[EPA/635/R-21/047]
IRIS Assessment Protocol
www.epa.gov/iris

Systematic Review Protocol for the Vanadium and Compounds (Oral Exposure) IRIS Assessment

(Preliminary Assessment Materials)

April 2021

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

DISCLAIMER

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

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

ADME	absorption, distribution, metabolism, and excretion
BMDL	benchmark dose lower confidence limit
$BW^{3/4}$	body-weight scaling to the 3/4 power
CAA	Clean Air Act
CAS	Chemical Abstracts Service
CASRN	Chemical Abstracts Service registry number
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CPAD	Chemical and Pollutant Assessment Division
CPHEA	Center for Public Health and Environmental Assessment
EPA	Environmental Protection Agency
GLP	good laboratory practices
HAP	hazardous air pollutant
HAWC	Health Assessment Workspace Collaborative
HEC	human equivalent concentration
HERO	Health and Environmental Research Online
IRIS	Integrated Risk Information System
LOAEL	lowest-observed-adverse-effect level
LOEL	lowest-observed-effect level
MeSH	Medical Subject Headings
MOA	mode of action
NMD	normalized mean difference
NOEL	no-observed-effect level
NTP	National Toxicology Program
OAR	Office of Air and Radiation
OECD	Organisation for Economic Co-operation and Development
OLEM	Office of Land and Emergency Management
ORD	Office of Research and Development
OSF	oral slope factor
PBPK	physiologically based pharmacokinetic
PECO	populations, exposures, comparators, and outcomes
PK	pharmacokinetic
POD	point of departure
RfD	oral reference dose
ROBINS-I	Risk of Bias in Nonrandomized Studies of Interventions
UF	uncertainty factor

AUTHORS | CONTRIBUTORS | REVIEWERS

Assessment Team

Erin Yost, Ph.D. (Assessment Manager)	U.S. EPA/ORD/CPHEA/CPAD
Kathleen Newhouse, M.S.	
Suryanarayana Vulimiri, Ph.D.	
Elizabeth Radke, Ph.D.	
Barbara Glenn, Ph.D.	
David Farrar, Ph.D.	

Executive Direction

Wayne E. Cascio, M.D. (CPHEA Director)	U.S. EPA/ORD/CPHEA
Samantha Jones, Ph.D. (CPHEA Associate Director)	
Kristina Thayer, Ph.D. (CPAD Director)	
James Avery, Ph.D. (CPAD Associate Director)	
Andrew Kraft, Ph.D. (CPAD Senior Science Advisor)	
Paul White, Ph.D. (CPAD Senior Science Advisor)	
Janice Lee, Ph.D. (CPAD Toxic Effects Assessment Branch Chief)	

Contributors and Production Team

Ryan Jones	HERO Director (U.S. EPA/ORD/CPHEA/CPAD)
Vicki Soto	Project Management Team (U.S. EPA/ORD/CPHEA/CPAD)
Dahnish Shams	Project Management Team (U.S. EPA/ORD/CPHEA/CPAD)
Maureen Johnson	CPHEA Webmaster (U.S. EPA/ORD/CPHEA)
Brittany Schulz	Reference Value Array Support (ORAU)
Courtney Lemeris	Tableau Support (ICF)

1. INTRODUCTION

The Integrated Risk Information System (IRIS) Program is undertaking a reassessment of the health effects of oral exposure to vanadium and compounds. An assessment of oral exposure to vanadium and compounds was identified as an Agency priority in December 2018 (https://www.epa.gov/iris/iris-program-outlook). The IRIS Program subsequently announced the initiation of a vanadium and compounds inhalation assessment in December 2019, which will be performed separately from the assessment of oral exposure.

IRIS assessments provide high quality, publicly available information on the toxicity of chemicals to which the public might be exposed. These science assessments are not regulations and do not constitute U.S. Environmental Protection Agency (EPA) policy. Science assessments such as these provide a critical part of the scientific foundation for subsequent risk assessment and risk management decisions made by EPA program and regional offices to protect public health. IRIS assessments are also used by states and local health agencies, Tribes, other federal agencies, international health organizations, and other external stakeholders.

A draft IRIS assessment plan (IAP) for oral exposure to vanadium and compounds (U.S. EPA, 2020a) was released for public comment and presented at a public science meeting on August 19, 2020 (https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=348792) to seek input on the problem formulation components of the assessment plan. The IAP specifies why oral exposure to vanadium and compounds was selected for evaluation, specifies the objectives and specific aims of the assessment, provides draft PECO (populations, exposures, comparators, and outcomes) criteria, and identifies key areas of scientific complexity. This assessment is being developed at the request of EPA's Office of Water, although other programs may have a use for this assessment, once finalized.

This protocol document incorporates the updated IAP content, including revisions based on public input and updated scoping needs, and presents the methods for conducting the systematic review and dose-response analysis for the assessment. Whereas the IAP describes *what* the assessment will cover, chemical-specific protocols describe *how* the assessment will be conducted (see Figure 1). The systematic review methods described in this protocol are based on the Office of Research and Development's *ORD Staff Handbook for Developing Integrated Risk Information System (IRIS) Assessments* (Version 1.0, November 2020, referred to as the "IRIS Handbook") (U.S. EPA, 2020b). These methods have been reviewed previously by the National Academy of Sciences (NASEM, 2018).

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

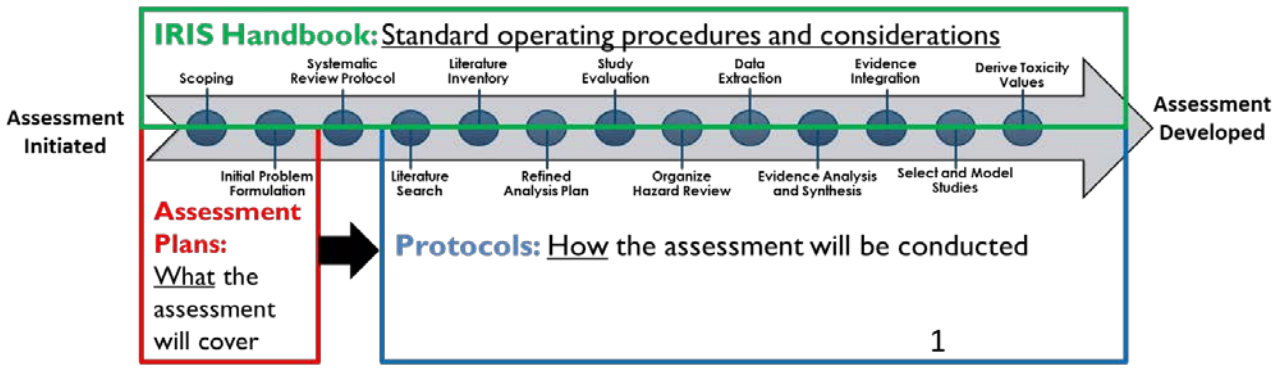


Figure 1. IRIS systematic review problem formulation and method documents.

2. SCOPING AND INITIAL PROBLEM FORMULATION SUMMARY

2.1. BACKGROUND

2.1.1. Sources, Production, and Use

Vanadium (V) is a naturally occurring metal that is the 22nd most abundant element in Earth's crust and is found in a variety of minerals and nearly all coal and petroleum crude oils (Kelley et al., 2017; ATSDR, 2012). Vanadium has important industrial applications: It is added as a ferrovanadium alloy to increase the strength of steel used for applications such as buildings, bridges, pipelines, and auto parts; it is used as a catalyst by the chemical industry; and it is used in vanadium redox flow batteries, which are a type of rechargeable battery used for large-scale storage of electricity. Worldwide demand for vanadium is expected to increase due to increasing demand for vanadium-containing steel in China and other industrialized countries and in vanadium redox flow batteries for alternative energy sources such as solar and wind (Kelley et al., 2017). The industrial importance of vanadium led to its designation as one of 35 "Critical Minerals" identified by the U.S. Department of the Interior in 2018, pursuant to an executive order (U.S. Department of the Interior, 2018).

The majority of vanadium production occurs in China, Russia, and South Africa. The U.S. is a net importer of vanadium-bearing raw materials but an exporter of vanadium products (Moskalyk and Alfantazi, 2003). Domestic production of vanadium in the U.S. occurs primarily through the recovery or recycling of vanadium from petroleum residues, utility ash, slags, and spent catalysts, with a minor amount produced as a byproduct of uranium mining (Polyak, 2020). Mining of vanadium has historically not been viable in the U.S. due to the relatively low vanadium content of naturally occurring deposits; however, scoping began in 2020 on the first vanadium mine in the U.S. on a vanadium-rich deposit in Nevada (Harvey, 2020).

In addition to its industrial uses, vanadium is also a micronutrient and has been demonstrated to be biochemically active. Vanadium is considered an essential element for certain bacteria, cyanobacteria, algae, and fungi, which are found to have vanadium-dependent enzymes. A functional role for vanadium in other species has not yet been identified, although vanadium deficiency has been described for birds, chickens, rats, guinea pigs, and goats. No symptoms of vanadium deficiency have been described for humans (Ścibior et al., 2020; Rehder, 2015). Although it remains an open question as to whether vanadium is essential to humans, it has been postulated that vanadium may play a role in human physiology due to its ubiquity in the environment, the binding affinity of vanadium for transport proteins such as transferrin and albumin, and the ability

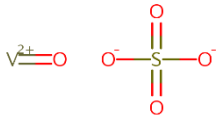
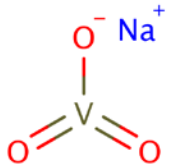
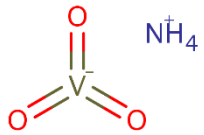
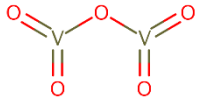
of vanadium ions to substitute for phosphate in a number of protein structures and to inhibit phosphatase and phosphorylase activity (Ścibior et al., 2020; Rehder, 2015).

Vanadium has been found to enhance the activity of insulin and mitigate the symptoms of diabetes and hypercholesterolemia, which is thought to occur due to vanadium-mediated phosphatase inhibition. Specifically, vanadium is hypothesized to inhibit protein tyrosine phosphatase 1B, which is responsible for inactivation of the insulin receptor, thereby activating the PI3K-Akt pathway that is responsible for the metabolism of carbohydrates and lipids (Crans, 2015; Crans et al., 2004). The insulin-enhancing effects of vanadium have been investigated in animal models (typically in streptozotocin-induced diabetic rodent models) and diabetic patients in clinical trials (Crans, 2015; Smith et al., 2008; Thompson and Orvig, 2006). Vanadium has also been investigated in vitro for putative antitumor effects (Evangelou, 2002). Vanadium-based pharmaceuticals have not been approved for clinical use, however, so these therapeutic applications remain investigational. The potential therapeutic applications of vanadium are outside the scope of the IRIS assessment, which is focused on potential adverse effects of vanadium exposure that are relevant to human health risk assessment. Studies on therapeutic applications, however, were inventoried as part of problem formulation (U.S. EPA, 2020a).

2.1.2. Physical and Chemical Properties

Vanadium has a complex chemistry, existing in the environment with four possible oxidation states (+2, +3, +4, +5) and a multitude of aqueous species including anions and cations (Gustafsson, 2019). Pure elemental vanadium does not exist naturally (Rehder, 2015; ATSDR, 2012). Table 1 lists the properties of elemental vanadium and the most common inorganic vanadium compounds that are used in toxicological studies, consisting of V^{+5} salts [sodium metavanadate ($NaVO_3$), sodium orthovanadate (Na_3VO_4), ammonium vanadate (NH_4VO_3)], V^{+4} salts [vanadyl sulfate ($VOSO_4$)], and vanadium pentoxide (V_2O_5). In aqueous solutions, these inorganic vanadium compounds form a spectrum of oxygen-containing ions that undergo redox, hydrolytic, and condensation reactions as a function of factors including pH, redox potential, concentration, and temperature, as demonstrated in the speciation diagrams in Figures 2–4. These three diagrams [from Gustafsson (2019) and Crans et al. (2004)] were selected as references for this protocol because they are based on equilibrium constants in solutions of relatively low ionic strength and ambient temperature, which reflects conditions in fresh surface waters and in typical laboratory drinking water studies.

Table 1. Chemical identity and physiochemical properties of selected vanadium compounds as curated by EPA's CompTox Chemicals Dashboard

Name	Elemental vanadium	Vanadyl sulfate	Sodium metavanadate	Ammonium metavanadate	Sodium orthovanadate	Vanadium pentoxide
CASRN	7440-62-2	27774-13-6	13718-26-8	7803-55-6	13721-39-6	1314-62-1
DTXSID[a]	2040282	4021428	3044336	1052533	2037269	2023806
Structure						
Molecular weight (g/mol)	50.942	163	121.928	116.978	183.907	181.878
Molecular formula	V	$VOSO_4$	$NaVO_3$	NH_4VO_3	Na_3VO_4	V_2O_5
Oxidation state		+4	+5	+5	+5	+5
Selected Synonym(s)	Vanadium	(Oxido)vanadium(2^+) sulfate; oxo(sulfato)vanadium; oxovanadium(IV) sulfate; vanadium oxide sulfate; vanadium oxosulfate; vanadium oxysulfate; vanadium sulfate; vanadic sulfate; vanadyl monosulfate; vanadin(IV) oxide sulfate	Sodium vanadate; sodium trioxidovanadate(1^-); sodium vanadium oxide; sodium vanadium trioxide; vanadic acid, monosodium salt; sodium vanadate(V)	Ammonium trioxovanadate(1^-); ammonium tris(oxido)vanadate(1^-); ammonium monovanadate; ammonium vanadate(V); vanadic acid, ammonium salt; ammonium vanadium oxide; ammonium vanadium trioxide	Trisodium tetraoxidovanadate(3^-); sodium vanadium oxide; trisodium vanadate; sodium vanadate(V); vanadic acid; trisodium salt	Vanadium oxide; mu-oxido[tetrakis(oxido)] divanadium; divanadium pentoxide; vanadic anhydride; vanadin(V) oxide; vanadium(V) oxide

1

Table 1. Chemical identity and physiochemical properties of selected vanadium compounds as curated by EPA's CompTox Chemicals Dashboard (continued)

Name	Elemental vanadium	Vanadyl sulfate	Sodium metavanadate	Ammonium metavanadate	Sodium orthovanadate	Vanadium pentoxide
Water solubility (mol/L)[b]	–	–	–	–	–	–
LogP: Octanol-Water[b]	–	–	–	–	–	–
Melting Point (°C)[b]	1.90e+3	–	630	–	858	690
Boiling Point (°C)[b]	3.00e+3	–		–	–	1.75e+3
Vapor Pressure (mmHg)[b]	–	–	–	–	–	–
Bioconcentration Factor[b]	4.36e+3	4.5	5.54	26.4	–	15.4

[a]DTXSIDs are unique substance identifiers used for curation by EPA's Distributed Structure-Searchable Toxicity (DSSTox) project.
[b]Experimental average values for physiochemical properties are shown here. Median values and ranges for physiochemical properties are also provided on EPA's Chemicals Dashboard at https://comptox.epa.gov/dashboard/. If no experimental values were available on the Chemicals Dashboard, "–" is shown.

The Pourbaix diagram in Figure 2 illustrates the redox reactions that occur as a function of pH and redox potential: V^{+5} species predominate under oxic conditions and high pH, V^{+4} species occur under suboxic conditions and low pH, and V^{+3} species occur under anoxic conditions. V^{+2} is not shown in the diagram as it is readily oxidized and unstable. As expected, based on those conditions, V^{+5} and V^{+4} are the prevailing vanadium species in most natural waters (Gustafsson, 2019). In laboratory studies, solutions of V^{+4} salts are stable at pH 3–4 but are readily oxidized to V^{+5} at pH 7 (Mutlu et al., 2017; Crans et al., 1995).

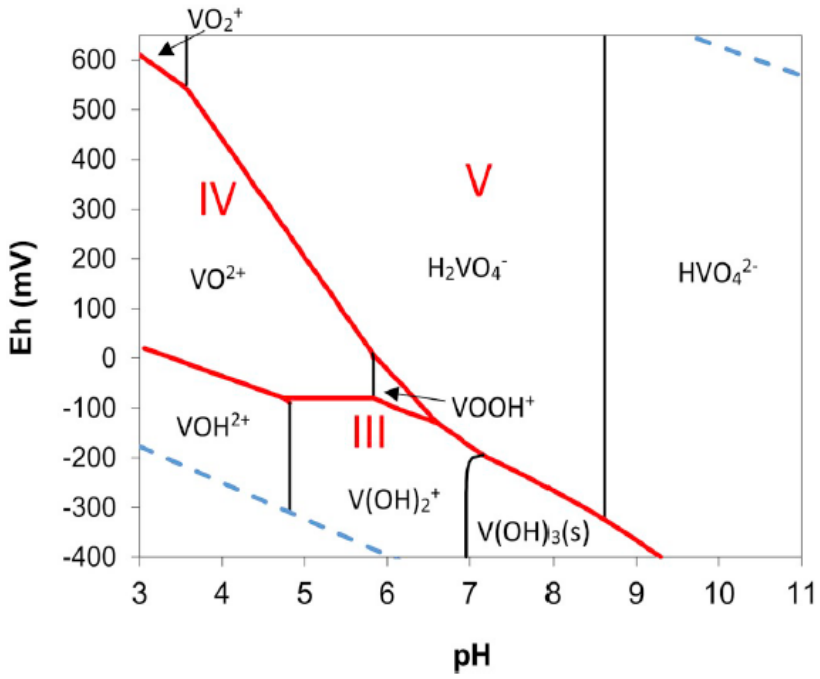


Figure 2. Predominance diagram showing aqueous speciation of vanadium as a function of pH and redox potential (Eh) as total dissolved V = 1 µM.
Temperature = 25°C, ionic strength = 0.01 M NaCl. The red solid lines separate the predominance fields of the three oxidation states III, IV, and V. The blue dashed lines represent the stability lines of water with respect to $H_2(g)$ (at low Eh) and $O_2(g)$ (at high Eh). Source: **Gustafsson (2019)**.

Figure 3 presents the aqueous speciation of V^{+5} as a function of pH and concentration. V^{+5} is present as monomeric species at low concentrations but forms oligomers at higher concentrations (dimers, tetramers, pentamers, and decamers), with decavanadates predominating at high concentrations and low pH. Although the V^{+5} oligomers are thermodynamically stable, they convert to monomeric species within milliseconds upon dilution (Crans et al., 1990). Figure 4 presents the aqueous speciation of V^{+4} as a function of pH and concentration. Aqueous solutions of V^{+4} are well characterized at low and high pH but are less characterized at neutral pH, in part because the free electron readily pairs and forms oligomeric/polymeric species and is not observable by electron paramagnetic resonance spectroscopy (Costa Pessoa, 2015).

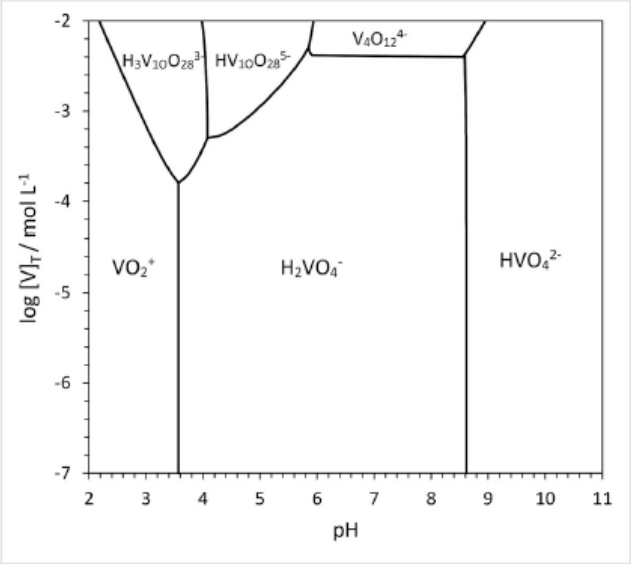


Figure 3. Predominance diagram showing aqueous speciation of V^{+5} as a function of pH and total molar concentration of vanadium. Temperature = 25°C, ionic strength = 0.01 M NaCl. Source: **Gustafsson (2019)**.

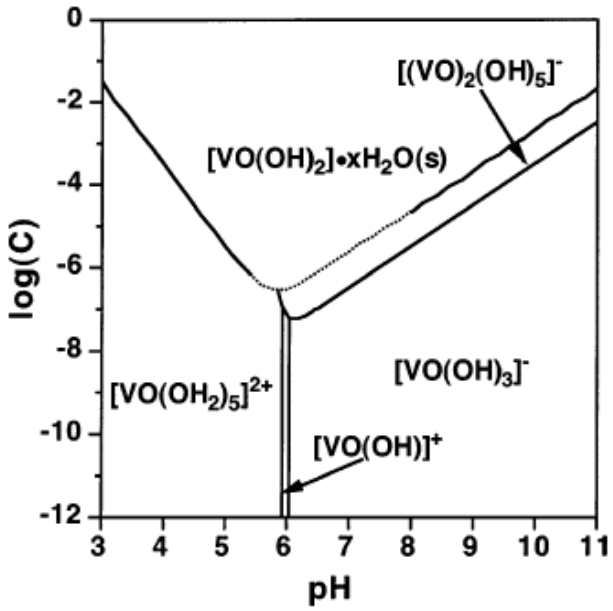


Figure 4. Predominance diagram showing aqueous speciation of V^{+4} as a function of pH and the total molar concentration of vanadium.
Temperature = 25°C, ionic strength for most constants = 0.1 M $LiOClO_4$. Source: **Crans et al. (2004)**. Developed based on equilibrium constants from **Henry et al. (1973)**, **Komura et al. (1977)**, and **Vilas Boas and Costa Pessoa (1987)**.

In the body, vanadium undergoes redox cycling and speciation driven by factors such as pH, local availability of reducing equivalents (e.g., glutathione-SH, NADPH), and complexation with

biomolecules (NTP, 2008; Byczkowski and Kulkarni, 1996; Nielsen, 1995). It is thought that vanadium ingested as V^{+5} (e.g., the oxoanions $H_2VO_4^-$ and HVO_4^{2-}) will partially reduce to V^{+4} (e.g., the oxocation VO_2^+) in the acidic conditions of the stomach, and then subsequently precipitate as a less soluble V^{+5} species [$VO(OH)_2$] in the more alkaline conditions of the intestines (Harrington et al., 2021; Treviño et al., 2019). The V^{+5} oxoanions ($H_2VO_4^-$ and HVO_4^{2-}) are absorbed more readily compared to VO_2^+ and $VO(OH)_2$. The absorption of vanadium following oral exposure is therefore expected to be influenced by the form of ingested vanadium and residence time, conditions in the gastrointestinal tract, and speed of conversion (Treviño et al., 2019; Nielsen, 1995). It is generally reported that V^{+5} is more toxic than V^{+4} (ATSDR, 2012; NTP, 2008), possibly due to differences in absorption. For instance, in a 14-day study in rats (Roberts et al., 2016), exposure to V^{+5} (as sodium metavanadate) led to higher blood and liver vanadium levels and greater toxicity compared to V^{+4} (as vanadyl sulfate); analysis of plasma from this study found that V^{+4} was the only species present regardless of the exposure compound, although it was unclear whether the conversion took place in vivo or during sample preparation (Harrington et al., 2021). This example illustrates that the form of ingested vanadium is an important determinant of absorption and toxicity, but in vivo speciation may be difficult or impossible to predict. Regarding the role of oligomers (e.g., decavanadates) in vanadium toxicity, it has been demonstrated in fish models (using intraperitoneal or intravenous injection) and in vitro that decavanadates are biologically active and have distinct interactions with cellular proteins compared to monovanadates (Aureliano, 2014; Aureliano and Crans, 2009); however, because decavanadates are rapidly converted to monovanadates upon dilution, it remains uncertain whether decavanadates reach the bloodstream after oral administration (Pessoa et al., 2015).

Vanadium is rarely incorporated into organic compounds, but commonly forms coordination complexes with organic and inorganic ligands. For instance, vanadium in crude oil is present as stable vanadyl^{+4}-porphyrin complexes (Gustafsson, 2019), and a large number of organic compounds including fulvic and humic acids can complex vanadium via oxygen groups (Huang et al., 2015). Vanadium in food is also likely present as complexes with proteins or other organic molecules. Vanadium coordination complexes with organic ligands have also been developed in the laboratory and studied since the 1960s as putative therapeutics for diabetes. The organic ligand in therapeutic vanadium coordination complexes acts to modify the pharmacokinetic properties of vanadium, increasing bioavailability and decreasing negative side effects such as gastrointestinal distress (Thompson and Orvig, 2006). Therapeutic vanadium coordination complexes break down to release the vanadium ion but may also enter the body intact by passive diffusion and form ternary complexes with cellular proteins, which may act to decrease the rate of clearance and increase the efficacy of the vanadium ion as an insulin enhancer (Pessoa et al., 2015; Thompson and Orvig, 2006).

2.1.3. Environmental Fate and Transport

Vanadium is naturally mobilized from Earth's crust by the chemical and mechanical weathering of rocks and by volcanic activity and biomass burning. The production of coal and petroleum also results in vanadium mobilization, and the combustion of fossil fuels is the biggest anthropogenic source of vanadium to the atmosphere (Schlesinger et al., 2017). Leachates from ores, slags, sewage sludge, fertilizers, and ash ponds and coal preparation wastes may contribute to anthropogenic release of vanadium into water and soil (ATSDR, 2012), although vanadium in wastes from the energy industry is commonly recovered and can be used for industrial applications, as indicated above. Vanadium contamination in soils has been observed in areas with vanadium mining and heavy industrial activity (Gustafsson, 2019).

In drinking water distribution systems, reservoirs of vanadium have been found in corrosion byproducts in lead and iron pipes. The source of vanadium in these corrosion byproducts is unclear but may be the result of the gradual precipitation of vanadium from the drinking water over time. This has led to concerns that disruption of these corrosion byproducts by chemical or physical processes could mobilize and increase vanadium levels in drinking water (Gerke et al., 2010; Gerke et al., 2009).

2.1.4. Potential for Human Exposure (Oral)

Vanadium is present at low concentrations in most foods, which serve as the major source of background vanadium exposure in the general population (ATSDR, 2012). Likely due to its natural abundance, vanadium is present in human breast milk, although at relatively low levels compared to other trace elements (Krachler et al., 2000). Vanadium is also included in some multivitamins and dietary supplements. The Institute of Medicine (IOM) Panel on Micronutrients derived a Tolerable Upper Intake Level of 26 µg V/kg-day for adult humans and stated that the risk of adverse effects resulting from intake of vanadium from food (6.5 to 18 µg V/day) or typical usage of dietary supplements (median 9 µg V/kg-day) was unlikely, whereas increased risk was likely to result from chronic intake of supplements containing larger doses of vanadium (e.g., doses of >100 mg/day are used in some human clinical trials) (IOM, 2001).

In 2016, EPA included vanadium on the drinking water Fourth Contaminant Candidate List (CCL 4), which is a list of contaminants that are not currently subject to national primary drinking water regulations but are known or anticipated to occur in public water systems (https://www.epa.gov/ccl/contaminant-candidate-list-4-ccl-4-0). Contaminants listed on the CCL may require regulation under the Safe Drinking Water Act (SDWA) if the Agency determines that the contaminant may have an adverse effect on the health of persons; the contaminant is known to occur or there is substantial likelihood that the contaminant will occur in public water systems with a frequency and at levels of public health concern; and in the sole judgment of the Administrator, regulation of the contaminant presents a meaningful opportunity for health risk reductions for persons served by public water systems (Safe Drinking Water Act, 2019). Vanadium (measured as

total vanadium; speciation and oxidation state were not determined) was monitored under EPA's Third Unregulated Contaminant Monitoring Rule (UCMR 3) from 2013 to 2015 and 3,625 of 4,922 public water systems (73.6%) detected vanadium at or above the minimum reporting level (2 µg/L). The data show that 163 of these public water systems (3.3%) had results above the reference concentration used in the UCMR 3 (21 µg/L)[1] (https://www.epa.gov/sites/production/files/2017-02/documents/ucmr3-data-summary-january-2017.pdf). In December 2018, an Integrated Risk Information System (IRIS) assessment of oral exposure to vanadium was identified by the EPA Office of Water as a priority for an IRIS assessment (https://www.epa.gov/iris/iris-program-outlook).

2.1.5. Previous Assessments of Oral Exposure to Vanadium and Compounds by the Environmental Protection Agency and Other Health Agencies

Existing human health reference values for vanadium and compounds from federal, state, and international agencies were searched in May 2020 as described in Appendix A and are depicted in Figure 5 (see Table 2 for a tabular summary, including derivation details of the displayed values; values with no derivation details are listed in Table 3). IRIS published a health effects assessment of vanadium and compounds in 1987, which includes a reference dose (RfD) for lifetime oral exposure to vanadium pentoxide (U.S. EPA, 1987). The RfD was based on an unpublished study by Stokinger et al. (1953) described in Patty's Industrial Hygiene and Toxicology (1981) in which an unspecified strain of rats was fed vanadium pentoxide over a lifetime at levels of 10 and 100 ppm vanadium. An RfD of 0.009 mg/kg-day for vanadium pentoxide was derived based on the no-observed-adverse-effect level (NOAEL) of 10 ppm vanadium (approximately 17.9 ppm vanadium pentoxide) for decreased hair cystine content. The RfD was calculated by assuming that rats eat food equivalent to 5% of their body weight and by applying an uncertainty factor (UF) of 100 (a factor of 10 for interspecies extrapolation and a factor of 10 to provide added protection for unusually sensitive individuals). IRIS also reviewed the carcinogenicity data available for vanadium and compounds and concluded that the weight-of-evidence classification for vanadium under the 1986 *Guidelines for Carcinogen Risk Assessment* (U.S. EPA, 1986) is Group D, not classifiable.

EPA also developed provisional peer-reviewed toxicity values (PPRTVs) for vanadium and its soluble inorganic compounds other than vanadium pentoxide in 2009, including a chronic provisional RfD (p-RfD) and subchronic p-RfD for vanadium (U.S. EPA, 2009). These values were based on kidney histopathology in a 6-month study by Boscolo et al. (1994), in which rats were given sodium metavanadate in drinking water at levels of 10 or 40 µg V/mL (first experiment) or 1 µg V/mL (second experiment); EPA estimated that these corresponded to doses of 0.12, 1.2, or

[1]The reference concentration for vanadium in drinking water used in the UCMR 3 was based on the ATSDR 1992 minimal risk level (MRL) of 0.003 mg/kg-day. The ATSDR 1992 *Toxicological Profile for Vanadium* is no longer publicly available and has been replaced by ATSDR (2012). The UCMR 3 reference concentration provides context around the detection of a particular contaminant above the minimum reporting level and does not constitute an "action level."

4.7 mg V/kg-day on the basis of default drinking water and body weight estimates. A subchronic p-RfD of 0.0007 mg/kg-day for vanadium was derived based on the NOAEL of 0.12 mg V/kg-day from the second experiment by adjusting upward by 0.1 mg/kg-day to account for likely background exposure to vanadium in diet and by applying a UF of 300 (a factor of 10 for interspecies extrapolation, a factor of 10 to protect unusually sensitive individuals, and a factor of 3 to account for database deficiencies). A chronic p-RfD of 0.00007 mg/kg-day for vanadium was derived from this same study by applying an additional UF of 10 to account for extrapolation to chronic exposure duration. This assessment also concluded that there was "Inadequate Information to Assess [the] Carcinogenic Potential" of vanadium based on the 2005 *Guidelines for Carcinogen Risk Assessment* (U.S. EPA, 2005a).

Since the publication of these prior assessments by EPA, new information on the health effects of vanadium and compounds has become available. The Agency for Toxic Substances and Disease Registry (ATSDR) 2012 *Toxicological Profile of Vanadium* concluded that increased blood pressure, hematological alterations, alterations in neurobehavioral tests, and developmental toxicity were the most sensitive outcomes in laboratory animal studies following intermediate duration (15- to 364-day) oral exposure to vanadium compounds, but noted that increased blood pressure and hematological effects were not consistently observed across animal studies at higher dose levels or in humans in a 12-week clinical trial (ATSDR, 2012). More recently, NTP has undertaken a series of studies in rats and mice on the health effects of oral (drinking water) exposure to vanadyl sulfate and sodium metavanadate, which include evaluation of a range of health outcomes and will provide additional information on the comparative toxicity of two common vanadium oxidation states. These include 14-day studies in rats and mice (Roberts et al., 2016), a 13-week study in mice, and an extended developmental toxicity study in rats in which F1 offspring are exposed from gestation day (GD) 6 through 13 weeks post-weaning. NTP's developmental and 13-week drinking water studies are expected to be posted in 2021 and interim results are currently available (https://ntp.niehs.nih.gov/ntp/results/pubs/posters/roberts_sot20190300.pdf).

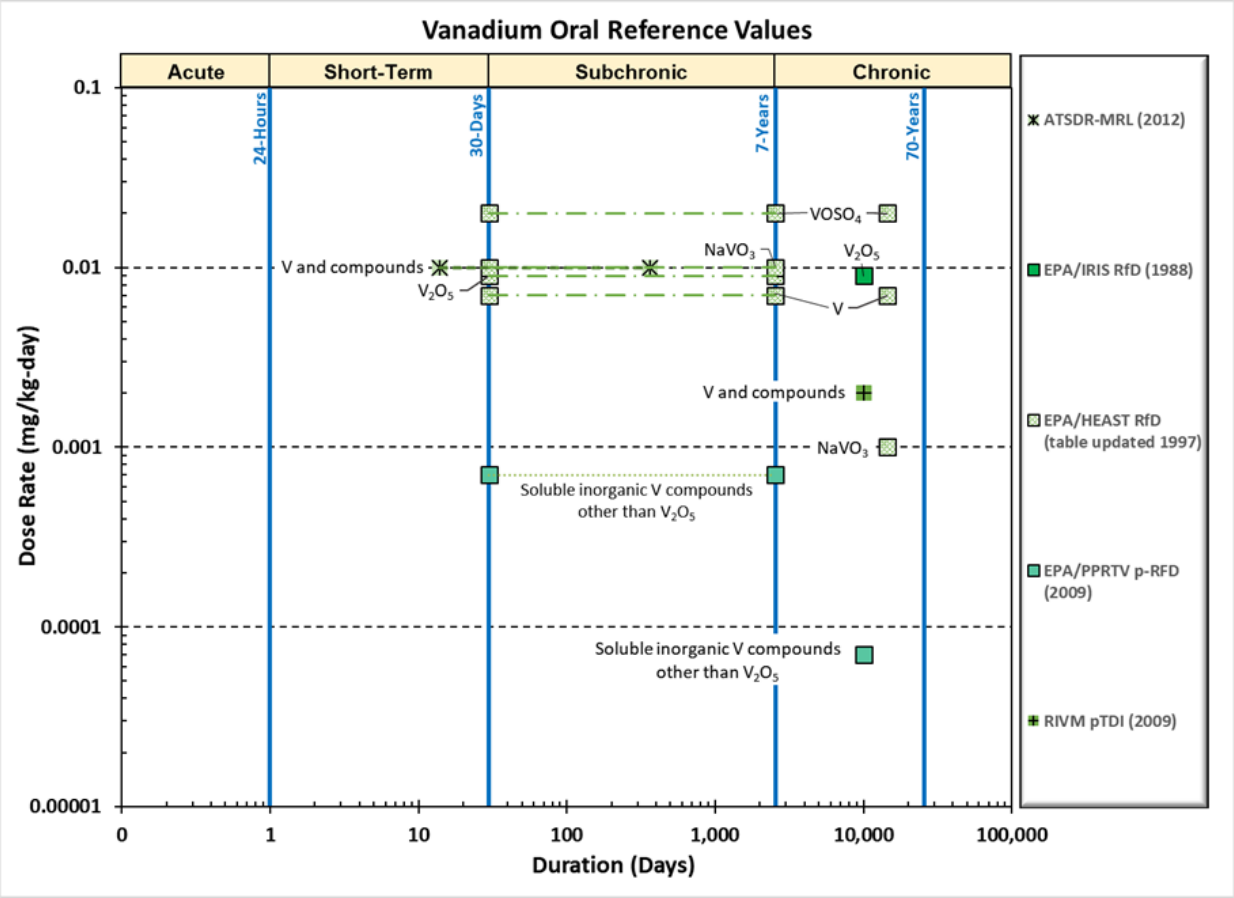


Figure 5. Available health effect reference values for oral exposure to vanadium compounds (current as of May 2020).

