EPA/635/R-18/155
IRIS Assessment Protocol
www.epa.gov/iris

Systematic Review Protocol for the Hexavalent Chromium IRIS Assessment (Preliminary Assessment Materials)

[CASRN 18540-29-9]

March 2019

Integrated Risk Information System
National Center for 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

ACGIH	American Conference of Governmental Industrial Hygienists
ACToR	Aggregated Computational Toxicology Resource
ADME	absorption, distribution, metabolism, and excretion
AIHA	American Industrial Hygiene Association
ATSDR	Agency for Toxic Substances and Disease Registry
BMD	benchmark dose
BMR	benchmark response
CAA	Clean Air Act
CalEPA	California Environmental Protection Agency
CASRN	Chemical Abstracts Service registry number
CCA	chromated copper arsenate
CCID	Chemical Classification Information Database
CCR	Canadian Categorization Results
CCRMP	Coordinated Chemicals Risk Management Programme Publications
CDAT	Chemical Data Access Tool
CEPA	Canadian Environmental Protection Act
CESAR	Canada's Existing Substances Assessment Repository
CHRIP	Chemical Risk Information Platform
CPSC	Consumer Product Safety Commission
Cr(III)	trivalent chromium
Cr(VI)	hexavalent chromium
CrO_4^{2-}	chromate
$Cr_2O_7^{2-}$	dichromate
DoCTER	Document Classification and Topic Extraction Resource
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECHA	European Chemicals Agency
EnviChem	Data Bank of Environmental Properties of Chemicals
EO	Executive Order
EPA	Environmental Protection Agency
ERPG	Emergency Response Planning Guidelines
ESIS	European Chemical Substances Information System
ESR	Existing Substances Regulation
FDA	Food and Drug Administration
GHS-J	Globally Harmonized System-Japan
GI	gastrointestinal
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HAP	hazardous air pollutant
HAWC	Health Assessment Workplace Collaborative
HERO	Health and Environmental Research Online
HPV	high production volume
HPVIS	High Production Volume Information System
HSDB	Hazardous Substances Data Bank
HSNO	Hazardous Substances and New Organisms
IAP	IRIS Assessment Plan
IARC	International Agency for Research on Cancer
IRIS	Integrated Risk Information System
IUCLID	International Uniform Chemical Information Database
IUR	inhalation unit risk
J-CHECK	Japan CHEmicals Collaborative Knowledge
JECDB	Japan Existing Chemical Data Base
LOAEL	lowest-observed-adverse-effect level
LOEL	lowest-observed-effect level
MOA	mode of action
NAP	National Academies Press
NATA	National-Scale Air Toxics Assessment
NCEA	National Center for Environmental Assessment
NCI	National Cancer Institute
NICNAS	National Industrial Chemicals Notification and Assessment Scheme
NIEHS	National Institute for Environmental Health Sciences
NIOSH	National Institute for Occupational Safety and Health
NIOSHTIC	National Institute for Occupational Safety and Health Technical Information Center
NMD	normalized mean difference
NOEL	no-observed-effect level
NSCEP	National Service Center for Environmental Publications
NTP	National Toxicology Program
OECD	Organisation for Economic Cooperation and Development
OEHHA	Office of Environmental Health Hazard Assessment
OPP	Office of Pesticide Programs

ORD	Office of Research and Development
OSF	oral slope factor
OSHA	Occupational Safety and Health Administration
PBPK	physiologically based pharmacokinetic
PEC	priority existing chemical
PECO	populations, exposures, comparators, and outcomes
PK	pharmacokinetic
POD	point of departure
RED	registration eligibility decision
REL	reference exposure level
RfC	reference concentration
RfD	reference dose
ROBINS-I	Risk of Bias in Nonrandomized Studies of Interventions
RoC	Report on Carcinogens
RTECS	Registry of Toxic Effects of Chemical Substances
SIDS	Screening Information Data Set
SRS	Substance Registry Services
TCEQ	Texas Commission on Environmental Quality
TSCA	Toxic Substances Control Act
TSCATS	Toxic Substances Control Act Test Submissions
UK	United Kingdom
UNEP	United Nations Environment Programme
WEEL	Workplace Environmental Exposure Level
WOS	Web of Science

AUTHORS | CONTRIBUTORS

Assessment Team

Catherine Gibbons (co-Assessment Manager)	U.S. EPA/ORD/NCEA
Alan Sasso (co-Assessment Manager)	
Xabier Arzuaga	
Johanna Congleton	
Barbara Glenn	
Rebecca Nachman	
Larissa Pardo (formerly with EPA)	
Elizabeth Radke	
Paul Reinhart	
Amina Wilkins	
Erin Yost	
David Lehmann	U.S. EPA/ORD/NHEERL

Contributors

Thomas Bateson	U.S. EPA/ORD/NCEA
Vincent Cogliano	
Laura Dishaw	
Leonid Kopylev	
Roman Mezencev	
Paul Schlosser	

Executive Direction

Tina Bahadori	NCEA Center Director
Mary Ross	NCEA Deputy Center Director
Emma Lavoie	NCEA Assistant Center Director for Scientific Support
Samantha Jones	NCEA Associate Director for Health (acting)
Kristina Thayer	NCEA/IRIS Division Director
James Avery	NCEA/IRIS Deputy Division Director (acting)
Andrew Kraft	NCEA/IRIS Associate Director for Science (acting)
Susan Rieth	NCEA/IRIS Quantitative Modeling Branch Chief

Production Team

Vicki Soto	Project Management Team
Dahnish Shams	Project Management Team
Maureen Johnson	NCEA Webmaster

Contractor Support

Kim Osborn	ICF International
Robyn Blain	
Michelle Cawley	
Alessandria Schumacher	

1. INTRODUCTION

The Integrated Risk Information System (IRIS) Program is undertaking a reassessment of the health effects of hexavalent chromium (Cr[VI]). Significant new epidemiologic and experimental animal toxicity information for Cr(VI) has become available since EPA's IRIS assessment for Cr(VI) was posted in 1998, including updates of occupational cohort studies (Proctor et al., 2016; Gibb et al., 2015) and a National Toxicology Program (NTP) bioassay that reported increased incidences of tumors in rats and mice exposed to Cr(VI) in drinking water (NTP, 2008). The dose-response information from epidemiologic and experimental animal studies published since 1998 could result in changes to current toxicity values. Cr(VI) was included on the December 2015 IRIS Program multiyear agenda (https://www.epa.gov/iris/iris-agenda) as a chemical having high priority for assessment development. It was also included in the December 2018 IRIS Program Outlook that provides an updated outlook of IRIS program activities (https://www.epa.gov/sites/production/files/2018-12/documents/iris_program_outlook_december_2018.pdf). Given the widespread exposure to Cr(VI) and the availability of studies that provide significant new health effects information, the IRIS Program is developing an updated assessment of Cr(VI).

