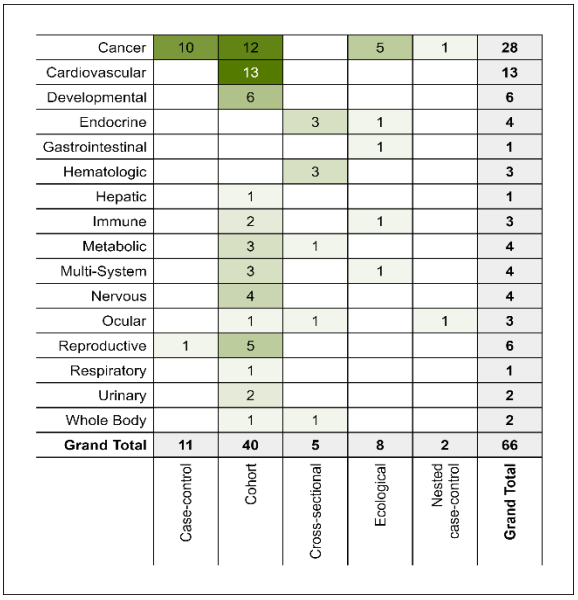
### A.3.3. Characterizing Animal and Epidemiological Studies

***Human Studies***

Literature Inventory

A survey of study designs and health systems assessed in the human studies that met PECO criteria and tabular summary of study design and findings is provided in Figure A-2. Among the 244 human studies, there were 178 randomized controlled trials that administered controlled quantities of oral nitrate or nitrite to identify potential health benefits; these studies were identified

and inventoried but will be considered supplemental material as the focus of this work is on potential adverse health effects due to exposure. The literature search also identified 66 observational epidemiology studies (n = 11 case-control, 2 nested case-control, 5 cross-sectional, 8 ecological, and 40 cohort) in which nitrate/nitrite exposure was evaluated using measurement in drinking water and/or food.



| | Case-control | Cohort | Cross-sectional | Ecological | Nested case-control | Grand Total |
|---|---|---|---|---|---|---|
| Cancer | 10 | 12 | | 5 | 1 | 28 |
| Cardiovascular | | 13 | | | | 13 |
| Developmental | | 6 | | | | 6 |
| Endocrine | | | 3 | 1 | | 4 |
| Gastrointestinal | | | | 1 | | 1 |
| Hematologic | | | 3 | | | 3 |
| Hepatic | | 1 | | | | 1 |
| Immune | | 2 | | 1 | | 3 |
| Metabolic | | 3 | 1 | | | 4 |
| Multi-System | | 3 | | 1 | | 4 |
| Nervous | | 4 | | | | 4 |
| Ocular | | 1 | 1 | | 1 | 3 |
| Reproductive | 1 | 5 | | | | 6 |
| Respiratory | | 1 | | | | 1 |
| Urinary | | 2 | | | | 2 |
| Whole Body | | 1 | 1 | | | 2 |
| **Grand Total** | **11** | **40** | **5** | **8** | **2** | **66** |

**Figure A-2. Survey of human studies that met PECO criteria summarized by study design and health systems assessed.**

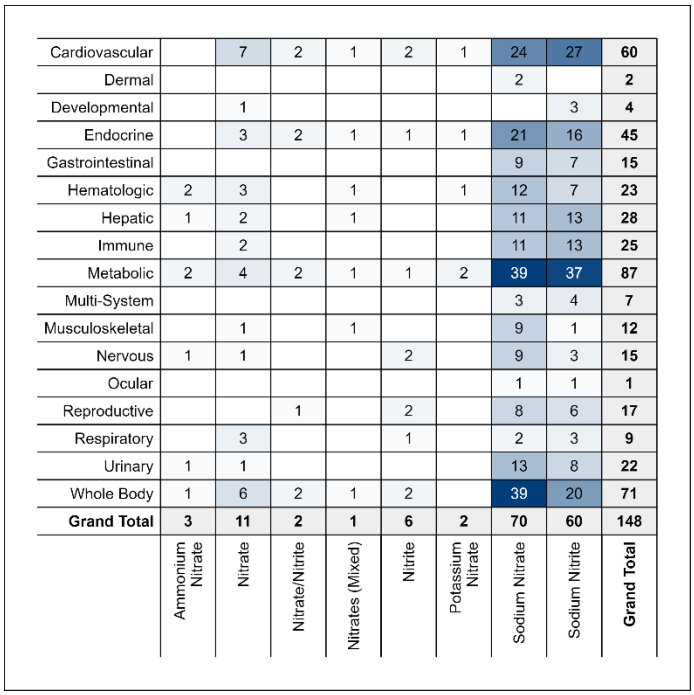
This is a thumbnail image of the interactive dashboard. The numbers in the heat map inset indicate the number of studies that investigated a health system within a study design. If a study evaluated multiple health outcomes, it is shown here multiple times.

### *Animal Studies*

Literature Inventory

A preliminary survey of study designs, species, form(s) of nitrate/nitrite evaluated, and health effects evaluated in the animal studies that met PECO criteria is provided in Figure A-3. The animal studies evaluated exposure to ammonium nitrate, potassium nitrate, sodium nitrate, sodium nitrite, and mixed or unspecified forms of nitrate/nitrite. There were 148 animal studies meeting PECO criteria, and many measured health endpoints in multiple categories. The number of studies for each health effect category shown in the heatmap may be larger than reported in section 5.1 due to the inclusion of additional endpoints (e.g., mRNA expression) along with those used to determine 'primary' health effect categories informed by each study. Most studies were conducted in rats and mice, but data were also available from one study of rabbits. Among the 148 studies, 27 studies administered multiple doses; in general, these study designs are preferred for toxicity value derivation over acute/short-term studies or studies that test a single dose level

(U.S. EPA, 2002), although there may be circumstances for which other study designs are more suitable.