Table 2. Details on derivation of the available health effect reference values for oral exposure to vanadium compounds[a]
(current as of May 2020; please consult citation source entities and other entities in Appendix Table A-1 for current values)

Reference value name[b]	Duration	Compound	Reference value (mg/kg-day)	Health effect	Point of departure	Qualifier	Source	Uncertainty factors[c]	Notes on derivation	Review status
EPA RfD (IRIS)[d]	Lifetime (chronic)	Vanadium pentoxide	0.009	Decreased cystine in hair of rats	0.89 mg/kg-day	NOAEL	Stokinger et al. (1953)	Total UF = 100 UF_A = 10 UF_H = 10	NOAEL Estimated[e]	Final (U.S. EPA, 1987)
EPA p-RFD (PPRTV)[f]	Subchronic	Vanadium and soluble inorganic compounds (excluding vanadium pentoxide)	0.0007	Kidney lesions in male rats exposed for 6 mos.	0.12 mg/kg-day 0.22 mg/kg-day	NOAEL $NOAEL_{ADJ}$	Boscolo et al. (1994)	Total UF = 300 UF_A = 10 UF_H = 10 UF_{DB} = 3	NOAEL Adjusted[g]	Provisional (U.S. EPA, 2009)
	Chronic		0.00007					Total UF = 3,000 UF_A = 10 UF_H = 10 UF_S = 10 UF_{DB} = 3		
EPA RfD (HEAST)[h]	Subchronic Chronic	Vanadium	0.007 0.007	Minor serum cholesterol changes in rats	0.7 mg/kg-day	NOAEL	Schroeder et al. (1970)	Total UF = 100 UF_A = 10 UF_H = 10	NOAEL Estimated[i]	Provisional (U.S. EPA, 1997)
	Subchronic Chronic	Vanadium sulfate	0.02 0.02		2.24 mg/kg-day	NOAEL				
	Subchronic	Vanadium pentoxide	0.009	Adopted IRIS chronic RfD	–	–	–	–	Adopted IRIS chronic RfD	
	Subchronic	Sodium meta-vanadate	0.01	Impaired kidney function in rats exposed for 3 mos.	1.3 mg/kg-day	NOAEL	Domingo et al. (1985)	Total UF = 100 UF_A = 10 UF_H = 10	NOAEL Conversion[j]	Provisional (U.S. EPA)

Table 2. Details on derivation of the available health effect reference values for oral exposure to vanadium compounds[a] (continued)

Reference value name[b]	Duration	Compound	Reference value (mg/kg-day)	Health effect	Point of departure	Qualifier	Source	Uncertainty factors[c]	Notes on derivation	Review status
EPA RfD (HEAST)[h] (continued)	Chronic		0.001					Total UF = 1,000 UF_A = 10 UF_H = 10 UF_S = 10		
ATSDR-MRL	Intermediate (15–365 days)	Vanadium and compounds	0.01	No change in blood pressure, body wt., or hematological or clinical chemistry parameters at highest dose in a 12-wk. study	0.5 mg/kg-day 0.12 mg/kg-day	NOAEL H_6O_8SV NOAEL V	Fawcett et al. (1997)	Total UF = 10 UF_H = 10	NOAEL V Calculated[k]	Final (ATSDR, 2012)
RIVM pTDI	Chronic	Vanadium and compounds	0.002	Develop-mental effects in rats	5 mg/kg-day 2.1 mg/kg-day	LOAEL NaO_3V LOAEL V	Domingo et al. (1986)	Total UF = 1,000 UF_A = 10 UF_H = 10 UF_L = 10	LOAEL V Calculated[l]	Provisional (Tiesjema and Baars, 2009)

[a]Health effect reference values listed in Table 2 are shown in Figure 5.
[b]ATSDR = Agency for Toxic Substances and Disease Registry; HEAST = Health Effects Assessment Summary Tables; MRL = minimal risk level; PPRTV = Provisional Peer-Reviewed Toxicity Value; RfD = reference dose; RIVM = *Rijksinstituut voor Volksgezondheid en Milieu,* The Netherlands Institute for Public Health and the Environment; TDI = tolerable daily intake.
[c]UF = uncertainty factor; subscripts indicate the type of UF that was applied. UFH – inter-human variability; UFA – animal to human variability; UFL – LOAEL to NOAEL adjustment; UFS – subchronic to chronic adjustment; UFDB – database uncertainty.

[d]This RfD has been adopted as a state value by the Nevada Division of Environmental Protection.
[e]The NOAEL was estimated based on the assumption that rats exposed to 10 ppm vanadium (17.85 ppm vanadium pentoxide) in food were consuming 5% of their body weight in food per day.
[f]The chronic p-RfD has been adopted as a state value by the Michigan Department of Environment, Great Lakes & Energy.
[g]The NOAEL was adjusted upward to account for possible additional vanadium exposure from the rats' basal diet.
[h]The chronic RfD for sodium metavanadate has been adopted by the Nevada Division of Environmental Protection.
[i]The NOAEL was estimated for rats exposed to 5 ppm vanadium in the form of vanadyl sulfate in drinking water.
[j]Rats were exposed to 10 ppm sodium metavanadate in their drinking water. Support documentation indicates that this exposure is equivalent to a dose rate of 0.55 mg vanadium/kg-day. While this is not explicitly stated anywhere in the text, 0.55 mg vanadium/kg-day equals 1.3 mg/kg-day sodium metavanadate, as per the following molecular weight conversion. Thus, 1.3 mg/kg-day was likely used as the point of departure:
NOAEL $NaVO_3$ = NOAEL V × $NaVO_3$ M.W./V molar mass = 0.55 mg V/kg-day × 121.928 g $NaVO_3$/mol/50.942 g V/mol = 1.3 mg $NaVO_3$/kg-day.
[k]NOAEL V = NOAEL $VOSO_4 \cdot 3H_2O$ × V molar mass/$VOSO_4 \cdot 3H_2O$ M.W. = 0.5 mg $VOSO_4 \cdot 3H_2O$/kg-day × 50.942 g V/mol/217.041 g $VOSO_4 \cdot 3H_2O$/mol = 0.12 mg V/kg-day.
[l]LOAEL V = LOAEL $NaVO_3$ × V molar mass/$NaVO_3$ M.W. = 5 mg NAO_3V/kg-day × 50.942 g V/mol/121.928 g $NaVO_3$/mol = 2.1 mg V/kg-day.

Table 3. Details on additional oral reference values lacking derivation descriptions[a]

Reference value name[b]	Duration	Compound	Reference value (mg/kg-day)	Health effect	Point of departure	Qualifier	Source	Uncertainty factors	Notes on derivation	Review status
TCEQ RfD	Chronic	Vanadium	0.0018	NR	NR	NR		NR	RfD developed with TCEQ's protocol (TCEQ, 2012)	Final (TCEQ, 2018)

[a]Health effect reference values listed in Table 3 are not shown in Figure 5 because they did not provide descriptions of how the value was derived.
[b]TCEQ = Texas Commission on Environmental Quality.

2.2. SCOPING SUMMARY

During scoping, the IRIS Program met with EPA program and regional offices that had interest in an IRIS assessment for vanadium and compounds to discuss specific assessment needs. Table 4 provides a summary of input from this outreach.

Table 4. EPA program and regional office interest in a reassessment of vanadium compounds

EPA program or regional office	Oral	Inhalation[a]	Statutes/Regulations	Anticipated uses/Interest
Office of Water	✓		Safe Drinking Water Act (SDWA) and Clean Water Act (CWA)	The SDWA requires EPA to list[b] contaminants that are currently not subject to any proposed or promulgated National Primary Drinking Water Regulation (NPDWR) but are known or anticipated to occur in public water systems, including vanadium. Contaminants listed on the CCL may require future regulation under SDWA. Under Section 304(a) of the CWA, EPA derives recommended ambient water quality criteria for the protection of human health. States and tribes may use these values or other values in their water quality standards to protect designated uses. Vanadium and compounds (oral) toxicological information may be used to address risk under the CWA and SDWA.

[a]The IRIS Program announced the initiation of a vanadium and compounds (inhalation) assessment in December 2019. A separate IAP will be released regarding the inhalation assessment.
[b]EPA's Final Contaminant Candidate List (CCL) 4 lists vanadium.

2.3. PROBLEM FORMULATION

Systematic review methods were used to identify a preliminary literature inventory for vanadium and compounds, using the literature search and screening methods described in Section 4. The ATSDR *Toxicological Profile for Vanadium* (ATSDR, 2012) was selected as the starting point for the literature search, and all references from the ATSDR document were retrieved and stored in EPA's Health and Environmental Research Online (HERO) database

(https://hero.epa.gov/hero/index.cfm/project/page/project_id/2357)[2] (see Section 4.1). Database searches were then conducted on March 28, 2019 by an EPA information specialist in three online databases (PubMed, Web of Science, Toxline)[3] and repeated on March 9, 2020 to identify records that had been published since the release of the 2012 ATSDR *Toxicological Profile for Vanadium* (see Section 4.2; database search strategies provided in Appendix B). Studies identified from the ATSDR document and the database searches were screened at the title/abstract level followed by full text screening (see Section 4.4), using PECO criteria (Populations, Exposures, Comparators, Outcomes; see Table 6) as a guide to identify relevant literature. Studies containing potentially relevant supplemental material were also tracked during the literature screening process (Table 7). Studies meeting PECO criteria were briefly summarized and are presented here using Tableau visualization software (https://www.tableau.com/) (see Section 4.5).

These methods were implemented in accordance with EPA Quality Assurance policies and procedures [Quality Policy Procedures[4] and CIO 2105.0 (formerly 5360.1 A2)[5]]. The results obtained from this systematic compilation of the evidence helped inform the specific aims and key science issues that will be the focus of the assessment.

2.4. LITERATURE INVENTORY RESULTS

The literature search and screening process identified 117 studies that met PECO criteria (38 epidemiological studies, 79 animal studies), and a total of 1,082 studies were tagged as potentially relevant supplemental material. No PBPK models for vanadium or vanadium compounds were identified.

This literature inventory summarizes the studies that met PECO criteria, but also includes human clinical trials and acute duration animal studies, which were tagged as potentially relevant supplemental material and may be used to supplement the evidence synthesis, as described in Section 3.2. The literature inventory has minor differences compared to the preliminary literature inventory presented in the IAP (U.S. EPA, 2020a) due to adjustments in the PECO criteria (see Section 3.2).

2.4.1. Human Studies Meeting PECO Criteria

A survey of study designs and health systems assessed in the human studies that met PECO criteria is provided in Figure 6, and a tabular summary is provided in Figure 7.

[2]EPA's HERO database provides access to the scientific literature behind EPA science assessments. The database includes more than 600,000 scientific references and data from the peer-reviewed literature used by EPA to develop its health assessment documents.

[3]The Toxline database was migrated to PubMed prior to the March 2020 literature search update, so the Toxline search was conducted only in March 2019.

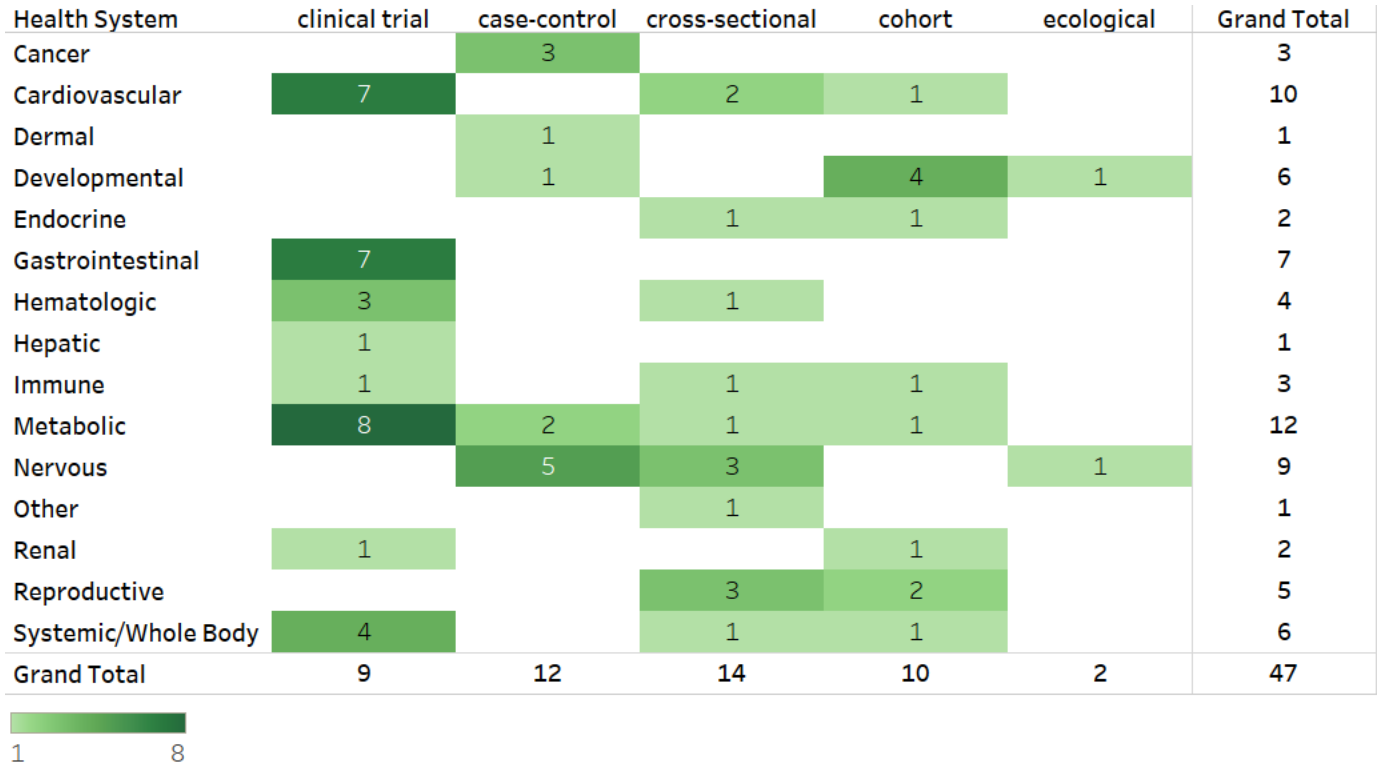
[4]U.S. Environmental Protection Agency Procedures for Quality Policy: https://www.epa.gov/quality/policies-and-procedures-about-quality-assurance-epa-organizations.

[5]*Policy and Program Requirements for the Mandatory Agency-Wide Quality System*: https://www.epa.gov/sites/production/files/2015-09/documents/epa_order_cio_21050.pdf.

The literature search identified 38 observational epidemiological studies, which evaluated the association of potentially adverse or beneficial health outcomes with total vanadium, but the specific form of vanadium was not determined. This included 36 studies (12 case-control, 14 cross-sectional, and 10 cohort) in the general population, pregnant women, infants, or children, in which vanadium exposure was evaluated using biomonitoring of blood (whole blood, plasma, or serum), urine, hair, seminal plasma, cerebrospinal fluid, saliva, or nails, but the route of exposure was unclear. Additionally, two ecological studies evaluated the association of human health outcomes with vanadium levels in soil, drinking water, or food.

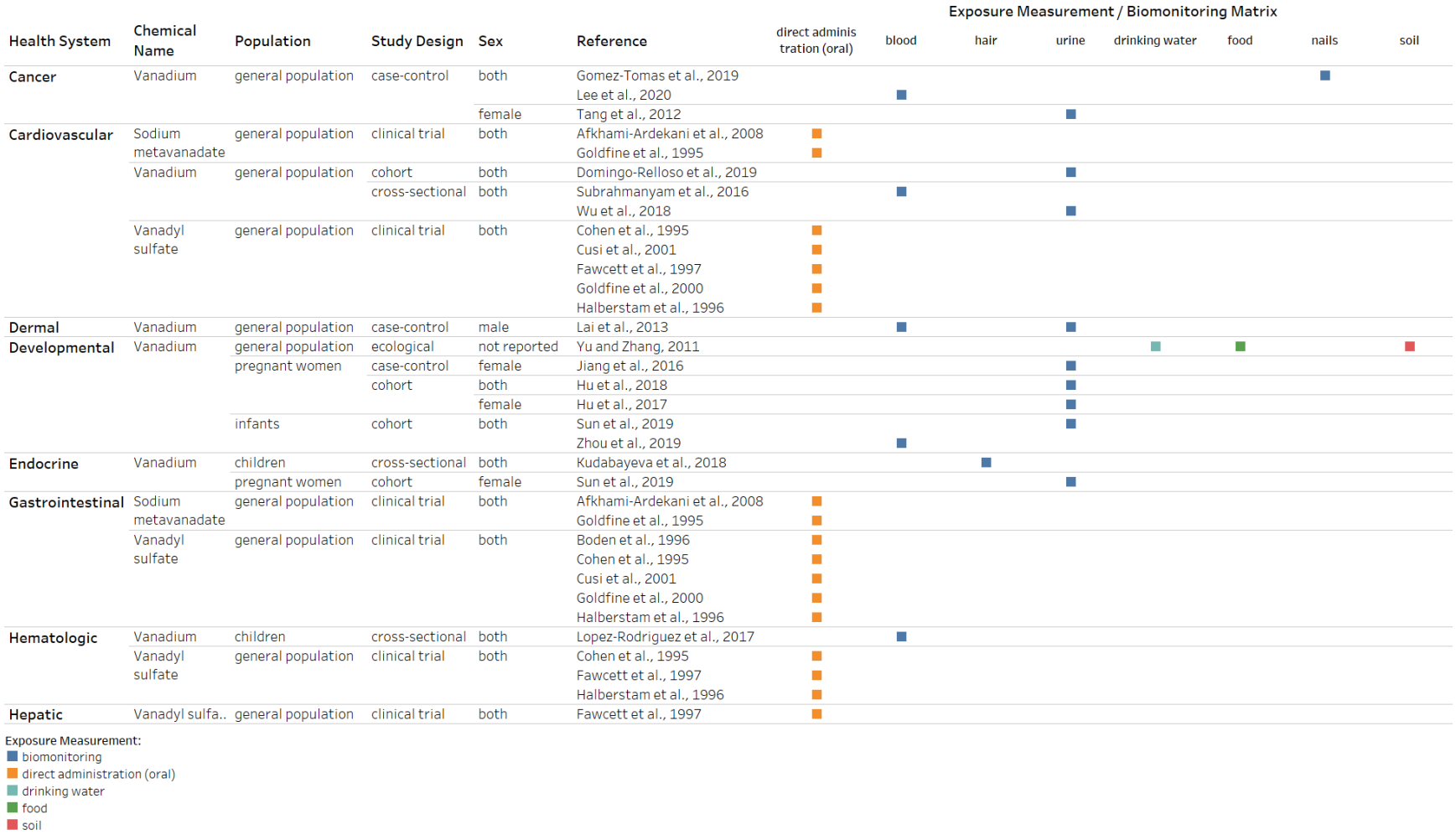
The literature search also identified nine clinical trials that administered vanadyl sulfate or sodium metavanadate directly to study participants.[6] As described in Section 3.2, these studies do not meet PECO criteria, but they are included in the literature inventory because they may be evaluated and used to supplement the evidence synthesis for certain endpoints that have evidence of adversity based on epidemiological and animal toxicological studies. Of the clinical trials, seven were conducted in diabetic patients for the purpose of evaluating the therapeutic effects of vanadium supplementation, with treatment durations of 2–6 weeks (Afkhami-Arekani et al., 2008; Cusi et al., 2001; Goldfine et al., 2000; Boden et al., 1996; Halberstam et al., 1996; Cohen et al., 1995; Goldfine et al., 1995); one evaluated effects of vanadyl sulfate supplementation on insulin sensitivity in seven healthy adults, with a treatment duration of 7 days (Jentjens and Jeukendrup, 2002), and one evaluated effects of vanadyl sulfate supplementation in 31 weight-training athletes, with a treatment duration of 12 weeks (Fawcett et al., 1997).

[6]Two additional clinical trials that evaluated "ammonium vanadyl tartrate" or "diammonium vanado-tartrate" are not summarized in this literature inventory but are tagged as potentially relevant supplemental information (Dimond et al., 1963; Somerville and Davies, 1962).



Health System	clinical trial	case-control	cross-sectional	cohort	ecological	Grand Total
Cancer		3				3
Cardiovascular	7		2	1		10
Dermal		1				1
Developmental		1		4	1	6
Endocrine			1	1		2
Gastrointestinal	7					7
Hematologic	3		1			4
Hepatic	1					1
Immune	1		1	1		3
Metabolic	8	2	1	1		12
Nervous		5	3		1	9
Other			1			1
Renal	1			1		2
Reproductive			3	2		5
Systemic/Whole Body	4		1	1		6
Grand Total	9	12	14	10	2	47

Figure 6. Survey of human studies that met PECO criteria by study design and health systems assessed. The numbers indicate the number of studies that investigated a particular health system, not the number of studies that observed an association with vanadium exposure. If a study evaluated multiple health outcomes, it is shown here multiple times. An interactive version of this figure that includes a more detailed description of study designs and results is available at the following URL: https://public.tableau.com/views/VanadiumEvidenceMapVisualizationsApril2021/ReadMe?:language=en&:display_count=y&publish=yes&:origin=viz_share_link



Health System	Chemical Name	Population	Study Design	Sex	Reference	Exposure Measurement / Biomonitoring Matrix: direct administration (oral)	blood	hair	urine	drinking water	food	nails	soil
Cancer	Vanadium	general population	case-control	both	Gomez-Tomas et al., 2019							■	
					Lee et al., 2020		■						
				female	Tang et al., 2012				■				
Cardiovascular	Sodium metavanadate	general population	clinical trial	both	Afkhami-Ardekani et al., 2008	■							
					Goldfine et al., 1995	■							
	Vanadium	general population	cohort	both	Domingo-Relloso et al., 2019				■				
			cross-sectional	both	Subrahmanyam et al., 2016		■						
					Wu et al., 2018				■				
	Vanadyl sulfate	general population	clinical trial	both	Cohen et al., 1995	■							
					Cusi et al., 2001	■							
					Fawcett et al., 1997	■							
					Goldfine et al., 2000	■							
					Halberstam et al., 1996	■							
Dermal	Vanadium	general population	case-control	male	Lai et al., 2013		■		■				
Developmental	Vanadium	general population	ecological	not reported	Yu and Zhang, 2011					■	■		■
		pregnant women	case-control	female	Jiang et al., 2016				■				
			cohort	both	Hu et al., 2018				■				
				female	Hu et al., 2017				■				
		infants	cohort	both	Sun et al., 2019				■				
					Zhou et al., 2019		■						
Endocrine	Vanadium	children	cross-sectional	both	Kudabayeva et al., 2018			■					
		pregnant women	cohort	female	Sun et al., 2019				■				
Gastrointestinal	Sodium metavanadate	general population	clinical trial	both	Afkhami-Ardekani et al., 2008	■							
					Goldfine et al., 1995	■							
	Vanadyl sulfate	general population	clinical trial	both	Boden et al., 1996	■							
					Cohen et al., 1995	■							
					Cusi et al., 2001	■							
					Goldfine et al., 2000	■							
					Halberstam et al., 1996	■							
Hematologic	Vanadium	children	cross-sectional	both	Lopez-Rodriguez et al., 2017		■						
	Vanadyl sulfate	general population	clinical trial	both	Cohen et al., 1995	■							
					Fawcett et al., 1997	■							
					Halberstam et al., 1996	■							
Hepatic	Vanadyl sulfa..	general population	clinical trial	both	Fawcett et al., 1997	■							

Figure 7. Tabular summary of study designs and exposure measurements used in human studies that met PECO criteria (continued on following page). An interactive version of this figure that includes a more detailed description of study designs and results is available at the following URL: https://public.tableau.com/views/VanadiumEvidenceMapVisualizationsApril2021/ReadMe?:language=en&:display_count=y&publish=yes&:origin=viz_share_link

						Exposure Measurement / Biomonitoring Matrix							
Health System	Chemical Name	Population	Study Design	Sex	Reference	direct administration (oral)	blood	cerebrospinal fluid	hair	semen	urine	saliva	soil
Immune	Vanadium	general population	cross-sectional	both	Pedro et al., 2019		■						
		infants	cohort	both	Zhou et al., 2019		■						
	Vanadyl sulfate	general population	clinical trial	both	Fawcett et al., 1997	■							
Metabolic	Sodium metavanadate	general population	clinical trial	both	Afkhami-Ardekani et al., 2008	■							
					Goldfine et al., 1995	■							
	Vanadium	general population	case-control	both	Li et al., 2017		■						
					Wang et al., 2014		■						
			cross-sectional	Null	Flores et al., 2011		■				■		
		pregnant women	cohort	female	Wang et al., 2020						■		
	Vanadyl sulfate	general population	clinical trial	both	Boden et al., 1996	■							
					Cohen et al., 1995	■							
					Cusi et al., 2001	■							
					Goldfine et al., 2000	■							
					Halberstam et al., 1996	■							
					Jentjens and Jeukendrup, 2002	■							
Nervous	Vanadium	general population	case-control	both	Roos et al., 2013			■					
					Squadrone et al., 2018		■						
				female	Naylor et al., 1984				■				
			cross-sectional	both	Kihira et al., 2015				■				
		children	case-control	both	Alqhazo and Rashaid, 2018				■				
					Skalny et al., 2017				■				
			cross-sectional	both	Blaurock-Busch et al., 2012				■				
				male	Tinkov et al., 2019		■		■				
			ecological	both	Zahran et al., 2012								■
Other	Vanadium	general population	cross-sectional	both	Inonu et al., 2019							■	
Renal	Vanadium	general population	cohort	both	Liu et al., 2020		■						
	Vanadyl sulfate	general population	clinical trial	both	Fawcett et al., 1997	■							
Reproductive	Vanadium	general population	cross-sectional	female	Zheng et al., 2015		■						
				male	Skalnaya et al., 2015					■			
					Wang et al., 2018					■	■		
		pregnant women	cohort	female	Jin et al., 2018						■		
					Zheng et al., 2014		■						
Systemic/Whole Body	Sodium metavanadate	general population	clinical trial	both	Afkhami-Ardekani et al., 2008	■							
					Goldfine et al., 1995	■							
	Vanadium	children	cross-sectional	both	Tascilar et al., 2011		■						
		pregnant women	cohort	female	Skalny et al., 2020				■				
	Vanadyl sulfate	general population	clinical trial	both	Cohen et al., 1995	■							
					Goldfine et al., 2000	■							

Exposure Measurement:
- biomonitoring
- direct administration (oral)
- soil

Figure 7 continued.

2.4.2. Animal Studies Meeting PECO Criteria

A survey of the types of vanadium compounds evaluated in animal studies that met PECO criteria is shown in Figure 8, and a survey of study designs, species, and health effects evaluated in the animal studies is provided in Figure 9. The animal studies evaluated exposure to ammonium metavanadate, sodium metavanadate, sodium orthovanadate, vanadyl sulfate, vanadium pentoxide, calcium orthovanadate, or calcium pyrovanadate. Of these, vanadyl sulfate and sodium metavanadate were the most frequently studied compounds. Two studies reported that animals were exposed to "ammonium vanadate" (Susić and Kentera, 1986) and "sodium vanadate" (Sun et al., 2014), which were inferred to be ammonium metavanadate and sodium metavanadate (respectively) based on the synonyms reported in Table 1 and are referred to accordingly here. One study reported that animals were exposed to "vanadium" or "vanadate," but the specific chemical form was unclear. Most studies were conducted in rats and mice, but data were also available in rabbits, cattle, goats, and sheep. Studies with acute exposure durations (<24 hours; 3 studies) do not meet PECO criteria, but are included in this literature inventory as they can be helpful to interpret findings from studies more directly informative for developing a chronic toxicity value; they will be considered potentially relevant supplemental material and will not undergo full study evaluation and data extraction.

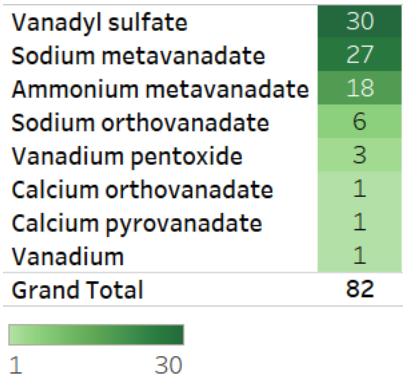


Figure 8. Survey of the vanadium compounds evaluated in the available animal studies, showing the number of studies that evaluated each vanadium compound. This includes acute studies, which did not meet PECO criteria but will be evaluated as potentially relevant supplemental information. If a study evaluated multiple types of vanadium compounds, it is shown here multiple times. An interactive version of this figure that includes a more detailed description of study designs and results is available at the following URL: https://public.tableau.com/views/VanadiumEvidenceMapVisualizationsApril2021/ReadMe?:language=en&:display_count=y&publish=yes&:origin=viz_share_link

Tabular summaries of the study designs and health effects evaluated in chronic, subchronic, and reproductive or developmental studies that tested multiple dose levels are provided in

Figures 10, 11, and 12, respectively.[7] 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 where other study designs are more suitable. Figures are organized by health outcomes evaluated.

[7]Dose levels shown in tabular summaries are those reported by the authors. For the assessment, doses reported as concentrations in food or drinking water (e.g., ppm, µg/mL) will be converted to mg/kg-day.

Health System	acute			short-term			subchronic				chronic				reproductive		developmental		Grand Total
	rat	mouse	sheep	rat	mouse	sheep	rat	mouse	cattle	goat	rat	mouse	cattle	goat	rat	mouse	rat	mouse	
Cancer							1				2	2							5
Cardiovascular				1	1		7		1	1	9		1				2		23
Dermal											1								1
Developmental																	6	2	8
Endocrine									1		1		1				1		4
Gastrointestinal			1		1		1												3
Hematologic				5			8		1		5								19
Hepatic	1		1	4	2		4	2	1		1				2	2	4		20
Immune				6	2		8	3	1		2						2		22
Metabolic				2	2		12		1	1	4		1		3		1		27
Musculoskeletal					1								1						2
Nervous				1	1		6				1						2		11
Renal			1	2	2		5	2	1		6				3	2	3		25
Reproductive								1						1	8	3	2		13
Respiratory				1	2		2				1						2		7
Systemic/Whole Body	2	1	1	11	3	1	23	2	1	1	8	3	1	1	4	2		1	61
Grand Total	2	1	1	11	3	1	29	3	1	1	15	3	1	1	9	3	9	3	82

1 23

Figure 9. Survey of animal studies that met PECO criteria by study design and species and health systems assessed. The numbers indicate the number of studies that investigated a particular health system, not the number of studies that observed an association with vanadium exposure. If a study evaluated multiple species, study designs, or health outcomes, it is shown here multiple times. An interactive version of this figure that includes a more detailed description of study designs and results is available at the following URL: https://public.tableau.com/views/VanadiumEvidenceMapVisualizationsApril2021/ReadMe?:language=en&:display_count=y&publish=yes&:origin=viz_share_link

Health System	Chemical Name	Species	Sex	Dosing Duration	All dose levels	Dose units	Reference	
Cancer	Ammonium metavanadate	mouse	male	30 wk	0, 10, 20	ppm V	Kingsnorth et al., 1986	
	Vanadyl sulfate	rat	female	180 d (28 d with 15 ppm, then with 25 ppm till 180 d)	0, 15,25	ppm V	Thompson et al., 1984	
Cardiovascular	Sodium metavanadate	rat	male	7 mon	0, 10, 40	ug V/mL	Carmignani et al., 1992	
				24 wk	0, 300, 3000	ppm	Susic and Kentera, 1988	
				210 d	0, 10, 40	ug V/mL	Boscolo et al., 1994	
		cattle	not reported	150 d	0, 3, 6, 9	mg V/kg	Pal et al., 2018	
	Vanadyl sulfate	rat	not reported	24 wk	0, 0.25, 1.2	mg/kg-d	Shah et al., 2016	
Dermal	Vanadium pentoxide	rat	male	75 d	0, 500, 1000	ppm V	Mountain et al., 1953	
				103 d	0, 100, 150	ppm V	Mountain et al., 1953	
Endocrine	Sodium metavanadate	cattle	not reported	150 d	0, 3, 6, 9	mg V/kg	Pal et al., 2018	
Hematologic	Sodium metavanadate	rat	male	24 wk	0, 300, 3000	ppm	Susic and Kentera, 1988	
	Vanadium pentoxide	rat	male	103 d	0, 100, 150	ppm V	Mountain et al., 1953	
	Vanadyl sulfate	rat	female	180 d (28 d with 15 ppm, then with 25 ppm till 180 d)	0, 15,25	ppm V	Thompson et al., 1984	
Immune	Vanadium pentoxide	rat	both	6 mon	0, 1, 100	mg V/L	Mravcova et al., 1993	
Metabolic	Sodium metavanadate	cattle	not reported	150 d	0, 3, 6, 9	mg V/kg	Pal et al., 2018	
	Vanadyl sulfate	rat	not reported	24 wk	0, 0.25, 1.2	mg/kg-d	Shah et al., 2016	
Musculoskeletal	Sodium metavanadate	cattle	not reported	150 d	0, 3, 6, 9	mg V/kg	Pal et al., 2018	
Renal	Sodium metavanadate	rat	male	7 mon	0, 10, 40	ug V/mL	Carmigani et al., 1992	
				24 wk	0, 300, 3000	ppm	Susic and Kentera, 1988	
				210 d	0, 10, 40	ug V/mL	Boscolo et al., 1994	

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

Figure 10. Summary of multidose chronic animal studies (continued on following page). An interactive version of this figure that includes a more detailed description of study designs and results is available at the following URL: https://public.tableau.com/views/VanadiumEvidenceMapVisualizationsApril2021/ReadMe?:language=en&:display_count=y&publish=yes&:origin=viz_share_link

Health System	Chemical Name	Species	Sex	Dosing Duration	All dose levels	Dose units	Reference	
Reproductive	Sodium metavanadate	goat	female	130 d	0, 2, 4, 6	ppm V	Tripathi et al., 2018	
Systemic/Whole Body	Ammonium metavanadate	mouse	male	30 wk	0, 10, 20	ppm V	Kingsnorth et al., 1986	
	Sodium metavanadate	rat	male	24 wk	0, 300, 3000	ppm	Susic and Kentera, 1988	
		cattle	not reported	150 d	0, 3, 6, 9	mg V/kg	Pal et al., 2018	
		goat	female	130 d	0, 2, 4, 6	ppm V	Tripathi et al., 2018	
	Sodium orthovanadate	rat	male	56 wk	0, 100, 200	ppm V	Steffen et al., 1981	
	Vanadium pentoxide	rat	male	75 d	0, 500, 1000	ppm V	Mountain et al., 1953	
				103 d	0, 100, 150	ppm V	Mountain et al., 1953	
	Vanadyl sulfate	rat	female	180 d (28 d with 15 ppm, then with 25 ppm till 180 d)	0, 15,25	ppm V	Thompson et al., 1984	

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

Figure 10 continued.