Preliminary materials for the Cr(VI) reassessment were released to the public in April and August 2014, and public meetings were held in June and October 2014 to seek input regarding the Cr(VI) assessment from the scientific community and interested parties (U.S. EPA, 2014b, c). The preliminary materials included a summary of the IRIS Program's scoping and problem formulation conclusions, information on the approaches used to identify pertinent literature, results of the literature search, approaches for selection of studies for hazard identification, and presentation of studies eligible for study evaluation in evidence tables and exposure-response arrays. A preliminary summary of toxicokinetic and mechanistic studies pertinent to the assessment was also presented.

The protocol is a new document adopted by the IRIS Program as part of its full implementation of systematic review [see presentation materials for the NAS Workshop "Review of Advances Made to the IRIS Process" (Bahadori and Thayer, 2018) and the NASEM (2018) report *Progress Toward Transforming the Integrated Risk Information System*]. The chemical-specific protocol describes *how* the assessment will be conducted, while the IRIS Assessment Plan (IAP), typically released early in the assessment process, describes *what* the assessment plans to cover based on scoping and problem formulation. As noted above, scoping and problem formulation documents were previously released for public comment; this protocol summarizes and updates those earlier materials (e.g., see Sections 1–4). Because development of the chromium assessment began before the introduction of these early stage documents to the IRIS process, EPA is

retroactively releasing the protocol, which presents the methods for conducting the systematic review and dose-response analysis, to provide similar public engagement steps and documentation as other assessments that started more recently. This protocol also includes specific aims and populations, exposures, comparators, and outcomes (PECO) criteria that were not a part of the 2014 preliminary materials but are now a part of IRIS Systematic Review materials. The IRIS Program posts assessment protocols on its website and in the Zenodo repository (https://zenodo.org/). Public comments will be considered as part of developing the draft assessment. This protocol documents the studies identified during the initial literature searches (U.S. EPA, 2014b, c) and updates to those literature searches. Additional literature search updates will be posted to the IRIS website when they are available.

2. SCOPING AND INITIAL PROBLEM FORMULATION SUMMARY

2.1. BACKGROUND

Elemental chromium is a Group 6 transition metal (atomic number 24 and atomic weight 52) on the periodic table, existing in nature in the form of various oxide minerals (Anger et al., 2005). It is present in the Earth's crust and has oxidation states ranging from −2 to +6, with the +3 (trivalent) and +6 (hexavalent) states being the most common (Losi et al., 1994). Chromium in the environment can originate from both natural and anthropogenic sources. Atmospheric releases of chromium from natural and anthropogenic sources are comparable in magnitude, while soil releases are mostly anthropogenic, and all water releases are anthropogenic (USGS, 1995). Conversion of Cr(VI) to Cr(III) may occur in the environment under reducing conditions (by ferrous iron, sulfides, and organic matter), while conversion of Cr(III) to Cr(VI) may occur under oxidizing conditions [by manganese oxide minerals; (Hausladen and Fendorf, 2017; McClain et al., 2017; Jardine et al., 2011; Cummings et al., 2007; Oze et al., 2007; Oze et al., 2004; Kim and Dixon, 2002; Fendorf et al., 2000; Fendorf, 1995)]. Most Cr(III) compounds are insoluble in water and immobile in soils (which helps inhibit oxidation), while Cr(VI) compounds are readily soluble in water and highly mobile and bioavailable (Fendorf et al., 2000; Fendorf, 1995). In addition to being stabilized by low solubility and mobility, Cr(III) compounds are more thermodynamically stable than Cr(VI) compounds under most pH values encountered in the environment (Fendorf, 1995).

Cr(VI) compounds are used for corrosion inhibition (including within water-cooling systems), pigment manufacturing (including textile dyeing and printing inks), metal finishing (chrome plating/electroplating), stainless steel production, leather manufacturing (leather tanning), refractories (linings for high-temperature industrial furnaces), drilling muds, pyrotechnics, chemical synthesis, and plastics (NIOSH, 2013b; NTP, 2011). Chromium compounds have been used in wood preservatives [as chromated copper arsenate (CCA) in pressure-treated wood; (ATSDR, 2012; Barnhart, 1997)]; however, this use began to decline in 2003 due to a voluntary phaseout of all residential uses of CCA pressure-treated wood (Bedinger, 2015; NTP, 2011).

Occupational exposures to Cr(VI) occur primarily from inhalation or dermal contact (NIOSH, 2013b), while general population exposures occur by inhalation of ambient air and ingestion of food and drinking water (NTP, 2011). Dermal exposure may also occur from using consumer products that contain chromium, such as some metals and wood or leather treated with chromium-containing compounds (ATSDR, 2012; NTP, 2011). According to data collected between 2013 and 2015 under EPA's Third Unregulated Contaminant Monitoring Rule (UCMR3), Cr(VI) has

been reported above the minimum reporting limit (0.03 µg/L) by approximately 90% of public water systems in the United States (U.S. EPA, 2014d). Ambient air concentrations of Cr(VI) in the United States typically range from 0.01 to 0.05 ng/m^3 (U.S. EPA, 2016) but have been measured at values above 1 ng/m^3 in urban and industrial areas (Oregon DEQ, 2016; Huang et al., 2014; CalEPA, 2004, 2003). Cr(VI) concentrations measured in air downwind of industrial facilities emitting Cr(VI) (such as chrome platers) have been found to be highly correlated with concentrations measured at the facilities (OAQPS, 2012; CalEPA, 2004, 2003).

2.1.1. Previous IRIS Assessment

EPA's 1998 IRIS assessment classified Cr(VI) as "Group A—known human carcinogen by the inhalation route of exposure" based on evidence of a causal relationship between inhalation of Cr(VI) and increased incidence of lung cancer in humans. An inhalation unit risk (IUR) for Cr(VI) of 1.2×10^{-2} per µg/m^3 was calculated based on increased incidence of lung cancer in chromate workers (Mancuso, 1997, 1975). The 1998 assessment concluded that the carcinogenicity of Cr(VI) "by the oral route of exposure cannot be determined and is classified as Group D." Accordingly, a cancer slope factor for ingested Cr(VI) was not derived.

EPA's 1998 IRIS assessment derived two inhalation reference concentrations (RfCs) for noncancer effects. An RfC of 8×10^{-3} µg/m^3 was derived based on nasal effects observed in an epidemiologic study of workers in chrome plating plants (Lindberg and Hedenstierna, 1983), and was specific to chromic acid mists and dissolved Cr(VI) aerosols. An additional RfC of 0.1 µg/m^3 was derived based on respiratory tract effects observed in subchronic duration rat studies (Malsch et al., 1994; Glaser et al., 1990), and was specific to Cr(VI) particulates. EPA's 1998 IRIS assessment also derived an oral reference dose (RfD) of 3×10^{-3} mg/kg-day for noncancer effects based on a no-observed-adverse-effect level (NOAEL) reported in a 1-year drinking water study in rats (MacKenzie et al., 1958).