| | Ammonium Nitrate | Nitrate | Nitrate/Nitrite | Nitrates (Mixed) | Nitrite | Potassium Nitrate | Sodium Nitrate | Sodium Nitrite | Grand Total |
|---|---|---|---|---|---|---|---|---|---|
| Cardiovascular | | 7 | 2 | 1 | 2 | 1 | 24 | 27 | **60** |
| Dermal | | | | | | | 2 | | **2** |
| Developmental | | 1 | | | | | | 3 | **4** |
| Endocrine | | 3 | 2 | 1 | 1 | 1 | 21 | 16 | **45** |
| Gastrointestinal | | | | | | | 9 | 7 | **15** |
| Hematologic | 2 | 3 | | 1 | | 1 | 12 | 7 | **23** |
| Hepatic | 1 | 2 | | 1 | | | 11 | 13 | **28** |
| Immune | | 2 | | | | | 11 | 13 | **25** |
| Metabolic | 2 | 4 | 2 | 1 | 1 | 2 | 39 | 37 | **87** |
| Multi-System | | | | | | | 3 | 4 | **7** |
| Musculoskeletal | | 1 | | 1 | | | 9 | 1 | **12** |
| Nervous | 1 | 1 | | | 2 | | 9 | 3 | **15** |
| Ocular | | | | | | | 1 | 1 | **1** |
| Reproductive | | | 1 | | 2 | | 8 | 6 | **17** |
| Respiratory | | 3 | | | 1 | | 2 | 3 | **9** |
| Urinary | 1 | 1 | | | | | 13 | 8 | **22** |
| Whole Body | 1 | 6 | 2 | 1 | 2 | | 39 | 20 | **71** |
| **Grand Total** | **3** | **11** | **2** | **1** | **6** | **2** | **70** | **60** | **148** |

**Figure A-3. Survey of animal studies that met PECO criteria by form of nitrate/nitrite administered and health systems.**

This is a thumbnail image of the interactive dashboard that is filterable by health system, form of nitrate/nitrite administered, and species. The numbers in the heat map inset indicate the number of studies that investigated a health system within form of nitrate/nitrite administered. If a study evaluated multiple health outcomes or presented several experiments, it is shown here multiple times.

### *Mechanistic Evidence*

Results from Database Search

There were 86 mechanistic studies tagged as supplemental material. Among these, the largest numbers of studies evaluated aspects of oxidative and nitrosative stress and hypoxia (n = 37); modulation of enzyme activity (n = 25); or nitric oxide mediated cell signaling (n = 21). Fewer (<20 studies) evaluated other mechanistic characteristics.

ToxCast and Tox21 High Throughput Screening Data

ToxCast and Tox21 high throughput screening data are available for each of the six forms of nitrate/nitrite considered here:

Sodium nitrate: (link)

Sodium nitrite: (link)

Potassium nitrate: (link)

Potassium nitrite: (link)
Ammonium nitrate: (link)
Calcium nitrate: (link)

Comparative Toxicogenomics Database

Nitrate and nitrite are included in the Comparative Toxicogenomics Database (CTDB). Below is a summary of the top interacting genes based on analysis of 257 and 150 studies presented in the CTDB, respectively (click here to see the entry for nitrates, and here to see the entry for nitrites, in the CTDB). Note, these studies were reviewed to identify any that were not otherwise retrieved from other sources (see Appendix D).

## A.4. CONCLUSIONS

The SEM used systematic review methods to identify PECO-relevant studies published from 2016–2022 (no date restriction for calcium nitrate) for six specified forms of nitrate/nitrite. There were 214 animal and human studies which evaluated effects of oral exposure to nitrate/nitrite, comprising 148 animal studies and 66 observational human studies. The animal studies and observational human studies, along with previously published studies as characterized in the ATSDR Toxicological Profile (ATSDR, 2017) and supporting information from the identified supplemental material including mechanistic and ADME information, should be sufficient to support hazard determination for the following health effect categories: cancer; cardiovascular; developmental; endocrine; hematopoietic; hepatic; metabolic; nervous; reproductive; urinary.

# APPENDIX B. SURVEY OF EXISTING TOXICITY VALUES

Table B-1 lists websites which are searched for relevant human health reference values, along with indications of the results of the search. In addition to these sources, the ToxVal database on the Chemicals Dashboard (https://comptox.epa.gov/dashboard/chemical_lists/TOXVAL_V5) is searched for both reference values and PODs as described in Appendix D.

**Table B-1. Sources searched for existing human health reference values**

| Source[a] | Query and/or link |
|---|---|
| ATSDR | http://www.atsdr.cdc.gov/toxprofiles/index.asp |
| | https://www.atsdr.cdc.gov/mrls/mrllist.asp |
| CalEPA | http://www.oehha.ca.gov/tcdb/index.asp |
| | https://www.arb.ca.gov/toxics/healthval/healthval.htm |
| DWSHA | https://www.epa.gov/sites/production/files/2018-03/documents/dwtable2018.pdf |
| Health Canada | https://www.canada.ca/en/services/health/publications/healthy-living.html |
| | http://publications.gc.ca/site/archivee-archived.html?url=http://publications.gc.ca/collections/collection_2012/sc-hc/H128-1-11-638-eng.pdf |
| | http://publications.gc.ca/site/archivee-archived.html?url=http://publications.gc.ca/collections/Collection/H46-2-96-194E.pdf |
| HEAST | http://epa-heast.ornl.gov/heast.php |
| | https://nepis.epa.gov/Exe/ZyPDF.cgi/2000O0GZ.PDF?Dockey=2000O0GZ.PDF |
| IRIS | http://www.epa.gov/iris/ |
| MI EGLE | https://www.michigan.gov/documents/deq/deq-rrd-chem-CleanupCriteriaTSD_527410_7.pdf |
| MDH | https://www.health.state.mn.us/communities/environment/risk/guidance/gw/table.html |
| NHMRC | https://www.nhmrc.gov.au/about-us/publications/australian-drinking-water-guidelines |
| NY DEC | https://www.dec.ny.gov/docs/remediation_hudson_pdf/techsuppdoc.pdf |
| OPP | https://iaspub.epa.gov/apex/pesticides/f?p=chemicalsearch:1 |
| PPRTV | https://www.epa.gov/pprtv/provisional-peer-reviewed-toxicity-values-pprtvs-assessments |
| RIVM | https://www.rivm.nl/bibliotheek/rapporten/711701092.pdf |
| | https://www.rivm.nl/bibliotheek/rapporten/711701025.pdf |