Health System	Chemical Name	Species	Sex	Dosing Duration	All dose levels	Dose units	Reference	
Cardiovascular	Ammonium metavanadate	rat	female	35 d	0, 3, 15, 30	mg V/kg	Wang et al., 2019	
	Sodium metavanadate	rat	male	3 mon	0, 5, 10, 50	ppm	Domingo et al., 1985	
		cattle	female	90 d	0, 2.5, 5	ppm V	Gupta et al., 2020	
	Sodium orthovanadate	rat	male	56 wk	0, 100, 200	ppm V	Steffen et al., 1981	
	Vanadyl sulfate	goat	not reported	84 d	0, 1, 2, 3	mg V/d	Zarqami et al., 2017	
Endocrine	Sodium metavanadate	cattle	female	90 d	0, 2.5, 5	ppm V	Gupta et al., 2020	
Hematologic	Sodium metavanadate	rat	female	10 wk	0, 50, 100	ppm V	Adachi et al., 2000	
		cattle	female	90 d	0, 2.5, 5	ppm V	Gupta et al., 2020	
Hepatic	Ammonium metavanadate	rat	female	5 wk	0, 3, 15, 30	mg V/kg	Wang et al., 2016	
				35 d	0, 3, 15, 30	mg V/kg	Wang et al., 2019	
	Sodium metavanadate	rat	female	10 wk	0, 50, 100	ppm V	Adachi et al., 2000	
			male	3 mon	0, 5, 10, 50	ppm	Domingo et al., 1985	
		cattle	female	90 d	0, 2.5, 5	ppm V	Gupta et al., 2020	
	Sodium orthovanadate	mouse	male	13 wk	0, 1, 10, 50	mg V/L	Sharma et al., 1981	
	Vanadyl sulfate	mouse	male	5 wk	0, 2, 10	mg/L	Villani et al., 2007	
					0, 10, 100, 50..	mg/L	Villani et al., 2007	

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

Figure 11. Summary of multidose subchronic animal studies (continued on following pages). An interactive version of this figure that includes a more detailed description of study designs and results is available at the following URL: https://public.tableau.com/views/VanadiumEvidenceMapVisualizationsApril2021/ReadMe?:language=en&:display_count=y&publish=yes&:origin=viz_share_link

Health System	Chemical Name	Species	Sex	Dosing Duration	All dose levels	Dose units	Reference	
Immune	Ammonium metavanadate	rat	female	35 d	0, 3, 15, 30	mg V/kg	Wang et al., 2019	
	Sodium metavanadate	rat	female	10 wk	0, 50, 100	ppm V	Adachi et al., 2000	
			male	3 mon	0, 5, 10, 50	ppm	Domingo et al., 1985	
		cattle	female	90 d	0, 2.5, 5	ppm V	Gupta et al., 2020	
	Sodium orthovanadate	mouse	male	13 wk	0, 1, 10, 50	mg V/L	Sharma et al., 1981	
	Vanadyl sulfate	mouse	male	5 wk	0, 2, 10	mg/L	Villani et al., 2007	
					0, 10, 100, 500, 1000	mg/L	Villani et al., 2007	
Metabolic	Ammonium metavanadate	rat	female	35 d	0, 3, 15, 30	mg V/kg	Wang et al., 2019	
	Sodium metavanadate	rat	male	3 mon	0, 5, 10, 50	ppm	Domingo et al., 1985	
		cattle	female	90 d	0, 2.5, 5	ppm V	Gupta et al., 2020	
	Vanadyl sulfate	goat	not reported	84 d	0, 1, 2, 3	mg V/d	Zarqami et al., 2017	
Nervous	Sodium metavanadate	rat	female	10 wk	0, 50, 100	ppm V	Adachi et al., 2000	
			male	8 wk	0, 4.1, 8.2, 16.4	mg/kg-d	Sanchez et al., 1998	
				12 wk	0, 0.5, 1.0, 2.0	g/L	Sun et al., 2017	
	Sodium orthovanadate	rat	male	56 wk	0, 100, 200	ppm V	Steffen et al., 1981	
Renal	Ammonium metavanadate	rat	female	35 d	0, 3, 15, 30	mg V/kg	Wang et al., 2019	
	Sodium metavanadate	rat	male	3 mon	0, 5, 10, 50	ppm	Domingo et al., 1985	
		cattle	female	90 d	0, 2.5, 5	ppm V	Gupta et al., 2020	
	Sodium orthovanadate	mouse	male	13 wk	0, 1, 10, 50	mg V/L	Sharma et al., 1981	
	Vanadium	rat	female	5 weeks	0, 3, 15, 30	mg V/kg	Wang et al., 2016	
	Vanadyl sulfate	mouse	male	5 wk	0, 2, 10	mg/L	Villani et al., 2007	
					0, 10, 100, 500, 1000	mg/L	Villani et al., 2007	

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

Figure 11 continued.

Health System	Chemical Name	Species	Sex	Dosing Duration	All dose levels	Dose units	Reference	
Reproductive	Vanadyl sulfate	mouse	male	5 wk	0, 2, 10	mg/L	Villani et al., 2007	■
					0, 10, 100, 500, 1000	mg/L	Villani et al., 2007	■
Respiratory	Ammonium metavanadate	rat	female	35 d	0, 3, 15, 30	mg V/kg	Wang et al., 2019	■
	Sodium metavanadate	rat	male	3 mon	0, 5, 10, 50	ppm	Domingo et al., 1985	■
Systemic/ Whole Body	Ammonium metavanadate	rat	female	35 d	0, 3, 15, 30	mg V/kg	Wang et al., 2019	■
	Sodium metavanadate	rat	female	10 wk	0, 50, 100	ppm V	Adachi et al., 2000	■
			male	3 mon	0, 5, 10, 50	ppm	Domingo et al., 1985	■
				8 wk	0, 4.1, 8.2, 16.4	mg/kg-d	Sanchez et al., 1998	■
		cattle	female	90 d	0, 2.5, 5	ppm V	Gupta et al., 2020	■
	Sodium orthovanadate	mouse	male	13 wk	0, 1, 10, 50	mg V/L	Sharma et al., 1981	■
	Vanadium	rat	male	60 d	0, 20, 40	mg/kg	Tubafard et al., 2010	■
	Vanadyl sulfate	mouse	male	5 wk	0, 2, 10	mg/L	Villani et al., 2007	■
					0, 10, 100, 500, 1000	mg/L	Villani et al., 2007	■
		goat	not reported	84 d	0, 1, 2, 3	mg V/d	Zarqami et al., 2017	■

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

Figure 11 continued.

Health System	Chemical Name	Species	Sex	Dosing Duration	All dose levels	Dose units	Reference	
Cardiovascular	Sodium metavanadate	rat	both	60 d (F0 male); 14 d premating + gestation + lactation (F0 female)	0, 5, 10, 20	mg/kg-d	Domingo et al., 1986	■
Developmental	Sodium metavanadate	rat	both	60 d (F0 male); 14 d premating + gestation + lactation (F0 female)	0, 5, 10, 20	mg/kg-d	Domingo et al., 1986	■
				GD6-GD14	0, 5, 10, 20	mg/kg	Paternain et al., 1987	■
	Sodium orthovanadate	mouse	both	GD0-18	0,7.5, 15, 30, 60	mg/kg	Sanchez et al., 1990	■
	Vanadyl sulfate	mouse	both	GD6-15	0, 37.5, 75, 150	mg/kg-d	Paternain et al., 1990	■
Hepatic	Sodium metavanadate	rat	both	60 d (F0 male); 14 d premating + gestation + lactation (F0 female)	0, 5, 10, 20	mg/kg-d	Domingo et al., 1986	■
	Sodium orthovanadate	mouse	female (dam)	GD0-18	0,7.5, 15, 30, 60	mg/kg	Sanchez et al., 1990	■
	Vanadyl sulfate	mouse	female (dam)	GD6-15	0, 37.5, 75, 150	mg/kg-d	Paternain et al., 1990	■
Immune	Sodium metavanadate	rat	both	60 d (F0 male); 14 d premating + gestation + lactation (F0 female)	0, 5, 10, 20	mg/kg-d	Domingo et al., 1986	■
Renal	Sodium metavanadate	rat	both	60 d (F0 male); 14 d premating + gestation + lactation (F0 female)	0, 5, 10, 20	mg/kg-d	Domingo et al., 1986	■
	Sodium orthovanadate	mouse	female (dam)	GD0-18	0,7.5, 15, 30, 60	mg/kg	Sanchez et al., 1990	■
	Vanadyl sulfate	mouse	female (dam)	GD6-15	0, 37.5, 75, 150	mg/kg-d	Paternain et al., 1990	■

Route of Exposure:
■ oral (gavage)

Figure 12. Summary of multidose reproductive and developmental animal studies (continued on following page). An interactive version of this figure that includes a more detailed description of study designs and results is available at the following URL: https://public.tableau.com/views/VanadiumEvidenceMapVisualizationsApril2021/ReadMe?:language=en&:display_count=y&publish=yes&:origin=viz_share_link

Health System	Chemical Name	Species	Sex	Dosing Duration	All dose levels	Dose units	Reference	
Reproductive	Sodium metavanadate	rat	both	60 d (F0 male); 14 d premating + gestation + lactation (F0 female)	0, 5, 10, 20	mg/kg-d	Domingo et al., 1986	■
			female (dam)	GD6-GD14	0, 5, 10, 20	mg/kg	Paternain et al., 1987	■
		mouse	male	64 d	0, 20, 40, 60, 80	mg/kg-d	Llobet et al., 1993	■
	Sodium orthovanadate	rat	female (dam)	mating-PND1	0, 0.25, 0.50	mg/mL	Ganguli et al., 1994a	■
		mouse	female (dam)	GD0-18	0,7.5, 15, 30, 60	mg/kg	Sanchez et al., 1990	■
	Vanadyl sulfate	mouse	female (dam)	GD6-15	0, 37.5, 75, 150	mg/kg-d	Paternain et al., 1990	■
Respiratory	Sodium metavanadate	rat	both	60 d (F0 male); 14 d premating + gestation + lactation (F0 female)	0, 5, 10, 20	mg/kg-d	Domingo et al., 1986	■
Systemic/ Whole Body	Sodium metavanadate	mouse	male	64 d	0, 20, 40, 60, 80	mg/kg-d	Llobet et al., 1993	■
	Sodium orthovanadate	mouse	female (dam)	GD0-18	0,7.5, 15, 30, 60	mg/kg	Sanchez et al., 1990	■

Route of Exposure:
■ oral (gavage)
■ oral (water)

Figure 12 continued.

2.4.3. Studies in Progress by the National Toxicology Program

The interim results of NTP's extended developmental study in rats and 13-week study in mice (currently available as a poster,[8] with complete results expected to be published in 2021) were also considered for problem formulation, as these studies were conducted by NTP following nomination by EPA and the National Institute of Environmental Health Sciences and are intended to address data gaps related to the oral toxicity of pentavalent and tetravalent vanadium compounds.[9]

In the developmental study, rat F1 offspring were initially exposed in utero and via breast milk, and then continued to receive the same dose levels as their mothers via drinking water for 13 weeks following weaning. Moribundity of F0 dams was observed during parturition and lactation in the 250 and 500 mg/L sodium metavanadate dose groups, with decreased maternal

[8] https://ntp.niehs.nih.gov/ntp/results/pubs/posters/roberts_sot20190300.pdf.
[9] https://ntp.niehs.nih.gov/getinvolved/nominate/summary/nm-n20806.html?utm_source=direct&utm_medium=prod&utm_campaign=ntpgolinks&utm_term=nm-n20806.

body weights in proportion to dose. F1 pups exposed to sodium metavanadate had decreased survival from postnatal days 1–10 in the 500 mg/L dose group, and F1 body weights at the end of the study were found to be decreased in males at doses ≥125 mg/L and in females in the 500 mg/L dose group. Conversely, no impacts on F0 or F1 survival or body weight were observed in rats exposed to vanadyl sulfate. Analysis of total vanadium concentrations in plasma and urine of a subset of F1 rats at the end of the exposure period in the developmental study indicated higher absorption of sodium metavanadate compared to vanadyl sulfate when consuming similar levels of vanadium, which may explain the differential toxicity between these two compounds. The analysis of clinical pathology, organ weight, and histopathology data from the developmental study is ongoing. Similarly, NTP's 13-week study in mice observed toxicity following exposure to sodium metavanadate but not vanadyl sulfate. Mice exposed to sodium metavanadate had decreased body weights (observed at doses of 500 mg/L in males and at 250 and 500 mg/L in females), decreased thymus weights (observed at doses of 250 mg/L in males and 500 mg/L in females), increased erythrocytes and reticulocytes (observed at 500 mg/L in males and females), and small decreases in hematocrit and hemoglobin.

2.4.4. Comparison with Studies Used in the 1987 IRIS Assessment

As described earlier in this document, the 1987 IRIS RfD of 0.009 mg/kg-day for vanadium pentoxide was based on a chronic (lifetime) NOAEL of 10 ppm vanadium (LOAEL of 100 ppm) for decreased hair cystine levels from the study in rats by Stokinger et al. (1953). Decreased hair cystine content is a biomarker that has been associated with certain pathological conditions in rodents and humans (Mountain et al., 1953) but has limited interpretation with respect to adversity and biological significance. Table 5 presents an overview of chronic health effects data that have become available since the 1987 IRIS vanadium health effects assessment. This table summarizes the study designs and NOELs/LOELs (reflecting only author-reported statistical significance) in the chronic animal studies from the current literature inventory that tested multiple dose levels of vanadium and that were not included in the 1987 IRIS assessment. Dose levels in this table are expressed as elemental vanadium to allow for comparison across compounds. The author-reported NOELs in these studies ranged from 1 to 100 ppm vanadium in drinking water and 3 to 6 ppm in diet. The author-reported LOELs ranged from 1 to 200 ppm vanadium in drinking water and 6 to 125.3 ppm vanadium in diet; the reported LOAEL was 0.078 mg/kg-day via oral gavage.

2.4.5. Literature Inventory Summary

The literature inventory includes a range of study designs and outcomes that are potentially useful for hazard identification or dose-response analysis for vanadium and compounds. On the basis of this survey, a refined evaluation plan was developed to focus on those health outcomes for which adequate evidence exists to develop conclusions about potential hazard (see Section 5).

Table 5. Summary of NOELs and LOELs from all multidose chronic animal studies that were not included in the 1987 IRIS health effects assessment of vanadium, with doses expressed as (A) parts-per-million (ppm) vanadium or (B) mg/kg-day vanadium. NOELs and LOELs are based on author-reported statistical significance. Results (bold italics) from Stokinger et al. (1953) (used to derive the 1987 IRIS RfD) are shown for reference. Studies are ordered from lowest to highest LOEL, followed by lowest to highest NOEL for studies that observed no effects within the tested dose range.

A.

Reference[a]	Chemical name	Route	Species (Strain)	NOEL (ppm vanadium)[b]	LOEL (ppm vanadium)[b]	Effects summary at LOEL
Pal et al. (2018)[c]	Sodium metavanadate	Diet	Cattle [Karan Fries (Tharparkar x Holstein Friesian) crossbred calves]	3	6	Increased insulin-like growth factor, increased total triiodothyronine (T3), increased total thyroxin (T4), increased bone alkaline phosphatase, decreased bone protein tyrosine phosphatase
Boscolo et al. (1994)[c]	Sodium metavanadate	Drinking water	Rat (Sprague-Dawley)	–	10/1[d]	Experiment 1: Increased systolic and diastolic blood pressure, plasma renin activity, plasma aldosterone, plasma aldosterone, urinary kallikrein and kinase I and II, and urinary potassium at 10 ppm Experiment 2: Increased systolic and diastolic blood pressure, decreased plasma aldosterone, decreased urinary kallikrein, decreased urinary calcium at 1 ppm
Carmignani et al. (1992)[c]	Sodium metavanadate	Drinking water	Rat (Sprague-Dawley)	–	10	Increased plasma renin activity, plasma aldosterone, aortic blood pressure; urine parameters (increased kallikrein levels, kininase I and II levels, enkephalinase levels)
Mravcová et al. (1993)[c]	Vanadium pentoxide	Drinking water	Rat (Wistar)	1	10	Increased spleen weight, decreased phagocytosis
Stokinger et al. (1953)[c]	***Vanadium pentoxide***	***Diet***	***Rat***	***10***	***100***	***Decreased hair cystine***
Susić and Kentera (1988)[e]	Sodium metavanadate	Diet	Rat (Long-Evans)	–	125.3	Decreased body weight, decreased cardiac output, increased total peripheral resistance (increased hematocrit and decreased plasma, blood and extracellular fluid volume observed at 1,253 ppm vanadium)

Table 5. Summary of NOELs and LOELs from all multidose chronic animal studies that were not included in the 1987 IRIS health effects assessment of vanadium, with doses expressed as (A) parts-per-million (ppm) vanadium or (B) mg/kg-day vanadium (continued)

Reference[a]	Chemical name	Route	Species (Strain)	NOEL (ppm vanadium)[b]	LOEL (ppm vanadium)[b]	Effects summary at LOEL
Steffen et al. (1981)[c]	Sodium orthovanadate	Drinking water	Rat (Sprague-Dawley)	–	100	Increased systolic blood pressure, increased relative heart weight (decreased body weight gain at 200 ppm vanadium)
Tripathi et al. (2018)[c]	Sodium metavanadate	Diet	Goat (Alpine × Beetal and Saanen × Beetal)	6	–	No change in final body weight, food intake, milk yield, or milk composition
Kingsnorth et al. (1986)[c]	Ammonium metavanadate	Drinking water	Mouse (CD-1)	20	–	No change in or survival or body weight gain

B.

Reference[a]	Chemical name	Route	Species (Strain)	NOEL (mg/kg-day vanadium)	LOEL (mg/kg-day vanadium)	Effects summary at LOEL
Shah et al. (2016)[f]	Vanadyl sulfate	Gavage	Rat	–	0.078	Increased serum triglycerides, increased total cholesterol, increased LDL-c, increased VLDL-c, decreased HDL-c, decreased plasma glucose, decreased serum insulin

[a]Carmignani et al. (1992) was published in a book containing proceedings of the 31st Congress of the EUROTOX. All other studies were published in peer-reviewed journals.

[b]1 ppm = 1 mg/kg diet or 1 mg/L drinking water.

[c]Studies by Boscolo et al. (1994), Pal et al. (2018), Carmignani et al. (1992), Mravcová et al. (1993), Stokinger et al. (1953), Steffen et al. (1981), Tripathi et al. (2018), and Kingsnorth et al. (1986) were interpreted as reporting dose levels for vanadium compounds in terms of elemental vanadium. Doses shown in this table are those reported by the authors.

[d]Boscolo et al. (1994) evaluated dose levels of 0, 10, and 40 µg/mL in Experiment 1, whereas Experiment 2 evaluated a single dose level (0 and 1 µg/mL).

[e]Susić and Kentera (1988) reported a LOEL of 300 ppm $NaVO_3$. This was converted to elemental vanadium using the following molecular weight conversion: LOEL V = LOEL $NaVO_3$ × V molar mass/$NaVO_3$ M.W. = 300 ppm $NaVO_3$ × 50.942 g V/mol/121.928 g $NaVO_3$/mol = 125.3 ppm V.

[f]Shah et al. (2016) reported a LOEL of 0.25 mg $VOSO_4$/kg-day. This was converted to elemental vanadium using the following molecular weight conversion: LOEL V = LOEL $VOSO_4$ × V molar mass/$VOSO_4$ M.W. = 0.25 mg $VOSO_4$/kg-day × 50.942 g V/mol/163 g $VOSO_4$/mol = 0.078 mg V/kg-day.

2.5. KEY SCIENCE ISSUES

The following key scientific issues were identified that warrant evaluation in this assessment.

- Key Science Issue #1: Consideration of potential toxicity and toxicokinetic differences across vanadium compounds.

  Differential absorption has been observed across inorganic vanadium compounds. For instance, as described earlier in this document, studies in progress by NTP preliminarily report that drinking water exposure to sodium metavanadate (V^{+5}) in rats led to higher levels of vanadium in plasma and urine as compared to vanadyl sulfate (V^{+4}) at similar vanadium exposure levels. This is consistent with reports that V^{+5} is absorbed more readily in the gastrointestinal tract compared to V^{+4} (Treviño et al., 2019; Nielsen, 1995). Absorption may be correlated with toxicity, as the effects reported by NTP in the preliminary report were more pronounced following exposure to sodium metavanadate compared to vanadyl sulfate. Given these apparent toxicokinetic differences across compounds, EPA plans to conduct separate toxicity evaluations for different vanadium compounds.

- Key Science Issue #2: Consideration of vanadium speciation.

  Available information indicates that vanadium in solution can readily convert between oxidation states and will form different spectra of species as a function of factors including pH, concentration, and redox potential. For instance, tetravalent vanadium in drinking water is stable at acidic pH but can convert to pentavalent species at neutral or basic pH (Mutlu et al., 2017). Given the apparent toxicokinetic (and, likely, toxicity) differences across vanadium compounds (see Key Science Issue #1), study evaluations will—to the extent possible—consider factors that could affect vanadium oxidation state and speciation in the available toxicity studies (e.g., pH of dosing solutions). Higher confidence will be placed in studies that have analytical confirmation of the vanadium species or oxidation state. Study evaluation considerations specific to vanadium are outlined in Sections 6.2.1 and 6.3.1.

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

This section outlines the specific aims and draft PECO criteria that will be used in developing the IRIS assessment. The overall objective of this assessment is to identify adverse health effects and characterize exposure-response relationships for these effects of vanadium and compounds to support development of oral toxicity values. This assessment will use systematic review methods to evaluate the epidemiological and toxicological literature for vanadium and compounds, including consideration of relevant mechanistic evidence. The evaluation conducted in this assessment will use relevant EPA guidance.[10]

3.1. SPECIFIC AIMS

- Identify epidemiological (i.e., human) and toxicological (i.e., experimental animal) literature reporting effects of exposure to vanadium compounds as outlined in the PECO criteria, and inventory literature that is potentially relevant to the specific aims (e.g., toxicokinetic, mechanistic). The ATSDR *Toxicological Profile for Vanadium* (ATSDR, 2012) will serve as the starting point for the literature search because it is the most recent review of health effects of vanadium and compounds published by a U.S. federal government agency that has undergone public comment and external peer review. Database searches will be conducted to identify records that have been published since the literature was last searched for the 2012 ATSDR *Toxicological Profile for Vanadium*.

- Conduct study evaluations (risk of bias and sensitivity) for individual epidemiological and toxicological studies and (if identified) PBPK models.

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

- Review and incorporate the available toxicokinetic and mechanistic information, as warranted to support assessment decisions. The toxicokinetic analyses will focus primarily on the key science issues identified in Section 2.5. The scope of the analysis of mechanistic information will be determined by the complexity and confidence in the phenotypic

[10]The EPA guidelines have been developed over time and address the state of the science at the time they were developed. Thus, evaluation methods may be updated as new science emerges, or when existing guidelines are updated. EPA guidance documents can be found at: http://www.epa.gov/iris/basic-information-about-integrated-risk-information-system#guidance/.

evidence in humans and animals, the likelihood of the analyses to affect evidence synthesis conclusions for human health, ability to inform dose-response extrapolation decisions, and the directness or relevance of the available model systems for understanding potential human health hazards.

- For each evidence stream (i.e., studies in humans, animal studies, and mechanistic or other supplemental studies, as appropriate and depending on data availability), synthesize the evidence across studies, assessing similar health outcomes using a narrative approach.

- For each health outcome, determine the strength of the evidence within and across evidence streams using structured frameworks to draw evidence integration judgments about the potential for vanadium and compounds exposure to be hazardous to humans for the oral route of exposure. Identify and discuss issues concerning potentially susceptible populations and life stages.

- Derive oral toxicity values [e.g., reference doses (RfDs), cancer risk estimates for oral exposure] as supported by the available data. The assessment will attempt to derive separate toxicity values for individual vanadium compounds or oxidation states (V^{+5}, V^{+4}) and an overall toxicity value for vanadium, as supported by the available data.

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

3.2. POPULATIONS, EXPOSURES, COMPARATORS, AND OUTCOMES CRITERIA

The PECO criteria are used to identify the evidence that addresses the specific aims of the assessment as well as to focus the search terms and inclusion/exclusion criteria in a systematic review. The PECO criteria for vanadium and compounds (Table 6) were based on (1) nomination of the chemicals for assessment, (2) discussions with scientists in the Office of Water to determine the scope of the assessment that would best meet Agency needs, (3) review of the health effects literature for vanadium and compounds to identify the health hazards potentially associated with oral exposure to vanadium and compounds and key areas of scientific complexity, and (4) public comments received on the IAP that was released in August 2020.

Minor revisions were made to the PECO criteria following the release of the IAP to better reflect the available literature inventory and the aims of the assessment. As described in Section 2.4.1, the available human evidence consists of observational studies among the general population using vanadium biomarkers to measure exposure and clinical trials of vanadium supplementation among healthy individuals or diabetics. The observational studies provide information about potential toxicity associated with vanadium measured in biological media including blood, urine, and hair, but the specific vanadium species or the route of exposure in the observational studies is not known. Although assessing the therapeutic effects of exposure to vanadium compounds is not the objective of this assessment, the clinical trials provide information

about biological effects in relation to an identified vanadium species, and the oral route of exposure is established. Most health systems evaluated in the clinical trials (e.g., metabolic, hematologic) also have data available from observational epidemiological studies (see Section 2.4.1, Figures 7 and 8). The clinical trials will therefore be considered as supplemental information but may be evaluated and used in the evidence synthesis when they can inform the evaluation of an endpoint that has evidence of adversity based on epidemiological or animal toxicological data. Additionally, animal toxicological studies evaluating vanadium salts as a therapeutic intervention (e.g., as an insulin enhancer in streptozotocin-induced diabetic animal models) or with acute exposure durations (i.e., <24 hours) will not be included in the PECO criteria but will be tracked as potentially relevant supplemental information.

Table 6. Populations, exposures, comparators, outcomes (PECO) criteria

PECO element	Evidence
Populations	**Human:** Any population and life stage (occupational or general population, including children, women of childbearing age, and other sensitive populations). **Animal:** Nonhuman mammalian animal species (whole organism) of any life stage (including preconception, in utero, lactation, peripubertal, and adult stages). Studies of transgenic animals will be tracked as mechanistic studies under "potentially relevant supplemental material."
Exposures	**Relevant forms:** Any forms of vanadium, other than vanadium coordination complexes with organic ligands developed for therapeutic research [e.g., bis(maltolato)oxyvanadium (VI)]. Those studies will be tracked as "potentially relevant supplemental information." **Human:** Any exposure to vanadium compound(s) via the oral route, including exposure via breastmilk. Studies will also be included if biomarkers of vanadium exposure are evaluated (e.g., measured vanadium levels in tissues or bodily fluids) but the exposure route is unclear. Clinical studies evaluating the therapeutic effects of vanadium supplementation will be tagged as "potentially relevant supplemental information" but will be included in the literature inventory and may be used in the evidence synthesis for endpoints that have evidence of adversity based on epidemiological and animal toxicological data. Other exposure routes, including inhalation, will be tagged as "potentially relevant supplemental information." **Animal:** Any exposure to vanadium compound(s) via the oral route, including exposure via breastmilk. Studies involving exposures to mixtures will be included only if they include an arm with exposure to vanadium compound(s) alone; otherwise, they will be tagged as potentially relevant supplemental material. Studies evaluating vanadium as a therapeutic intervention in animal models of disease (e.g., as an insulin enhancer in streptozotocin-induced diabetic animal models, or as a modulator of lipid metabolism in animals fed a high fat diet) will be tagged as potentially relevant supplemental material unless they also include normal control and vanadium treatment groups (e.g., wild-type animals, normal diet). Acute studies (<24 hours) will be included in the literature inventory as they can be helpful to interpret findings from studies more directly informative for developing a chronic toxicity value; however, these studies will be tagged as potentially relevant supplemental material and will not undergo study evaluation or full data extraction.

Table 6. Populations, exposures, comparators, outcomes (PECO) criteria (continued)

PECO element	Evidence
Comparators	**Human:** A comparison or referent population exposed to lower levels (or no exposure/exposure below detection limits), or exposure for shorter periods of time. Worker surveillance studies, however, are considered to meet PECO criteria even if no statistical analyses using a referent group are presented. Case reports or case series of more than 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." **Animal:** A concurrent control group exposed to vehicle-only treatment or untreated control.
Outcomes	All health outcomes (both cancer and noncancer). In general, endpoints related to clinical diagnostic criteria, disease outcomes, histopathological examination, or other apical/phenotypic outcomes are considered to meet PECO criteria and prioritized for evidence synthesis over outcomes such as biochemical measures.
PK/PBPK models	Studies describing pharmacokinetic (PK) or physiologically based pharmacokinetic (PBPK) models for any form of vanadium will be included. **Classical Pharmacokinetic (PK) or Dosimetry Model Studies:** Classical PK or dosimetry modeling usually divides the body into just one or two compartments, which are not specified by physiology, where movement of a chemical into, between, and out of the compartments is quantified empirically by fitting model parameters to ADME (absorption, distribution, metabolism, and excretion) data. This category is for papers that provide detailed descriptions of PK models, that are not a PBPK model. Note: ADME studies often report classical PK parameters, such as bioavailability (fraction of an oral dose absorbed), volume of distribution, clearance rate, or half-live(s). If a paper only provides such results in tables with minimal description of the underlying model or software (i.e., uses standard PK software without elaboration), including "noncompartmental analysis," it should be listed only as a supplemental material ADME study. **Physiologically based Pharmacokinetic (PBPK) 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.

On the basis of feedback received in public comments, the "relevant forms" of vanadium in the PECO criteria were broadened to include any forms of vanadium rather than focusing on inorganic forms. Vanadium coordination complexes with organic ligands that were developed for therapeutic research [e.g., bis(maltolato)oxyvanadium (VI)], however, will not be a primary focus of the assessment because the ligands are different from those that occur in the environment and may have different toxicokinetics and toxicity and because the evaluation of vanadium as a therapeutic intervention is not the primary focus of this assessment. These studies will be tracked as potentially relevant supplemental material as described Table 7.

Table 7. Major categories of "potentially relevant supplemental material"

Category	Evidence
Mechanistic studies	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 any route of exposure, includes transgenic models), ex vivo, and in silico studies. Genotoxicity tests are considered "mechanistic." Studies where the chemical is used as a laboratory reagent generally do not need to be tagged (e.g., as a chemical probe used to measure antibody response).
Nonmammalian model systems	Studies in nonmammalian model systems, e.g., fish, birds, *C. elegans*.
Nonoral route of administration	Studies in which humans or animals (whole organism) were exposed via a nonoral route (e.g., inhalation, injection, dermal exposure). This categorization generally does not apply to epidemiological studies where the exposure route may be unclear; such studies are considered to meet PECO criteria when oral exposure is plausible. Studies evaluating inhalation exposure to vanadium are under evaluation in a separate IRIS assessment.
ADME and toxicokinetic	Toxicokinetic (ADME) studies are primarily controlled experiments, where defined exposures usually occur by intravenous, oral, inhalation, or dermal routes, and the concentration of particles, a chemical, or its metabolites in blood or serum, other body tissues, or excreta are then measured. These data are used to estimate the amount absorbed (A), distributed (D), metabolized (M), and/or excreted/eliminated (E) through urine, breath, feces. • The most informative studies involve measurements over time such that the initial increase and subsequent concentration decline is observed, preferably at multiple exposure levels. However, data collected from multiple tissues or excreta at a single time-point also inform distribution. • ADME data can also be collected from human subjects who have had environmental or workplace exposures that are not quantified or fully defined. To be useful, however, such data must involve either repeated measurements over a period when exposure is known (e.g., is zero because previous exposure ended) ***or*** time- and subject-matched tissue or excreta concentrations (e.g., plasma and urine, or maternal and cord blood). • ADME data, especially metabolism and tissue partition coefficient information, can be generated using in vitro model systems. Although in vitro data may not be as definitive as in vivo data, these studies should also be tracked as ADME. For large evidence bases it may be appropriate to separately track the in vitro ADME studies. *Studies describing environmental fate and transport or metabolism in bacteria are not tagged as ADME.
Exposure characteristics (no health outcome assessment)	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).

Table 7. Major categories of "potentially relevant supplemental material" (continued)

Category	Evidence
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. Animal studies evaluating vanadium alloys and complexes of vanadium with other metals will also be included in this category. This categorization generally does not apply to epidemiological studies where source and form of vanadium exposure might be unclear; such studies are tracked as meeting PECO criteria when oral exposure is plausible.
Case reports	Case reports describing health outcomes after exposure will be tracked as potentially relevant supplemental information when the number of subjects is ≤3.
Records with no original data	Records that do not contain original data, such as other agency assessments, informative scientific literature reviews, editorials or commentaries.
Conference abstracts/abstract only	Records that do not contain sufficient documentation to support study evaluation and data extraction.
Acute animal studies	Animal studies with acute exposure durations (defined as less than 24 hours) that otherwise meet PECO criteria.
Vanadium coordination complexes with organic ligands developed for therapeutic research	Studies of vanadium coordination complexes with organic ligands developed for therapeutic research [e.g., bis(maltolato)oxyvanadium (VI)] that otherwise meet PECO criteria.
Clinical trials (human)	Clinical trials evaluating the therapeutic effects of vanadium supplementation.
Intervention studies (animal models)	Studies that evaluate vanadium as a therapeutic intervention in animal models of disease (e.g., streptozotocin-induced diabetic rodents) or in animals fed modified diets (e.g., high fat or high sucrose diet).

In addition to the PECO criteria, studies containing supplemental material that are potentially relevant to the specific aims will be tracked during the literature screening process. Table 7 presents major categories of supplemental material. The criteria are used to tag studies during screening and to prioritize studies for consideration in the assessment on the basis of likelihood to impact assessment conclusions. Studies may be tagged to one or more of these categories; and in some cases, studies that met PECO criteria were also tagged as containing supplemental material.