2.2. SCOPING SUMMARY

During scoping, the IRIS Program met with EPA program and regional offices that had interest in an IRIS assessment for Cr(VI) to discuss specific assessment needs. As discussed in the April 2014 preliminary materials document (U.S. EPA, 2014b), the scope of the IRIS assessment will be limited to potential health effects by the inhalation and oral routes of exposure. EPA's Office of Pesticide Programs (OPP) previously evaluated the dermal exposure pathway in its reregistration eligibility decision (RED) for CCA pesticides (U.S. EPA, 2008c), and no priority needs related to dermal exposure were identified by other EPA program and regional offices. Table 1 provides a summary of EPA offices, programs, and regions that have interest in the assessment and what their specific needs are.

Table 1. EPA program and regional office interest in a reassessment of Cr(VI)

EPA program or regional office	Oral	Inh.	Statutes/regulations and anticipated uses/interest
OLEM EPA Regions 1–10	✓	✓	CERCLA and RCRA Cr(VI) has been identified as a contaminant of concern at numerous contaminated waste sites, including more than 100 NPL sites. CERCLA authorizes EPA to conduct short- or long-term cleanups at Superfund sites and later recover cleanup costs from potentially responsible parties under Section 107. Cr(VI) toxicological information may be used to make risk determinations for response actions (e.g., short-term removals, long-term remedial response actions, RCRA Corrective Action).
OW	✓		SDWA Currently, the EPA drinking water standard of 0.1 mg/L is for total chromium (Federal Register, 2010). The SDWA requires EPA to periodically review the NPDWR for each contaminant and revise the regulation, if appropriate. Cr(VI) toxicological information may be used to inform risk determinations associated with revisiting the NPDWR. Chromium is listed under the NPDWR.

CERCLA = Comprehensive Environmental Response, Compensation, and Liability Act; RCRA = Resource Conservation and Recovery Act; Inh. = inhalation; NPDWR = National Primary Drinking Water Regulation; NPL = National Priority List; OLEM = Office of Land and Emergency Management; OW = Office of Water; SDWA = Safe Drinking Water Act.

2.3. PROBLEM FORMULATION

Problem formulation information pertaining to the reassessment of Cr(VI) was included in the preliminary materials documents released to the public in April and August 2014 (U.S. EPA, 2014b, c); two public meetings were held in June and October 2014 to obtain public input on these materials.

As discussed in the April 2014 preliminary materials document (U.S. EPA, 2014b), EPA consulted federal, state, and international agency health assessments published since the U.S. EPA (1998b) *IRIS Toxicological Review of Hexavalent Chromium* to identify studies and scientific issues that may impact the reassessment of Cr(VI). EPA has continued to consult other agency health assessments following the 2014 public meetings. These health agencies, and information regarding the basis of any protective exposure values or health determinations, are presented in Tables 2 to 4. Based on prior health agency assessments of Cr(VI) described in Tables 2 and 4, the health effects of primary interest for evaluation in the current IRIS assessment are respiratory and gastrointestinal (GI) effects. These health agencies also identify other potential target systems of possible interest to the current IRIS assessment; these are discussed in Section 3.1 (U.S. EPA, 2014b).

Table 2. Cr(VI) values for inhalation exposure (µg/m³) from U.S. federal and state agencies[a] and international bodies (in reverse chronological order)

Reference	Value (µg/m³)	Time adjustment	Chemical note	Endpoints/basis
TCEQ (2014)	0.0043	Lifetime/chronic	Particulate compounds	Excess lung cancer mortality risk of 1×10^{-5}, using risk value derived from Gibb et al. (2000a) and Crump et al. (2003).
	0.066	Lifetime/chronic	Particulate compounds	Respiratory effect (increased relative lung weight after 90 days of exposure) in rats (Glaser et al., 1985).
	0.39	Acute	Particulate compounds	Respiratory effect (increased relative lung weight after 30 days of exposure) in rats (Glaser et al., 1990).
IPCS (2013)	0.03	Lifetime/chronic	Cr(VI) salts	Respiratory effects in rats (Glaser et al., 1990).
	0.005	Lifetime/chronic	Chromium trioxide, chromic acid	Upper respiratory effects in humans (Lindberg and Hedenstierna, 1983).
NIOSH (2013a)	0.2	8-hour TWA, 40-hour workweek	All Cr(VI) compounds	Lung cancer and nonmalignant respiratory effects. Based on analysis of Baltimore cohort data by Park et al. (2004).
ATSDR (2012)	0.005	Chronic	Dissolved aerosols and mists	Upper respiratory effects (nasal irritation/ulceration, mucosal atrophy, and decreases in spirometric parameters), based on Lindberg and Hedenstierna (1983).
	N/A	Chronic	Particulates	Insufficient data
	0.005	Intermediate	Dissolved aerosols and mists	Upper respiratory effects (nasal irritation/ulceration, mucosal atrophy, and decreases in spirometric parameters), based on Lindberg and Hedenstierna (1983).
	0.3	Intermediate	Particulates	Respiratory tract (lung) and other effects. Based on quantitative analysis of rat studies (Glaser et al., 1990; Glaser et al., 1985) performed by Malsch et al. (1994).
OEHHA (2008)	0.2	Chronic	Soluble compounds	Respiratory effect (bronchoalveolar hyperplasia) in rats (Glaser et al., 1990).
	0.002	Chronic	Chromic trioxide (as chromic acid mist)	Respiratory effects in humans (Lindberg and Hedenstierna, 1983).

Reference	Value (µg/m^3)	Time adjustment	Chemical note	Endpoints/basis
OSHA (2006)	5	8-hour TWA	All Cr(VI) compounds	Lung cancer and nasal tissue damage. Based on quantitative analysis of Baltimore cohort data by Gibb et al. (2000b) and Gibb et al. (2000a).
RIVM (2001)	0.0025	Chronic	Inhalable dust	Excess lifetime lung cancer risk of 1×10^{-4}, based on analysis of human occupational studies by the 1987 and 1994 World Health Organization air quality guidelines.[b]
U.S. EPA (1998b)	0.008	Lifetime/chronic	Chromic acid mists/dissolved chromium aerosols	Effects in the nasal cavity. Based on Lindberg and Hedenstierna (1983).
	0.1	Lifetime/chronic	Cr(VI) particulates	Respiratory effects. Based on quantitative analysis of rat studies (Glaser et al., 1990; Glaser et al., 1985) performed by Malsch et al. (1994).

N/A = not applicable; TWA = time-weighted average.
[a]Selected values from states known by U.S. EPA to have derived independent values; most states typically adopt values from U.S. EPA.
[b]Risk value rationale and studies unchanged in WHO (2000).