| Source[a] | Query and/or link |
|---|---|
| TCEQ | https://www.tceq.texas.gov/remediation/trrp/trrppcls.html |
| WHO | http://www.who.int/ipcs/publications/ehc/en/ |

[a]ATSDR = Agency for Toxic Substances and Disease Registry; CalEPA = California Environmental Protection Agency; DWSHA = Drinking Water Standards and Health Advisories; HEAST = Health Effects Assessment Summary Tables; IRIS = Integrated Risk Information System; MDH = Minnesota Department of Health; MI EGLE = Michigan Department of Environment, Great Lakes & Energy; NHMRC = National Health and Medical Research Council; NY DEC = New York State Department of Environmental Conservation; OPP = Office of Pesticide Programs; PPRTV = Provisional Peer-Reviewed Toxicity Values; RIVM = *Rijksinstituut voor Volksgezondheid en Milieu,* The Netherlands Institute for Public Health and the Environment; TCEQ = Texas Commission on Environmental Quality; WHO = World Health Organization.

# APPENDIX C. LITERATURE SEARCH STRATEGIES

**Table C-2. Results of initial literature search**

| Database | Search terms | Number of citations[a] |
|---|---|---|
| Web of Science (WoS)<br><br>Dates covered: 1/1/2018–8/17/2022<br><br>Search date: 8/17/2022 | TS=("14797-55-8" OR "14797-65-0" OR "13446-48-5" OR "7631-99-4" OR "7632-00-0" OR "7758-09-0" OR "7757-79-1" OR "6484-52-2" OR "6484-52-2" OR "nitrate" OR "nitrates" OR "nitrite" OR "nitrites" OR "sodium nitrate" OR "sodium nitrates" OR "sodium nitrite" OR "sodium nitrites" OR "potassium nitrate" OR "potassium nitrates" OR "potassium nitrite" OR "potassium nitrites" OR "ammonium nitrate" OR "ammonium nitrates") AND PY=(2018–2022) | 48,417 |
| Web of Science (WoS)<br><br>Dates covered: 1/1/2016–12/31/2017<br><br>Search date: 1/25/2023 | TS=("14797-55-8" OR "14797-65-0" OR "13446-48-5" OR "7631-99-4" OR "7632-00-0" OR "7758-09-0" OR "7757-79-1" OR "6484-52-2" OR "6484-52-2" OR "13477-34-4" OR "10124-37-5" OR "nitrate" OR "nitrates" OR "nitrite" OR "nitrites" OR "sodium nitrate" OR "sodium nitrates" OR "sodium nitrite" OR "sodium nitrites" OR "potassium nitrate" OR "potassium nitrates" OR "potassium nitrite" OR "potassium nitrites" OR "ammonium nitrate" OR "ammonium nitrates" OR "calcium nitrate") AND PY=(2016–2017) | 16,681 |
| PubMed<br><br>Dates covered: 1/1/2018–8/17/2022<br><br>Search date: 8/17/2022 (Updated on 8/29/2023) | (("14797-55-8"[tw] OR "14797-65-0"[tw] OR "13446-48-5"[tw] OR "7631-99-4"[tw] OR "7632-00-0"[tw] OR "7758-09-0"[tw] OR "7757-79-1"[tw] OR "6484-52-2"[tw] OR "6484-52-2"[tw] OR "nitrate"[tw] OR "nitrates"[tw] OR "nitrite"[tw] OR "nitrites"[tw] OR "sodium nitrate"[tw] OR "sodium nitrates"[tw] OR "sodium nitrite"[tw] OR "sodium nitrites"[tw] OR "potassium nitrate"[tw] OR "potassium nitrates"[tw] OR "potassium nitrite"[tw] OR "potassium nitrites"[tw] OR "ammonium nitrate"[tw] OR "ammonium nitrates"[tw]) AND ("2018"[Date - Publication] : "3000"[Date - Publication])) | 22,172 |
| PubMed | (("14797-55-8"[tw] OR "14797-65-0"[tw] OR "13446-48-5"[tw] OR "7631-99-4"[tw] OR "7632-00-0"[tw] OR "7758-09-0"[tw] OR "7757-79-1"[tw] OR "6484-52-2"[tw] OR | 8,393 |