It is important to emphasize that being tagged as supplemental material does not mean the study is excluded from consideration in the assessment. The initial screening-level distinctions between a study meeting the PECO criteria and a supplemental study are often made for practical reasons, and the tagging structure in Table 7 is designed to ensure the supplemental studies are categorized for easy retrieval during the course of developing the assessment. Studies that meet the PECO criteria are those that are most likely to be used to derive toxicity values and will thus undergo individual-level study evaluation and data extraction, as described in the protocol. For

evidence-rich topics, this is most likely to be animal and epidemiological evidence. For most IRIS assessments, identifying all available pharmacokinetic models is also considered critical and thus those are generally included in the PECO criteria. In contrast, the impact on the assessment conclusions of individual studies tagged as supporting material is often difficult to assess during the screening phase of the assessment. Studies tagged as supplemental may (1) become critical to the interpretation of other evidence at individual-level study evaluation (e.g., genotoxicity studies when conducting a cancer MOA analysis is needed); (2) may be a single study that contributes to a well-accepted scientific conclusion and does not need to be evaluated and summarized at the individual study level [e.g., dioxin as an aromatic hydrocarbon receptor (AhR) agonist]; (3) provide key references for preparing certain sections in an IRIS assessment (e.g., background information on sources, production, or use; overview of toxicokinetics); or (4) provide context for the decision to conduct the assessment or for the assessment conclusions (e.g., information on pathways and levels of exposure).

4. LITERATURE SEARCH AND SCREENING STRATEGIES

4.1. USE OF EXISTING ASSESSMENTS

The ATSDR *Toxicological Profile for Vanadium* (ATSDR, 2012) was selected as the starting point for the literature search because it is the most recent review of health effects of vanadium and compounds published by a U.S. federal government agency that has undergone public comment and external peer review. All references from the 2012 ATSDR *Toxicological Profile for Vanadium* were extracted by an EPA information specialist and stored in the Health and Environmental Research Online (HERO) database (https://hero.epa.gov/hero/index.cfm/project/page/project_id/2357).[11]

4.2. LITERATURE SEARCH STRATEGIES

4.2.1. Database Searches

Database searches were conducted to identify records that had been published since the writing of the ATSDR *Toxicological Profile for Vanadium*. The databases listed below were searched for records published between 2010 and 2020. The start date was selected as 2010 as a precaution to capture records published near the last literature search date for the citations in the ATSDR document.[12]

- PubMed (National Library of Medicine)
- Web of Science (Thomson Reuters)
- Toxline (National Library of Medicine)[13]

The database searches focused only on the chemical name (and synonyms or trade names) with no additional limits. The search terms were based on previous vanadium review efforts by IRIS and were reviewed carefully to ensure that a wide array of vanadium compounds were

[11]EPA's HERO database provides access to the scientific literature behind EPA science assessments. The database includes more than 600,000 scientific references and data from the peer-reviewed literature used by EPA to develop its health assessment documents.
[12]Personal correspondence with ATSDR indicated that the final literature update for the 2012 *Toxicological Profile for Vanadium* was conducted in August 2011.
[13]The Toxline database was migrated to PubMed prior to the March 2020 literature search update, so the Toxline search was conducted only in March 2019.

encompassed. Because each database has its own search architecture, the resulting search strategy was tailored to account for each database's unique search functionality. The detailed search strategies are presented in Appendix B. Literature searches were conducted using EPA's HERO database,[14] with no language restrictions applied.

Because the number of records retrieved was large, records were imported into SWIFT Review software [https://www.sciome.com/swift-review/; see also Howard et al. (2016)] to identify those most likely applicable to human health. In brief, SWIFT Review has preset literature search filters developed by information specialists that can be applied to separate studies that may present a health outcome from those that likely do not (e.g., exposure only, analytical methods). The filters function like a typical search strategy, where studies are tagged as belonging to a certain category based on terms appearing in title, abstract, keyword, or medical subject headings (MeSH) fields The records identified in the literature search for vanadium were filtered using tags in SWIFT Review for lines of evidence (human, animal, in vitro). The details of the search strategies that underlie the filters are available at https://hawcprd.epa.gov/media/attachment/SWIFT-Review_Search_Strategies.pdf. Studies not retrieved using these filters were not considered further. Studies that included one or more of the search terms in the title, abstract, keyword, or MeSH fields were exported as a RIS file for screening in DistillerSR,[15] as described in Section 4.4.

The database searches will be updated throughout draft development to identify literature published during the course of review. The last full literature search update will be conducted less than 1 year before the planned release of the draft document for public comment. The results returned (i.e., the number of "hits" from each electronic database or other literature source), including the results of any literature search updates, are documented in the literature flow diagrams (see Section 4.4.2), which also reflect the literature screening decisions. The IRIS Program takes extra steps to ensure identification of pertinent studies by encouraging the scientific community and the public to identify additional studies and ongoing research and by considering late breaking studies that would impact the credibility of the conclusions, even during the review process.[16] Studies identified after peer review begins will be considered for inclusion only if they meet the PECO criteria and could fundamentally alter the assessment's primary conclusions.

4.2.2. Other Resources Consulted

The literature search strategies described above are designed to be broad, but like any search strategy, studies may be missed (e.g., studies published before 2010 that were not included in the ATSDR document; cases where the specific chemical is not mentioned in title, abstract, or keyword content; "gray" literature that is not indexed in the databases listed above). Thus, in

[14]Health and Environmental Research Online: https://hero.epa.gov/hero/.
[15]DistillerSR is a web-based systematic review software used to screen studies: https://www.evidencepartners.com/products/distillersr-systematic-review-software.
[16]IRIS "stopping rules": https://www.epa.gov/sites/production/files/2014-06/documents/iris_stoppingrules.pdf.

addition to the database searches, the approaches outlined below will be used to identify studies that may have been missed based on the literature search (see Appendix C for search methods). Records that appear to meet the PECO criteria will be uploaded into DistillerSR, annotated with respect to source of the record, and screened. Searching of these sources will be summarized to include the source type or name, the search string (when applicable), number of results present within the resource, and the URL (when available and applicable).

- Reference list of studies screened as meeting the PECO criteria after full-text review are reviewed at the title level.

- Reference lists of finalized IRIS and PPRTV assessments of vanadium and any published journal review articles specifically focused on human health.

- References from EPA's Toxicity Values database (ToxValDB), accessed via EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard/), to identify studies or assessments that present point of departure (POD) information. ToxValDB collates publicly available toxicity dose-effect related summary values typically used in risk assessments, many of which are from gray literature and are not available in databases such as Pub Med or Web of Science. These include POD data collected from data sources within EPA's ACToR (Aggregated Computational Toxicology Resource) and ToxRefDB (Toxicity Reference Database), and no-observed and lowest-observed (adverse) effect levels (NOEL, NOAEL, LOEL, LOAEL) data extracted from repeated dose toxicity studies submitted under European Union (EU) REACH regulation (Registration, Evaluation, Authorisation and Restriction of Chemicals). Also included are RfDs from EPA's IRIS and dose descriptors from EPA's PPRTV documents. Acute toxicity information is extracted from a number of different sources, including: OECD eChemPortal, ECHA (European Chemicals Agency), NLM (National Library of Medicine) HSDB (Hazardous Substances Data Bank), ChemIDplus via EPA TEST (Toxicity Estimation Software Tool), and the EU JRC (Joint Research Centre) AcutoxBase. Data from the EU COSMOS project (Integrated In Silico Models for the Prediction of Human Repeated Dose Toxicity of COSMetics to Optimise Safety) have also been included in ToxValDB. Although many of the resources included in the "Other Sources Consulted" list are represented in ToxValDB, they are also manually searched because most of the ToxValDB entries have not undergone quality control to ensure accuracy or completeness and might not include recent studies.

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

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

- National Toxicology Program (NTP) Chemical Effects in Biological Systems (CEBS) database of study results and research projects (https://manticore.niehs.nih.gov/cebssearch).

- The Organisation for Economic Cooperation and Development (OECD) Screening Information DataSet (SIDS) High Production Volume Chemicals https://www.echemportal.org/echemportal/substance-search

- EPA's ECOTOX database (https://cfpub.epa.gov/ecotox/index.cfm) for the chemical(s) of interest.

- ToxCast or Tox21 high throughput screening information accessed via EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard/). These data can be used to generate mechanistic insight, predict health outcomes using appropriate models, and potentially inform dose-response modeling. Their importance for outcome prediction and dose-response modeling depends on the context, size, and quality of retrieved results and the lack of availability of other data typically used for these purposes.

- References identified by technical consultants, during peer-review, and during public comment periods (when applicable).

4.3. INCLUSION OF NONPUBLIC AND NONPEER-REVIEWED DATA

IRIS assessments rely mainly on publicly accessible, peer-reviewed studies. However, it is possible that unpublished data directly relevant to the PECO criteria may be identified during assessment development. In these instances, EPA will try to get permission to make the data publicly available (e.g., in HERO); data that cannot be made publicly available are not used in IRIS assessments. In addition, on rare occasions, considering the type of report, EPA may obtain external peer review if the owners of the data are willing to have the study details and results made publicly accessible (U.S. EPA, 2015). This independent, contractor-driven, peer review would include an evaluation of the study similar to that for peer review of a journal publication. The contractor would identify and select two or three scientists knowledgeable in scientific disciplines relevant to the topic as potential peer reviewers. Persons invited to serve as peer reviewers would be screened for conflict of interest. In most instances, the peer review would be conducted by letter review. The study authors would be informed of the outcome of the peer review and given an opportunity to clarify issues or provide missing details. The study and its related information, if used in the IRIS assessment, would become publicly available. In the assessment, EPA would acknowledge that the document underwent external peer review managed by EPA, and the names of the peer reviewers would be identified. In certain cases, the IRIS Program will conduct an assessment for utility and data analysis based on having access to a description of study methods and raw data that have undergone rigorous quality assurance/quality control review (e.g., ToxCast/Tox21 data, results of NTP studies) but that have not yet undergone external peer review.

Unpublished data from personal author communication can supplement a peer-reviewed study provided the information is made publicly available. If such ancillary information is acquired,

it will be documented in the Health Assessment Workspace Collaborative (HAWC) or on the HERO project page (depending on the nature of the information received).

4.4. LITERATURE SCREENING STRATEGY

This screening strategy was used to identify the literature inventory described in Section 2.4 and will be used in subsequent literature search updates. The PECO criteria are used to determine inclusion or exclusion of a reference as a primary source of health effects data or a published PBPK model. In addition to the inclusion of studies that meet the PECO criteria, studies containing supplemental material that is potentially relevant to the specific aims are tracked during the screening process using the categories described in Section 3.2. Although not considered to directly meet PECO criteria, these studies are not strictly excluded unless otherwise specified. Unlike studies that meet PECO criteria, supplemental studies may not be subject to systematic review unless specifically defined questions are identified that focus the mechanistic (or other) analysis to inform the specific aims (see Section 3.1).

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

For citations with no abstract, articles are screened on the basis of all or some of the following: title relevance, page numbers (articles two pages in length or less may be assumed to be conference reports, editorials, or letters), and PubMed MeSH (Medical Subject Headings; e.g., a study might not be considered further if there are no human health- or biology-related MeSH terms). Screening conflicts are resolved by discussion among the primary screeners with consultation by a third reviewer or technical advisor (if needed) to resolve any remaining disagreements. Eligibility status of non-English studies is assessed using the same approach, although online translation tools may be used to assess eligibility at the title and abstract level.

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

The results of this screening process are posted on the project page for this assessment in the HERO database (https://hero.epa.gov/hero/index.cfm/project/page/project_id/2357) and "tagged" with appropriate category descriptors (e.g., studies meeting PECO criteria, potentially

relevant supplemental material, excluded). Results are annotated and reported in a literature flow diagram (see Section 4.4.2, Figure 13).

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

4.4.1. Multiple Publications of the Same Data

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

4.4.2. Literature Flow Diagram

A literature flow diagram summarizing literature inventory results is provided in Figure 13, below.

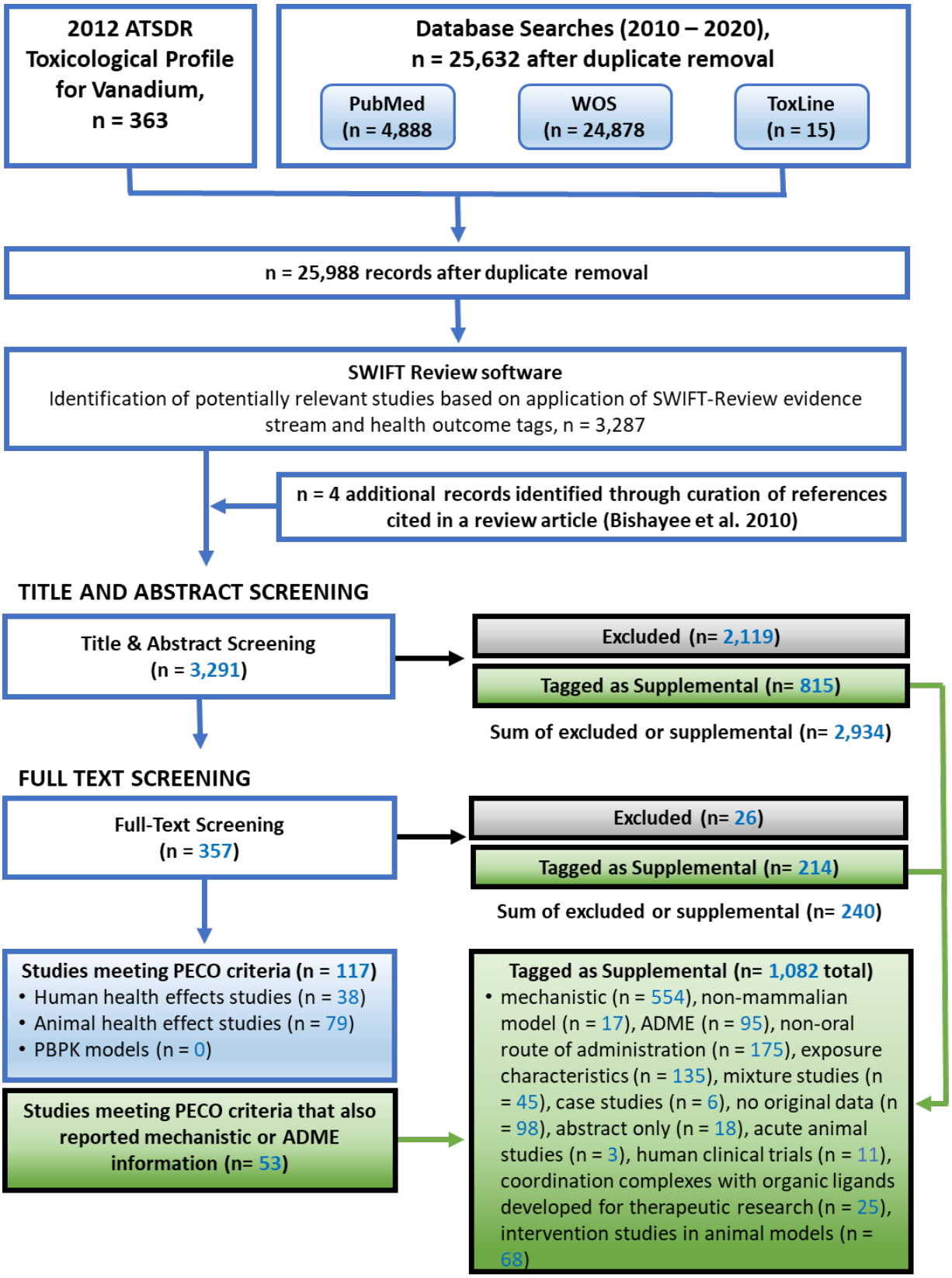


Figure 13. Literature search flow diagram for vanadium and compounds.

4.5. SUMMARY-LEVEL LITERATURE INVENTORIES

During full text screening, studies that met PECO criteria are briefly summarized using DistillerSR. For animal studies, the following information is captured: chemical form, study type [acute (<24 hours), short term (1–30 days), subchronic (30–90 days), chronic (>90 days), reproductive, developmental], duration and timing of treatment, route, species, strain, sex, dose or concentration levels tested, dose or concentration units, health system and specific endpoints assessed, and a brief summary of findings at the health system level [null, no-observed-effect level (NOEL), or lowest-observed-effect level (LOEL) based on author-reported statistical significance with an indication of which specific endpoints were affected]. For human studies, the following information is summarized: chemical form, population type (e.g., general population-adult, occupational, pregnant women, infants and children), study type (e.g., controlled trial, cross-sectional, cohort, case-control), short free text description of study population, sex, major route of exposure (if known), description of how exposure was assessed, health system and specific outcome assessed, and a summary of findings at the health system level based on author-reported statistical significance (null or an indication of any associations found and a description of how the exposure was quantified in the analysis). Studies are extracted into DistillerSR by one team member and checked by at least one other team member. These study summaries are referred to as literature inventories and are presented using Tableau visualization software (https://www.tableau.com/). These literature inventories facilitate subsequent review of individual studies or sets of studies by topic-specific experts.

Inventories may also be created for other categories of studies that were tagged as "potentially relevant supplemental material" during screening, including mechanistic studies (e.g., in vitro or in silico models), ADME studies, and other studies that do not meet the specific PECO criteria but that may still be relevant to the research question(s). Here, the objective is to create an inventory of studies that can be tracked and further summarized as needed—for example, by model system, key characteristic [e.g., of carcinogens; Smith et al. (2016)], mechanistic endpoint, or key event—to support analyses of critical mechanistic questions that arise at various stages of the systematic review (see Section 9.2 for a description of the process for determining the specific questions and pertinent mechanistic studies to be analyzed). ADME data and related information can be critical to the next steps of prioritizing or evaluating individual PECO-specific studies and will be reviewed by subject-matter experts early in the assessment process. Note that PBPK models are considered to meet PECO criteria while ADME and toxicokinetic-related studies are most commonly tracked as potentially relevant supplemental material. Any inventories of potentially relevant supplemental material created for this assessment will be visualized in HAWC and made publicly available.

5. REFINED EVALUATION PLAN

The purpose of the refined evaluation plan is to describe any refinements to the set of studies meeting the PECO criteria to be carried forward to study evaluation and help determine which studies tagged as "potentially relevant supplemental material" may need to be considered in the assessment. The information identified through screening and creation of the inventory (e.g., the types of exposure and outcome measures and analyses conducted in the studies) is used to guide selection of issues and considerations used to identify the studies that will move forward to study evaluation. For some assessments, additional refinement criteria may include prioritization beyond what is specified in the IAP such as focusing on specific exposure levels, routes of exposure, or toxic metabolites as identified by absorption, distribution, metabolism, and excretion (ADME) studies. The refined evaluation plan also serves as the basis for the selection or grouping of outcomes/endpoints for review from among a set of related measures.

The vanadium (oral) IRIS assessment will focus on those health outcomes for which adequate evidence exists to develop conclusions about potential hazard, based on the literature inventory (Section 2.4). Specifically, it is clear that in the absence of additional studies there will not be sufficient evidence to draw conclusions about gastrointestinal,[17] dermal,[18] musculoskeletal,[19] respiratory,[20] endocrine,[21] or effects tagged as "other."[22] Thus, unless more evidence becomes available, studies on these health outcomes will not undergo study evaluation or evidence synthesis to inform hazard characterization, and the information relating to those effects will be briefly summarized at the literature inventory level. Animal toxicological studies reporting effects tagged as "Systemic/Whole Body" (body weight, food/water consumption, mortality) that

[17]Gastrointestinal distress (e.g., diarrhea) was reported as a side effect in clinical trials of oral vanadium supplementation and in some animal studies. This side effect was reported qualitatively and has limited utility for risk assessment, so will not be a focus of the IRIS assessment.

[18]Two studies evaluated dermal effects (one epidemiological study evaluating androgenic alopecia and one animal study evaluating hair cystine content).

[19]Two animal toxicological studies evaluated musculoskeletal effects (histopathology or phosphatases in bone).

[20]Six animal toxicological studies evaluated respiratory effects (lung weight, histopathology, or collagen). Lung weight was decreased in a 14-day drinking water study conducted by NTP and was identified by the authors as a potentially novel effect of vanadium oral exposure. Other studies did not observe an effect on lung weight, although decreased lung collagen was reported in a multigenerational study in rats.

[21]The available data on endocrine effects consists of two studies in cattle and two epidemiological studies evaluating thyroid hormones or thyroid gland, two rat studies evaluating adrenal weight, and an evaluation of adrenal hormones in a human clinical trial. Although changes in thyroid hormone levels were reported, there was no clear pattern of effect, and it is unlikely that the existing data could be used to support a hazard conclusion.

[22]One epidemiological study evaluated periodontal disease.

do not evaluate any other health systems[23] also will not undergo study evaluation or evidence synthesis but can be considered to help interpret findings for other outcomes and will be summarized at the literature inventory level. Studies evaluating the other health systems identified in the literature inventory (listed below) will proceed to study evaluation:

- Cancer
- Cardiovascular
- Developmental
- Hematologic
- Hepatic
- Immune
- Metabolic
- Nervous
- Renal
- Reproductive (male and female)

[23]Five animal toxicological studies evaluated body weights, food/water consumption, or mortality but did not evaluate any other health systems.

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

The general approach for evaluating PECO-relevant primary health effect studies of all study types is described in Section 6.1. The specifics of applying the approach, however, differ; thus, they are described separately for epidemiological studies, animal toxicological studies, and human clinical trials[24] in Sections 6.2, 6.3, and 6.4, respectively. Different approaches are used for evaluation of PBPK models (see Section 6.5) and mechanistic studies (see Sections 6.6 and 9.2).

6.1. STUDY EVALUATION OVERVIEW FOR HEALTH EFFECT STUDIES

Key concerns for the review of epidemiological studies, animal toxicological studies, and human clinical trials are risk of bias, which is the assessment of internal validity (factors that affect the magnitude or direction of an effect in either direction) and insensitivity (factors that limit the ability of a study to detect a true effect; low sensitivity is a bias toward the null when an effect exists). Reporting quality is evaluated to determine the extent the available information allows for evaluating these concerns. The study evaluations are aimed at discerning the expected magnitude of any identified limitations (focusing on limitations that could substantively change a result) and considering the expected direction of the bias. Conflict of interest is not explicitly evaluated as the evaluations of risk of bias and sensitivity are designed to encompass the primary aspects of methodological design that could engender concern, irrespective of the sponsoring entity. The study evaluation considerations described below can be refined to address a range of study designs, health effects, and chemicals. The general approach for reaching an overall judgment for the study (or a specific analysis in a study) regarding confidence in the reliability of the results is illustrated in Figure 14.

[24]Human clinical trials will undergo study evaluation only if they are included in the evidence synthesis, as described in Section 3.2.

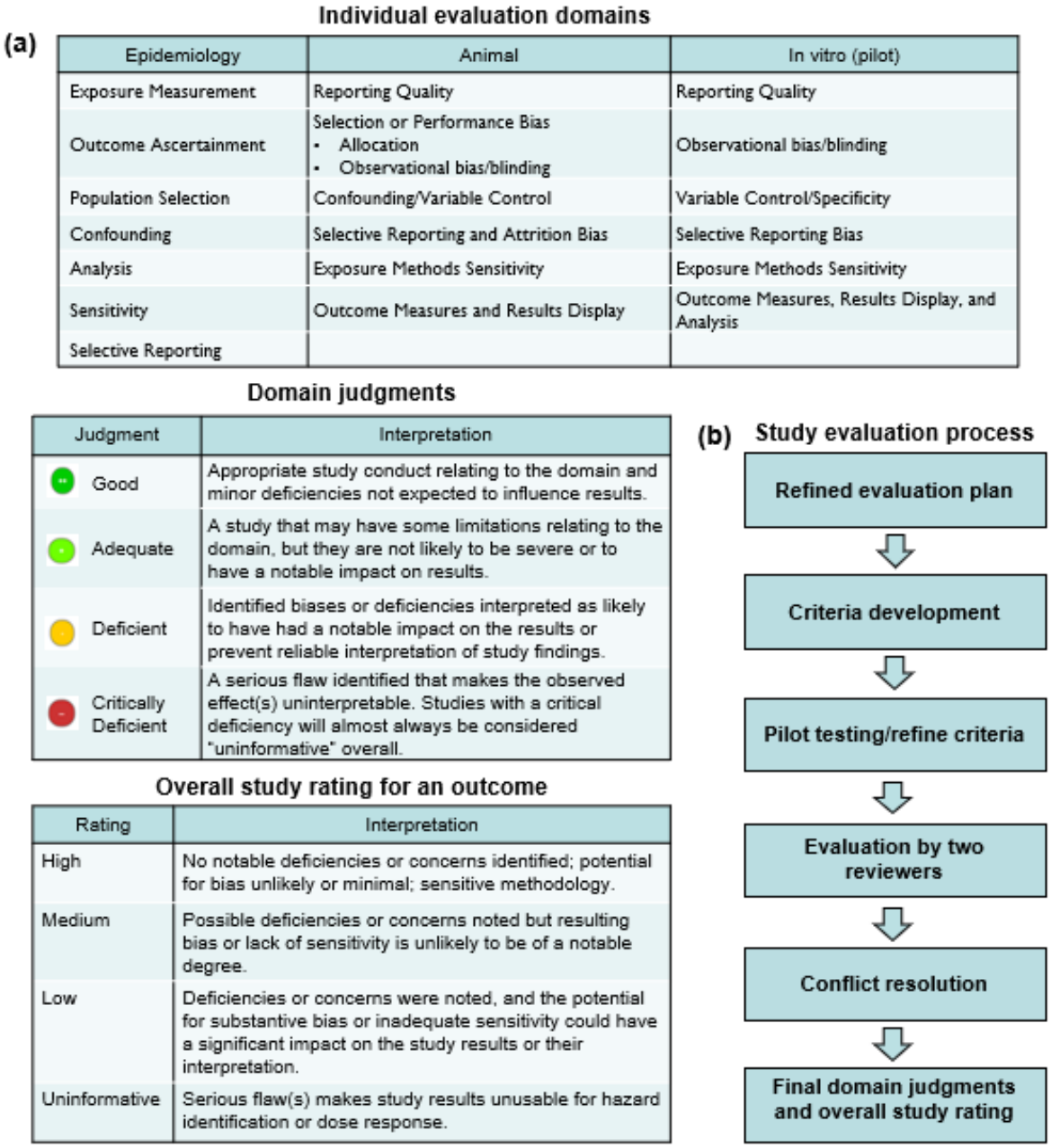


Figure 14. Overview of IRIS study evaluation process (a) An overview of the evaluation process. (b) The evaluation domains and definitions for ratings (i.e., domain and overall judgments, performed on an outcome-specific basis).

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

Considerations for evaluating studies will be developed in consultation with topic-specific technical experts and existing guidance documents when available, including EPA guidance for

carcinogenicity, neurotoxicity, reproductive toxicity, and developmental toxicity (U.S. EPA, 2005a, 1998, 1996, 1991). The independent evaluations include a pilot phase to assess and refine the evaluation process. During this phase, decisions will be compared and a consensus reached between reviewers, and when necessary, differences will be resolved by discussion between the reviewers, the chemical assessment team, or technical experts. As reviewers examine a group of studies, additional chemical-specific knowledge or methodological concerns may emerge, and a second pass may become necessary. Refinements to the study evaluation process made during the pilot phase and subsequent implementation will be acknowledged as updates to the protocol.

For studies that examine more than one outcome, the evaluation process will be performed separately for each outcome because the utility of a study can vary for different outcomes. If a study examines multiple endpoints for the same outcome,[25] evaluations may be performed at a more granular level if appropriate, but these measures may still be grouped for evidence synthesis.

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

For each outcome in a study,[26] reviewers will reach a consensus judgment of *Good*, *Adequate*, *Deficient*, *Not reported*, or *Critically deficient* for each evaluation domain. If a consensus is not reached, a third reviewer will perform conflict resolution. It is important to stress that these evaluations are performed in the context of the study's utility for identification of individual hazards. While limitations specific to the usability of the study for dose-response analysis are useful to note for later decisions, they do not contribute to the study confidence classifications. These categories are applied to each evaluation domain for each study as follows:

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

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

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

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

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

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

- *Critically deficient* reflects a judgment that the study conduct introduced a serious flaw that makes the study uninterpretable. Studies with a determination of critically deficient in an evaluation domain will almost always be considered overall "uninformative" and thus not used for hazard identification or dose-response analysis, but they may be used to highlight potential research gaps. Given this potential for exclusion, this classification is used infrequently and with extreme care; methodological limitations warranting this classification are defined a priori on an exposure- and outcome-specific basis and are inherently severe enough to warrant exclusion on the basis of a single critical deficiency. Serious flaws that do not warrant study exclusion will be classified as *Deficient*.

Once the evaluation domains have been rated, the identified strengths and limitations will be considered to reach a study confidence classification of *high*, *medium*, or *low* confidence, or *uninformative* for each specific health outcome. This classification is based on the reviewer judgments across the evaluation domains and includes consideration of the likely impact the noted deficiencies in bias and sensitivity, or inadequate reporting, have on the results. There are no pre-defined weights for the domains, and the reviewers are responsible for applying expert judgment to determine the impact of identified limitations on the overall study confidence classification for a given health outcome. The classifications, which reflect a consensus judgment between reviewers, are defined as follows:

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

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

- *Low* confidence: A substandard study where deficiencies or concerns are noted, and the potential for bias or inadequate sensitivity could have a significant impact on the study

results or their interpretation. Typically, *low* confidence studies have a *deficient* evaluation for one or more domains, although some *medium* confidence studies may have a *deficient* rating in domain(s) considered to have less influence on the magnitude or direction of effect estimates. Generally, *low* confidence results are given less weight compared to *high* or *medium* confidence results during evidence synthesis and integration (see Section 10.1, Tables 14 and 15) 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. Studies rated as *low* confidence only because of sensitivity concerns about bias toward the null would require additional consideration during evidence synthesis. Observing an effect in these studies may increase confidence, assuming the study is otherwise well conducted (see Section 9).

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

Study evaluation determinations reached by each reviewer and the consensus judgment between reviewers will be documented in EPA's version of Health Assessment Workspace Collaborative (HAWC), a free and open-source web-based software application.[27] Final study evaluations housed in HAWC, including the rationale supporting the individual domain and overall study evaluation determinations, will be made available when the draft is publicly released. The study confidence classifications and their rationales will be carried forward and considered as part of evidence synthesis (see Section 9), to aid in the interpretation of results across studies.

6.2. EPIDEMIOLOGY STUDY EVALUATION

Evaluation of epidemiological 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, while study sensitivity is typically concerned with identifying the latter.

The principles and framework used for evaluating epidemiological studies are adapted from the principles in the Cochrane Risk of Bias in Nonrandomized Studies of Interventions [ROBINS-I; (Sterne et al., 2016)], modified to address environmental and occupational exposures. Core and prompting questions, presented in Table 8, are used to collect information to guide evaluation of each domain. Core questions represent key concepts, whereas the prompting questions help the reviewer focus on relevant details under each key domain. Exposure- and outcome-specific criteria

[27]HAWC is a modular web-based interface to facilitate development of human health assessments of chemicals: https://hawcproject.org/portal/.

to use during evaluation of studies will be developed using the core and prompting questions and refined during a pilot phase with engagement from topic-specific experts. The types of information that may be the focus of those criteria are listed in Table 9. Epidemiological study evaluation considerations specific to vanadium are described in Section 6.2.1.

Table 8. Questions to guide the development of criteria for each domain in epidemiological studies

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

Table 8. Questions to guide the development of criteria for each domain in epidemiological studies (continued)

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
Outcome ascertainment Does the outcome measure reliably distinguish the presence or absence (or degree of severity) of the outcome?	For all: • Is outcome ascertainment likely to be affected by knowledge of, or presence of, exposure (e.g., consider access to health care, if based on self-reported history of diagnosis)? For case-control studies: • Is the comparison group without the outcome (e.g., controls in a case-control study) based on objective criteria with little or no likelihood of inclusion of people with the disease? For mortality measures: • How well does cause of death data reflect occurrence of the disease in an individual? How well do mortality data reflect incidence of the disease? For diagnosis of disease measures: • Is the diagnosis based on standard clinical criteria? If it is based on self-report of the diagnosis, what is the validity of this measure? For laboratory-based measures (e.g., hormone levels): • Is a standard assay used? Does the assay have an acceptable level of interassay variability? Is the sensitivity of the assay appropriate for the outcome measure in this study population?	Is there a concern that any outcome misclassification is nondifferential, differential, or both? What is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations require customization to the outcome **Good** • High certainty in the outcome definition (i.e., specificity and sensitivity), minimal concerns with respect to misclassification. • Assessment instrument was validated in a population comparable to the one from which the study group was selected. **Adequate** • Moderate confidence that outcome definition was specific and sensitive, some uncertainty with respect to misclassification but not expected to greatly change the effect estimate. • Assessment instrument was validated but not necessarily in a population comparable to the study group. **Deficient** • Outcome definition was not specific or sensitive. • Uncertainty regarding validity of assessment instrument. **Critically deficient** • Invalid/insensitive marker of outcome. • Outcome ascertainment is very likely to be affected by knowledge of, or presence of, exposure. Note: Lack of blinding should not be automatically construed to be critically deficient.
Participant selection Is there evidence that selection into or out of the study (or analysis	For longitudinal cohort: • Did participants volunteer for the cohort based on knowledge of exposure or preclinical disease symptoms? Was entry into the cohort or continuation in the cohort related to exposure and	Were differences in participant enrollment and follow-up evaluated to assess bias?	These considerations may require customization to the outcome. This could include determining what study designs effectively allow analyses of associations appropriate to the outcome measures (e.g., design to capture incident vs. prevalent cases, design to capture early pregnancy loss). **Good** • Minimal concern for selection bias based on description of recruitment process

Table 8. Questions to guide the development of criteria for each domain in epidemiological studies (continued)

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
sample) was jointly related to exposure and to outcome?	outcome? For occupational cohort: • Did entry into the cohort begin with the start of the exposure? • Was follow-up or outcome assessment incomplete, and if so, was follow-up related to both exposure and outcome status? • Could exposure produce symptoms that would result in a change in work assignment/work status ("healthy worker survivor effect")? For case-control study: • Were controls representative of population and periods from which cases were drawn? • Are hospital controls selected from a group whose reason for admission is independent of exposure? • Could recruitment strategies, eligibility criteria, or participation rates result in differential participation relating to both disease and exposure? For population-based survey: • Was recruitment based on advertisement to people with knowledge of exposure, outcome, and hypothesis?	If potential for bias is a concern, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)? Were appropriate analyses performed to address changing exposures over time in relation to symptoms? Is there a comparison of participants and nonparticipants to address whether differential selection is likely?	(e.g., selection of comparison population, population-based random sample selection, recruitment from sampling frame including current and previous employees). • Exclusion and inclusion criteria specified and would not induce bias. • Participation rate is reported at all steps of study (e.g., initial enrollment, follow-up, selection into analysis sample). If rate is not high, there is appropriate rationale for why it is unlikely to be related to exposure (e.g., comparison between participants and nonparticipants or other available information indicates differential selection is not likely). **Adequate** • Enough of a description of the recruitment process to be comfortable that there is no serious risk of bias. • Inclusion and exclusion criteria specified and would not induce bias. • Participation rate is incompletely reported but available information indicates participation is unlikely to be related to exposure. **Deficient** • Little information on recruitment process, selection strategy, sampling framework or participation OR aspects of these processes raise the potential for bias (e.g., healthy worker effect, survivor bias). **Critically deficient** • Aspects of the processes for recruitment, selection strategy, sampling framework, or participation result in concern that selection bias resulted in a large impact on effect estimates (e.g., convenience sample with no information about recruitment and selection, cases and controls are recruited from different sources with different likelihood of exposure, recruitment materials stated outcome of interest and potential participants are aware of or are concerned about specific exposures).
Confounding Is confounding of the effect of the exposure likely?	Is confounding adequately addressed by considerations in: • Participant selection (matching or restriction)?	If potential for bias is a concern, what is the predicted direction or	These considerations require customization to the exposure and outcome, but this may be limited to identifying key covariates. **Good** • Conveys strategy for identifying key confounders. This may include a priori

Table 8. Questions to guide the development of criteria for each domain in epidemiological studies (continued)

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

Table 8. Questions to guide the development of criteria for each domain in epidemiological studies (continued)

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
			the outcomes and exposure levels) are not presented. • Strategy of evaluating confounding is unclear or is not recommended [e.g., only based on statistical significance criteria or stepwise regression {forward or backward elimination)]. **Critically deficient** • Includes variables in the models that are colliders or intermediates in the causal pathway, indicating that substantial bias is likely from this adjustment; or Confounding is likely present and not accounted for, indicating that all of the results were most likely due to bias. Presenting a progression of model results with adjustments for different potential confounders, if warranted.
Analysis Does the analysis strategy and presentation convey the necessary familiarity with the data and assumptions?	• Are missing outcome, exposure, and covariate data recognized and, if necessary, accounted for in the analysis? • Does the analysis appropriately consider variable distributions and modeling assumptions? • Does the analysis appropriately consider subgroups of interest (e.g., based on variability in exposure level or duration or susceptibility)? • Is an appropriate analysis used for the study design? • Is effect modification considered, based on considerations developed a priori? Does the study include additional analyses addressing potential biases or limitations (i.e., sensitivity analyses)?	If potential for bias is a concern, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations may require customization to the outcome. This could include the optimal characterization of the outcome variable and ideal statistical test (e.g., Cox regression). **Good** • Use of an optimal characterization of the outcome variable. • Quantitative results presented (effect estimates and confidence limits or variability in estimates) (i.e., not presented only as a *p*-value or "significant"/"not significant"). • Descriptive information about outcome and exposure provided (where applicable). • Amount of missing data noted and addressed appropriately (discussion of selection issues—missing at random vs. differential). • Where applicable, for exposure, includes LOD (and percentage below the LOD), and decision to use log transformation. • Includes analyses that address robustness of findings, e.g., examination of exposure-response (explicit consideration of nonlinear possibilities, quadratic, spline, or threshold/ceiling effects included, when feasible); relevant sensitivity analyses; effect modification examined based only on a priori rationale with

Table 8. Questions to guide the development of criteria for each domain in epidemiological studies (continued)

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
			sufficient numbers. • No deficiencies in analysis evident. Discussion of some details may be absent (e.g., examination of outliers). **Adequate** Same as Good, except: • Descriptive information about exposure provided (where applicable) but may be incomplete; might not have discussed missing data, cutpoints, or shape of distribution. • Includes analyses that address robustness of findings (examples in Good), but some important analyses are not performed. **Deficient** • Does not conduct analysis using optimal characterization of the outcome variable. • Descriptive information about exposure levels not provided (where applicable). • Effect estimate and *p*-value presented, without standard error or confidence interval. • Results presented as statistically "significant"/"not significant." **Critically deficient** • Results of analyses of effect modification examined without clear a priori rationale and without providing main/principal effects (e.g., presentation only of statistically significant interactions that were not hypothesis driven). Analysis methods are not appropriate for design or data of the study.