Table 3. Cr(VI) cancer risk values for inhalation exposure from U.S. federal and state agencies[a] and international bodies (in reverse chronological order)

Reference	Risk factor $(\mu g/m^3)^{-1}$	Rationale
TCEQ (2014)	Unit risk factor: 2.28×10^{-3} (particulate compounds)	Linearly extrapolated lung cancer risk based on a weighted average of Gibb et al. (2000a) (Baltimore cohort) and Crump et al. (2003) (Painesville cohort).
IPCS (2013)	Occupational exposure risk: 6×10^{-3}	Linearly extrapolated lung cancer risk based on Gibb et al. (2000a).
	Environmental exposure risk: 4×10^{-2}	
IARC (2012).	*Carcinogenic to humans* (Group 1)[b]	Lung cancer, based on multiple evidence streams. Positive associations between Cr(VI) exposure and cancer of the nose and nasal sinuses also cited.
NTP (2011)	Known to be human carcinogen[b]	Cancers of the lung and sinonasal cavity, based on studies in humans.
CalEPA (2011)	0.16 (95% upper confidence: 0.35)	Linearly extrapolated lung cancer risk based on Gibb et al. (2000a).
	1×10^{-2} (lower bound)	Linearly extrapolated lung cancer risk based on Luippold et al. (2003).
WHO (2000)	4×10^{-2}	Linearly extrapolated lung cancer risk based on multiple human occupational studies.
U.S. EPA (1998b)	Inhalation unit risk: 1.2×10^{-2}	Linearly extrapolated lung cancer risk based on Mancuso (1997, 1975).

[a]Selected values from states known by U.S. EPA to have derived independent values; most states typically adopt values from U.S. EPA.
[b]Agency does not derive a quantitative risk factor.

Table 4. Cr(VI) values for oral exposure from U.S. federal and state agencies[a] and international bodies (in reverse chronological order)

Reference	Risk value or limit	Rationale[b]
Health Canada (2016)	Maximum acceptable concentration: 50 µg/L	Cancer precursor, mouse small intestine hyperplasia
TCEQ (2016)	RfD: 3.1×10^{-3} mg/kg-day	Cancer precursor, mouse small intestine hyperplasia
IPCS (2013)	Tolerable daily intake: 9×10^{-4} mg/kg-day	Mouse small intestine noncancer effects
ATSDR (2012)	Chronic MRL: 9×10^{-4} mg/kg-day	Mouse small intestine noncancer effects
	Intermediate MRL: 5×10^{-3} mg/kg-day	Hematological effects (rat data at 22 days)
CalEPA (2011)	Cancer PHG: 0.02 µg/L	1×10^{-6} cancer risk using OSF of 0.5 $(mg/kg\text{-}day)^{-1}$ (mouse small intestine tumors)
	Noncancer PHG: 2 µg/L	Liver noncancer effects (rats)
SWRCB (2014); CDPH (2013)	Proposed MCL: 10 µg/L Note: invalidated [see CA State Water Board (2017) fact sheet]	Cancer risk [see CalEPA (2011)]
NJ DEP (2009)	Soil remediation criterion: 1 ppm soil concentration	1×10^{-6} cancer risk using OSF of 0.5 $(mg/kg\text{-}day)^{-1}$ (mouse small intestine tumors)
U.S. EPA (2008a, 2008b)	OSF: 0.791 $(mg/kg\text{-}day)^{-1}$	Upper-bound cancer risk estimate (mouse small intestine tumors)
Values based on science or rules published prior to 2008 National Toxicology Program study		
FDA (2013)	Allowable level in bottled water: 0.1 mg/L (or 100 ppb) total chromium	Not specified
U.S. EPA [Federal Register (2010)]	MCL: 100 ppb (total chromium)	Allergic dermatitis[c]
WHO (2003)	50 µg/L	Provisional value (nonspecific)
RIVM (2001)	5×10^{-3} mg/kg-day	Provisional noncancer effects, based on no-effect level [rats; (MacKenzie et al., 1958)]

Reference	Risk value or limit	Rationale[b]
U.S. EPA (1998b)	RfD: 3×10^{-3} mg/kg-day	No effect level for noncancer effects [rats; (MacKenzie et al., 1958)]

MCL = maximum contaminant level; MRL = minimal risk level; OSF = oral slope factor; PHG = public health goal.
[a]Selected values from states known by U.S. EPA to have derived independent values; most states typically adopt values from U.S. EPA (based on un-speciated total chromium).
[b]All values based on mouse data from NTP (2008), unless otherwise noted.
[c]Based on rule promulgated in 1991 (National Primary and Secondary Drinking Water Regulations, 56 FR 3526, 1-30-91 and 54 FR 22062, 5-22-89).

3. ASSESSMENT APPROACH, SPECIFIC AIMS, AND DRAFT POPULATIONS, EXPOSURES, COMPARATORS, AND OUTCOMES (PECO) CRITERIA

3.1. ASSESSMENT APPROACH

The overall objective of this assessment is to identify adverse health effects and characterize exposure-response relationships for the effects of Cr(VI) to support the development of toxicity values. This assessment uses systematic review methods to evaluate the epidemiological and toxicological literature for Cr(VI); relevant mechanistic evidence is also considered. The evaluations conducted in this assessment are consistent with relevant EPA guidance.[1]

The specific approach taken to the reassessment of the health effects of Cr(VI) was based on input received during scoping, a survey of the health effects of Cr(VI) previously identified by government health agencies (including EPA) and international health organizations, as well as consideration of the physicochemical properties of Cr(VI). As discussed in the preliminary materials released in 2014 (U.S. EPA, 2014b, c), the IRIS assessment will include evaluations of the evidence relevant to all cancer outcomes, and will evaluate noncancer effects for the following potential target systems: respiratory, gastrointestinal, hepatic, hematological, immunological, reproductive, and developmental. As discussed further below, for cancer and nasal irritation via the inhalation route, the systematic review will focus on data that may improve the quantitative dose-response analysis, conducted in EPA's 1998 IRIS assessment, for these outcomes.

3.1.1. Evaluation of the Potential Carcinogenicity of Inhaled Cr(VI)

EPA's 1998 IRIS assessment classified Cr(VI) as "Group A—known human carcinogen by the inhalation route of exposure" based on evidence of a causal relationship between inhalation of Cr(VI) and increased incidence of lung cancer in humans. The same conclusion has since been reached by other federal and state health agencies and international organizations (TCEQ, 2014; IPCS, 2013; NIOSH, 2013b; IARC, 2012; CalEPA, 2011; NTP, 2011; OSHA, 2006). Therefore, as discussed in the preliminary materials released in 2014 (U.S. EPA, 2014b, c), this assessment will focus on the review of the evidence for lung cancer to identify studies not included in the 1998 assessment that might improve the quantitative dose-response analysis for human lung cancer.