| Database | Search terms | Number of citations[a] |
|---|---|---|
| Dates covered: 1/1/2016–12/31/2017<br><br>Search date: 1/25/2023 | "6484-52-2"[tw] OR "13477-34-4"[tw] OR "10124-37-5"[tw] OR "nitrate"[tw] OR "nitrates"[tw] OR "nitrite"[tw] OR "nitrites"[tw] OR "sodium nitrate"[tw] OR "sodium nitrates"[tw] OR "sodium nitrite"[tw] OR "sodium nitrites"[tw] OR "potassium nitrate"[tw] OR "potassium nitrates"[tw] OR "potassium nitrite"[tw] OR "potassium nitrites"[tw] OR "ammonium nitrate"[tw] OR "ammonium nitrates"[tw] OR "calcium nitrate") AND ("2016"[Date - Publication] : "2017"[Date - Publication])) | |
| TOXNET<br><br>Dates covered: 1/1/2016–12/05/2017 | @SYN0+@AND+@OR+(nitrate+nitrates+nitrite+nitrites+@TERM+@rn+14797-55-8+@TERM+@rn+14797-65-0+@TERM+@rn+7631-99-4+@TERM+@rn+7757-79-1+@TERM+@rn+6484-52-2+@TERM+@rn+7632-00-0+@TERM+@rn+7758-09-0)+@RANGE+yr+2016+2017+@NOT+@org+"nih+reporter"<br><br>@SYN0+@AND+@OR+(nitrate+nitrates+nitrite+nitrites+@TERM+@rn+14797-55-8+@TERM+@rn+14797-65-0+@TERM+@rn+7631-99-4+@TERM+@rn+7757-79-1+@TERM+@rn+6484-52-2+@TERM+@rn+7632-00-0+@TERM+@rn+7758-09-0)+@RANGE+yr+2017+2017+@NOT+@org+"nih+reporter" | 1,992 |
| TOTAL: | Merged reference sets (After removal of duplicates) | 73,395 |

[a]The numbers in this document are current as of October 6, 2023, but are subject to slight changes due to ongoing deduplication efforts.

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

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

Review of the citation reference lists is typically done manually because they are not available in a file format (e.g., RIS) that permits uploading into screening software applications. Manual review entails scanning the title, study summary, or study details as presented in the resource for those that appear to meet the PECO criteria. Any records identified that were not identified from the other sources are annotated with respect to source and screened as outlined in Section 3.2.

## D.2. EUROPEAN CHEMICALS AGENCY

A search of the ECHA registered substances database was conducted using the chemical names. The registration dossier associated with the chemical name was retrieved by navigating to and clicking the eye-shaped view icon displayed in the chemical summary panel. The general information page and all subpages included under the Toxicological Information tab were reviewed to identify any human or animal health effects information from 2016 onwards that would be eligible for inclusion based on PECO criteria.

## D.3. EPA CHEMVIEW

The EPA ChemView database (U.S. EPA, 2019a) using the chemical CASRN was searched. The prepopulated CASRN match and the "Information Submitted to EPA" output option filter were selected before generating results. If results were available, the square-shaped icon under the "Data Submitted to EPA" column was selected, and the following records were included:

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

- TSCA Section 8(d) (Health and safety studies)
- TSCA Section 8(e) (Substantial risk)
- FYI (Voluntary documents)

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

## D.4. NTP CHEMICAL EFFECTS IN BIOLOGICAL SYSTEMS

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

## D.5. ECOTOX DATABASE

EPA's ECOTOX Knowledgebase (https://cfpub.epa.gov/ecotox/search.cfm) was searched using the chemical names. Results were refined to terrestrial mammalian studies by selecting the terrestrial tab at the top of the search page and sorting the results by species group. Results were reviewed to verify that it was not already identified from the database search (or searches of "other sources consulted") search prior to moving forward to screening.

## D.6. EPA COMPTOX CHEMICAL DASHBOARD VERSION TO RETRIEVE A SUMMARY OF ANY TOXCAST OR TOX21 HIGH THROUGHPUT SCREENING INFORMATION

Version 3.0.9 of the CompTox Chemicals Dashboard (U.S. EPA, 2019b) was accessed for high throughput screening (HTS) data by searching the Dashboard by CASRN. Next, the "Bioactivity" section was selected and the availability of ToxCast/Tox21 HTS data for active and inactive assays was examined in the "TOXCAST: Summary" tab. If active assays were reported, the figure was copied for presentation in the SEM. This figure presents (i) scatterplot of scaled assay responses vs. AC50 values for each active assay endpoint, and (ii) cytotoxicity limit as a vertical line.

More detailed information on the results of ToxCast and Tox21 assays are available in the CompTox Chemicals Dashboard section "ToxCast/Tox21," which includes chemical analysis data, dose-response data and model fits, and "flags" assigned by an automated analysis, which might suggest false positivity/negativity or indicate other anomalies in the data. This information is not summarized further for the purposes of the SEM, which is focused on identifying the extent of available evidence.

## D.7. COMPARATIVE TOXICOGENOMICS DATABASE

This CTDB database (http://ctdbase.org/) was searched using the chemical names in the "keyword search" with pulldown menu set to "Chemicals." The reference list of studies reporting gene/protein interactions with the query chemical were compared to existing references in HAWC. Unique references screened according to PECO and supplemental material criteria.

**Table D-1. Summary table for other sources search results**

| Source[a] | Source address | Search terms | Search date | Total unique number of results not already identified in literature search | Records found to be PECO-relevant |
|---|---|---|---|---|---|
| Review of reference lists from existing assessments (final or publicly available draft) or journal review articles that focused on human health | OEHAA 2018; EFSA 2017 (Sodium nitrate); EFSA 2017 (Sodium nitrite); various review articles | NA | NA | 21 | 5 |
| EPA CompTox (Computational Toxicology Program) Chemicals Dashboard (ToxVal) | | Results from human health, oral/ingestion route of exposure: pod, toxicity value, lethality effect level | 5/25/2023 | | |
| | Nitrate: https://comptox.epa.gov/dashboard/chemical/hazard/DTXSID5024217 | ATSDR MRL; IRIS NOAEL; RSL RfD; IRIS RfD; OW RfD | | 0 | 0 |
| | Nitrite ion: https://comptox.epa.gov/dashboard/chemical/hazard/DTXSID5024219 | ATSDR MRL; IRIS NOEL; HEAST NOEL; RSL RfD; IRIS RfD; OW RfD; DOD MEG; HEAST RfD | | 0 | 0 |
| | Sodium nitrate: https://comptox.epa.gov/dashboard/chemical/details/DTXSID6020937 | COSMOS HNEL; COSMOS LEL; ECHA IUCLID NOAEL; ChemIDplus LD50; ECHA IUCLID LD50 | | 0 | 0 |
| | Sodium nitrite: https://comptox.epa.gov/dashboard/chemical/hazard/DTXSID0020941 | COSMOS LEL; COSMOS HNEL; HESS NOEL; ECHA IUCLID NOEL; ECHA IUCLID NOAEL; ECHA IUCLID LOAEL; EFSA BMDL; ChemIDplus LD50; ECHA IUCLID LD50 | | 0 | 0 |