Table 8. Questions to guide the development of criteria for each domain in epidemiological studies (continued)

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
Selective reporting Is there reason to be concerned about selective reporting?	• Were results provided for all the primary analyses described in the methods section? • Is there appropriate justification for restricting the amount and type of results that are shown? • Are only statistically significant results presented?	If potential for bias is a concern, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations generally do not require customization and may have fewer than four levels. **Good** • The results reported by study authors are consistent with the primary and secondary analyses described in a registered protocol or methods paper. **Adequate** • The authors described their primary (and secondary) analyses in the methods section and results were reported for all primary analyses. **Deficient** • Concerns were raised based on previous publications, a methods paper, or a registered protocol indicating that analyses were planned or conducted that were not reported, or that hypotheses originally considered to be secondary were represented as primary in the reviewed paper. • Only subgroup analyses were reported, suggesting that results for the entire group were omitted. • Only statistically significant results were reported.
Sensitivity Is there a concern that sensitivity of the study is not adequate to detect an effect?	• Is the exposure range adequate to detect associations and exposure-response relationships? • Was the appropriate population included? • Was the length of follow-up adequate? Is the time/age of outcome ascertainment optimal given the interval of exposure and the health outcome? • Are there other aspects related to risk of bias or otherwise that raise concerns about sensitivity?		These considerations may require customization to the exposure and outcome and may have fewer than four levels. Some study features that affect study sensitivity may have already been included in the other evaluation domains. Other features that have not been addressed should be included here. Some examples include: **Adequate** • The range of exposure levels provides adequate variability to evaluate the relevant associations. • The population was exposed to levels expected to have an impact on response. • The study population was sensitive to the development of the outcomes of interest (e.g., ages, life stage, sex). • The timing of outcome ascertainment was appropriate given expected latency

Table 8. Questions to guide the development of criteria for each domain in epidemiological studies (continued)

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
			for outcome development (i.e., adequate follow-up interval). • The study was adequately powered to observe an association based on underlying population sensitivity and exposure contrasts. • No other concerns raised regarding study sensitivity. **Deficient** • Concerns were raised about the issues described for *adequate* that are expected to notably decrease the sensitivity of the study to detect associations for the outcome.

Table 9. Information relevant to evaluation domains for epidemiological studies

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

6.2.1. Epidemiological Study Evaluation Considerations Specific to Vanadium

The criteria that will be used for evaluating vanadium exposure in epidemiological studies are summarized in Table 10. Biomarker measurements of vanadium from urine, blood, hair, or toenails will be considered relevant to either acute or long-term continuous exposure. Concentrations in hair or toenails may reflect exposures during the previous several months based on their rate of growth, although the relevant period has not been investigated for vanadium (Gutiérrez-González et al., 2019). Toenail vanadium was strongly correlated with vanadium in hair ($r = 0.61$) in a small study (Raińska et al., 2005). Validated reference values are available for hair, blood, plasma, and urine using inductively coupled mass spectrometry (ICP-MS) (Goulle et al., 2005). Quality control procedures include the use of certified reference material (urine or hair) or

nail reference material generated by individual laboratories, recovery analysis, procedural blanks, duplicate samples, or spike samples. Sample mass has been associated with concentrations measured in toenails; therefore, correction methods are necessary. Vanadium concentrations in toenails were found to be inversely associated with age and positively associated with alcohol consumption. Therefore, these factors may be confounders of associations for some outcomes.

Well-established and sensitive methods for measurement of vanadium concentrations include measurement using graphite furnace atomic absorption spectrometry (GF-AAS; with a preconcentration procedure), isotope dilution mass spectrometry (ID-MS), ICP-MS, and neutron activation analysis (NAA) with radiochemical separation. Detection limits of these methods have been summarized previously (ATSDR, 2012). Because toxic properties of vanadium species differ, measurements that report vanadium species are preferred to measurements of total vanadium. If only total vanadium were measured in the sample media, the exposure measurement domain would be rated deficient and the overall study confidence would be determined to be *low*.

Table 10. Criteria for evaluating exposure measurements in epidemiological studies of vanadium

Rating	Criteria
Good	Evidence that exposure was consistently assessed using well-established analytical methods. Well-established and sensitive methods include measurement of vanadium using GF-AAS (with a preconcentration procedure); isotope dilution mass spectrometry (ID-MS); ICP-MS; and NAA with radiochemical separation. And all of the following: • Exposure was assessed in a relevant time window (i.e., temporality is established and sufficient latency occurred prior to disease onset) for development of the outcome. • There is evidence that a sufficient number of the exposure data measurements are above the limit of quantification for the assay. • Details on quality control provided include measures to avoid contamination in sampling, sample handling and storage of blood and urine samples, and sample mass (minimum 10 mg with adjustment for mass) for toenails. QA statistics on precision and accuracy reported. • There is sufficient specificity/sensitivity and range or variation in exposure measurements that would minimize potential for exposure measurement error and misclassification by allowing exposure classifications to be differentiated (i.e., can reliably categorize participants into groups such as high vs. low exposure).

Table 10. Criteria for evaluating exposure measurements in epidemiological studies of vanadium (continued)

Rating	Criteria
Adequate	Evidence that exposure was consistently assessed using methods described in *Good*, but there were some concerns about quality control measures or other potential for nondifferential misclassification. And all of the following: • Exposure was assessed in a relevant time window for development of the outcome. • There is evidence that a sufficient number of the exposure data measurements are above the limit of quantification for the assay. • The laboratory analysis included some data on standard quality control measures with demonstrated precision and accuracy. • There is sufficient specificity/sensitivity and range or variation in exposure measurements that would minimize potential for exposure measurement error and misclassification by allowing exposure classifications to be differentiated (i.e., can reliably categorize participants into groups such as high vs. low exposure).
Deficient	Any of the following: • Only total vanadium in the sample media is reported. • There is a lack of detail on the analytical methods that reduces the ability to assess exposure misclassification. • There is some concern, but no direct evidence, that the exposure was assessed using methods that have not been validated or empirically shown to be consistent with methods that directly measure exposure. • Exposure was assessed in a relevant time window(s) for development of the outcome, but there could be some concern about the potential for bias due to reverse causation between exposure and outcome even though there is no direct evidence that it is present. • There is some concern over insufficient specificity/sensitivity and range or variation in exposure measurements that may result in considerable exposure measurement error and misclassification when exposure classifications are compared (i.e., data do not lend themselves to reliably categorize participants into groups such as high vs. low exposure, or there is considerable uncertainty in exposure values that do not allow for confidence in the examination of small per-unit changes in continuous exposures).

Table 10. Criteria for evaluating exposure measurements in epidemiological studies of vanadium (continued)

Rating	Criteria
Critically deficient	Any of the following: • Exposure was assessed in a time window that is unknown or not relevant for development of the outcome. This could be due to clear evidence of bias due to reverse causation between exposure and outcome, or other concerns such as the lack of temporal ordering of exposure and disease onset, insufficient latency, or having exposure measurements that are not reliable measures of exposure during the etiologic window(s). • Direct evidence that bias was likely since the exposure was assessed using methods with poor validity. • Evidence of differential exposure misclassification (e.g., differential recall of self-reported exposure). • There is evidence that an insufficient number of the exposure data measurements were above the limit of quantification for the assay.

6.3. EXPERIMENTAL ANIMAL STUDY EVALUATION

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

Table 11 provides the standard domains and core questions along with some basic considerations for guiding the evaluation. Some domain considerations will need to be tailored to the chemical and endpoint/outcome, while others are generalizable across assessments (e.g., considerations for reporting quality). Assessment teams work with subject matter experts to develop the assessment-specific considerations. These specific considerations are determined prior to performing study evaluation, although they may be refined as the study evaluation proceeds (e.g., during pilot testing). Animal toxicological study evaluation considerations specific to vanadium are described in Section 6.3.1.

Table 11. Questions to guide the development of criteria for each domain in experimental animal toxicological studies

Evaluation concern	Domain—core question	Prompting questions	General considerations
Reporting quality	**Reporting quality** Does the study report information for evaluating the design and conduct of the study for the endpoint(s)/outcome(s) of interest? *Notes:* *Reviewers will reach out to study authors for missing information on pH, characterization of vanadium speciation in drinking water, or both, and gavage studies since this information is considered important for exposure characterization. Authors will also be contacted for other missing information when studies are considered key for hazard evaluation, dose-response.* *This domain is limited to reporting. Other aspects of the exposure methods, experimental design, and endpoint evaluation methods are evaluated using the domains related to risk of bias and study sensitivity.*	Does the study report the following? **Critical information** necessary to perform study evaluation: • Species, test article name, levels and duration of exposure, route (e.g., oral, inhalation), qualitative or quantitative results for at least one endpoint of interest **Important information** for evaluating the study methods: • Test animal: strain, sex, source, and general husbandry procedures • Exposure methods: source, purity, method of administration, pH or speciation analysis of vanadium dosing solutions, or both (drinking water and gavage studies) • Experimental design: frequency of exposure, animal age and life stage during exposure and at endpoint/outcome evaluation • Endpoint evaluation methods: assays or procedures used to measure the endpoints/outcomes of interest	• These considerations typically do not need to be refined by assessment teams, although in some instances the **important information** may be refined depending on the endpoints/outcomes of interest or the chemical under investigation. • A judgment and rationale for this domain should be given for the study. Typically, these will not change regardless of the endpoints/outcomes investigated by the study. In the rationale, reviewers should indicate whether the study adhered to GLP, OECD, or other testing guidelines. o **Good:** All **critical and important information** is reported or inferable for the endpoints/outcomes of interest. o **Adequate:** All **critical information** is reported but some **important information** is missing. The missing information, however, is not expected to significantly impact the study evaluation. o **Deficient:** All **critical information** is reported but **important information** is missing that is expected to significantly reduce the ability to evaluate the study. o **Critically deficient:** Study report is missing any pieces of **critical information**. Studies that are Critically Deficient for reporting are Uninformative for the overall rating and not considered further for evidence synthesis and integration.

Table 11. Questions to guide the development of criteria for each domain in experimental animal toxicological studies (continued)

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

Table 11. Questions to guide the development of criteria for each domain in experimental animal toxicological studies (continued)

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

Table 11. Questions to guide the development of criteria for each domain in experimental animal toxicological studies (continued)

Evaluation concern		Domain—core question	Prompting questions	General considerations
Risk of Bias (continued)	**Confounding/variable control**	**Confounding** Are variables with the potential to confound or modify results controlled for and consistent across all experimental groups? *Note:* *See Section 6.3.1 for vanadium-specific considerations for this domain.*	For each study: • Are there differences across the treatment groups (e.g., coexposures, vehicle, diet, palatability, husbandry, health status) that could bias the results? • If differences are identified, to what extent are they expected to impact the results?	These considerations may need to be refined by assessment teams, as the specific variables of concern can vary by experiment or chemical. A judgment and rationale for this domain should be given for each cohort or experiment in the study, noting when the potential for confounding is restricted to specific endpoints/outcomes. • **Good:** Outside of the exposure of interest, variables that are likely to confound or modify results appear to be controlled for and consistent across experimental groups. • **Adequate:** Some concern that variables that were likely to confound or modify results were uncontrolled or inconsistent across groups but are expected to have a minimal impact on the results. • **Deficient:** Notable concern that potentially confounding variables were uncontrolled or inconsistent across groups and are expected to substantially impact the results. • **Critically deficient:** Confounding variables were presumed to be uncontrolled or inconsistent across groups and are expected to be a primary driver of the results.

Table 11. Questions to guide the development of criteria for each domain in experimental animal toxicological studies (continued)

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

Table 11. Questions to guide the development of criteria for each domain in experimental animal toxicological studies (continued)

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity	**Exposure methods sensitivity**	**Chemical administration and characterization** Did the study adequately characterize exposure to the chemical of interest and the exposure administration methods? *Notes:* *See Section 6.3.1 for vanadium-specific considerations for this domain.* *Consideration of the appropriateness of the route of exposure is not evaluated at the individual study level. Relevance and utility of the routes of exposure are considered in the PECO criteria for study inclusion and during evidence synthesis.*	For each study: • Does the study report the source and purity or composition (e.g., identity and percent distribution of different isomers) of the chemical? If not, can the purity or composition be obtained from the supplier (e.g., as reported on the website)? • Was independent analytical verification of the test article purity and composition performed? • Did the authors take steps to ensure the reported exposure levels were accurate? • Are there concerns about the methods used to administer the chemical (e.g., gavage volume)? • If necessary, based on consideration of chemical-specific knowledge (e.g., instability in solution, volatility) or exposure design (e.g., the frequency and duration of exposure), or both, were chemical concentrations in the dosing solutions or diet analytically confirmed?	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). A judgment and rationale for this domain should be given for each cohort or experiment in the study. • **Good:** Chemical administration and characterization is complete (i.e., source, purity, and analytical verification of the test article are provided). There are no concerns about the composition, stability, or purity of the administered chemical or the specific methods of administration. • **Adequate:** Some uncertainties in the chemical administration and characterization are identified but these are expected to have minimal impact on interpretation of the results (e.g., source and vendor-reported purity are presented, but not independently verified; purity of the test article is suboptimal but not concerning). • **Deficient:** Uncertainties in the exposure characterization are identified and expected to substantially impact the results (e.g., source of the test article is not reported; levels of impurities are substantial or concerning; deficient administration methods, such as a gavage volume considered too large for the species or life stage at exposure). • **Critically deficient:** Uncertainties in the exposure characterization are identified and there is reasonable certainty that the results are largely attributable to factors other than exposure to the chemical of interest (e.g., identified impurities are expected to be a primary driver of the results).

Table 11. Questions to guide the development of criteria for each domain in experimental animal toxicological studies (continued)

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity (continued)	**Exposure methods sensitivity (continued)**	**Exposure timing, frequency and duration** Was the timing, frequency, and duration of exposure sensitive for the endpoint(s)/outcome(s) of interest?	For each endpoint/outcome or grouping of endpoints/outcomes in a study: • Does the exposure period include the critical window of sensitivity? • Were the duration and frequency of exposure sensitive for detecting the endpoint of interest?	Considerations for this domain are highly variable depending on the endpoint(s)/outcome(s) of interest and must be refined by assessment teams. A judgment and rationale for this domain should be given for each endpoint/outcome or group of endpoints/outcomes investigated in the study. • **Good:** The duration and frequency of the exposure was sensitive, and the exposure included the critical window of sensitivity (if known). • **Adequate:** The duration and frequency of the exposure was sensitive, and the exposure covered most of the critical window of sensitivity (if known). • **Deficient:** The duration or frequency of the exposure, or both, are not sensitive and did not include the majority of the critical window of sensitivity (if known). These limitations are expected to bias the results toward the null. • **Critically deficient:** The exposure design was not sensitive and is expected to strongly bias the results toward the null. The rationale should indicate the specific concern(s).

Table 11. Questions to guide the development of criteria for each domain in experimental animal toxicological studies (continued)

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity (continued)	**Outcome measures and results display**	**Endpoint sensitivity and specificity** Are the procedures sensitive and specific for evaluating the endpoint(s)/outcome(s) of interest? *Note:* • Sample size alone is not a reason to conclude an individual study is critically deficient. • Considerations related to adjustments/ corrections to endpoint measurements (e.g., organ weight corrected for body weight) are addressed under results presentation.	For each endpoint/outcome or grouping of endpoints/outcomes in a study: • Are there concerns regarding the sensitivity, specificity, or validity of the protocols? • Are there serious concerns regarding the sample size? • Are there concerns regarding the timing of the endpoint assessment?	Considerations for this domain are highly variable depending on the endpoint(s)/outcome(s) of interest and must be refined by assessment teams. A judgment and rationale for this domain should be given for each endpoint/outcome or group of endpoints/outcomes investigated in the study. Examples of potential concerns include: • Selection of protocols that are insensitive or nonspecific for the endpoint of interest. • Evaluations did not include all treatment groups (e.g., only control and high dose). • Use of unreliable methods to assess the outcome. • Assessment of endpoints at inappropriate or insensitive ages, or without addressing known endpoint variation (e.g., due to circadian rhythms, estrous cyclicity). • Decreased specificity or sensitivity of the response due to the timing of endpoint evaluation, as compared to exposure (e.g., short-acting depressant or irritant effects of chemicals; insensitivity due to prolonged period of nonexposure prior to testing).

Table 11. Questions to guide the development of criteria for each domain in experimental animal toxicological studies (continued)

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity (continued)	**Outcome measures and results display (continued)**	**Results presentation** Are the results presented in a way that makes the data usable and transparent?	For each endpoint/outcome or grouping of endpoints/outcomes in a study: • Does the level of detail allow for an informed interpretation of the results? • Are the data analyzed, compared, or presented in a way that is inappropriate or misleading?	Considerations for this domain are highly variable depending on the outcomes of interest and must be refined by assessment teams. A judgment and rationale for this domain should be given for each endpoint/outcome or group of endpoints/outcomes investigated in the study. Examples of potential concerns include: • Nonpreferred presentation (e.g., developmental toxicity data averaged across pups in a treatment group, when litter responses are more appropriate; presentation of absolute organ-weight data when relative weights are more appropriate). • Failing to present quantitative results either in tables or figures. • Pooling data when responses are known or expected to differ substantially (e.g., across sexes or ages). • Failing to report on or address overt toxicity when exposure levels are known or expected to be highly toxic. • Lack of full presentation of the data (e.g., presentation of mean without variance data, concurrent control data are not presented).

Table 11. Questions to guide the development of criteria for each domain in experimental animal toxicological studies (continued)

Evaluation concern	Domain—core question	Prompting questions	General considerations
Overall confidence	**Overall confidence** Considering the identified strengths and limitations, what is the overall confidence rating for the endpoint(s)/outcome(s) of interest? *Note:* *Reviewers 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, the confidence may be increased.*	For each endpoint/outcome or grouping of endpoints/outcomes in a study: • Were concerns (i.e., limitations or uncertainties) related to the reporting quality, risk of bias, or sensitivity identified? • If yes, what is their expected impact on the overall interpretation of the reliability and validity of the study results, including (when possible) interpretations of impacts on the magnitude or direction of the reported effects?	The overall confidence rating considers the likely impact of the noted concerns (i.e., limitations or uncertainties) in reporting, bias, and sensitivity on the results. A confidence rating and rationale should be given for each endpoint/outcome or group of endpoints/outcomes investigated in the study. Confidence ratings are described above (see Section 6.1.1).

OECD = Organisation for Economic Cooperation and Development.

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

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

6.3.1. Animal Toxicology Study Evaluation Considerations Specific to Vanadium

Vanadium speciation chemistry in animal toxicological studies will be considered in the "Chemical administration and characterization" domain (Table 12). The highest confidence will be placed in studies that report the form of vanadium that was used and have analytical chemistry data indicating the vanadium species present in the exposure media (established analytical methods for vanadium are described in Section 6.2.1). For drinking water and gavage studies, it is also important that the pH of the dosing solutions is appropriate for ensuring the stability of the species being evaluated, as described previously in Section 2.12. In particular, V^{+4} compounds (e.g., vanadyl sulfate) must be prepared at pH 3–4 to ensure stability, because they will be readily oxidized to V^{+5} as the pH approaches neutral (Harrington et al., 2021; Mutlu et al., 2017), and V^{+5} compounds (e.g., sodium metavanadate) can convert to V^{+4} at low pH (Harrington et al., 2021).

If information is not provided on the form of vanadium used (e.g., sodium metavanadate, vanadyl sulfate) or on the chemistry of dosing solutions (i.e., pH or a speciation analysis), study authors will be contacted for this information and allowed 4 weeks to respond. Any information obtained through personal correspondence with the authors must be made public to be used in the assessment. If this information cannot be obtained, the study will be rated Deficient in the "Chemical administration and characterization" domain and *Low* confidence overall. There are fewer concerns about speciation chemistry in dietary exposure studies, but evaluations will consider whether the preparation of the diet may have affected the stability of the vanadium compounds.

Table 12. Vanadium-specific criteria for evaluating the "Chemical administration and characterization" domain in animal toxicological studies

Rating	Criteria
Good	Study reports the form of vanadium, and analytical chemistry data are provided indicating the species present in the exposure media. Efforts were made to ensure the stability of vanadium species in the exposure media (e.g., pH of dosing solutions is reported and is in the appropriate range to ensure the stability of the species being tested).
Adequate	Study reports the form of vanadium, and efforts were made to ensure the stability of vanadium species in the exposure media (e.g., pH of dosing is reported and is in the appropriate range for the species being tested), but there is no analytical confirmation of the species present in the exposure media.
Deficient	Study does not report the form of vanadium (e.g., reports exposure to "vanadium," "vanadate," or "vanadium salt" but does not specify which compound). Or study reports the form of vanadium but does not indicate that efforts were made to ensure the stability of vanadium species in the dosing media (e.g., pH of dosing solutions is not reported or is inappropriate), and there is no analytical confirmation of the species present in the exposure media.

The potential for decreased palatability will be considered in the "Confounding/variable control" domain for any studies that test higher dose levels. Dose-related decreases in palatability have been observed in NTP's drinking water studies at sodium metavanadate and vanadyl sulfate

concentrations greater than 250 mg/L (Roberts et al., 2016). The "Confounding/variable control" domain will also consider specific concerns about overt toxicity in drinking water and gavage studies that use sodium orthovanadate because this compound may produce a solution with a high pH. Mutlu et al. (2017) reported that the pH of sodium orthovanadate solutions in tap water increased from 9.4 to 11.5 with increasing concentrations from 125 to 2,000 mg/L. Oral dosing solutions above pH 9 are not recommended because they may cause side effects including diarrhea, vomiting, necrosis, and pain (Turner et al., 2011). Therefore studies testing sodium orthovanadate in this dose range will be considered to have the potential for confounding unless it is indicated that the pH was adjusted toward neutral.

6.4. HUMAN CLINICAL TRIAL STUDY EVALUATION

The evidence base relevant to the oral route of exposure to vanadium includes several clinical trials evaluating sodium metavanadate or vanadyl sulfate as interventions for diabetes-associated parameters (e.g., body mass index, cholesterol) or as a supplement in healthy individuals. As discussed in the PECO criteria (Section 3.2), clinical trials that may inform the hazard identification for endpoints evaluated in the epidemiological or animal toxicological evidence will be evaluated for risk of bias and sensitivity.

For evaluation of these studies, we will explore using the Cochrane risk of bias tool for randomized trials (RoB 2) with signaling and prompting questions and guidance tailored to the appropriate study design [i.e., individually randomized, parallel-group trials or randomized cross-over designs (Sterne et al., 2019)]. The tool includes five domains that address the types of bias that may affect the results of randomized trials. These domains are bias arising from the randomization process; bias due to deviations from intended interventions; bias due to missing outcome data; bias in measurement of the outcome; and bias in selection of the reported result. Answers to the signaling questions are inputs to an algorithm that results in domain-specific judgments and an overall RoB judgment for the trial. The possible overall judgments in the tool are low risk of bias, some concerns, and high risk of bias. Although an algorithm is used to arrive at an overall judgment, the RoB 2.0 guidance stresses that the evaluator should verify and change the judgments if they determine this is appropriate. In addition, a risk of bias determination should be focused on whether the identified issues within a domain lead to a "risk of material bias" that affect the reliability of the study conclusions. In addition to the ROB 2.0 domains, the sensitivity and selective reporting domains from the epidemiological study evaluation tool will be included in the evaluation.

For the purposes of the risk of bias evaluation for this assessment, these tool-based judgments are considered equivalent to the overall confidence-based judgments used in the IRIS study evaluation tool (Table 13).

Table 13. Cochrane RoB 2.0 tool-based judgments and the equivalent confidence-based judgments used in the IRIS study evaluation tool

Cochrane RoB 2.0	IRIS study evaluation confidence ratings
Low Risk of Bias	*High* Confidence
Some Concerns for Risk of Bias	*Medium* Confidence
High Risk of Bias	*Low* Confidence

6.5. PHYSIOLOGICALLY BASED PHARMACOKINETIC MODEL DESCRIPTIVE SUMMARY AND EVALUATION

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

No PBPK models for vanadium and compounds were identified in the survey of the literature. If the comprehensive literature search or updates to that initial search identify any PBPK models, they will be evaluated in accordance with the Quality Assurance Project Plan for PBPK models (U.S. EPA, 2020c).

6.6. MECHANISTIC STUDY EVALUATION

As described in Section 4.4, the initial literature screening identifies sets of other potentially informative studies, including mechanistic studies, as "potentially relevant supplemental information." Mechanistic information includes any experimental measurement related to a health outcome that informs the biological or chemical events associated with phenotypic effects; these measurements can improve understanding of the mechanisms involved in the biological effects following exposure to a chemical but are not generally considered by themselves adverse outcomes. Mechanistic data are reported in a diverse array of observational and experimental studies across species, model systems, and exposure paradigms, including in vitro, in vivo (by various routes of exposure), ex vivo, and in silico studies. Chapters 9 and 10 outline an approach for the consideration of information from mechanistic studies where the specific analytical approach is targeted to the assessment needs depending on the extent and nature of the human and animal evidence.

Individual study-level evaluations of mechanistic endpoints are not typically pursued. To undergo a full reporting quality, risk of bias, and sensitivity evaluation of every identified study that may report mechanistic information before the relevant toxicity pathways have been identified or

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

If a subset of individual mechanistic studies is identified for evaluation, the study evaluation considerations will differ depending on the type of endpoints, study designs, and model systems or populations evaluated. It should be noted that because the evaluation process is outcome specific, overall confidence classifications for human or animal studies that have already been determined will not automatically apply to mechanistic endpoints if reported in the same study; a separate evaluation of the mechanistic endpoints should be performed as the utility of a study may vary for the different outcomes reported. Developing specific considerations requires a familiarity with the studies to be evaluated and cannot be conducted in the absence of knowledge of the relevant study designs, measurements, and analytic issues. Knowledge of issues related to the hazards and the outcomes identified in the revised evaluation plan is also important for developing specific evaluation considerations. One challenge is that novel methodologies for studying mechanistic evidence are continuously being developed and implemented and often no "standard practices" exist.

The evaluation of mechanistic studies applies similar principles as those described above for the evaluation of epidemiological and experimental animal studies. Table 14 provides the standard domains and core questions for the evaluation of studies conducted in in vitro test systems, along with some basic considerations for guiding the evaluation. The evaluation process focuses on assessing aspects of the study design and conduct through three broad types of evaluations: reporting quality, risk of bias, and study sensitivity. Some domain considerations are tailored to the chemical and to the assay(s) or endpoint(s) being evaluated. Assessment teams work with subject matter experts to develop specific considerations. These specific considerations are determined prior to performing study evaluation, although they may be refined as the study evaluation proceeds (e.g., during pilot testing). Assessment- or assay-specific considerations are documented and made publicly available in the assessment.

Table 14. Pilot testing domains, questions, and general considerations to guide the evaluation of in vitro studies

Evaluation concern	Domain—core question	Prompting questions	General considerations
Reporting quality	**Reporting quality** Does the study report information for evaluating the design and conduct of the study for the assay(s) or endpoint(s) of interest? *Notes:* *Reviewers should reach out to authors to obtain missing information when studies are considered key for hazard evaluation or dose-response.* *This domain is limited to reporting. Other aspects of the exposure methods, experimental design, and endpoint evaluation methods are evaluated using the domains related to risk of bias and study sensitivity.*	Does the study report the following? **Critical information** necessary to perform study evaluation: • Cell/tissue type(s) or test system, test material/chemical name, description of vehicle, concentration and duration of treatments, qualitative or quantitative results from at least one endpoint investigated. **Important information** for evaluating the study methods: • Test system: cell/tissue source (and verification of cell type, if demonstrated to be prone to contamination); cell passage number, cell counts or density/confluence at treatment and analysis; incubation conditions (e.g., temperature, CO_2/O_2 concentration, humidity level); media composition (e.g., serum, antibiotics) and source; other measures taken to avoid contamination (e.g., mycoplasma testing). • Exposure and design: Purity and source of chemical and vehicle, method and timing of administration, timepoints of data collection. • Endpoint evaluation methods: description of the endpoints measured and test assays used (sample size and replicates are considered under "outcome evaluation," paralleling what is done for in vivo studies).	These considerations typically do not need to be refined by assessment teams, although in some instances the **important information** may be refined depending on the assay or endpoints of interest or the chemical under investigation. A judgment and rationale for this domain should be given for the study. Typically, these will not change regardless of the assays used or endpoints investigated by the study. In the rationale, reviewers should indicate whether the study adhered to GLP, OECD, or other testing guidelines. • **Good**: All **critical and important information** is reported or inferable for the assay or endpoints of interest. • **Adequate**: All **critical information** is reported but some **important information** is missing. The missing information, however, is not expected to significantly impact the study evaluation. • **Deficient**: All **critical information** is reported but **important information** is missing that is expected to significantly reduce the ability to evaluate the study. • **Critically deficient**: Study report is missing any pieces of **critical information**. Studies that are *critically deficient* for reporting are *uninformative* for the overall rating and not considered further for evidence synthesis and integration.

Evaluation concern		Domain—core question	Prompting questions	General considerations

Table 14. Pilot testing domains, questions, and general considerations to guide the evaluation of in vitro studies (continued)

Evaluation concern		Domain—core question	Prompting questions	General considerations
Risk of bias	**Observational bias/blinding**	**Observational bias/blinding** Did the study implement measures to reduce observational bias? Considerations will vary depending on the specific assay/model system being used.	For each assay or endpoint or grouping of endpoints in a study: • Did the study take steps to minimize observational bias during analysis (e.g., blinding/coding of slides or plates for analysis, collection of data from randomly selected fields)? • If not, did the study use a design or approach for which such procedures can be inferred? • Were the assays evaluated using automated approaches (e.g., microplate readers) that reduce concern for observational bias? • What is the expected impact of failure to implement (or report implementation) of these methods/procedures on results?	These considerations typically do not need to be refined by the assessment teams. (Note that it can be useful for teams to identify highly subjective measures of endpoints/outcomes where observational bias may strongly influence results prior to performing evaluations.) A judgment and rationale for this domain should be given for each assay or endpoint or group of endpoints investigated in the study. • **Good:** Measures to reduce observational bias were described (e.g., specific mention of blinding or coding of slides for analysis) OR observational bias not a concern because of use of automated/computer-driven systems or standard laboratory kits. • **Adequate:** Measures for reducing observational bias (e.g., blinding) can be inferred or were reported but described incompletely; impact on results is expected to be minor. • **Not reported:** Measures to reduce observational bias were not described. o **(Interpreted as *adequate*)** The potential concern for bias was mitigated because protocol cited includes

Evaluation concern		Domain—core question	Prompting questions	General considerations
				description of requirements for blinding/coding or impact on results is expected to be minor. o **(Interpreted as *deficient*)** No protocol cited; the potential impact on the results is major (e.g., endpoint measures are highly subjective). • **Critically deficient**: Strong evidence for observational bias that could have impacted the results.
Risk of bias (continued)	**Variable control and specificity**	**Variable Control** Are all introduced variables with the potential to affect the results of interest controlled for and consistent across experimental groups?	For each study: • Are there concerns regarding the negative (untreated and/or vehicle) controls used? Were negative controls run concurrently? If known, do the results in the negative control groups differ significantly from expected background or historical incidence for the assay(s) of interest? • If applicable, was the assay signal normalized to account for nonbiological differences across replicates and exposure groups? • Are there any known or presumed differences across treatment groups (e.g., coexposures, culture conditions, variations in reagent production lots) that could bias the results? If differences are identified, to what extent are they expected to impact the results?	These considerations may need to be refined by assessment teams, as the specific variables of concern can vary by experiment or chemical. A judgment and rationale for this domain should be given for each experiment in the study, noting when the potential for confounding is restricted to specific assays or endpoints. • **Good**: Outside of the exposure of interest, variables that are likely to impact results appear to be controlled for and consistent across experimental groups. • **Adequate:** Some concern that variables that were likely to impact results were uncontrolled or inconsistent across groups but are expected to have a minimal impact on the results. • **Deficient:** Notable concern that important study variables were uncontrolled or inconsistent across groups and are expected to substantially impact the results. • **Critically deficient:** Influential study variables were presumed to be uncontrolled or inconsistent across groups and are expected to be a primary driver of the results.