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

3.1.2. Evaluation of the Effects of Inhaled Cr(VI) on the Nasal Cavity

In the 1998 assessment (U.S. EPA, 1998b), EPA concluded that a number of occupational epidemiological studies demonstrated an association between inhalation of Cr(VI) and upper respiratory irritation and atrophy. Based on EPA's 1998 evaluation of the literature and the determination that the effects of Cr(VI) on the nasal cavity have been well established [e.g., OSHA (2006), U.S. EPA (2014c)], EPA will not reevaluate the qualitative evidence for an association between Cr(VI) exposure and nasal irritation/atrophy. Rather, the review of the evidence for nasal effects will focus on identifying studies that might improve quantitative dose-response analysis for this outcome. This decision to focus the systematic review on studies useful for an improved dose-response analysis is an update from the preliminary materials released in 2014 (U.S. EPA, 2014b, c).

For noncancer effects occurring in the respiratory tract beyond the nasal cavity (bronchopulmonary), and for systemic effects, both hazard identification and dose-response will be evaluated.

3.1.3. Toxicokinetics of Cr(VI)

The absorption and metabolism of Cr(VI) are topics that have been thoroughly reviewed in previous health agency documents (IPCS, 2013; NIOSH, 2013a; ATSDR, 2012; CalEPA, 2011; OSHA, 2006). Briefly, chromium exists in multiple oxidation states, but the hexavalent and trivalent states are most prevalent biologically. Following oral or inhalation exposure (and prior to systemic absorption), Cr(VI) can be reduced to Cr(III) within the GI tract or the respiratory tract, respectively. If reduced to the trivalent state prior to uptake, chromium is poorly absorbed by cells and is not toxic. However, chromium in the hexavalent state can be readily absorbed by cells lining the GI or respiratory tract. After systemic absorption, Cr(VI) will continue to reduce to Cr(III) within cells and tissues in the body. Only total chromium (Cr[VI] + Cr[III]) can be accurately measured in biological tissues and excreta. This has implications for how human epidemiological studies are evaluated for exposure, and how absorption, distribution, metabolism, or excretion (ADME) studies are screened and inventoried.

The route of exposure affects the local and systemic distribution of chromium because Cr(VI) will pass through different fluids and tissues of varying reduction capacity depending on the site of absorption. Orally ingested Cr(VI) is likely to be absorbed in the GI tract and liver (both of which will reduce Cr[VI] to Cr[III]). Due to the first-pass effect, less Cr(VI) may be available for absorption to systemic circulation and other tissues following oral ingestion. Inhaled Cr(VI) is likely to be absorbed in the respiratory tract and distributed to systemic circulation as Cr(VI) because less extracellular reduction may occur. Cr(VI) administered by injection (intravenous or intraperitoneal) or intratracheal instillation bypasses mechanisms that reduce and dampen systemic Cr(VI) absorption and distribution. As a result, the toxicological effects induced by Cr(VI) at both portal-of-entry and systemic tissues differ by exposure route. Exposures to Cr(VI) via oral

and inhalation routes will be considered more toxicologically relevant than other routes of exposure (e.g., dermal, injection, or intratracheal). Criteria for the screening of studies that include consideration of route of exposure are described in Section 3.3.

Extrapolating Cr(VI) dose-response data from animals to humans is complex in light of these toxicokinetic properties (IPCS, 2013; ATSDR, 2012). The reassessment will consider the available Cr(VI) toxicokinetic models for the quantitative analysis of toxicity data. As a result, physiologically based pharmacokinetic (PBPK) models will undergo study evaluation.

3.2. SPECIFIC AIMS

The aims of the assessment are to:

- Identify epidemiological (i.e., human) and toxicological (i.e., experimental animal) literature reporting effects of exposure to Cr(VI) as outlined in the PECO. The assessment will include evaluations of the evidence relevant to all cancer outcomes and will evaluate noncancer effects for the following potential target systems: respiratory, GI, hepatic, hematological, immunological, reproductive, and developmental. The systematic review will focus on identifying data from inhalation exposures that are useful for deriving quantitative estimates for lung cancer and nasal effects rather than revisiting the qualitative identification of hazard for these outcomes.

- Evaluate mechanistic events associated with exposure to Cr(VI) that inform the development or progression of the health effects identified in humans and animals. The scope of these analyses will be determined by the complexity and confidence in the evidence in humans and animals, likelihood to impact evidence synthesis conclusions for human health, and the directness or relevance of the model systems for understanding potential human health hazards. The primary focus will be on the analysis of mechanistic evidence for cancer and noncancer effects of the GI tract following oral exposures to Cr(VI). Because the hazard identification of lung cancer and nasal effects will not be revisited, the mechanistic analyses for these health effects will focus on evidence that may affect the dose-response assessment.

- Conduct study evaluations (risk of bias and sensitivity) for individual epidemiological and toxicological studies and PBPK models as defined by the scoping decisions described in Section 3.1.

- Extract data on relevant health outcomes from selected epidemiological and toxicological studies based on the study evaluations.

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

- For each health outcome, express strength of evidence conclusions from across studies (or subsets of studies) separately for studies in humans and animals. If studies informing mechanisms were synthesized, then mechanistic evidence from either human or animal studies will be integrated with the health effects evidence.

- For each health outcome, integrate strength of evidence conclusions across evidence streams (human and animal) to conclude whether a substance is hazardous to humans. Identify and discuss issues concerning potentially susceptible populations and life stages. Biological support from mechanistic studies and nonmammalian model systems will be considered based on the iterative prioritization approach outlined in the PECO.

- Derive toxicity values (e.g., RfDs, RfCs, cancer risk estimates) as supported by the available data. Apply toxicokinetic and dosimetry modeling to account for interspecies differences.

- 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.

3.3. PECO CRITERIA

The PECO is used to identify the evidence that addresses the specific aims of the assessment and to focus the literature screening, including the inclusion/exclusion criteria, in a systematic review. The PECO criteria for Cr(VI) (see Table 5) are based on (1) nomination of the chemical for assessment, (2) discussions with scientists in EPA program and regional offices to determine the scope of the assessment that will best meet Agency needs, (3) preliminary review of the health effects literature for Cr(VI) (primarily reviews and authoritative health assessment documents) to identify the major health hazards associated with exposure to Cr(VI) and key areas of scientific complexity, and (4) input received during public discussion of preliminary materials released to the public in 2014.

In addition to the PECO criteria, studies containing supplemental material that are potentially relevant to the specific aims of the assessment were tracked during the literature screening process. Although these studies did not meet PECO criteria, they were not excluded from further consideration. The categories used to track studies as "potentially relevant supplemental material" during screening and to prioritize these studies for consideration in the assessment based on likelihood to impact evidence synthesis conclusions for human health are described in Section 4.3.