| Source[a] | Source address | Search terms | Search date | Total unique number of results not already identified in literature search | Records found to be PECO-relevant |
|---|---|---|---|---|---|
| | | ToxRefDB LEL, NEL, LOAEL, NOAEL (based on NTP 2001 report) | | | |
| | Potassium nitrate: https://comptox.epa.gov/dashboard/chemical/details/DTXSID4029692 | DOE Wildlife Benchmark; COSMOS HNEL; COSMOS LEL; ECHA IUCLID NOAEL; ChemIDplus LD50; ECHA IUCLID LD50 | | 0 | 0 |
| | Potassium nitrite: https://comptox.epa.gov/dashboard/chemical/hazard/DTXSID5042320 | ECHA IUCLID NOEL; ECHA IUCLID NOAEL; ECHA IUCLID LOAEL; ChemIDplusLD50; ECHA IUCLID LD50 | | 0 | 0 |
| | Ammonium nitrate: https://comptox.epa.gov/dashboard/chemical/hazard/DTXSID2029668 | ECHA IUCLID NOAEL; ChemIDplus LD50; ECHA IUCLID LD50 | | 0 | 0 |
| | Calcium nitrate: https://comptox.epa.gov/dashboard/chemical/hazard/DTXSID1039719 | ECHA IUCLID NOAEL; ChemIDplus LD50 | | 0 | |
| ECHA, Chemical Registration Dossiers | | | 5/26/2023 | | |
| | Sodium nitrate: https://echa.europa.eu/registration-dossier/-/registered-dossier/15423/1/1 | EC number: 231-554-3 | | 0 | 0 |
| | Sodium nitrite: https://echa.europa.eu/registration-dossier/-/registered-dossier/14890 | EC number: 231-555-9 | | 0 | 0 |

| Source[a] | Source address | Search terms | Search date | Total unique number of results not already identified in literature search | Records found to be PECO-relevant |
|---|---|---|---|---|---|
| | Potassium nitrate: https://echa.europa.eu/registration-dossier/-/registered-dossier/15481 | EC number: 231-818-8 | | 0 | 0 |
| | Potassium nitrite: No dossier available | EC number: 231-832-4 | | 0 | 0 |
| | Ammonium nitrate: https://echa.europa.eu/registration-dossier/-/registered-dossier/15999 | EC number: 229-347-8 | | 0 | 0 |
| | Calcium nitrate: https://echa.europa.eu/registration-dossier/-/registered-dossier/15487 | EC number: 233-332-1 | | 1 | 0 |
| EPA ChemView | https://chemview.epa.gov/chemview?tf=0&ch=14797-55-8_10124-37-5_13477-34-4_14797-55-8_7631-99-4_7757-79-1_14797-65-0_7632-00-0_7758-09-0&su=256737574985&as=31098&ac=115166378999&ma=4-11-1981377-4_16848473-4_16848474-4_49007566&gs=&tds=0&tdl=100&tas1=1&tas2=asc&tas3=undefined&tss= | Nitrate; nitrite; nitrite ion; potassium nitrate; potassium nitrite; sodium nitrate; sodium nitrite; ammonium nitrate; calcium nitrate | 5/25/2023 | 0 | 0 |
| NTP CEBS | https://manticore.niehs.nih.gov/cebssearch/ | | 5/24/2023 | | |
| | https://cebs.niehs.nih.gov/cebs/test_article/7631-99-4 | Sodium nitrate: Only test article purity | | 0 | 0 |
| | https://cebs.niehs.nih.gov/cebs/test_article/7632-00-0 | Sodium nitrite: Link to NTP 2001 study | | 0 | 0 |

| Source[a] | Source address | Search terms | Search date | Total unique number of results not already identified in literature search | Records found to be PECO-relevant |
|---|---|---|---|---|---|
| | https://cebs.niehs.nih.gov/cebs/test_article/7757-79-1 | Potassium nitrate: Only test article purity | | 0 | 0 |
| OECD Echem Portal | https://hpvchemicals.oecd.org/UI/Search.aspx | Potassium nitrate; potassium nitrite; sodium nitrate; sodium nitrite; ammonium nitrate; calcium nitrate | 5/24/2023 | 0 | 0 |
| ECOTOX Database | https://cfpub.epa.gov/ecotox/search.cfm | Nitrate; nitrite (terrestrial, mammalian studies only) | 5/24/2023 | 0 | 0 |
| Comparative Toxicogenomics Database (CTDB) | http://ctdbase.org/ | Nitrate; nitrite; potassium nitrate; potassium nitrite; sodium nitrate; sodium nitrite; ammonium nitrate; calcium nitrate | 5/25/2023 | 39 | 0 |
| **TOTAL (after de-duplication)** | | | | 56 | 4 |

[a]PECO = populations, exposures, comparators, and outcomes; NA = not applicable; POD = point of departure; ECHA = European Chemicals Agency; NTP CEBS = National Toxicology Program Chemical Effects in Biological Systems; OECD = Organisation for Economic Co-operation and Development.

# REFERENCES


ATSDR (Agency for Toxic Substances and Disease Registry). (2017). Toxicological profile for nitrate and nitrite [ATSDR Tox Profile]. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service. https://www.atsdr.cdc.gov/toxprofiles/tp204.pdf.