Table 14. Pilot testing domains, questions, and general considerations to guide the evaluation of in vitro studies (continued)

Evaluation concern		Domain—core question	Prompting questions	General considerations
Risk of bias (continued)	Variable control and specificity (continued)	**Specificity** Did the study address features inherent to the test system or experiment or physicochemical properties of the test substance(s) that have the potential to affect the results for the endpoint(s) of interest independent of the effect of the test chemical on those endpoint(s)?	For each study: • Did the test compound induce cytotoxicity (or were the levels used sufficient to induce cytotoxicity in related systems) to a degree that is expected to affect interpretation of results? • Are there concerns regarding the need for positive controls (e.g., concerns that the effects of interest may be inhibited or otherwise not manifest in the test system)? If one was used, was the selected positive test substance appropriate and was the intended positive response induced? If known, do the results in the positive control groups differ significantly from expected background or historical incidence? • Can the test article interfere with a given assay (e.g., auto-fluoresces or inhibits enzymatic processes necessary for assay signals)?	These considerations may need to be refined by assessment teams, as the specific variables of concern can vary by experiment or chemical. A judgment and rationale for this domain should be given for each experiment in the study, noting when the potential to affect results is restricted to specific assays or endpoints. • **Good:** Outside of the exposure of interest, features of the test system or chemical properties that are likely to modify or interfere with test results appear to be controlled for and consistent across experimental groups. • **Adequate:** Some concern that features of the test system or chemical properties likely to modify or interfere with results were uncontrolled or inconsistent across groups but are expected to have a minimal impact on the results. • **Deficient:** Notable concern that features of the test system or chemical properties were uncontrolled or inconsistent across groups and are expected to substantially impact the results. • **Critically deficient:** Features of the test system or chemical properties were not accounted for and presumed to be uncontrolled or inconsistent across groups and are expected to be a primary driver of results.

Table 14. Pilot testing domains, questions, and general considerations to guide the evaluation of in vitro studies (continued)

Evaluation concern		Domain—core question	Prompting questions	General considerations
Risk of bias (continued)	**Selective reporting**	**Selective Reporting** Did the study present results, quantitatively or qualitatively, for all prespecified assays or endpoints and replicates described in the methods?	For each study: • 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)? (Note that this domain does not consider the appropriateness of the analysis or results presentation.)	These considerations typically do not need to be refined by assessment teams. A judgment and rationale for this domain should be given for each assay or endpoint in the study. • **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 are available from supplemental material. If results omissions are identified, the authors provide an explanation and these are not expected to impact the interpretation of the results. • **Adequate:** Quantitative or qualitative results are reported for most prespecified assays or endpoints (explicitly stated or inferred), exposure groups, and evaluation timepoints. Omissions are not explained but are not expected to significantly impact the interpretation of the results. • **Deficient:** Quantitative or qualitative results are missing for many prespecified assays or endpoints (explicitly stated or inferred), exposure groups, and evaluation timepoints; omissions are not explained and may significantly impact the interpretation of the results. • **Critically Deficient:** Extensive results omissions are identified, preventing comparisons of results across treatment groups.

Table 14. Pilot testing domains, questions, and general considerations to guide the evaluation of in vitro studies (continued)

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity	**Exposure methods**	**Chemical administration and characterization** Did the study adequately characterize exposure to the chemical of interest and the exposure administration methods?	For each study: • Are there concerns regarding the purity or composition (e.g., identity and percent distribution of different isomers) of the test material/chemical? If so, can the purity or composition be obtained from the supplier (e.g., as reported on the website)? • Was independent analytical verification of the test article purity and composition performed? If not, is this a significant concern for this substance? • Are there concerns about the stability of the test chemical in the vehicle or culture media (e.g., pH, solubility, volatility, adhesion to plastics) that were not corrected for (e.g., observed precipitate formation, enclosed chambers not used for testing volatile chemicals)? • Are there concerns about the preparation or storage conditions of the test substance? • Are there concerns about the methods used to administer the chemical?	It is essential that these criteria are considered, and potentially refined, by assessment teams, as the specific variables of concern can vary by chemical (e.g., stability may be an issue for one chemical but not another). A judgment and rationale for this domain should be given for each experiment in the study. • **Good:** Chemical administration and characterization are 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. • **Adequate:** Some uncertainties in the chemical administration and characterization are identified but these are expected to have minimal impact on interpretation of the results (e.g., source and vendor-reported purity are presented, but not independently verified; purity of the test article is suboptimal but not concerning. • **Deficient:** Uncertainties in the exposure characterization are identified and expected to substantially impact the results (e.g., source of the test article is not reported, levels of impurities are substantial or concerning, deficient administration methods were used). • **Critically deficient:** Uncertainties in the exposure characterization are identified and there is reasonable certainty that the results are largely attributable to factors other than exposure to the chemical of interest (e.g., identified impurities are expected to be a primary driver of the results).

Table 14. Pilot testing domains, questions, and general considerations to guide the evaluation of in vitro studies (continued)

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity (continued)	Exposure methods (continued)	**Exposure timing, frequency, and duration** Were the timing, frequency, and duration of exposure sensitive for the assay/model system of interest?	Considerations will vary depending on the specific assay/model system used, but may include the following for each assay or endpoint or grouping of endpoints in a study: • Were steps taken to determine the appropriate concentration range of the test article in the test system? Are there concerns that the amount of test article administered may not have reached a sufficient concentration to induce an effect? • Was the exposure duration sufficient to cause a measurable impact on the endpoint of interest (in the absence of a positive control)? • Was the doubling time considered in the frequency of dosing, timing of culture, or duration in culture at treatment? • Was the confluency at treatment appropriate? Are there concerns that the cells were quiescent/senescent or growth inhibited due to confluence?	Considerations for this domain are highly variable depending on the assay/model system of interest and must be refined by assessment teams. A judgment and rationale for this domain should be given for each assay or endpoint or group of endpoints investigated in the study. • **Good:** The duration and frequency of the exposure were sensitive, and the exposure concentration(s) were sufficient. • **Adequate:** The duration and frequency of the exposure were sensitive, and the exposure concentration(s) are presumed to have been sufficient. • **Deficient:** The duration or frequency of the exposure was not sensitive and did not include appropriate exposure concentrations. These limitations are likely to bias the results toward the null. • **Critically deficient:** The exposure design was not sensitive and is expected to strongly bias the results toward the null. The rationale should indicate the specific concern(s).

Table 14. Pilot testing domains, questions, and general considerations to guide the evaluation of in vitro studies (continued)

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity (continued)	Outcome measures, results display, and analysis	**Endpoint sensitivity** Are the procedures sensitive and specific for evaluating the endpoint(s) of interest?	For each endpoint or grouping of endpoints in a study: • Was the endpoint assessment methodology consistent with accepted guidelines or established criteria for the assay(s)/endpoint measures used in the study? • 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). • Is the cell/tissue type selected for the study appropriate and sensitive (e.g., is it routinely used) for measuring the endpoints of interest for the target organ system of interest? Are there known variations in cellular signaling unique to the model system that could influence the possibility of detecting the effect(s) of interest? • Are there concerns about the number of replicates or sample size in the study? • Are there concerns regarding the timing of the endpoint assessment?	Considerations for this domain are highly variable depending on the assay or endpoint(s) of interest and must be refined by assessment teams. A judgment and rationale for this domain should be given for each assay or endpoint or group of endpoints investigated in the study. Examples of potential concerns include: • Selection of protocols that are insensitive or nonspecific for the endpoint of interest. • Use of unreliable methods to assess the outcome. • Assessment of endpoints in insensitive cells or tissues.

Table 14. Pilot testing domains, questions, and general considerations to guide the evaluation of in vitro studies (continued)

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity (continued)	Outcome measures, results display, and analysis (cont.)	**Results presentation and analysis** Are the results presented in a way that makes the data usable and transparent?	For each assay/endpoint or grouping of endpoints in a study: • Does the level of detail allow for an informed interpretation of the results? • Are the data analyzed, compared, or presented in a way that is inappropriate or misleading? Flag potentially inappropriate statistical comparisons for further review.	Considerations for this domain are highly variable, depending on the endpoints of interest, and must be refined by assessment teams. A judgment and rationale for this domain should be given for each assay or endpoint or group of endpoints investigated in the study. Examples of potential concerns include: • Nonpreferred presentation (e.g., only presenting data normalized to controls). • Failing to present quantitative results. • Pooling data when responses are known or expected to differ substantially (e.g., across sexes or ages). • Averaging technical replicates rather than independent replicates. • Lack of full presentation of the data (e.g., presentation of mean without variance data; concurrent control data are not presented).

Table 14. Pilot testing domains, questions, and general considerations to guide the evaluation of in vitro studies (continued)

Evaluation concern	Domain—core question	Prompting questions	General considerations
Overall confidence	**Overall confidence** Considering the identified strengths and limitations, what is the overall confidence rating for the assay(s) or endpoint(s) of interest? *Note:* *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: • Were concerns (i.e., limitations or uncertainties) related to the reporting quality, risk of bias, or sensitivity identified? • If yes, what is their expected impact on the overall interpretation of the reliability and validity of the study results, including (when possible) interpretations of impacts on the magnitude or direction of the reported effects?	The overall confidence rating considers the likely impact of the noted concerns (i.e., limitations or uncertainties) in reporting, bias, and sensitivity on the results. A confidence rating and rationale should be given for each assay or endpoint or group of endpoints investigated in the study. Confidence rating definitions are described above (see Section 6.1).

1

7. ORGANIZING THE HAZARD REVIEW

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

Table 15 lists some questions that may be asked of the evidence to assist with this decision. These questions extend from considerations and decisions made during development of the refined evaluation plan to include review of the concerns raised during individual study evaluations, as well as the direction and magnitude of the study-specific results. Resolution of these questions will then inform critical decisions about the organization of the hazard evaluation and what studies may be useful in dose-response analyses.

Table 15. Querying the evidence to organize syntheses for human and animal evidence

Evidence	Questions	Follow-up questions
ADME	Are absorption, distribution, metabolism, or excretion different by route of exposures studied, life stage when exposure occurred, or dosing regimens used?	Will separate analyses be needed by route of exposure or by methods of dosing within a route of exposure (e.g., are large differences expected between gavage and dietary exposures)? Which life stages and what dosing regimens are more relevant to human exposure scenarios?
	Is there toxicity information for metabolites that also should be evaluated for hazard?	What exposures will be included in the evaluation?
	Is the parent chemical or metabolite also produced endogenously?	
Outcomes	What outcomes are reported in studies? Are the data reported in a comparable manner across studies (similar output metrics at similar levels of specificity, such as adenomas and carcinomas quantified separately)?	At what level (hazard, grouped outcomes, or individual outcomes) will the synthesis be conducted? What commonalities will the outcomes be grouped by: • health effect, • exposure levels, • functional or population-level consequences (e.g., endpoints all ultimately leading to decreased fertility or impaired cognitive function), or • involvement of related biological pathways? How well do the assessed human and animal outcomes relate within a level of grouping?
	Are there interrelated outcomes? If so, consider whether some outcomes are more useful or of greater concern than others.	
	Does the evidence indicate greater sensitivity to effects (at lower exposure levels or severity) in certain subgroups (by age, sex, ethnicity, life stage)? Should the hazard evaluation include a subgroup analysis?	
	Does incidence or severity of an outcome increase with duration of exposure or a particular window of exposure? What exposure time frames are relevant to development or progression of the outcome?	
	Is there mechanistic evidence that informs any of the outcomes and how might they be grouped together?	
	How robust is the evidence for specific outcomes? • What outcomes are reported by both human and animal studies and by one or the other? Were different animal species and sexes (or other important population-level differences) tested? • In general, what are the study confidence conclusions of the studies (*high*, *medium*, *low*, not informative) for the different outcomes? Is there enough evidence from *high* and *medium* confidence studies for particular outcomes to draw conclusions about causality?	What outcomes should be highlighted? Should the others be synthesized at all? Would comparisons by specific limitations be informative?

Table 15. Querying the evidence to organize syntheses for human and animal evidence (continued)

Evidence	Questions	Follow-up questions
Dose-response	Did some outcomes include better coverage of exposure ranges that may be most relevant to human exposure than others?	What outcomes and study characteristics are informative for development of toxicity values?
	Does the study have multiple dose levels for which you can evaluate a dose-response gradient? Are there outcomes with study results of sufficient similarity (e.g., an established linkage in a biological pathway) to allow examination or calculation of common measures of effect across studies? Do the mechanistic data identify surrogate or precursor outcomes that are sufficient for dose-response analysis?	
	Are there subgroups that exhibit responses at lower exposure levels than others?	
	Are there findings from ADME studies that could inform data-derived extrapolation factors, link toxicity observed via different routes of exposure, or link effects between humans and experimental animals?	Is there a common internal dose metric that can be used to compare species or routes of exposure?

8. DATA EXTRACTION OF STUDY METHODS AND RESULTS

Data extraction and content management will be carried out using HAWC. Data extraction elements that may be collected from epidemiological and animal toxicological studies are available at the following URL: https://hawcprd.epa.gov/media/attachment/Data_extraction_fields_for_epidemiology_and_animal_toxicological_studies.docx. The content of the data extraction may be revised following the identification of the studies included in the review as part of a pilot phase to assess the data extraction workflow. Not all studies that meet the PECO criteria go through data extraction. Studies evaluated as being *uninformative* are not considered further and would, therefore, not undergo data extraction. In addition, outcomes that are determined to be less relevant during refinement of PECO criteria may not go through data extraction or may have only minimal data extraction. The same may be true for *low* confidence studies if sufficient *medium* and *high* confidence studies are available. All findings are considered for extraction, regardless of statistical significance, although the level of extraction for specific outcomes within a study may differ (i.e., ranging from a narrative to full extraction of dose-response effect size information). Similarly, decisions about data extraction for *low* confidence studies are typically made during implementation of the protocol based on consideration of the quality and extent of the available evidence. The version of the protocol released with the draft assessment will outline how *low* confidence studies were treated for extraction and evidence synthesis.

The data extraction results for included studies will be presented in the assessment and made available for download from HAWC in Excel format when the draft is publicly released. [NOTE: The following browsers are fully supported for accessing HAWC: Google Chrome (preferred), Mozilla Firefox, and Apple Safari. There are errors in functionality when viewed with Internet Explorer.] Data extraction will be performed by one member of the evaluation team and checked by one or two other members. Discrepancies in data extraction will be resolved by consultation with a third member of the evaluation team. Once the data have been verified, they will be "locked" to prevent accidental changes. Digital rulers, such as WebPlotDigitizer (https://automeris.io/WebPlotDigitizer), are used to extract numerical information from figures. Use of digital rulers is documented during extraction.

As previously described, routine attempts will be made to obtain information missing from human and animal health effect studies, if it is considered influential during study evaluations (see Section 6) or when it can provide information required to conduct a meta-analysis (e.g., missing group size or variance descriptors such as standard deviation or confidence interval). Missing data

from individual mechanistic (e.g., in vitro) studies will generally not be sought. Outreach to study authors will be documented and considered unsuccessful if researchers do not respond to email or phone requests within 1 month of the attempt to contact.

8.1. STANDARDIZING REPORTING OF EFFECT SIZES

In addition to providing quantitative outcomes in their original units for all study groups, results from outcome measures will be transformed, when possible, to a common metric to help compare distinct but related outcomes that are measured with different scales. These standardized effect size estimates facilitate systematic evaluation and evidence integration for hazard identification, whether meta-analysis is feasible for an assessment (see Section 9.1). Based on metrics across the available studies, a common metric may be used and the calculation will be presented in the assessment.

For epidemiological studies, the typical approach is to extract adjusted statistical estimates when possible, rather than unadjusted or raw estimates.

It is important to consider the variability associated with effect size estimates, with stronger studies generally showing more precise estimates. Effect size estimation can be affected, however, by such factors as variances that differ substantially across treatment groups, or by lack of information to characterize variance, especially for animal studies in biomedical research (Vesterinen et al., 2014).

8.2. STANDARDIZING ADMINISTERED DOSE LEVELS/CONCENTRATIONS

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

Exposure levels will be converted to vanadium equivalents. For example, a study of sodium metavanadate that reports the dose as mg $NaVO_3$/kg-day will be converted to mg V/kg-day using a molecular weight conversion. Unless otherwise reported by study authors, the background level in experimental animal studies is assumed 0 ppm (0 mg/kg-day).

9. SYNTHESIS WITHIN LINES OF EVIDENCE

The evidence synthesis provides the foundation for evidence integration, which is a distinct, but related, process described in Section 10. The syntheses of separate lines of evidence (i.e., human, animal, and mechanistic evidence) described in this section will directly inform the integration across the lines of evidence to draw overall conclusions for each of the assessed human health effects (described in Section 10). The phrase "evidence integration" used here is analogous to the phrase "weight of evidence" used in some other assessment processes (EFSA, 2017; U.S. EPA, 2017; NRC, 2014; U.S. EPA, 2005a).

For each potential human health effect or smaller subset of related outcomes, EPA will separately synthesize the available phenotypic human and animal evidence pertaining to that potential health effect. Generally, evidence will be synthesized separately for different vanadium compounds or oxidation states. Mechanistic evidence also will be considered in targeted analyses conducted before, during, and after developing syntheses of the phenotypic human and animal evidence. The results of the mechanistic evidence analyses will be used to inform key uncertainties, depending on the extent and nature of the human and animal evidence. Thus, the human and animal evidence syntheses (or the lack of phenotypic data in humans and animals) help determine the approach to be taken in synthesizing the available mechanistic evidence. In this way, the mechanistic synthesis might range from a high-level summary (or detailed analysis) of potential mechanisms of action to specific, focused questions needed to address key uncertainties unaddressed by the phenotypic human and animal evidence (e.g., shape of the dose-response curve at low doses, applicability of the animal evidence to humans, or addressing susceptible populations).

Each synthesis will provide a summary discussion of the available evidence that addresses considerations regarding causation. These considerations are adapted from considerations for causality introduced by Austin Bradford Hill (Hill, 1965): consistency, exposure-response relationship, strength of the association, temporal relationship, biological plausibility, coherence, and "natural experiments" in humans [Table 16; see additional discussion in U.S. EPA (2005a) and U.S. EPA (1994)]. Importantly, the evidence synthesis process explicitly considers and incorporates the conclusions from the individual study evaluations (see Section 6).

Table 16. Information most relevant to describing primary considerations informing causality during evidence syntheses

Consideration	Description of the consideration and its application in IRIS syntheses
Study confidence	Description: Incorporates decisions about study confidence within each of the considerations. Application: In evaluating the evidence for each of the causality considerations described in the following rows, syntheses will consider study confidence decisions. *High* confidence studies carry the most weight. Syntheses will consider specific limitations and strengths of studies and how they inform each consideration.
Consistency	Description: Examines the similarity of results (e.g., direction, magnitude) across studies. Application: Syntheses will evaluate the homogeneity of findings on a given outcome or endpoint across studies. When inconsistencies exist, the syntheses consider whether results were "conflicting" (i.e., unexplained positive and negative results in similarly exposed human populations or in similar animal models) or "differing" (i.e., mixed results explained by differences between human populations, animal models, exposure conditions, or study methods) (U.S. EPA, 2005a) on the basis of analyses of potentially important explanatory factors such as: • Confidence in studies' results, including study sensitivity (e.g., some study results that appear to be inconsistent may be explained by potential biases or other attributes that affect sensitivity). • Exposure, including route (if applicable) and administration methods, levels, duration, timing with respect to outcome development, and exposure assessment methods (i.e., in epidemiological studies). • Specificity and sensitivity of the endpoint for evaluating the health effect in question (e.g., functional measures can be more sensitive than organ weights). • Populations or species, including consideration of potential susceptible groups or differences across life stage at exposure or endpoint assessment. • Toxicokinetic information explaining observed differences in responses across route of exposure, other aspects of exposure, species, or life stages. The interpretation of consistency will emphasize biological significance, to the extent that it is understood, over statistical significance (see additional discussion in Section 9.4). Statistical significance from suitably applied tests (this may involve consultation with an EPA statistician) adds weight when biological significance is not well understood. Consistency in the direction of results increases confidence in that association even in the absence of statistical significance. In some cases, it may be helpful to consider the potential for publication bias and to provide context to interpretations of consistency.[a]

Table 16. Information most relevant to describing primary considerations informing causality during evidence syntheses (continued)

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

Table 16. Information most relevant to describing primary considerations informing causality during evidence syntheses (continued)

Consideration	Description of the consideration and its application in IRIS syntheses
Mechanistic evidence related to biological plausibility	Description: There are multiple uses for mechanistic information (see Section 9.2), and this consideration overlaps with "coherence." This examines the biological support (or lack thereof) for findings from the human and animal health effect studies and becomes more impactful on the hazard conclusions when notable uncertainties in the strength of those sets of studies exist. These analyses can also improve understanding of dose- or duration-related development of the health effect. In the absence of human or animal evidence of apical health endpoints, the synthesis of mechanistic information may drive evidence integration conclusions (when such information is available). Application: Syntheses can evaluate evidence on precursors, biomarkers, or other molecular or cellular changes related to the health effect(s) of interest to describe the likelihood that the observed effects result from exposure. This will be an analysis of existing evidence and not simply whether a theoretical pathway can be postulated. This analysis may not be limited to evidence relevant to the PECO criteria but may also include evaluations of biological pathways (e.g., for the health effect; established for other, possibly related, chemicals). The synthesis will consider the sensitivity of the mechanistic changes and the potential contribution of alternative or previously unidentified mechanisms of toxicity.
Natural experiments	Description: Specific to epidemiological studies and rarely available, this examines effects in populations that have experienced well-described, pronounced changes in chemical exposure (e.g., lead exposures before and after banning of lead in gasoline). Application: Compared to other observational designs, natural experiments have the benefit of dividing people into exposed and unexposed groups without their influencing their own exposure status. During synthesis, associations in *medium* and *high* confidence natural experiments can substantially reduce concerns about residual confounding.

PECO = populations, exposures, comparators, and outcomes.

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

Data permitting, the syntheses will also discuss analyses relating to potential susceptible populations.[28] These analyses will be based on knowledge about the health outcome or organ system affected, demographics, genetic variability, life stage, health status, behaviors or practices, social determinants, and exposure to other pollutants (see Table 17). This information will be used to describe potential susceptibility among specific populations or subgroups in a separate section (see Section 10.3) summarizing across lines of evidence and hazards to inform hazard identification and dose-response analyses.

Table 17. Individual and social factors that may increase susceptibility to exposure-related health effects

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

9.1. SYNTHESES OF HUMAN AND ANIMAL HEALTH EFFECTS EVIDENCE

The syntheses of the human and animal health effect evidence will focus on describing aspects of the evidence that best inform causal interpretations, including the exposure context examined in the sets of studies. These syntheses (or the lack of data within these lines of evidence) help determine the approach to be taken in synthesizing the available mechanistic evidence (see Section 9.2). The mechanistic synthesis might range from a high-level summary of potential mechanisms of action to specific, focused questions needed to address key uncertainties identified from the human and animal syntheses and integration (e.g., shape of dose-response at low doses, applicability of the animal evidence to humans, addressing susceptible populations).

[28]Various terms have been used to characterize populations that may be at increased risk of developing health effects from exposure to environmental chemicals, including "susceptible," "vulnerable," and "sensitive." Further, these terms have been inconsistently defined across the scientific literature. The term susceptibility is used in this protocol to describe populations at increased risk, focusing on biological (intrinsic) factors and social and behavioral determinants that can modify the effect of a specific exposure. Certain factors resulting in higher exposures to specific groups (e.g., proximity, occupation, housing), however, may not be analyzed to describe potential susceptibility among specific populations or groups.

Evidence synthesis will be based primarily on studies of *high* and *medium* confidence. *Low* confidence studies may be used, if few or no studies with higher confidence are available, to help evaluate consistency, or if the study designs of the *low* confidence studies address notable uncertainties in the set of *high* or *medium* confidence studies on a given health effect. If *low* confidence studies are used, a careful examination of risk bias and sensitivity with potential impacts on the evidence synthesis conclusions will be included in the narrative.

As previously described, these syntheses will articulate the strengths and the weaknesses of the available evidence organized around the considerations described in Table 16, as well as issues that stem from the evaluation of individual studies (e.g., concerns about bias or sensitivity). If possible, results across studies will be compared using graphs and charts or other data visualization strategies. The analysis will typically include examination of results stratified by any or all of the following: study confidence classification (or specific issues within confidence evaluation domains); population or species; exposures [e.g., level, patterns (intermittent or continuous), duration, intensity]; sensitivity (e.g., low vs. high); and other factors that may have been identified in the refined evaluation plan (e.g., sex, life stage, or other demographic). The number of studies and the differences encompassed by the studies will determine the extent to which specific types of factors can be examined to stratify study results. Additionally, for both the human and animal evidence syntheses, if supported by the available data, additional analyses across studies (such as meta-analysis) may also be conducted.

9.2. MECHANISTIC INFORMATION

The synthesis of mechanistic information informs the integration of health effect evidence for both hazard identification (i.e., biological plausibility or coherence of the available human or animal evidence, inferences regarding human relevance, or the identification of susceptible populations and life stages across the human and animal evidence) and dose-response evaluation. Therefore, the synthesis of the mechanistic data focuses on the evidence most likely to be useful for augmenting the human or animal health effect evidence. Based on the identified gaps in understanding, the mechanistic synthesis may focus on providing information on precursor events, a biological understanding of how effects develop or are related, the human relevance of animal results, or identifying likely susceptible populations and life stages. This means that, for example, if extensive *high* confidence human or animal evidence is available, the need to synthesize all available mechanistic evidence will be diminished. In these cases, the synthesis will focus on the analysis and interpretation of smaller sets of mechanistic studies that specifically address controversial issues to resolve, such as those related to applicability of animal evidence to humans when the human evidence is weak, or the shape of the dose-response at low exposure levels when this understanding is highly uncertain and data informing this uncertainty exist.

The evidence available to describe mechanistic events or MOAs (U.S. EPA, 2005a) is typically aggregated from numerous studies, often involving a diverse range of exposure paradigms

and models, as well as a wide spectrum of diverse endpoints. In addition, a chemical may operate through multiple mechanistic pathways (U.S. EPA, 2005a). Similarly, multiple mechanistic pathways might interact to cause an adverse effect. In contrast to the defined scope of the evaluation and syntheses of PECO-specific human or animal health effect studies, the potential utility and interpretation of mechanistic information can be quite broad and hard to define. Thus, to be pragmatic and provide clear and transparent syntheses of the most useful information, the mechanistic syntheses for most health outcomes will focus on a subset of the most relevant mechanistic studies. It should be stressed, however, that the process of evaluating mechanistic information differs fundamentally from evaluations of the other evidence streams. More specifically, the mechanistic analysis for any specific substance depends on evaluating the confidence that the relevant data are consistent with a plausible biological understanding of how a chemical exposure might generate an adverse outcome, rather than focusing on evaluations of individual studies.

To identify the focused set(s) of studies for use in analyses of critical mechanistic questions, the synthesis applies a phased approach that progressively focuses the scope of the mechanistic information to be considered. This stepwise focusing, which begins during the literature search and screening steps based on problem formulation decisions, depends primarily on the potential hazard signals that arise from the human and/or animal health effect studies, or from mechanistic studies that signal potential hazards that have not been examined in health effect studies. Examples of the focused questions or scenarios triggering these mechanistic evaluations, as well as when during the systematic review they are likely to apply, are listed in Table 18. While the specific methods for evaluating the sets of studies relevant to each question will vary, some general considerations are provided below.

Table 18. Preparation for the analysis of mechanistic evidence

Assessment stages of identifying mechanistically relevant information	Examples of evidence to review and key considerations
Scoping and problem formulation materials	• For the chemical under review, identify existing chemical-specific MOAs from other agency assessments or review articles. If summary information is lacking, are there structurally similar chemicals that are better studied mechanistically? • Are there indications that a specific mechanistic analysis will be warranted? For example, are there recognized areas of scientific controversy or predefined assessment questions that are already known to require a mechanistic evaluation (e.g., chemicals with a potential mutagenic MOA)? o If so, consider whether additional, targeted literature searches would be informative. o If mechanistic information relevant to a key scientific controversy or to address a mutagenic MOA is lacking, consider whether inferences can be drawn from structure-activity relationships or other "data-poor" approaches. • What is the active moiety of the agent? Are there metabolites that should be considered? Are there indications that the purity is critically important? Is the chemical endogenously produced?
Literature inventory of toxicokinetic, ADME, and physicochemical information	• Based on ADME differences across species, does information exist that suggests a lack of relevance of the animal exposure scenarios to human situations? Is there evidence that the active moiety would not be expected to reach the target tissue(s) in some species? • Are there metabolic pathways involved that may indicate greater sensitivity at a particular life stage or in susceptible human populations? • If a validated PBPK model is available, *revisit any decisions to focus on specific routes of exposure and consider the use of alternative exposure markers.*
Literature inventories of human, animal, and mechanistic information (including all in vitro and in silico studies)	• Which human health hazards (both cancer and noncancer) appear to be well studied in the mechanistic inventory? For cancer, which key characteristics of carcinogens are indicated by the database? o Are there mechanistic studies on an organ system, hazard, or key characteristic that were not examined by human or animal studies meeting the PECO criteria? *If so, consider evidence mapping or similar approaches to highlight these knowledge gaps.* • Are there mechanistic endpoints identified from human and animal studies meeting PECO criteria that could be added to the mechanistic inventory?

Table 18. Preparation for the analysis of mechanistic evidence (continued)

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

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

The information collected (e.g., in sortable inventories) is used to identify studies available for consideration in addressing the specific gaps in understanding identified as critical to address through the application of the questions in Table 18, including postulated mechanistic pathways or MOAs that may be involved in the toxicity of the chemical. Subsequently, from the studies available to potentially address the identified gaps in understanding, the synthesis will focus on those considered most impactful to the specified evaluation based on study design characteristics (which may or may not encompass all studies relevant for a particular question), with a transparent documentation of the rationale for the focusing. As the potential influence of the information provided by these studies can vary depending on the hazard question(s) or the associated mechanistic events or pathways, the level of rigor will also depend on their potential impact of increased understanding to hazard identification or dose-response decisions, and may range from

overviews of potential mechanisms or cursory insights drawn from sets of unanalyzed results to detailed evaluations of a subset of the most relevant mechanistic studies.

Although the application of this approach cannot be predefined, for the small subsets of studies that best address the key mechanistic questions, the synthesis prioritizes studies based on their toxicological relevance to answering the specific question (e.g., model system, specificity of the assay for the effect of interest). The path for focusing the mechanistic database will be documented in the updated protocol released with the draft assessment.

More rigorous analyses are particularly important when the set(s) of studies available to inform influential mechanistic conclusions are inconsistent and potentially conflicting, or when the studies include experiments that directly challenge the necessity of proposed mechanistic relationships between exposure and an apical effect (e.g., altering a receptor-mediated pathway through chemical intervention or using knock-out animals). More detailed analyses may also be useful when it is apparent that the study design aspects in the available studies are likely to have significant flaws or introduce important uncertainties (e.g., potential shortcomings identified during the evaluation of exposure methods may be clarified using mechanistic studies). In some instances, additional literature searches may be warranted, targeting mechanistic events or biological pathways that are not specific to one chemical.

For the more rigorous mechanistic analyses, the review is facilitated by pathway-based organizational methods and established evidence evaluation frameworks. These approaches provide transparency and objectivity to the integration and interpretation of mechanistic events and pathways anchored to the specific questions that have been identified (e.g., anchored to a specific health effect) across diverse sets of relevant data (e.g., human, animal, and in vitro studies). The mechanistic analyses inform the evidence integration across lines of evidence, as well as the dose-response analyses, that are described in Sections 11 and 12. Examples of how mechanistic information can inform these steps are summarized in Table 19.

Table 19. Examples of iterative questions and considerations that focus the synthesis and application of mechanistic information for evidence integration and dose-response analysis

Systematic review step	Mechanistic synthesis triggers and example actions
Human and animal evidence syntheses (see Chapter 9)	• Did the sets of studies report findings that appear to be biologically related to the health effects of interest? *Consider whether these findings might serve as precursors informing an association between exposure and effect; if there are notable uncertainties in the set of studies (e.g., they are all low confidence), consider a focused analysis of precursors to inform strength of evidence; if the data amenable to dose-response analysis are weak or only at high exposure levels, consider evaluating the precursor data for quantitative analysis.* • Do the results appear to differ by categories that indicate the apparent presence of susceptible populations (e.g., across demographics, species, strains, sexes, or life stages)? *Consider analyses to better characterize the sources and impact of potential susceptibilities that might be explained by mechanistic information (e.g., due to genetic polymorphisms or metabolic deficiencies).* • Are there other key uncertainties or data gaps that were identified during the analyses of the sets of available human or animal health effect studies? If so, does the literature inventory of mechanistic studies indicate that there are likely to be a reasonable number of studies on the topic? *If yes, a focused analysis of these studies may be informative. If no, consider whether an additional focused search of mechanistic information might be worthwhile (i.e., to identify other informative studies that were not captured by the initial PECO criteria).*
Integration within the human and animal evidence (see Section 11.1): Information relating to biological plausibility	• Are there notable uncertainties in the sets of human or animal health effect studies for which related mechanistic information is available? An understanding of mechanistic pathways [e.g., by identifying mechanistic precursor events linked qualitatively or quantitatively to apical health effect(s)] can increase the strength of the evidence integration conclusions.

Table 19. Examples of iterative questions and considerations that focus the synthesis and application of mechanistic information for evidence integration and dose-response analysis (continued)

Systematic review step	Mechanistic synthesis triggers and example actions
Integration across lines of evidence (see Section 11.2): Considering human relevance of animal findings	• When human evidence is lacking or has results that differ from animals, is there evidence that the mechanisms underlying the effects in animals operate in humans*? Analyses of the mechanisms underlying the animal response in relation to those presumed to operate in humans, or the suitability of the animal models to a specific human health outcome, can inform the extent to which the animal response is likely to be directly relevant to humans.* • The analysis will focus on evaluations of the following issues. The extent of the analysis will vary depending on the impact of the animal evidence on the conclusions. o Evidence for a plausible mechanistic pathway or MOA, within which the key events and relationships are evaluated regarding the likelihood of similarities (e.g., in presence or function) across species. o Coherence of mechanistic changes observed in exposed humans (or a demonstrated lack of changes that would be expected, e.g., that are known to be linked to the health effect) with animal evidence of mechanistic/toxicological changes. o ADME information describing similarities across species, primarily relating to distribution (e.g., to the likely target tissue) and metabolism (particularly if a metabolite is known to be more/less toxic).
Evidence integration across lines of evidence (see Sections 11.2 and 11.3): Characterizing potential susceptible populations or life stages	• A mechanistic understanding of how a health outcome develops, even without a full MOA, can clarify characteristics of important events (e.g., their presence or sensitivity across life stages or across genetic variations) and helps identify susceptible populations. • Identification of life stages or groups likely to be at greatest risk can clarify hazard descriptions and identify key data gaps including whether the most susceptible populations or life stages have been adequately tested. If a proposed mechanistic pathway or MOA indicates a sensitive population or life stage in humans, consider whether the appropriate analogous exposures and populations or life stages were adequately represented in the human or animal database. • When there is evidence of susceptibilities, but specific studies addressing these susceptibilities are unavailable for quantitative analysis, susceptibility data may support refined human variability UFs or probabilistic uncertainty analyses.