Table 5. Populations, exposures, comparators, and outcomes (PECO) criteria

PECO element	Evidence
Populations	**Human:** Any population and life stage (occupational or general population, including children and other potentially sensitive populations). **Animal:** Nonhuman mammalian animal species (whole organism) of any life stage (including preconception, in utero, lactation, peripubertal, and adult stages).
Exposures	**Human:** Any exposure to Cr(VI), including occupational exposures, via oral or inhalation routes. Exposures by the inhalation and oral routes may be assessed based on administered dose or concentration, biomonitoring data (e.g., urine, blood, or other specimens), environmental or occupational-setting measures (e.g., air, water, dust levels), or job title or residence. Some relevant forms of compounds containing Cr(VI) (18540-29-9) are listed below: • Chromic acid (H_2CrO_4 [7738-94-5] and $H_2Cr_2O_7$ [13530-68-2]) • Salts of the chromate (CrO_4^{2-}) and dichromate ($Cr_2O_7^{2-}$) anions: Sodium chromate (7775-11-3), sodium dichromate (10588-01-9), sodium dichromate dihydrate (7789-12-0), potassium chromate (7789-00-6), potassium dichromate (7778-50-9) • Chromium(VI) trioxide (commonly referred to as chromium oxide [1333-82-0]) • Calcium chromate (13765-19-0) **Animal:** Any exposure to Cr(VI) via oral or inhalation routes based on administered dose or concentration. Cr(VI) may be administered orally via gavage or ad libitum in diet or drinking water. Cr(VI) may be administered by inhalation via whole-body or nose-only systems. Relevant forms of Cr(VI) are listed above. Animal studies involving exposures to mixtures will be included only if they include exposure to Cr(VI) alone.
Comparators	**Human:** A comparison or referent population exposed to lower levels (or no exposure/exposure below detection limits) of Cr(VI), or exposure to Cr(VI) for shorter periods of time. **Animal:** A concurrent control group exposed to vehicle-only treatment or an untreated control.
Outcomes	All cancer outcomes are considered; noncancer health outcomes are considered for the following potential target systems: respiratory, GI, hepatic, hematological, immunological, reproductive, or developmental effects. As discussed above, EPA anticipates that a systematic review for other health effect categories (e.g., nephrotoxicity, neurotoxicity) will not be undertaken unless a significant amount of new evidence is identified.
PBPK models	Studies describing PBPK models for Cr(VI) will be included.

4. LITERATURE SEARCH AND SCREENING STRATEGIES

4.1. LITERATURE SEARCH STRATEGIES

Literature search strategies were developed using key terms and words related to the PECO criteria. Relevant subject headings and text-words were crafted into a search strategy that was designed to maximize the sensitivity and specificity of the search results. The search strategy was run, and the results were assessed to ensure that all previously identified relevant primary studies were retrieved in the search. Because each database has its own search architecture, the resulting search strategy was tailored to account for the unique search functionality of each database.

The following databases were searched:

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

Searches were not restricted by publication date, and no language restrictions were applied. Web of Science results were limited using the research areas filter. All Web of Science research areas identified in the search results were prioritized by a technical advisor as high priority (e.g., toxicology), low priority (e.g., chemistry), and not relevant (e.g., forestry). Literature searches were conducted in bibliographic databases as described in Appendix A and uploaded to EPA's Health and Environmental Research Online (HERO) database.[2]

Additional relevant literature not found through database searching was sought by:

- Manually searching citations from review articles and studies considered to meet PECO criteria after screening ("included" studies).
- Searches of gray literature, including primary studies that are not indexed in databases of peer-reviewed literature (e.g., technical reports from government agencies or scientific research groups; unpublished laboratory studies conducted by industry; working papers from research groups or committees; and white papers), or other nontypical searches. Gray literature is typically identified by searching the EPA Chemistry Dashboard

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

(https://comptox.epa.gov/dashboard) during problem formulation, by engaging with technical experts, and during solicitation of Agency, interagency, and public comment at multiple steps in the IRIS process.

- "Backward" searches (to identify articles cited by included studies, reviews, or prior assessments by other agencies).

The initial search was performed in January 2013, and literature search updates were conducted in July 2013, February 2014, April 2015, April 2016, May 2017, December 2017, and is current through May 2018. The literature search will be updated throughout draft development to identify literature published during the course of the review. The last full literature search update will be conducted less than one 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, which also reflect the literature screening decisions (see Section 4.3).

The IRIS Program takes extra steps to ensure identification of pertinent studies by (1) encouraging the scientific community and the public to identify additional studies and ongoing research; (2) searching for publicly available data submitted under the Toxic Substances Control Act and the Federal Insecticide, Fungicide, and Rodenticide Act; and (3) considering late-breaking studies that would impact the credibility of the conclusions, even during the review process.[3] Studies identified after peer review begins will only be considered for inclusion if they meet the PECO criteria and may fundamentally alter the assessment's conclusions.

4.2. NON-PEER-REVIEWED DATA

IRIS assessments rely mainly on publicly accessible, peer-reviewed studies. However, it is possible that gray literature (i.e., studies that are not reported in the peer-reviewed literature) directly relevant to the PECO may be identified during assessment development (e.g., good laboratory practice [GLP] studies submitted to EPA, dissertations, etc.). In this case, if the data substantially affect assessment decisions or conclusions (i.e., potential to impact the PECO statement, hazard conclusions, or dose-response analysis), EPA can obtain external peer review if the owners of the data are willing to have the study details and results made publicly accessible. This independent, contractor-driven peer review would include an evaluation of the study, similar to a peer review of a journal publication. The contractor would identify and select two to 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 prior to confirming their service. In most instances, the peer review would be conducted by letter review. The study authors would be informed of the outcome of the peer review and given an opportunity

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

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 the EPA, and the names of the peer reviewers would be identified. In certain cases, IRIS 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 (e.g., raw) data from personal author communication can supplement a peer-reviewed study if the information is made publicly available (typically through documentation in HERO).

4.3. SCREENING PROCESS

The PECO criteria were 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 PECO criteria, the exclusion criteria noted below were applied, while also tagging studies as appropriate to allow for later retrieval, dependent on assessment needs:

- Studies that were previously determined not to be pertinent, as described in the 2014 Supplemental Materials (U.S. EPA, 2014b, c);
- Study materials that have not been peer reviewed, unless they are expected to have a substantial impact on the assessment (as described in Section 4.2);
- Records that do not contain original data, such as other agency assessments, informative scientific literature reviews, grant submissions (from the National Institutes of Health [NIH] reporter database), editorials, or commentaries;
- Chromium compounds that did not meet PECO criteria (e.g., metal chromates; animal studies of exposures to mixtures containing Cr[VI]);
- Ecology studies;
- Studies appearing as abstracts only (e.g., conference abstracts); and
- Non-English studies in which the titles and abstracts (when available) did not suggest direct relevance to the PECO or specific aims.