Bannach-Brown, A; Przybyła, P; Thomas, J; Rice, ASC; Ananiadou, S; Liao, J; Macleod, MR. (2018). Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error (pp. 1-26). bioRxiv. http://dx.doi.org/10.1101/255760.

Blessinger, T; Davis, A; Chiu, WA; Stanek, J; Woodall, GM; Gift, J; Thayer, KA; Bussard, D. (2020). Application of a unified probabilistic framework to the dose-response assessment of acrolein. Environ Int 143: 105953. http://dx.doi.org/10.1016/j.envint.2020.105953.

Bosch, HM; Rosenfield, AB; Huston, R; Shipman, HR; Woodward, FL. (1950). Methemoglobinemia and Minnesota well supplies. J Am Water Works Assoc 42: 161-170.

Bryan, NS; Loscalzo, J. (2011). Nitrite and nitrate in human health and disease. In NS Bryan; J Loscalzo (Eds.). New York, NY: Humana Press. http://dx.doi.org/10.1007/978-1-60761-616-0.

Bryan, NS; Petrosino, JF. (2017). Nitrate-reducing oral bacteria: Linking oral and systemic health. Chapter 3. In NS Bryan; J Loscalzo (Eds.), Nitrate and nitrite in human health and disease (2nd ed.). Boston, MA: Humana Press.

CalEPA (California Environmental Protection Agency). (2018). Public health goals for nitrate and nitrite in drinking water. Sacramento, CA: Office of Environmental Health Hazard Assessment. https://oehha.ca.gov/media/downloads/water/chemicals/phg/nitratephg051118.pdf.

CEC (Commission of the European Communities). (1992). Food science and techniques: Reports of the scientific committee for food. (EUR13913). Luxembourg, Luxembourg. http://ec.europa.eu/food/fs/sc/scf/reports/scf_reports_26.pdf.

Coggan, AR; Baranauskas, MN; Hinrichs, RJ; Liu, Z; Carter, SJ. (2021). Effect of dietary nitrate on human muscle power: a systematic review and individual participant data meta-analysis [Review]. JISSN 18: 66. http://dx.doi.org/10.1186/s12970-021-00463-z.

Coggan, ARR; Thies. (2020). Simultaneous Pharmacokinetic Analysis of Nitrate and its Reduced Metabolite, Nitrite, Following Ingestion of Inorganic Nitrate in a Mixed Patient Population. Pharm Res 37: 235-235. http://dx.doi.org/10.1007/s11095-020-02959-w.

Cohen, AM; Hersh, WR; Peterson, K; Yen, PY. (2006). Reducing workload in systematic review preparation using automated citation classification. J Am Med Inform Assoc 13: 206-219. http://dx.doi.org/10.1197/jamia.M1929.

Desimone, LA. (2009). Quality of water from domestic wells in principal aquifers of the United States, 1991-2004. In Geological Survey Scientific Investigations (pp. 139). (SIR 2008-5227). Reston, VA: U.S. Department of the Interior, U.S. Geological Survey. http://pubs.usgs.gov/sir/2008/5227/.

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.

(EFSA), EFSA. (2017a). Re-evaluation of sodium nitrate (E 251) and potassium nitrate (E 252) as food additives. EFSA J 15: e04787. http://dx.doi.org/10.2903/j.efsa.2017.4787.

(EFSA), EFSA. (2017b). Re-evaluation of potassium nitrite (E 249) and sodium nitrite (E 250) as food additives. EFSA J 15: e04786. http://dx.doi.org/10.2903/j.efsa.2017.4786.

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

FDA. Title 21: Food and drugs: Chapter I: Food and Drug Administration, Department of Health and Human Services: Subchapter B: Food for human consumption: Part 165-- Beverages, 21 CFR § 165 699-719 (2021). https://www.ecfr.gov/current/title-21/chapter-I/subchapter-B/part-165

Fields, S. (2004). Global nitrogen: Cycling out of control. Environ Health Perspect 112: A556-A563.

Gonzalez, A; Hyde, E; Sangwan, N; Gilbert, JA; Viirre, E; Knight, R. (2016). Migraines are correlated with higher levels of nitrate-, nitrite-, and nitric oxide-reducing oral microbes in the American gut project cohort. mSystems 1. http://dx.doi.org/10.1128/mSystems.00105-16.

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. http://dx.doi.org/10.1177/003591576505800503.

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

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

Howard, BE; Phillips, J; Tandon, A; Maharana, A; Elmore, R; Mav, D; Sedykh, A; Thayer, K; Merrick, BA; Walker, V; Rooney, A; Shah, RR. (2020). SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation. Environ Int 138: 105623. http://dx.doi.org/10.1016/j.envint.2020.105623.

Hutson, SS, ., B.,arber, N. L.,., K.,enney, J. F.,., L.,insey, K. S.,., L.,umia, D. S.,., and Maupin, M. A. (2004). Estimated use of water in the united states in 2000. USGS 1268.

IARC (International Agency for Research on Cancer). (2010). Ingested nitrate and nitrite, and cyanobacterial peptide toxins [IARC Monograph] (pp. v-vii, 1-412). Lyon, France. http://monographs.iarc.fr/ENG/Monographs/vol94/mono94.pdf.

IPCS (International Programme on Chemical Safety). (2005). SIDS Initial Assessment Report For SIAM 20. Sodium Nitrite. http://www.inchem.org/documents/sids/sids/7632000.pdf.

JECFA (Joint FAO/WHO Expert Committee on Food Additives). (1995). Evaluation of certain food additives and contaminants: Forty-fourth report of the Joint FAO/\NHO Expert Committee on Food Additives. (BIOSIS/96/25016). Geneva, Switzerland: World Health organization.