Table 19. Examples of iterative questions and considerations that focus the synthesis and application of mechanistic information for evidence integration and dose-response analysis (continued)

Systematic review step	Mechanistic synthesis triggers and example actions
Dose-response analysis (see Chapter 13): Biological understanding, including the identification of precursor events	• A biological understanding of mechanistic events/MOAs, including the identification of precursor events in humans and the exposure conditions expected to result in these effects, can inform the use of: o Particular dose-response models (e.g., models integrating data across several related outcomes or incorporating toxicokinetic knowledge). o Proximal measures of exposure (e.g., external vs. internal metrics). o Surrogate endpoints (e.g., use of well-established precursors in lieu of direct observation of apical endpoints). o Improved characterization of responses (e.g., combination of related outcomes, such as benign and malignant tumors resulting from the same MOA).

PECO = populations, exposures, comparators, and outcomes; MOA = mode of action; ADME = absorption, distribution, metabolism, and excretion; UF = uncertainty factor.

10. INTEGRATION ACROSS LINES OF EVIDENCE

For the analysis of human health outcomes that might result from chemical exposure, IRIS assessments draw integrated conclusions across human, animal, and mechanistic evidence (see Section 9). During evidence integration, a two-step, sequential process will be used, as follows (and depicted in Figure 15):

- Step 1: Judgments regarding the strength of the evidence from the available human and animal studies are made in parallel, but separately. These judgments incorporate mechanistic evidence (or MOA understanding) that informs the biological plausibility and coherence of the available human or animal health effect studies. Note that at this stage, the animal evidence judgment does not yet consider the human relevance of that evidence.

- Step 2: The animal and human evidence judgments are combined to draw an overall conclusion(s) that incorporates inferences drawn on the basis of information on the human relevance of the animal evidence, coherence across the human and animal evidence, and susceptibility. Without evidence to the contrary, the human relevance of animal findings is assumed.

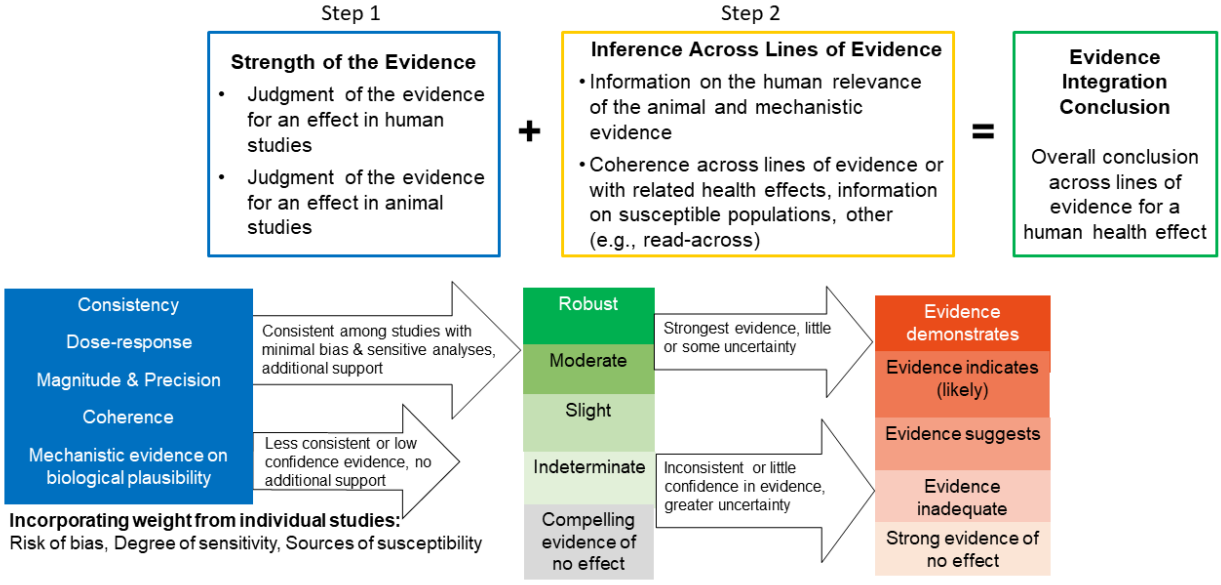


Figure 15. Process for evidence integration.

The decision points within the structured two-step evidence integration process will be summarized in an evidence profile table for each health effect category (see Table 20) in support of

the evidence integration narrative. Human and animal evidence judgments from Step 1 and the overall evidence integration conclusion from Step 2 are reached using decision frameworks (see Sections 10.1 and 10.2 for details) that are based on considerations originally described by Austin Bradford Hill (Hill, 1965). This process is similar to that used by the Grading of Recommendations Assessment, Development and Evaluation [GRADE; (Morgan et al., 2016; Guyatt et al., 2011; Schünemann et al., 2011)], which arrives at an overall integration conclusion based on consideration of the body of evidence. As described in Section 9, the human, animal, and mechanistic syntheses serve as inputs to the evidence integration decisions; thus, the major conclusions from these syntheses will be summarized in the evidence profile table (see Table 20) supporting the evidence integration narrative. The evidence profile table summarizes the judgments and their evidence basis for each step of the structured evidence integration process. Separate sections are included for human and animal evidence judgments, inference across streams, and the overall evidence integration conclusion. The table presents the key information from the evidence that informs each judgment.

Table 20. Evidence profile table template

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

Table 20. Evidence profile table template (continued)

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

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

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

[c]Study confidence, based on evaluation of risk of bias and study sensitivity (see Section 6), and information on susceptibility will be considered when evaluating the other factors that increase or decrease certainty (e.g., consistency). Notably, lack of findings in studies deemed insensitive neither increases nor decreases certainty. Typically, *medium* confidence in only a single study is not a factor that increases certainty, whereas *high* confidence in a single, extensive or rigorous study (e.g., a guideline study) is such a factor.

10.1. INTEGRATION WITHIN THE HUMAN AND ANIMAL EVIDENCE

As summarized above, prior to drawing overall evidence integration conclusions about whether a chemical is likely to cause particular health effect(s) in humans given relevant exposure circumstances, judgments are drawn regarding the strength of evidence for the available human and animal evidence, separately. If relevant mechanistic evidence in exposed humans and animals (or their cells) was synthesized, this line of evidence will be integrated with the evidence from health effects studies. The considerations outlined in Table 16 (see Section 9) are evaluated in the context of how they impact the strength of evidence (see Table 21), and the judgments are reached using the structured frameworks explained in Tables 22 and 23 (for human and animal evidence, respectively). They are summarized in tabular format using the template in Table 20 to transparently convey expert judgments made throughout the evidence synthesis and integration processes. The evidence profile table allows for consistent documentation of the supporting rationale for each decision. At least two independent reviewers will independently assess judgments made for evidence synthesis with differences resolved by discussion to reach a consensus.

Table 21. Considerations that inform judgments regarding the strength of the human and animal evidence

Consideration	Increased evidence strength (of the human or animal evidence)	Decreased evidence strength (of the human or animal evidence)
The structured categories and criteria in Tables 22 and 23 will guide the application of strength of evidence judgments for an outcome or health effect. Evidence synthesis scenarios that do not warrant an increase or decrease in evidence strength will be considered "neutral" and do not need to be described in Table 21		
Risk of bias; sensitivity (across studies)	• An evidence base of *high* or *medium* confidence studies increases strength.	• An evidence base of mostly *low* confidence studies decreases strength. An exception to this is an evidence base of studies where the primary issues resulting in *low* confidence are related to insensitivity. This may increase evidence strength in cases where an association is identified because the expected impact of study insensitivity is towards the null. • Decisions to increase strength for other considerations in this table should generally not be made if there are serious concerns for risk of bias.
Consistency	• Similarity of findings for a given outcome (e.g., of a similar magnitude, direction) across independent studies or experiments increases strength,[a] particularly when consistency is observed across populations (e.g., location) or exposure scenarios in human studies, and across laboratories, populations (e.g., species), or exposure scenarios (e.g., duration; route; timing) in animal studies.	• Unexplained inconsistency (conflicting evidence) decreases strength. Generally, strength should not be decreased if discrepant findings can be reasonably explained by study confidence conclusions; variation in population or species, sex, or life stage; exposure patterns (e.g., intermittent or continuous); levels (low or high); or duration or intensity.
Strength (effect magnitude) and precision	• Evidence of a large magnitude effect (considered either within or across studies) can increase strength. Effects of a concerning rarity or severity can also increase strength, even if they are of a small magnitude. • Precise results from individual studies or across the set of studies increases strength, noting that biological significance is prioritized over statistical significance.	• Strength may be decreased if effect sizes that are small in magnitude are concluded not to be biologically significant, or if there are only a few studies with imprecise results.

Table 21. Considerations that inform judgments regarding the strength of the human and animal evidence (continued)

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

Table 21. Considerations that inform judgments regarding the strength of the human and animal evidence (continued)

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

[a]Publication bias has the potential to result in strength of evidence judgments that are stronger than would be merited if the entire body of research were available. However, the existence of publication bias can be difficult to determine (see Section 9.4.3 for additional discussion). If strong evidence of publication bias exists for an outcome, the increase in evidence strength resulting from considering the consistency of the evidence across studies may be reduced.

For human and animal evidence, the analyses of each consideration in Table 21 will be used to develop a strength of evidence judgment. Tables 22 and 23 provide the criteria that will guide how to draw the judgments for each health effect, and the terms that will be used to summarize those judgments. These terms are applied to human and animal evidence separately; and, within the human and animal lines of evidence, separate judgments may be applied to different vanadium compounds or oxidation states. Briefly, the terms *Robust* and *Moderate* are standardized characterizations for judgments that the relevant effect(s) observed in humans or animals results from exposure to vanadium and compounds; these two terms are differentiated by the quantity and quality of information available to rule out alternative explanations for the results. For example, repeated observations of effects by independent studies examining various aspects of exposure or response (e.g., different exposure settings, dose levels or patterns, populations or species, and related endpoints) will result in a stronger strength of evidence judgment. The term *Slight* indicates situations in which there is some evidence indicating an association within the evidence stream, but substantial uncertainties in the data exist to prevent judgments that the relevant effect(s) observed in humans or animals can be reliably attributed to exposure to vanadium and compounds. *Indeterminate* reflects evidence stream judgments when no studies are available, or situations when the evidence is inconsistent or primarily of *low* confidence. *Compelling evidence of no effect* represents a situation in which extensive evidence across a range of populations and exposures has identified no effects/associations. This scenario is seldom used because it requires a high degree of confidence in the conduct of individual studies, including consideration of study sensitivity, and comprehensive assessments of health outcomes and life stages of exposure.

Table 22. Framework for evidence judgments from studies in humans

Within-stream strength-of-evidence judgment	Description
Robust (⊕⊕⊕) …evidence in human studies *(strong signal of effect with little residual uncertainty)*	A set of *high* or *medium* confidence independent studies reporting an association between the exposure and the health outcome, with reasonable confidence that alternative explanations, including chance, bias, and confounding, can be ruled out across studies. The set of studies is primarily consistent, with reasonable explanations when results differ; and an exposure response gradient is demonstrated. Additional supporting evidence, such as associations with biologically related endpoints in human studies (coherence) or large estimates of risk or severity of the response, may increase confidence but are not required. Mechanistic evidence from exposed humans, if available, may add support informing considerations such as exposure response, temporality, coherence, and MOA, thus, raising the level of certainty to robust for a set of studies that otherwise would be described as moderate.
Moderate (⊕⊕⊙) …evidence in human studies *(signal of effect with some uncertainty)*	A smaller number of studies (at least one *high* or *medium* confidence study with supporting evidence), or with some heterogeneous results, that do not reach the degree of confidence required for robust. For multiple studies, there is primarily consistent evidence of an association, but there may be some uncertainty due to potential chance, bias, or confounding. For a single study, there is a large magnitude or severity of the effect, or a dose-response gradient, or other supporting evidence, and there are no serious residual methodological uncertainties. Supporting evidence could include associations with related endpoints, including mechanistic evidence from exposed humans, if available, based on considerations such as exposure response, temporality, coherence, and MOA.
Slight (⊕⊙⊙) …evidence in human studies *(signal of effect with large amount of uncertainty)*	One or more studies reporting an association between exposure and the health outcome, where considerable uncertainty exists. In general, the evidence is limited to a set of consistent *low* confidence studies, or higher confidence studies with unexplained heterogeneity. Supporting coherent evidence is sparse. Biological support from mechanistic evidence in exposed humans may also be independently interpreted as slight. This also includes scenarios where there are serious residual uncertainties across studies (these uncertainties typically relate to exposure characterization or outcome ascertainment, including temporality) in a set of largely consistent *medium* or *high* confidence studies. This category serves primarily to encourage additional study where evidence does exist that might provide some support for an association, but for which the evidence does not reach the degree of confidence required for moderate.
Indeterminate (⊙⊙⊙) …evidence in human studies *(signal cannot be determined for or against an effect)*	No studies available in humans or situations when the evidence is highly inconsistent and primarily of *low* confidence. In addition, this may include situations where higher confidence studies exist, but unexplained heterogeneity exists, and there are additional outstanding concerns such as effect estimates of *low* magnitude, uninterpretable patterns with respect to exposure levels, or uncertainties or methodological limitations that result in an inability to discern effects from exposure. A set of largely null studies could be concluded to be indeterminate if the evidence does not reach the level required for compelling evidence of no effect.

Table 22. Framework for evidence judgments from studies in humans (continued)

Within-stream strength-of-evidence judgment	Description
Compelling evidence of no effect (- - -) …in human studies *(strong signal for lack of an effect with little uncertainty)*	Several high confidence studies showing null results (for example, an odds ratio of 1.0), ruling out alternative explanations including chance, bias, and confounding with reasonable confidence. Each of the studies should have used an optimal outcome and exposure assessment and adequate sample size (specifically for higher exposure groups and for susceptible populations). The set as a whole should 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 life stages.

Table 23. Framework for evidence judgments from studies in animals

Within-stream strength-of-evidence judgment	Description
Robust (⊕⊕⊕) …evidence in animals *(strong signal of effect with little residual uncertainty)*	The set of *high* or *medium* confidence experiments includes consistent findings of adverse or toxicologically significant effects across multiple laboratories, exposure routes, experimental designs (e.g., a subchronic study and a two-generation study), or species, and the experiments can reasonably rule out the potential for nonspecific effects (e.g., resulting from toxicity) to have resulted in the findings. Any inconsistent evidence (evidence that cannot be reasonably explained by the respective study design or differences in animal model) is from a set of experiments of lower confidence. At least two of the following additional factors in the set of experiments support a causal association: coherent effects across multiple related endpoints (may include mechanistic endpoints); an unusual magnitude of effect, rarity, age at onset, or severity; a strong dose-response relationship; or consistent observations across animal life stages, sexes, or strains. Alternatively, mechanistic data in animals (in vivo or in vitro) that address the above considerations or that provide experimental support for an MOA that defines a causal relationship with reasonable confidence may raise the level of certainty to robust for evidence that otherwise would be described as moderate or, exceptionally, slight or indeterminate.

Table 23. Framework for evidence judgments from studies in animals (continued)

Within-stream strength-of-evidence judgment	Description
Moderate (⊕⊕⊙) …evidence in animals *(signal of effect with some uncertainty)*	A set of evidence that does not reach the degree of certainty required for robust, but which includes at least one *high* or *medium* confidence study and information strengthening the likelihood of a causal association. Although the results are largely consistent, notable uncertainties remain. However, while inconsistent evidence and/or evidence indicating nonspecific effects (e.g., maternal toxicity at doses causing developmental effects) may exist, it is not sufficient to reduce or discount the level of concern regarding the positive findings from the supportive experiments or it is from a set of experiments of lower confidence. The set of experiments supporting the effect provide additional information supporting a causal association, such as consistent effects across laboratories or species; coherent effects across multiple related endpoints (may include mechanistic endpoints); an unusual magnitude of effect, rarity, age at onset, or severity; a strong dose-response relationship; and/or consistent observations across exposure scenarios (e.g., route, timing, duration), sexes, or animal strains. Mechanistic data in animals (in vivo or in vitro) that address the above considerations or that provide information supporting an association between exposure and effect with reasonable confidence may raise the level of certainty to moderate for evidence that otherwise would be described as slight.
Slight (⊕⊙⊙) …evidence in animals *(signal of effect with large amount of uncertainty)*	Scenarios in which there is a signal of a possible effect, but the evidence is conflicting or weak. Most commonly, this includes situations where only *low* confidence experiments are available and supporting coherent evidence is sparse. It also applies when one *medium* or *high* confidence experiment is available without additional information strengthening the likelihood of a causal association (e.g., corroboration within the same study or from other studies). Lastly, this includes scenarios in which there is evidence that would typically be characterized as moderate, but inconsistent evidence (evidence that cannot be reasonably explained by the respective study design or differences in animal model) from a set of experiments of higher confidence (may include mechanistic evidence) exists. Strong biological support from mechanistic studies in exposed animals or animal cells may also be independently interpreted as slight. Notably, to encourage additional research, it is important to describe situations for which evidence does exist that might provide some support for an association but is insufficient for a conclusion of moderate.
Indeterminate (⊙⊙⊙) …evidence of the effect under review in animals *(signal cannot be determined for or against an effect)*	No animal studies were available, the available endpoints are not informative to the hazard question under evaluation, or the evidence is highly inconsistent and primarily of *low* confidence. In addition, this may include situations where higher confidence studies exist, but there is unexplained heterogeneity and additional concerns such as small effect sizes (given what is known about the endpoint) or a lack of dose-dependence. A set of largely null studies could be concluded to be indeterminate if the evidence does not reach the level required for compelling evidence of no effect.

Table 23. Framework for evidence judgments from studies in animals (continued)

Within-stream strength-of-evidence judgment	Description
Compelling evidence of no effect (- - -) …in animals *(strong signal for lack of an effect with little uncertainty)*	A set of *high* confidence experiments examining a reasonable spectrum of endpoints relevant to a type of toxicity that demonstrate a lack of biologically significant effects across multiple species, both sexes, and a broad range of exposure levels. The data are compelling in that the experiments have examined the range of scenarios across which health effects in animals could be observed, and an alternative explanation (e.g., inadequately controlled features of the studies' experimental designs; inadequate sample sizes) for the observed lack of effects is not available. The experiments were designed to specifically test for effects of interest, including suitable exposure timing and duration, post exposure latency, and endpoint evaluation procedures, and to address potentially susceptible populations and life stages. Mechanistic data in animals (in vivo or in vitro) that address the above considerations or that provide information supporting the lack of an association between exposure and effect with reasonable confidence may provide additional support to this judgment.

10.2. OVERALL EVIDENCE INTEGRATION CONCLUSIONS

The second stage of evidence integration combines animal and human evidence judgments while also considering mechanistic information on the human relevance of the animal evidence, relevance of the mechanistic evidence to humans (especially in cases where animal evidence is lacking), coherence across lines of evidence, and information on susceptible populations and life stages. Based on the integration across lines of evidence, this stage culminates in an evidence integration narrative that summarizes the conclusions regarding each potential health effect (i.e., each noncancer health effect and specific type of cancer, or broader grouping of related outcomes as defined in the evaluation plan). For each health effect, this narrative will include a summary of the strength of each line of evidence and an overall conclusion across the lines of evidence, with exposure context provided. The first sentence of the evidence integration narrative should include the summary conclusion, and, for evaluations of carcinogenicity, include the cancer descriptor (U.S. EPA, 2005a). Table 24 describes the five evidence integration conclusion levels, the integration conclusion language associated with each level, and the types of evidence that fit each level. The five integration conclusion levels reflect the differences in the amount and quality of the data that inform the evaluation of whether exposure may cause the health effect(s) under specified exposure conditions.

Table 24. Conclusions for the evidence integration narrative

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

Table 24. Conclusions for the evidence integration narrative (continued)

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

Table 24. Conclusions for the evidence integration narrative (continued)

Evidence integration conclusion[a] in narrative	Evidence integration conclusion level	Explanation and example scenarios[b]
The currently available **evidence is inadequate** to assess whether [chemical] may cause [health effect] in humans under relevant exposure circumstances.	**Evidence inadequate**	This conveys either a lack of information or an inability to interpret the available evidence for [health effect]. On an assessment-specific basis, a single use of this "inadequate" conclusion level might be used to characterize the evidence for multiple health effect categories (i.e., all health effects that were examined and did not support other conclusion levels).[g] • This conclusion level is used if there is *indeterminate* human and animal evidence. • This conclusion level is also used with *slight* animal evidence and *compelling evidence of no effect* human evidence. • This conclusion level could also be used with *slight*-to-*robust* animal evidence and *indeterminate* human evidence if strong mechanistic information indicated that the animal evidence is unlikely to be relevant to humans. A conclusion of **inadequate** is not a determination that the agent does not cause the indicated health effect(s). It simply indicates that the available evidence is insufficient to reach conclusions.
Strong evidence supports no effect in humans under relevant exposure circumstances. This conclusion is based on studies of [humans or animals] that assessed [exposure or dose] levels of [range of concentrations].	**Strong evidence supports no effect**	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 life stages of exposure relevant to the heath effect of interest. • This conclusion level is used if there is *compelling evidence of no effect* in human studies *and compelling evidence of no effect* to *indeterminate* in animals. • This conclusion level is also used if there is *indeterminate* human evidence and *compelling evidence of no effect* in animal models concluded to be relevant to humans. • This conclusion level could also be used with *compelling evidence of no effect* in human studies and *moderate* to *robust* animal evidence if strong mechanistic information indicated that the animal evidence is unlikely to be relevant to humans.

[a]Evidence integration conclusions 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 allow for evaluations only at the category of health effects examined; this will more frequently be the case for conclusion levels of "evidence suggests" and "evidence inadequate."

[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 life stage 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," will be 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 pathway 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.

For evaluations of carcinogenicity, consistent with EPA's cancer guidelines (U.S. EPA, 2005a), one of EPA's standardized cancer descriptors will be used as a shorthand characterization of the evidence integration narrative, describing the overall potential for carcinogenicity. These 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. More than one descriptor can be used when a chemical's effects differ by exposure level or route (U.S. EPA, 2005a). In some cases, mutagenicity will also be evaluated (e.g., when there is evidence of carcinogenicity), because it influences the approach to dose-response assessment and subsequent application of adjustment factors for exposures early in life (U.S. EPA, 2005a, b).

For each cancer subtype, an evidence integration narrative will be provided as described above, and an appropriate descriptor will be selected as described in the EPA cancer guidelines. If a systematic review of more than one cancer type was conducted, the conclusion for the cancer type(s) with the highest confidence will be used as the basis for the standardized cancer descriptor. When considering evidence on carcinogenicity across human and animal evidence, consistent with EPA guidance (U.S. EPA, 2005a), site concordance is not required. The cancer descriptor and evidence integration narrative, including application of the MOA framework, will also consider the conditions of carcinogenicity, including exposure (e.g., route; level) and susceptibility (e.g., genetics; life stage), as the data allow (Farland, 2005; U.S. EPA, 2005a, b).

10.3. HAZARD CONSIDERATIONS FOR DOSE-RESPONSE

This section provides a transition from hazard identification to the dose-response section, highlighting (1) information that will 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 toxicokinetic information, and (4) the identification of biologically based BMR levels. The pool of outcomes and study-specific endpoints will be discussed to identify which categories of effects and study designs are considered the strongest and most appropriate for quantitative assessment of a given health effect. Health effects that were analyzed in relation to exposure levels within or closer to the range of exposures encountered in the environment are particularly informative. When there are multiple endpoints for an organ/system, considerations for characterizing the overall impact on this organ/system will be discussed. For example, if there are multiple histopathological alterations relevant to liver function changes, liver necrosis may be selected as the most representative endpoint to consider for dose-response analysis. This section may review or clarify which endpoints or combination of endpoints in each organ/system characterize the overall effect for dose-response analysis. For cancer types, consideration will be given to the overall risk of multiple types of tumors. Multiple tumor types (if applicable) will be discussed, and a rationale given for any grouping.

Biological considerations that are important for dose-response analysis (e.g., that could help with selection of a BMR) will be discussed. The impact of route of exposure on toxicity to different organs/systems will be examined, if appropriate. The existence and validity of PBPK models or toxicokinetic information that may allow the estimation of internal dose for route-to-route extrapolation will be presented. In addition, mechanistic evidence presented in Section 9 that will influence the dose-response analyses will be highlighted, for example, evidence related to susceptibility or potential shape of the dose-response curve (i.e., linear, nonlinear, or threshold model). Mode(s) of action will be summarized including any interactions between them relevant to understanding overall risk. Some 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?

This section will also draw from Sections 9 and 10 to describe the evidence (i.e., human, animal, mechanistic) regarding populations and life stages susceptible to the hazards identified and factors that increase risk of the hazards. This section should include a discussion of the populations that may be, in general, susceptible to the health effects identified to be hazards of exposure to the

assessed chemical, even if there are no specific data on effects of exposure to that chemical in the potentially susceptible population. Background information about biological mechanisms or ADME, as well as biochemical and physiological differences among life stages may be used to guide the selection of populations and life stages to consider. At a minimum, particular consideration will be given to infants and children, pregnant women, and women of childbearing age. Evidence on factors that contribute to some population groups having increased responses to chemical exposure or factors that contribute to increases in exposure or dose will be summarized and evaluated with respect to patterns across studies pertinent to consistency, coherence, and the magnitude and direction of effect measures. Relevant factors may include intrinsic factors (e.g., age, sex, genetics, health status, behaviors); extrinsic factors (e.g., socioeconomic status, access to health care); and differential exposure levels or frequency (e.g., occupation-related exposure, residential proximity to locations with greater exposure intensity).

The section will consider options for using data related to susceptible populations to impact dose-response analysis. In particular, an attempt will be made to highlight where it might be possible to develop separate risk estimates for a specific population or life stage or determine whether evidence is available to select a data-derived uncertainty factor (UF).

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

Selection of specific data sets for dose-response assessment and performance of the dose-response assessment are conducted after hazard identification is complete and involve database- and chemical-specific biological judgments. A number of EPA guidance 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 guidance.

For IRIS assessments, dose-response assessments are typically performed for both noncancer and cancer hazards following chronic exposure[29] to the chemical of interest, if supported by existing data. For noncancer hazards, an oral reference dose (RfD) is derived. An RfD is an estimate, with uncertainty spanning perhaps an order of magnitude, of an exposure to the human population (including susceptible subgroups) that is likely to be without an appreciable risk of deleterious health effects over a lifetime (U.S. EPA, 2002, §4.2). RfDs may also be derived for cancer effects in cases where 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) (see Section 11.2.3).

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, an oral slope factor (OSF) facilitates estimation of human cancer risks. An OSF is a plausible upper-bound lifetime cancer risk from chronic ingestion of a chemical per unit of mass consumed per unit body weight, per day (mg/kg-day). In contrast with reference values (RfVs), an OSF can be used in conjunction with exposure information to predict cancer risk at a given dose.

As discussed in Section 2 ("Scoping and Initial Problem Formulation") of this protocol, the IRIS assessment will have the goal of developing oral toxicity value(s) (RfD, OSF, or both) for vanadium and compounds. The assessment will attempt to derive separate toxicity values for

[29]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).

individual vanadium compounds or oxidation states (i.e., V^{+4} and V^{+5}) as well as an overall toxicity value for vanadium, as supported by the available data.

11.1. SELECTING STUDIES FOR DOSE-RESPONSE ASSESSMENT

The dose-response assessment begins with a review of the important health effects highlighted in the hazard identification step (see Section 10), particularly among the studies of highest quality and that exemplify the study attributes summarized in Table 25. This review also considers 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.

Among the studies that support hazard conclusions, those that are most useful for dose-response analysis generally have at least one exposure level in the region of the dose-response curve near the benchmark response (the response level to be used for deriving toxicity values), to minimize low-dose extrapolation, and more exposure levels and larger sample sizes overall (U.S. EPA, 2012). Preference will be given to studies that characterize exposure to vanadium according to the criteria outlined in Sections 6.2.1 and 6.3.1. In addition to these more general considerations, specific issues that may impact the feasibility of dose-response modeling for individual data sets are described in more detail in the *Benchmark Dose Technical Guidance* (U.S. EPA, 2012).

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 (see Section 7), semiquantitative analysis may be feasible (e.g., via NOAEL/LOAEL). Studies of low sensitivity may be less useful if they fail to detect a true effect or yield points of departure with wide confidence limits, but such studies would be considered for inclusion in a meta-analysis.

Table 25. Attributes used to evaluate studies for derivation of toxicity values

Study attributes		Considerations	
		Human studies	**Animal studies**
Study confidence		*High* or *medium* confidence studies are highly preferred over *low* confidence studies. The available *high* and *medium* confidence studies are further differentiated on the basis of the study attributes below, and a reconsideration of the specific limitations identified and their potential impact on dose-response analyses.	
Rationale for choice of species		Human data are preferred over animal data to eliminate interspecies extrapolation uncertainties (e.g., in toxicodynamics, relevance of specific health outcomes to humans).	Animal studies provide supporting evidence when adequate human studies are available and are considered principal studies when adequate human studies are not available. For some hazards, studies of particular animal species known to respond similarly to humans would be preferred over studies of other species.
Relevance of exposure paradigm	Exposure route	Studies involving human environmental exposures (oral, inhalation).	Studies by a route of administration relevant to human environmental exposure are preferred. A validated toxicokinetic 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.	

Table 25. Attributes used to evaluate studies for derivation of toxicity values (continued)

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 concentrations vs. target concentrations) and that evaluate vanadium speciation or ensure an appropriate pH range for the species tested, or both, are preferred. Relevant internal dose measures may facilitate extrapolation to humans, as would availability of a suitable animal PBPK model in conjunction with an animal study reported in terms of administered exposure.
Measurement of health outcome(s)	Studies that can reliably distinguish the presence or absence (or degree of severity) of the outcome are preferred. Outcome ascertainment methods using generally accepted or standardized approaches are preferred.	
	Studies with individual data are preferred in general. Examples include to characterize experimental variability more realistically and to characterize overall incidence of individuals affected by related outcomes.	
	Among several relevant health outcomes, preference is generally given to those with greater biological significance.	
Study size and design	Preference is given to studies using designs reasonably expected to have power to detect responses of suitable magnitude.[b] This does not mean that studies with substantial responses but low power would be ignored, but that they should be interpreted in light of a confidence interval or variance for the response. Studies that address changes in the number at risk (through decreased survival, loss to follow-up) are preferred.	

[a]An exposure or other variable that is associated with both exposure and outcome but is not an intermediary between the two.
[b]Power is an attribute of the design and population parameters, based on a concept of repeatedly sampling a population; it cannot be inferred post hoc using data from one experiment (Hoenig and Heisey, 2001).

11.2. CONDUCTING DOSE-RESPONSE ASSESSMENTS

EPA uses a two-step approach for dose-response assessment that distinguishes analysis of the dose-response data in the range of observation from any inferences about responses at lower environmentally relevant exposure levels (U.S. EPA, 2012; 2005a, §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 point of departure (POD); see Section 11.2.1 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 11.2.2 and 11.2.3, 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 reference values, IRIS assessments typically derive a candidate value from each suitable data set, whether for human or animal (see Section 11.1). Evaluating these candidate values grouped within a particular organ/system yields a single organ/system-specific value for each organ/system under consideration. Next, evaluation of these organ/system-specific 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 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.

11.2.1. Dose-response Analysis in the Range of Observation

For conducting a dose-response assessment, toxicodynamic ("biologically based") modeling can be used when there are sufficient data to ascertain the mode of action and quantitatively support model parameters that represent rates and other quantities associated with the key precursor events of the mode of action.

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

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

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

The response level at which the POD is calculated is guided by the severity of the endpoint. If linear extrapolation is used, selection of a response level corresponding to the point of departure is not highly influential, so standard values near the low end of the observable range are generally used (for example, 10% extra risk for cancer bioassay data, 1% for epidemiological data, lower for rare cancers). Nonlinear approaches account for 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. 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. The point of departure is the 95% lower bound on the dose associated with the selected response level.

EPA has developed standard approaches for determining the relevant dose to be used in the dose-response modeling in the absence of appropriate toxicokinetic 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; 1991, §3.2).

- Doses are standardized to equivalent human terms to facilitate comparison of results from different species. Oral doses are scaled allometrically using $mg/kg^{3/4}$-day as the equivalent dose metric across species. Allometric scaling pertains to equivalence across species, not across life stages, and is not used to scale doses from adult humans or mature animals to infants or children (U.S. EPA, 2011a; 2005a, §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 case of vanadium, unless PBPK studies are identified in future searches no attempt will be made to convert inhalation study data to oral PODs.

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

11.2.2. Extrapolation: Slope Factors

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); see Section 11.2.3 below.

Differences in susceptibility may warrant derivation of multiple slope factors, with separate estimates for susceptible populations and life stages (U.S. EPA, 2005a, b). If appropriate chemical-specific data on susceptibility from early life exposures are available, these data are used to develop cancer risk values that specifically address any potential for differential potency in early life stages (U.S. EPA, 2005a, b). If such data are not available, the evidence synthesis and integration analyses support a mutagenic MOA for carcinogenicity, and the extrapolation approach is linear, the dose-response assessment should indicate that, in the development of risk estimates, the default *age-dependent adjustment factors* should be used with the cancer slope factor or unit risk and age-specific estimates of exposure (U.S. EPA, 2005a, b). The derivation of an OSF for vanadium and compounds conducted as part of the current assessment will be performed consistent with EPA guidance.

11.2.3. 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, §3.3.4); in general, the reference value is based not on tumor incidence, but on a key precursor event in the MOA that is necessary for tumor formation.

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 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 toxicokinetics or toxicodynamics. If available, a biologically based model that adjusts fully for toxicokinetic and toxicodynamic differences across species may be used. Otherwise, the POD is standardized to equivalent human terms or is based on toxicokinetic or dosimetry modeling, which may range from detailed chemical-specific to default approaches (U.S. EPA, 2014, 2011a), and a factor of $10^{1/2}$ (rounded to 3) is applied to account for the remaining uncertainty involving toxicokinetic and toxicodynamic 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 toxicodynamics or toxicokinetics is considered (U.S. EPA, 2014, 2002).[30, 31] When sufficient data are available, an intraspecies UF either less than or greater than 10-fold may be justified (U.S. EPA, 2002). This factor may be reduced if the POD is derived from or adjusted specifically for susceptible individuals [not for a general population that includes both susceptible and nonsusceptible individuals; (U.S. EPA, 2002, §4.4.5; 1998, §4.2; 1996, §4; 1994, §4.3.9.1; 1991, §3.4)]. When the use of such data or modeling is not supported, a UF with a default value of 10 is considered.

- *LOAEL to NOAEL*: If a POD is based on a LOAEL, 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, 1991).

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

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

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

[31]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.

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 an RfV is divided by the product of these factors. U.S. EPA (2002, §4.4.5) recommends that any composite factor that exceeds 3,000 represents excessive uncertainty and recommends against relying on the associated RfV. The derivation of an RfD for vanadium and compounds conducted as part of the current assessment will be performed consistent with EPA guidance summarized above. As previously mentioned, this assessment will attempt to derive separate RfDs for individual vanadium compounds or oxidation states (i.e., V^{+4} and V^{+5}) and an overall RfD for vanadium, as supported by the available data.