In addition to the inclusion of studies that meet PECO criteria, studies containing supplemental material that is potentially relevant to the specific aims were tracked during the screening process (see Section 4.4.2). Although not considered to directly meet PECO criteria, these studies were not strictly excluded unless otherwise specified. Unlike studies that meet PECO

criteria, supplemental studies may not be subject to systematic review unless predefined questions are identified that focus the mechanistic (or other) analysis are added to the specific aims and PECO criteria. Studies that were categorized as "potentially relevant supplemental material" include the following:

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

- ADME studies: Studies designed to capture information regarding absorption, distribution, metabolism, and excretion, including toxicokinetic studies (e.g., studies describing quantitative models or data for Cr[VI] reduction kinetics in biological media [e.g., gastric juice, red blood cells, lung, and GI tract epithelial cells]). Such information may be helpful in updating or revising the parameters used in existing PBPK models.

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

- Susceptible populations: Studies that identify potentially susceptible subgroups, such as studies that focus on a specific demographic, life stage, or genotype. (These are categorized under "Mechanistic studies.").

- Related to included studies: Versions of other studies (e.g., updated cohort analyses) that meet PECO criteria.

- Human case reports or case series: In most cases, case reports and case series will be tracked as potentially relevant supplemental information.

- Routes of exposure not pertinent to PECO: Studies using dermal, injection, or intratracheal administration.

- Acute duration exposures: Animal studies of acute or short-term (less than 28 days) exposure duration.

4.3.1. 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 appeared to meet the PECO criteria using a structured form in DRAGON (ICF Consulting, 2018). For non-English studies, if the title and abstract were written in English, the eligibility status of these studies was assessed using the same approach. For citations with no abstract, articles were screened based on title relevance and page numbers (articles two pages in length or less may be assumed to be

conference reports, editorials, or letters). All screening conflicts were resolved by a technical advisor.

Studies not meeting PECO criteria but identified as "potentially relevant supplemental material" were categorized (i.e., tagged) during the title and abstract screening process (further described in Section 4.4). Conflict resolution is not required during the screening process to identify supplemental information (i.e., tagging by a single screener is sufficient to identify the study as potentially relevant supplemental material that may be considered during draft development).

To ensure all relevant references were identified in the initial screening, the excluded materials were reviewed to identify misclassified studies meeting PECO criteria or potentially relevant supplemental material that may have been missed during the database searches. A subset of excluded studies was prioritized for a second round of screening using text analytics. Supervised clustering and machine learning using ICF's Document Classification and Topic Extraction Resource (DoCTER) was conducted to ensure that all mechanistic studies were identified. Supervised clustering is a form of semi-supervised machine learning that uses seeds or known-to-be-relevant studies. DoCTER includes multiple text analytic algorithms (K-means and non-negative matrix factorization) that can be used to find studies with titles and abstracts that are similar to "seed studies" previously identified as relevant (Varghese et al., 2017). These algorithms create a user-defined number of clusters based on keyword similarities in the title and abstract, and each algorithm is broadly accepted in the text analytics scientific field. Machine learning uses similar algorithms, but requires a robust training set to predict the likelihood that a given unclassified study is relevant. For this effort, both supervised clustering and machine learning were used to prioritize a set of studies to rescreen. Training data and seeds were derived from the 806 studies classified as mechanistic in the first round of screening. Results were rescreened for relevance to mechanistic endpoints. In addition to tagging studies as mechanistic, screeners were also directed to tag any additional supporting studies or health effect studies that were identified using the text analytics prioritization methods described here.

Following the efforts to identify misclassified mechanistic studies and the literature search updates described above, ICF identified 1,288 on-topic mechanistic references for screening. These references were further screened using title and abstract information by two independent EPA staff members, followed by conflict resolution if screening results were different. Due to the large number of studies, it was necessary to develop deprioritization criteria to begin to set aside studies that are potentially less impactful to the assessment of mechanistic events. These studies were tagged so that they may be accessed later in the mechanistic analysis if needed.

The following types of mechanistic studies were deprioritized for further screening:

- Studies that were misidentified as on-topic during the first round of screening (e.g., studies that did not include Cr[VI] or other oxidation states of chromium)

- References only containing an abstract (i.e., conference abstracts)
- Book chapters and reviews
- Untranslated foreign language articles
- Studies that only report chromium detection methods
- Studies in less common model systems (e.g., plants, marine mammals)
- Studies that are only relevant to a health effect not being evaluated (e.g., nephrotoxicity)

In addition, many studies were identified that used Cr(VI) as a positive control for new assay validation or that were co-exposures (e.g., to investigate the antioxidant properties of a new compound). Most of these studies did not contain information useful for the mechanistic analysis of Cr(VI) and were deprioritized. However, studies were retained for full-text review if there was any indication that they might be useful for mechanistic understanding or might report mechanistically relevant information regarding a health effect not reported in human or animal studies (e.g., neurotoxicity). Studies were categorized and tagged based on the above criteria using DistillerSR to record why each was deprioritized. This allows the assessors to revisit certain study categories if deemed important later in the assessment process.

The mechanistic references that were prioritized for further consideration were categorized by endpoint type using DistillerSR. Prioritized endpoints included studies relevant to cancer or effects on the GI, respiratory, reproductive, developmental, hepatic, immune, or hematological systems. Mechanistic references were also categorized if relevant to one or more of the 10 key characteristics of carcinogens (Smith et al., 2016), (intracellular) ADME, and/or contained pathology findings. References were also tagged with the following: study type (in vivo, ex vivo, in vitro), presence of "omics" data, relevance to a certain species based on whole organism or cell type, and reported data using an acellular system. These tags allowed further prioritization and organization for the next phase of screening.

Mechanistic references may be processed through an additional round of title and abstract-based categorization to further assist with prioritization (for example, in vivo studies may be categorized by route of exposure). This will allow additional narrowing of the mechanistic studies of highest interest before the full text review and quality evaluation steps.

4.3.2. Full-Text-Level Screening

Records that were not excluded based on the title and abstract advanced to full-text review. Full-text copies of these potentially relevant records were retrieved, stored in the HERO database, and independently assessed by two screeners to confirm eligibility according to the PECO criteria. Screening conflicts were 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 advanced to full-text review were also tagged as "potentially relevant supplemental material" as appropriate.

The results of this screening process have been posted on the project page for this assessment in the HERO database (https://hero.epa.gov/hero/index.cfm/project/page/project_id/2233), and studies have been "tagged" with appropriate category descriptors (e.g., included, "potentially relevant supplemental material," excluded). Results have also been annotated and reported in a literature flow diagram (see Figure 1).

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.