Kapil, V; Haydar, SM; Pearl, V; Lundberg, JO; Weitzberg, E; Ahluwalia, A. (2013). Physiological role for nitrate-reducing oral bacteria in blood pressure control. Free Radic Biol Med 55: 93-100. http://dx.doi.org/10.1016/j.freeradbiomed.2012.11.013.

Kulinkina, AV; Mohan, VR; Francis, MR; Kattula, D; Sarkar, R; Plummer, JD; Ward, H; Kang, G; Balraj, V; Naumova, EN. (2016). Seasonality of water quality and diarrheal disease counts in urban and rural settings in south India. Sci Rep 6: 20521. http://dx.doi.org/10.1038/srep20521.

Kuper F, T, il HP. (1995). Subchronic toxicity experiments with potassium nitrite in rats. In E Council of (Ed.), Health aspects of nitrate and its metabolites (particularly nitrite):

Proceedings, international workshop, Bilthoven (Netherlands), 8-10 November 1994 (pp. 195-212). Strasbourg, France: Council of Europe Press.

Lin, YJ; Cheng, CJ; Chen, JW; Lin, Z. (2020). Incorporating exogenous and endogenous exposures into dietary risk assessment of nitrates and nitrites in vegetables: A probabilistic integrated toxicokinetic modeling approach. J Agric Food Chem 68: 1079-1090. http://dx.doi.org/10.1021/acs.jafc.9b06720.

Maekawa, A; Ogiu, T; Onodera, H; Furuta, K; Matsuoka, C; Ohno, Y; Odashima, S. (1982). Carcinogenicity studies of sodium nitrite and sodium nitrate in F-344 rats. Food Chem Toxicol 20: 25-33. http://dx.doi.org/10.1016/s0278-6915(82)80005-7.

Mattila, M; Niinistö, S; Takkinen, HM; Tapanainen, H; Reinivuo, H; Åkerlund, M; Suomi, J; Ahonen, S; Ilonen, J; Toppari, J; Knip, M; Veijola, R; Virtanen, SM. (2020). Maternal nitrate and nitrite intakes during pregnancy and risk of islet autoimmunity and type 1 diabetes: The DIPP cohort study. J Nutr 150: 2969-2976. http://dx.doi.org/10.1093/jn/nxaa250.

Mensinga, TT; Speijers, GJ; Meulenbelt, J. (2003). Health implications of exposure to environmental nitrogenous compounds [Review]. Toxicol Rev 22: 41-51. http://dx.doi.org/10.2165/00139709-200322010-00005.

Michigan DEQ (Michigan Department of Environmental Quality). (2015). Chemical update worksheet: Nitrite (DD). Lansing, MI. https://www.michigan.gov/-/media/Project/Websites/egle/Documents/Programs/RRD/Remediation/Rules---Criteria/Chemical-Update-Worksheets/N---Z/Nitrite-Datasheet.pdf?rev=278b35ed9dca4a99b76e89b9fc22bab1.

Mina, A; Stathopoulos, J; Sinanian, T; Mcneice, L; Holmes, D; Fletcher, KL; Bottero, E; Banukumar, S; Vazquez, S. (2021). Comparison of different methods used in drugs of abuse for sample validity testing including pH methods, specific gravity methods, TECO™ Drug Adulteration Test Strip and oxidant assay. 2: 550-566. http://dx.doi.org/10.1515/almed-2021-0026.

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.

NDEP (Nevada Division of Environmental Protection). (2020). Nevada Division of Environmental Protection basic comparison levels: 2020 update. https://ndep.nv.gov/uploads/documents/ndep-bcls-august-2020.pdf.

NRC (National Research Council). (2009). Science and decisions: Advancing risk assessment. Washington, DC: National Academies Press. http://dx.doi.org/10.17226/12209.

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). (2001a). Toxicology and carcinogenesis studies of sodium nitrite (CAS No. 7632-00-0) in F344/N rats and B6C3F1 mice (drinking water studies) (pp. 280). (TR 495). Research Triangle Park, NC: U.S. Department of Health and Human Services, National Institutes of Health, National Toxicology Program. https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/PB2001107676.xhtml.

NTP (National Toxicology Program). (2001b). Toxicology and carcinogenesis studies of sodium nitrite (CASRN 7632-00-0) in F344/N rats and B6C3F1 mice (drinking water studies) [NTP](pp. 7-273). (ISSN 0888-8051 NTP TR 495; NIH Publication No. 01-3954). Research Triangle Park, NC. http://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr495.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). (2021). Report on carcinogens, fifteenth edition. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service. http://dx.doi.org/10.22427/NTP-OTHER-1003.

Pokorny L, M, aturana I, B.,ortle WH. (2006). Sodium Nitrate and Nitrite. In Kirk-Othmer Encyclopedia of Chemical Technology (5th ed.). New York: John Wiley & Sons. http://dx.doi.org/10.1002/0471238961.1915040916151115.a01.pub2.

SCF (Scientific Committee on Food). (1997). Reports of the Scientific Committee for Food. Luxembourg: European Commission. https://food.ec.europa.eu/system/files/2020-12/sci-com_scf_reports_38.pdf.

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).

Sellars, S; Nunes, V. (2021). Synthetic nitrogen fertilizer in the U.S. farmdoc daily 11.

Speijers, GJA; van Went, GF; van Apeldoorn, ME; Montizaan, GK; Janus, JA; Canton, JH; van Gestel, CAM; van der Heijden, CA; Heijna-Merkus, E; Knapp, AGA; Luttik, R; de Zwart, D. (1989). Integrated criteria document nitrate; effects. Appendix to RIVM Report No. 758473012. Bilthoven, Netherlands: National Institute of Public Health and Environmental Protection. https://www.rivm.nl/publicaties/integrated-criteria-document-nitrate.

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.