REFERENCES

Afkhami-Arekani, M; Karimi, M; Mohammadi, SM; Nourani, F. (2008). Effect of sodium metavanadate supplementation on lipid and glucose metabolism biomarkers in type E diabetic patients. Malaysian Journal of Nutrition 14: 113-119.

ATSDR (Agency for Toxic Substances and Disease Registry). (2012). Toxicological profile for vanadium [ATSDR Tox Profile]. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service. http://www.atsdr.cdc.gov/toxprofiles/tp58.pdf

Aureliano, M. (2014). Decavanadate contribution to vanadium biochemistry: In vitro and in vivo studies. Inorganica Chim Acta 420: 4-7. http://dx.doi.org/10.1016/j.ica.2013.10.010

Aureliano, M; Crans, DC. (2009). Decavanadate (V10O286-) and oxovanadates: Oxometalates with many biological activities. J Inorg Biochem 103: 536-546. http://dx.doi.org/10.1016/j.jinorgbio.2008.11.010

Boden, G; Chen, X; Ruiz, J. (1996). Effects of vanadyl sulphate on carbohydrate and lipid metabolism in patients with non-insulin dependent diabetes mellitus. Metabolism 45: 1130-1135.

Boscolo, P; Carmignani, M; Volpe, AR; Felaco, M; Del Rosso, G; Porcelli, G; Giuliano, G. (1994). Renal toxicity and arterial hypertension in rats chronically exposed to vanadate. Occup Environ Med 51: 500-503. http://dx.doi.org/10.1136/oem.51.7.500

Byczkowski, JZ; Kulkarni, AP. (1996). Pro-oxidant biological effects of inorganic component of petroleum: vanadium and oxidative stress. Wright-Patterson AFB, OH: Armstrong Laboratory, Occupational and Environmental Health Directorate.

Carmignani, M; Volpe, AR; Porcelli, G; Boscolo, P; Preziosi, P. (1992). Chronic exposure to vanadate as factor of arterial hypertension in the rat: toxicodynamic mechanisms. Arch Toxicol Suppl 15: 117-120. http://dx.doi.org/10.1007/978-3-642-77260-3_15

Cohen, N; Halberstam, M; Shlimovich, P; Chang, CJ; Shamoon, H; Rossetti, L. (1995). Oral vanadyl sulfate improves hepatic and peripheral insulin sensitivity in patients with non-insulin-dependent diabetes mellitus. J Clin Invest 95: 2501-2509. http://dx.doi.org/10.1172/JCI117951

Costa Pessoa, J. (2015). Thirty years through vanadium chemistry. J Inorg Biochem 147: 4-24. http://dx.doi.org/10.1016/j.jinorgbio.2015.03.004

Crans, DC. (2015). Antidiabetic, Chemical, and Physical Properties of Organic Vanadates as Presumed Transition-State Inhibitors for Phosphatases. J Org Chem 80: 11899-11915. http://dx.doi.org/10.1021/acs.joc.5b02229

Crans, DC; Mahrooftahir, M; Keramidas, AD. (1995). Vanadium chemistry and biochemistry of relevance for use of vanadium compounds as antidiabetic agents. Mol Cell Biochem 153: 17-24. http://dx.doi.org/10.1007/BF01075914

Crans, DC; Rithner, CD; Theisen, LA. (1990). Application of time-resolved vanadium-51 2D NMR for quantitation of kinetic exchange pathways between vanadate monomer, dimer, tetramer, and pentamer. J Am Chem Soc 112: 2901-2908. http://dx.doi.org/10.1021/ja00164a009

Crans, DC; Smee, JJ; Gaidamauskas, E; Yang, L. (2004). The chemistry and biochemistry of vanadium and the biological activities exerted by vanadium compounds. Chem Rev 104: 849-902. http://dx.doi.org/10.1021/cr020607t

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

Cusi, K; Cukier, S; Defronzo, RA; Torres, M; Puchulu, FM; Redondo, JC. (2001). Vanadyl sulfate improves hepatic and muscle insulin sensitivity in type 2 diabetes. J Clin Endocrinol Metab 86: 1410-1417. http://dx.doi.org/10.1210/jcem.86.3.7337

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

Dimond, EG; Caravaca, J; Benchimol, A. (1963). Vanadium: Excretion, toxicity, lipid effect in man. Am J Clin Nutr 12: 49-53. http://dx.doi.org/10.1093/ajcn/12.1.49

Domingo, JL; Llobet, JM; Tomas, JM; Corbella, J. (1985). Short-term toxicity studies of vanadium in rats. J Appl Toxicol 5: 418-421. http://dx.doi.org/10.1002/jat.2550050616

Domingo, JL; Paternain, JL; Llobet, JM; Corbella, J. (1986). Effects of vanadium on reproduction, gestation, parturition and lactation in rats upon oral administration. Life Sci 39: 819-824. http://dx.doi.org/10.1016/0024-3205(86)90460-1

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

Evangelou, AM. (2002). Vanadium in cancer treatment. Crit Rev Oncol Hematol 42: 249-265. http://dx.doi.org/10.1016/S1040-8428(01)00221-9

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

Fawcett, JP; Farquhar, SJ; Thou, T; Shand, BI. (1997). Oral vanadyl sulphate does not affect blood cells, viscosity or biochemistry in humans. Pharmacol Toxicol 80: 202-206.

Gerke, TL; Scheckel, KG; Maynard, JB. (2010). Speciation and distribution of vanadium in drinking water iron pipe corrosion by-products. Sci Total Environ 408: 5845-5853. http://dx.doi.org/10.1016/j.scitotenv.2010.08.036

Gerke, TL; Scheckel, KG; Schock, MR. (2009). Identification and distribution of vanadinite (Pb-5(V5+O4)3Cl) in lead pipe corrosion by-products. Environ Sci Technol 43: 4412-4418. http://dx.doi.org/10.1021/es900501t

Goldfine, AB; Patti, ME; Zuberi, L; Goldstein, BJ; Leblanc, R; Landaker, EJ; Jiang, ZY; Willsky, GR; Kahn, CR. (2000). Metabolic effects of vanadyl sulfate in humans with non-insulin-dependent diabetes mellitus: in vivo and in vitro studies. Metabolism 49: 400-410. http://dx.doi.org/10.1016/S0026-0495(00)90418-9

Goldfine, AB; Simonson, DC; Folli, F; Patti, ME; Kahn, CR. (1995). Metabolic effects of sodium metavanadate in humans with insulin-dependent and noninsulin-dependent diabetes mellitus in vivo and in vitro studies. J Clin Endocrinol Metab 80: 3311-3320. http://dx.doi.org/10.1210/jcem.80.11.7593444

Goulle, JP; Mahieu, L; Castermant, J; Neveu, N; Bonneau, L; Laine, G; Bouige, D; Lacroix, C. (2005). Metal and metalloid multi-elementary ICP-MS validation in whole blood, plasma, urine and hair - Reference values. Forensic Sci Int 153: 39-44. http://dx.doi.org/10.1016/j.forsciint.2005.04.020

Gustafsson, JP. (2019). Vanadium geochemistry in the biogeosphere -speciation, solid-solution interactions, and ecotoxicity. Appl Geochem 102: 1-25. http://dx.doi.org/10.1016/j.apgeochem.2018.12.027

Gutiérrez-González, E; García-Esquinas, E; de Larrea-Baz, NF; Salcedo-Bellido, I; Navas-Acien, A; Lope, V; Gómez-Ariza, JL; Pastor, R; Pollán, M; Pérez-Gómez, B. (2019). Toenails as biomarker of exposure to essential trace metals: A review. Environ Res 179: 108787. http://dx.doi.org/10.1016/j.envres.2019.108787

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

Halberstam, M; Cohen, N; Shlimovich, P; Rossetti, L; Shamoon, H. (1996). Oral vanadyl sulfate improves insulin sensitivity in NIDDM but not in obese nondiabetic subjects. Diabetes 45: 659-666. http://dx.doi.org/10.2337/diab.45.5.659

Harrington, JM; Hainesa, LG; Levine, KE; Liyanapatirana, C; Essader, AS; Fernando, RA; Robinson, VG; Roberts, GK; Stout, MD; Hooth, MJ; Waidyanatha, S. (2021). Internal dose of vanadium in rats following repeated exposure to vanadyl sulfate and sodium orthovanadate via drinking water. Toxicol Appl Pharmacol Pre-Proof: 115395. http://dx.doi.org/10.1016/j.taap.2021.115395

Harvey, JD. (2020). The Bureau of Land Management begins scoping for proposed Gibellini vanadium mine project southeast of Eureka, Nevada. Available online at https://www.blm.gov/press-release/bureau-land-management-begins-scoping-proposed-gibellini-vanadium-mine-project

Henry, RP; Mitchell, PCH; Prue, JE. (1973). Hydrolysis of the oxovanadium(IV) ion and the stability of its complexes with the 1,2-dihydroxybenzenato(2–) ion. J Chem Soc, Dalton Trans (11): 1156-1159. http://dx.doi.org/10.1039/DT9730001156

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

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

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

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

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

Huang, J, enHow; Huang, F; Evans, L, es; Glasauer, S. (2015). Vanadium: Global (bio)geochemistry. Chem Geol 417: 68-89. http://dx.doi.org/10.1016/j.chemgeo.2015.09.019

IOM (Institute of Medicine). (2001). Arsenic, boron, nickel, silicon, and vanadium. In Dietary reference intakes for vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc. Washington, DC: National Academy Press. http://www.nap.edu/openbook.php?record_id=10026&page=502

Jentjens, RL; Jeukendrup, AE. (2002). Effect of acute and short-term administration of vanadyl sulphate on insulin sensitivity in healthy active humans. Int J Sport Nutr Exerc Metab 12: 470-479. http://dx.doi.org/10.1123/ijsnem.12.4.470

Kelley, KD; Scott, CT; Polyak, DE; Kimball, BE. (2017). Vanadium. Critical mineral resources of the United States—Economic and environmental geology and prospects for future supply: U.S. Geological Survey Professional Paper 1802 (pp. U1–U36). Reston, VA: U.S. Geological Survey. http://dx.doi.org/10.3133/pp1802U

Kingsnorth, AN; Lamuraglia, GM; Ross, JS; Malt, RA. (1986). Vanadate supplements and 1,2-dimethylhydrazine induced colon cancer in mice: increased thymidine incorporation without enhanced carcinogenesis. Br J Cancer 53: 683-686.

Komura, A; Hayashi, M; Imanaga, H. (1977). Hydrolytic behavior of oxovanadium(IV) ions. Bull Chem Soc Jpn 50: 2927-2931. http://dx.doi.org/10.1246/bcsj.50.2927

Kowalska, M. (1988). The effect of vanadium on lung collagen content and composition in two successive generations of rats. Toxicol Lett 41: 203-208.

Krachler, M; Prohaska, T; Koellensperger, G; Rossipal, E; Stingeder, G. (2000). Concentrations of selected trace elements in human milk and in infant formulas determined by magnetic sector field inductively coupled plasma-mass spectrometry. Biol Trace Elem Res 76: 97-112. http://dx.doi.org/10.1385/BTER:76:2:97

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

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

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

Moskalyk, RR; Alfantazi, AM. (2003). Processing of vanadium: A review. Miner Eng 16: 793-805. http://dx.doi.org/10.1016/S0892-6875(03)00213-9

Mountain, JT; Delker, LL; Stokinger, HE. (1953). Studies in vanadium toxicology; Reduction in the cystine content of rat hair. AMA Arch Ind Hyg Occup Med 8: 406-411.

Mravcová, A; Jírová, D; Jancí, H; Lener, J. (1993). Effects of orally administered vanadium on the immune system and bone metabolism in experimental animals. Sci Total Environ 134: 663-669. http://dx.doi.org/10.1016/S0048-9697(05)80069-5

Mutlu, E; Cristy, T; Graves, SW; Hooth, MJ; Waidyanatha, S. (2017). Characterization of aqueous formulations of tetra- and pentavalent forms of vanadium in support of test article selection in toxicology studies. Environ Sci Pollut Res Int 24: 405-416. http://dx.doi.org/10.1007/s11356-016-7803-x

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

Nielsen, FH. (1995). Vanadium in mammalian physiology and nutrition [Review]. Met Ions Biol Syst 31: 543-573.

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). (2008). Chemical information review document for oral exposure to tetravalent and pentavalent vanadium compounds: Supporting nomination for toxicological evaluation by the National Toxicology Program [NTP]. Research Triangle Park, NC. https://ntp.niehs.nih.gov/ntp/htdocs/chem_background/exsumpdf/niehs_vanadium_compounds_508.pdf

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

Pal, RP; Mani, V; Tripathi, D; Kumar, R; Kewalramani, NJ. (2018). Influence of Feeding Inorganic Vanadium on Growth Performance, Endocrine Variables and Biomarkers of Bone Health in

Crossbred Calves. Biol Trace Elem Res 182: 248-256. http://dx.doi.org/10.1007/s12011-017-1095-y

Pessoa, JC; Garribba, E; Santos, MFA; Santos-Silva, T. (2015). Vanadium and proteins: Uptake, transport, structure, activity and function. Coord Chem Rev 301-302: 49-86. http://dx.doi.org/10.1016/j.ccr.2015.03.016

Polyak, DE. (2020). Mineral commodity summaries 2020: Vanadium data sheet (pp. 180-181). Reston, VA: U.S. Geological Survey.

Rainńska, E; Biziuk, M; Sarbu, C; Szczepaniak, K; Frontasyeva, MV; Culicov, O; Bode, P; Astel, A. (2005). Assessment of phosphatic fertilizer production impact on occupational staff based on NAA of hair, nails, and inhald particles. J Environ Sci Health A Tox Hazard Subst Environ Eng 40: 2137-2152. http://dx.doi.org/10.1080/10934520500234635

Rehder, D. (2015). The role of vanadium in biology. Metallomics 7: 730-742. http://dx.doi.org/10.1039/c4mt00304g

Roberts, GK; Stout, MD; Sayers, B; Fallacara, DM; Hejtmancik, MR; Waidyanatha, S; Hooth, MJ. (2016). 14-Day Toxicity Studies of Tetravalent and Pentavalent Vanadium Compounds in Harlan Sprague Dawley Rats and B6C3F1/N Mice via Drinking Water Exposure. Toxicology Reports 3: 531-538. http://dx.doi.org/10.1016/j.toxrep.2016.05.001

Safe Drinking Water Act. Title XIV of the Public Health Service Act Safety of Public Water Systems (Safe Drinking Water Act) as amended through Pub. L. No. 116-92, (2019). https://www.epa.gov/sdwa/title-xiv-public-health-service-act-safety-public-water-systems-safe-drinking-water-act

Schlesinger, WH; Klein, EM; Vengosh, A. (2017). Global biogeochemical cycle of vanadium [Review]. Proc Natl Acad Sci USA 114: E11092-E11100. http://dx.doi.org/10.1073/pnas.1715500114

Schroeder, HA; Mitchener, M; Nason, AP. (1970). Zirconium, niobium, antimony, vanadium and lead in rats: life term studies. J Nutr 100: 59-68. http://dx.doi.org/10.1093/jn/100.1.59

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

Ścibior, A; Pietrzyk, Ł; Plewa, Z; Skiba, A. (2020). Vanadium: Risks and possible benefits in the light of a comprehensive overview of its pharmacotoxicological mechanisms and multi-applications with a summary of further research trends. J Trace Elem Med Biol 61: 126508. http://dx.doi.org/10.1016/j.jtemb.2020.126508

Shah, SZH; Naveed, AK; Rashid, A. (2016). Effects of oral vanadium on glycaemic and lipid profile in rats. J Pak Med Assoc 66: 1592-1596.

Smith, DM; Pickering, RM; Lewith, GT. (2008). A systematic review of vanadium oral supplements for glycaemic control in type 2 diabetes mellitus. QJM 101: 351-358. http://dx.doi.org/10.1093/qjmed/hcn003

Smith, MT; Guyton, KZ; Gibbons, CF; Fritz, JM; Portier, CJ; Rusyn, I; DeMarini, DM; Caldwell, JC; Kavlock, RJ; Lambert, PF; Hecht, SS; Bucher, JR; Stewart, BW; Baan, RA; Cogliano, VJ; Straif, K. (2016). Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis [Review]. Environ Health Perspect 124: 713-721. http://dx.doi.org/10.1289/ehp.1509912

Somerville, J; Davies, B. (1962). Effect of vanadium on serum cholesterol. Am Heart J 64: 54-56. http://dx.doi.org/10.1016/0002-8703(62)90091-1

Steffen, RP; Pamnani, MB; Clough, DL; Huot, SJ; Muldoon, SM; Haddy, FJ. (1981). Effect of prolonged dietary administration of vanadate on blood pressure in the rat. Hypertension 3: I173-I178. http://dx.doi.org/10.1161/01.HYP.3.3_Pt_2.I173

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

Stokinger, HE. (1981). The Metals: Vanadium, V. In GD Clayton; FE Clayton (Eds.), Patty's industrial hygiene and toxicology: Volume 2A: Toxicology (3rd rev ed., pp. 2013-2033). New York, NY: John Wiley and Sons.

Stokinger, HE; Wagner, WD; Mountain, JT; Stacksill, FR; Dobrogorski, OJ; Keenan, RG. (1953). Unpublished results [Cited in Patty's Industrial Hygiene and Toxicology, 3rd ed., 1981]. Cincinnati, OH: Division of Occupational Health.

Sun, L; Shi, DJ; Gao, XC; Mi, SY; Yu, Y; Han, Q. (2014). The protective effect of vanadium against diabetic cataracts in diabetic rat model. Biol Trace Elem Res 158: 219-223. http://dx.doi.org/10.1007/s12011-014-9925-7

Susić, D; Kentera, D. (1986). Effect of chronic vanadate administration on pulmonary circulation in the rat. Respiration 49: 68-72. http://dx.doi.org/10.1159/000194861

Susić, D; Kentera, D. (1988). Dependence of the hypertensive effect of chronic vanadate administration on renal excretory function in the rat. J Hypertens 6: 199-204.

TCEQ (Texas Commission on Environmental Quality). (2012). TCEQ guidelines to develop toxicity factors. (Revised RG-442). Austin, TX. http://www.tceq.texas.gov/publications/rg/rg-442.html

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

Thompson, KH; Orvig, C. (2006). Vanadium in diabetes: 100 years from Phase 0 to Phase I. J Inorg Biochem 100: 1925-1935. http://dx.doi.org/10.1016/j.jinorgbio.2006.08.016

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

Treviño, S; Díaz, A; Sánchez-Lara, E; Sanchez-Gaytan, BL; Perez-Aguilar, JM; González-Vergara, E. (2019). Vanadium in Biological Action: Chemical, Pharmacological Aspects, and Metabolic Implications in Diabetes Mellitus [Review]. Biol Trace Elem Res 188: 68-98. http://dx.doi.org/10.1007/s12011-018-1540-6

Tripathi, D; Mani, V; Pal, RP. (2018). Effect of vanadium supplementation on production performance, nutrient utilization, plasma mineral concentration, and mineral balance in lactating goats. Biol Trace Elem Res 188: 412-418. http://dx.doi.org/10.1007/s12011-018-1426-7

Turner, PV; Pekow, C; Vasbinder, MA; Brabb, T. (2011). Administration of substances to laboratory animals: Equipment considerations, vehicle selection, and solute preparation [Editorial]. J Am Assoc Lab Anim Sci 50: 614-627.

U.S. Department of the Interior. (2018). Final list of critical minerals 2018. Fed Reg 83: 23295-23296.

U.S. EPA (U.S. Environmental Protection Agency). Health effects assessment summary tables for superfund (HEAST): Sodium Metavanadate (CASRN 13718-26-8). https://epa-heast.ornl.gov/heast.php

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

U.S. EPA (U.S. Environmental Protection Agency). (1987). Integrated Risk Information System (IRIS): Vanadium pentoxide (CASRN 1314-62-1) [EPA Report]. Washington, DC. https://cfpub.epa.gov/ncea/iris/iris_documents/documents/subst/0125_summary.pdf#nameddest=rfd

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

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

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

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

U.S. EPA (U.S. Environmental Protection Agency). (1997). 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] (pp. 1-89). (EPA/630/R-95/001F). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. http://www.epa.gov/risk/guidelines-neurotoxicity-risk-assessment

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

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

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

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

U.S. EPA (U.S. Environmental Protection Agency). (2009). Provisional peer-reviewed toxicity values for vanadium and its soluble inorganic compounds other than vanadium pentoxide (CASRN 7440-62-2 and others): Derivation of subchronic and chronic oral RfDs [EPA Report]. (EPA/690/R-09/070F). Cincinnati, OH. https://cfpub.epa.gov/ncea/pprtv/documents/Vanadium.pdf

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

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

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

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

U.S. EPA (U.S. Environmental Protection Agency). (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). (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). (2020a). IRIS assessment plan for oral exposure to vanadium and compounds (scoping and problem formulation materials). (EPA/635/R-20/112). Washington, DC. https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=348792

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

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

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

Vilas Boas, LV; Costa Pessoa, J. (1987). Vanadium. In G Wilkinson; RD Gillard; JA McCleverty (Eds.), Comprehensive coordination chemistry: The synthesis, reactions, properties & applications of coordination compounds Vol 3, Main group and early transition elements. New York, NY: Pergamon Press.

APPENDICES

APPENDIX A. SURVEY OF EXISTING VANADIUM ORAL TOXICITY VALUES

Table A-1 lists websites which were searched for relevant human health reference values for vanadium and compounds, along with indications of the results of the search. In addition to these sources, the ToxValDB on EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard/chemical_lists/TOXVAL_V5) was also searched for additional reference values that were not captured by other sources. When values were identified for vanadium, they are shown in Figure 5 and described in Table 2 if details were provided on how the values were derived. When values were identified from sources that did not provide derivation details, they are described in Table 3 but not shown in Figure 5. The values in these tables are current as of May 2020.

Table A-1. Sources searched for human health reference values for vanadium

Source[a]	Search results	Query and/or link
ATSDR	See Table 2	http://www.atsdr.cdc.gov/toxprofiles/index.asp
		https://www.atsdr.cdc.gov/mrls/mrllist.asp
CalEPA	No values found	http://www.oehha.ca.gov/tcdb/index.asp
		https://www.arb.ca.gov/toxics/healthval/healthval.htm
DWSHA	No values found	https://www.epa.gov/sites/production/files/2018-03/documents/dwtable2018.pdf
Health Canada	No values found	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	See Table 2	http://epa-heast.ornl.gov/heast.php
		https://nepis.epa.gov/Exe/ZyPDF.cgi/2000O0GZ.PDF?Dockey=2000O0GZ.PDF
IRIS	See Table 2	http://www.epa.gov/iris/
ITER	2 records found; no unique values	https://toxnet.nlm.nih.gov/newtoxnet/iter.htm

Table A-1. Sources searched for human health reference values for vanadium (continued)

Source[a]	Search results	Query and/or link
MDH	No values found	https://www.health.state.mn.us/communities/environment/risk/guidance/gw/table.html
MI EGLE	PPRTV value was adopted as state value (see Table 2)	https://www.michigan.gov/documents/deq/deq-rrd-chem-CleanupCriteriaTSD_527410_7.pdf
NHMRC	No values found	https://www.nhmrc.gov.au/about-us/publications/australian-drinking-water-guidelines
NY DEC	No values found	https://www.dec.ny.gov/docs/remediation_hudson_pdf/techsuppdoc.pdf
OPP	No search results found	https://iaspub.epa.gov/apex/pesticides/f?p=chemicalsearch:1
PPRTV	See Table 2	https://www.epa.gov/pprtv/provisional-peer-reviewed-toxicity-values-pprtvs-assessments
RIVM	See Table 2	https://www.rivm.nl/bibliotheek/rapporten/711701092.pdf
	No values found	https://www.rivm.nl/bibliotheek/rapporten/711701025.pdf
TCEQ	See Table 3	https://www.tceq.texas.gov/remediation/trrp/trrppcls.html
WHO	Environmental Health Criteria document available; no reference values found	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; ITER = International Toxicity Estimates for Risk; 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 B. ELECTRONIC DATABASE SEARCH STRATEGIES

Table B-1. Database search strategy

Source	Search strategy	Number of records
ATSDR *Toxicological Profile for Vanadium* (2012)	References pulled from ATSDR document	363
WOS 3/28/2019 3/9/2020	((TS="Ammonium metavanadate" OR TS="Ammonium monovanadate" OR TS="Ammonium trioxovanadate" OR TS="Monosodium trioxovanadate" OR TS="Oxosulfatovanadium pentahydrate" OR TS="Sodium metavanadate" OR TS="Sodium o-vanadate" OR TS="Sodium orthovanadate" OR TS="Sodium pervanadate" OR TS="Sodium tetraoxovanadate" OR TS="Sodium trioxovanadate" OR TS="Sodium vanadate" OR TS="Trisodium orthovanadate" OR TS="Trisodium tetraoxovanadate" OR TS="Trisodium vanadate" OR TS="Vanadic sulfate" OR TS="vanadium" OR TS="Vanadyl sulfate" OR TS="Vanadic" OR TS="Vanadin" OR TS="sodium peroxyvanadate" OR TS="Vanadyl sulfate pentahydrate" OR TS="Ammonium vanadate" OR TS="Divanadium trioxide" OR TS="Sodium hexavanadate") AND PY=(2010-2019)) ((TS="Sodium tetravanadate" OR TS="Sodium vanadite" OR TS="Sulfovanadic acid" OR TS="vanadium salt" OR TS="Tetrachlorovanadium" OR TS="Trichlorooxo vanadium" OR TS="Trichlorooxovanadium" OR TS="Trichlorooxovanadium oxide" OR TS="Vanadic acid" OR TS="Vanadic oxide" OR TS="Vanadious" OR TS="Vanadosulfuric acid" OR TS="Vanadyl chloride" OR TS="Vanadyl trichloride" OR TS="Divanadium pentaoxide" OR TS="Divanadium pentoxide" OR TS="Vanadic acid anhydride" OR TS="Vanadic anhydride" OR TS="Vanadin(V) oxide" OR TS="Vanadium dust" OR TS="Vanadium fume" OR TS="Vanadium oxide" OR TS="Vanadium pentaoxide" OR TS="Vanadium pentoxide") AND PY=(2010-2019)) ((TS="Vanadium" AND (TS="chloride" OR TS="dichloride" OR TS="oxide" OR TS="oxychloride" OR TS="oxytrichloride" OR TS="sesquioxide" OR TS="sulfate" OR TS="sulphate" OR TS="tetrachloride" OR TS="trichloride" OR TS="trioxide")) AND PY=2010-2019)	24,878

Table B-1. Database search strategy (continued)

Source	Search strategy	Number of records
PUBMED 3/28/2019 3/9/2020	(((7440-62-2[rn] OR 00J9J9XKDE[rn] OR 27774-13-6[rn] OR 6DU9Y533FA[rn] OR 13718-26-8[rn] OR 13721-39-6[rn] OR 7803-55-6[rn] OR FL85PX638G[rn] OR 12439-96-2[rn] OR "Ammonium metavanadate"[tw] OR "Ammonium monovanadate"[tw] OR "Ammonium trioxovanadate"[tw] OR "Monosodium trioxovanadate"[tw] OR "Oxosulfatovanadium pentahydrate"[tw] OR "Sodium metavanadate"[tw] OR "Sodium o-vanadate"[tw] OR "Sodium orthovanadate"[tw] OR "Sodium pervanadate"[tw] OR "Sodium tetraoxovanadate"[tw] OR "Sodium trioxovanadate"[tw] OR "Sodium vanadate"[tw] OR "Trisodium orthovanadate"[tw] OR "Trisodium tetraoxovanadate"[tw] OR "Trisodium vanadate"[tw] OR "Vanadic sulfate"[tw] OR vanadium[tw] OR "Vanadyl sulfate"[tw] OR Vanadic[tw] OR Vanadin[tw] OR "sodium peroxyvanadate"[tw] OR "Vanadyl sulfate pentahydrate"[tw] OR 16785-81-2[rn] OR 12436-28-1[rn] OR 12058-74-1[rn] OR 64082-34-4[rn] OR 10580-52-6[rn] OR 7718-98-1[rn] OR 1314-34-7[rn] OR 7632-51-1[rn] OR 11115-67-6[rn] OR 7727-18-6[rn] OR "Ammonium vanadate"[tw] OR "Divanadium trioxide"[tw] OR "Sodium hexavanadate"[tw] OR "Sodium tetravanadate"[tw] OR "Sodium vanadite"[tw] OR "Sulfovanadic acid"[tw] OR "vanadium salt"[tw] OR Tetrachlorovanadium[tw] OR "Trichlorooxo vanadium"[tw] OR Trichlorooxovanadium[tw] OR "Trichlorooxovanadium oxide"[tw] OR "Vanadic acid"[tw] OR "Vanadic oxide"[tw] OR Vanadious[tw] OR "Vanadosulfuric acid"[tw] OR "Vanadyl chloride"[tw] OR "Vanadyl trichloride"[tw] OR 1314-62-1[rn] OR "Divanadium pentaoxide"[tw] OR "Divanadium pentoxide"[tw] OR "Vanadic acid anhydride"[tw] OR "Vanadic anhydride"[tw] OR "Vanadin(V) oxide"[tw] OR "Vanadium dust"[tw] OR "Vanadium fume"[tw] OR "Vanadium oxide"[tw] OR "Vanadium pentaoxide"[tw] OR "Vanadium pentoxide"[tw]) OR (Vanadium[tw] AND (chloride[tw] OR dichloride[tw] OR oxide[tw] OR oxychloride[tw] OR oxytrichloride[tw] OR sesquioxide[tw] OR sulfate[tw] OR sulphate[tw] OR tetrachloride[tw] OR trichloride[tw] OR trioxide[tw]))) AND ("2010"[PDAT] : "3000"[PDAT]))	4,888

Table B-1. Database search strategy (continued)

Source	Search strategy	Number of records
TOXLINE 3/28/2019	@SYN0+@AND+@OR+(@TERM+@rn+7440-62-2+@TERM+@rn+27774-13-6+@TERM+@rn+13718-26-8+@TERM+@rn+13721-39-6+@TERM+@rn+7803-55-6+@TERM+@rn+12439-96-2+@TERM+@rn+16785-81-2+@TERM+@rn+12436-28-1+@TERM+@rn+12058-74-1+@TERM+@rn+64082-34-4+@TERM+@rn+10580-52-6+@TERM+@rn+7718-98-1+@TERM+@rn+1314-34-7+@TERM+@rn+7632-51-1+@TERM+@rn+11115-67-6+@TERM+@rn+7727-18-6+@TERM+@rn+1314-62-1)+@RANGE+yr+2010+2019+@NOT+@org+pubmed+pubdart+nih @SYN0+@AND+@OR+(FL85PX638G+6DU9Y533FA+00J9J9XKDE+"Ammonium+metavanadate"+"Ammonium+monovanadate"+"Ammonium+trioxovanadate"+"Monosodium+trioxovanadate"+"Oxosulfatovanadium+pentahydrate"+"Sodium+metavanadate"+"Sodium+o-vanadate"+"Sodium+orthovanadate"+"Sodium+pervanadate"+"Sodium+tetraoxovanadate"+"Sodium+trioxovanadate"+"Sodium+vanadate"+"Trisodium+orthovanadate"+"Trisodium+tetraoxovanadate"+"Trisodium+vanadate"+"Vanadic+sulfate"+vanadium+"Vanadyl+sulfate"+Vanadic+Vanadin+"sodium+peroxyvanadate"+"Vanadyl+sulfate+pentahydrate"+"Ammonium+vanadate"+"Divanadium+trioxide"+"Sodium+hexavanadate"+"Sodium+tetravanadate"+"Sodium+vanadite"+"Sulfovanadic+acid"+"vanadium+salt"+"Trichlorooxo+vanadium"+Tetrachlorovanadium+Trichlorooxovanadium+"Trichlorooxovanadium+oxide"+"Vanadic+acid"+"Vanadium+dust"+"Vanadium+fume"+"Vanadium+oxide"+"Vanadium+pentaoxide"+"Vanadium+pentoxide"+"Vanadic+oxide"+Vanadious+"Vanadosulfuric+acid"+"Vanadyl+chloride"+"Vanadyl+trichloride"+"Divanadium+pentaoxide"+"Divanadium+pentoxide"+"Vanadic+acid+anhydride"+"Vanadic+anhydride"+"Vanadin+V+oxide")+@RANGE+yr+2010+2019+@NOT+@org+pubmed+pubdart+nih @SYN0+@AND+vanadium+@OR+(chloride+dichloride+oxide+oxychloride+oxytrichloride+sesquioxide+sulfate+sulphate+tetrachloride+trichloride+trioxide)+@RANGE+yr+2010+2019+@NOT+@org+pubmed+pubdart+nih	15
TOTAL	25,988 unique items were discovered using this search strategy.	25,988

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

Review of reference lists from existing assessments (final or publicly available draft), journal reviews articles and studies considered relevant to the PECO criteria on the basis of 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 not identified from the other sources are formatted in a RIS file format, imported into DistillerSR, annotated with respect to source, and screened as outlined in Section 4.4. For tracking assessments or reviews, the name of the source citation and the number of records imported into DistillerSR are noted. The reference list of any study included in the literature inventory was reviewed manually to identify titles that appeared relevant to the PECO criteria. These citations are tracked in a spreadsheet, compared against the literature base to determine if they were unique to the project, and then added to DistillerSR to be screened at the title and abstract stage for PECO relevance.

EPA's Toxicity Values Database (ToxValDB) (searched via EPA's CompTox Chemicals Dashboard)

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

European Chemicals Agency (ECHA)

The ECHA registered substances database is searched using the CASRN. The registration dossier associated with the CASRN is 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 are downloaded in PDF (Portable Document Format), including all nested reports having unique URLs. In addition, the data are extracted from each dossier page and used to populate an Excel tracking sheet. Extracted fields include data from the general information page regarding the registration type and publication dates, and on a typical study summary page the primary fields reported in the administrative data, data source, and effect

levels sections. Each study summary results in more than one row in the tracking sheet if more than one data source or effect level is reported.

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

EPA ChemView

EPA's ChemView database (https://chemview.epa.gov/chemview) is searched using the CASRN. The prepopulated CASRN match and the "Information Submitted to EPA" output option filter selected before generating results. If results are available, the square-shaped icon under the "Data Submitted to EPA" column is selected, and the following records are 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 ecotoxicological and physical and chemical property entries are excluded. When results are available, extractors navigate into each record until a substantial risk report link is identified and saved as a PDF file. If the report cannot be saved, due to file corruption or broken links, the record is excluded during full-text review as "unable to obtain record." Most substantial risk reports contained multiple document IDs, so citations are 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 is generated, the study is moved forward to DistillerSR where it is screened according to PECO and supplemental material criteria.

NTP Chemical Effects in Biological Systems

This database is searched using the CASRN (https://manticore.niehs.nih.gov/cebssearch). All non-NTP data are 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.

OECD Echem Portal

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

- OECD HPV
- OECD SIDS IUCLID
- SIDS United Nations Environment Programme (UNEP)

ECOTOX Database

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

ToxCast or Tox21 high throughput screening information (searched via EPA's CompTox Chemicals Dashboard)

EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard/) is searched using CASRN to access high throughput screening (HTS) data from ToxCast or Tox21. For each chemical, the "Bioactivity" section is selected and the availability of ToxCast/Tox21 HTS data for active and inactive assays is examined in the "TOXCAST: Summary" tab.