Sterne, JAC; Savovic, J; Page, MJ; Elbers, RG; Blencowe, NS; Boutron, I; Cates, CJ; Cheng, HY; Corbett, MS; Eldridge, SM; Emberson, JR; Hernan, MA; Hopewell, S; Hrobjartsson, A; Junqueira, DR; Juni, P; Kirkham, JJ; Lasserson, T; Li, T; Mcaleenan, A; Reeves, BC; Shepperd, S; Shrier, I; Stewart, LA; Tilling, K; White, IR; Whiting, PF; Higgins, JPT. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 366: l4898. http://dx.doi.org/10.1136/bmj.l4898.

TCEQ (Texas Commission on Environmental Quality). (2023). TRRP protective concentration levels: May 2023 PCL and supporting tables. Austin, TX: Texas Commission on Environmental Quality :: TCEQ. Retrieved from https://www.tceq.texas.gov/remediation/trrp/trrppcls.html

Til, HP; Falke, HE; Kuper, CF; Willems, MI. (1988). Evaluation of the oral toxicity of potassium nitrite in a 13-week drinking-water study in rats. Food Chem Toxicol 26: 851-859. http://dx.doi.org/10.1016/0278-6915(88)90026-9.

U.S. EPA (U.S. Environmental Protection Agency). (1987). Nitrite IRIS summary. https://iris.epa.gov/ChemicalLanding/&substance_nmbr=78.

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). (1990). Drinking Water Criteria Document for Nitrate/Nitrite [EPA Report]. (TR-1242-60B). Cincinnati, OH.

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

U.S. EPA (U.S. Environmental Protection Agency). (1991b). Integrated risk information system (IRIS): Chemical assessment summary: Nitrate; CASRN 14797-55-8. Washington, DC: U.S.

Environmental Protection Agency. Integrated Risk Information System. https://cfpub.epa.gov/ncea/iris/iris_documents/documents/subst/0076_summary.pdf.

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). (1997). Health effects assessment summary tables: FY 1997 update [EPA Report]. (EPA540R97036). Washington, DC: U.S. Environmental Protection Agency, Office of Emergency and Remedial Response. http://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=2000O0GZ.txt.

U.S. EPA (U.S. Environmental Protection Agency). (1998). 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). (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). (2011). 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). (2012). 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). (2014). Guidance for applying quantitative data to develop data-derived extrapolation factors for interspecies and intraspecies extrapolation [EPA Report]. (EPA/100/R-14/002F). Washington, DC: Risk Assessment Forum, Office of the Science Advisor. https://www.epa.gov/sites/production/files/2015-01/documents/ddef-final.pdf.

U.S. EPA (U.S. Environmental Protection Agency). (2015). 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). Chemical Contaminant Summaries for the Third Six-Year Review of Existing National Primary Drinking Water Regulations [EPA Report]. Washington, DC: U.S. Environmental Protection Agency, Office of Water. https://www.epa.gov/sites/production/files/2016-12/documents/810r16004.pdf.

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

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

U.S. EPA (U.S. Environmental Protection Agency). (2019a). ChemView [Database]. Retrieved from https://chemview.epa.gov/chemview

U.S. EPA (U.S. Environmental Protection Agency). (2019b). CompTox Chemicals Dashboard [Database]. Research Triangle Park, NC. Retrieved from https://comptox.epa.gov/dashboard

U.S. EPA (U.S. Environmental Protection Agency). (2020a). 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). (2020b). 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). (2021). CompTox chemicals dashboard. Washington, DC. Retrieved from https://comptox.epa.gov/dashboard

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.

Walton, G. (1951). Survey of literature relating to infant methemoglobinemia due to nitrate-contaminated water. Am J Public Health Nations Health 41: 986-996. http://dx.doi.org/10.2105/AJPH.41.8_Pt_1.986.

Ward, MH. (2009). Too much of a good thing? Nitrate from nitrogen fertilizers and cancer [Review]. Rev Environ Health 24: 357-363. http://dx.doi.org/10.1515/reveh.2009.24.4.357.

Ward, MH; deKok, TM; Levallois, P; Brender, J; Gulis, G; Nolan, BT; VanDerslice, J. (2005). Workgroup report: Drinking-water nitrate and health—Recent findings and research needs. Environ Health Perspect 113: 1607-1614. http://dx.doi.org/10.1289/ehp.8043.

Water and Air Quality Bureau, H, ealthy Environments and Consumer Safety Branch, H.,ealth Canada. (2013). Guidelines for Canadian drinking water quality: Guideline technical document: Nitrate and nitrite. In Guidelines for Canadian Drinking Water Quality. Ottawa, Ontario: Minister of Health. http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/nitrate_nitrite/index-eng.php.

WHO (World Health Organization). (2003). Nitrate and nitrite in drinking-water: Background document for development of WHO Guidelines for drinking-water quality. (WHO/SDE/WSH/04.03/56). Geneva, Switzerland. http://apps.who.int/iris/handle/10665/75380.

WHO (World Health Organization). (2016). Nitrate and nitrite in drinking-water: Background document for development of WHO guidelines for drinking-water quality. (WHO/FWC/WSH/16.52). Geneva, Switzerland. http://www.who.int/water_sanitation_health/dwq/chemicals/nitrate-nitrite-background-jan17.pdf.

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.

Zeilmaker, MJ; Bakker, MI; Schothorst, R; Slob, W. (2010). Risk assessment of N-nitrosodimethylamine formed endogenously after fish-with-vegetable meals. Toxicol Sci 116: 323-335. http://dx.doi.org/10.1093/toxsci/kfq093.

Zeilmaker, MJ; Meulenbelt, J; Kortboyer, JM; Slob, W. (1996). Safety assessment of nitrate: Mathematical modeling of nitrite formation in man and its application in the risk assessment of nitrate. (Report No. 235802002). Bilthoven, Netherlands: National Institute of Public Health and the Environment (RIVM).