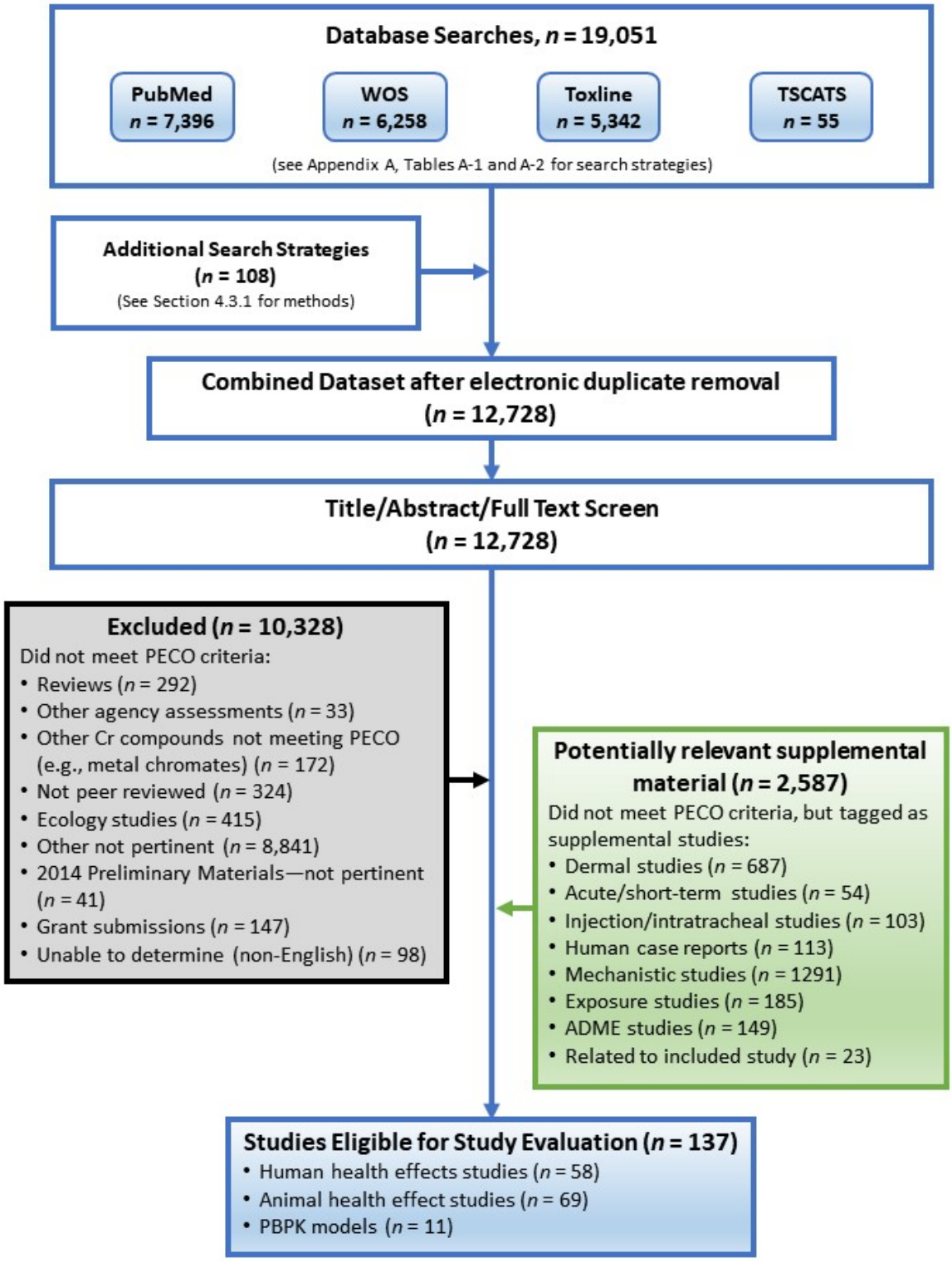


Figure 1. Literature search flow diagram for Cr(VI).

4.3.3. Multiple Publications of the Same Data

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

4.4. SUMMARY-LEVEL LITERATURE INVENTORIES

During title/abstract or full-text level screening, studies tagged based on PECO eligibility were further categorized based on features such as evidence type (human, animal, mechanistic, PBPK, etc.), health outcome(s), and/or endpoint measure(s) included in the study. Literature inventories for studies meeting PECO criteria were created to develop summary-level, sortable lists that include some basic study design information (e.g., study population, exposure information such as doses administered or biomarkers analyzed, age/life stage[4] of exposure, endpoints examined, etc.). These literature inventories facilitate subsequent review of individual studies or sets of studies by topic-specific experts.

4.4.1. Studies Meeting PECO Criteria

The preliminary materials released in 2014 (U.S. EPA, 2014b, c) presented evidence tables for the human and animal studies determined to be eligible for study evaluation. Following the 2014 public meetings, these data tables were maintained in Microsoft Word format and were revised to correct errors identified by public commenters, EPA staff, and contractors. During this revision process, additional data were added to the tables (both from studies already contained in the tables and studies found in subsequent literature searches or public submissions). The summary-level information in these tables was used as an inventory to prioritize data migration to the Health Assessment Workplace Collaborative (HAWC; see Section 8), initiate HAWC study entries, and identify subject matter experts for performing study evaluations. Depending on study confidence (see Section 6) and data type, data from the inventories were migrated to HAWC. Any studies identified as meeting the PECO criteria since the start of HAWC migration will be entered directly into HAWC (and will not be added to the Microsoft Word inventory tables).

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

4.4.2. Potentially Relevant Supplemental Material

Inventories were also created for 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 studies on endpoints or routes of exposure 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 and mechanistic data (and related information) can be critical to the next steps of prioritizing or evaluating individual PECO-specific studies, and thus these studies were reviewed by subject matter experts early in the assessment process. ADME and mechanistic inventories released in 2014 (U.S. EPA, 2014c) were revised to correct errors identified by public commenters and will continue to be updated with new studies during assessment development. (Note: PBPK models are typically considered to meet PECO criteria, while ADME and toxicokinetic-related studies are most commonly tracked as potentially relevant supplemental material). Cr(VI) ADME studies will continue to be sorted into the following categories: (1) animal and human in vivo (oral, inhalation, intratracheal, intravenous, intraperitoneal, subcutaneous, and multi-route), (2) quantitative in vitro/ex vivo (gastric and red blood cell), (3) mechanistic distribution/reduction (multiple system types), and (4) human biomonitoring. Summary information, such as species, tissues examined, level of time-course sampling, and Cr(VI) reducing capacities, will continue to be extracted from these studies. Mechanistic studies have been sorted according to the screening criteria outlined in Section 4.3 to facilitate the analysis of mechanistic events.

5. REFINED EVALUATION PLAN

The refined evaluation plan describes refinements made to the set of studies that met PECO criteria and are to be carried forward to study evaluation. The process also helps determine which studies tagged as "potentially relevant supplemental material" may need to be considered in the assessment. Refinements were based on (1) input from public comments on the preliminary materials released in 2014 (U.S. EPA, 2014b, c), (2) literature screening and creation of the inventories of studies meeting PECO criteria and potentially relevant supplemental material by EPA staff and contractors, and (3) review of the inventories by subject matter experts. The refined evaluation plan also identifies the endpoints, grouped by outcomes, that will be the primary focus of the outcome-specific evaluations. These specifications will aid in implementing the endpoint-specific study evaluation criteria (see Section 6).

5.1. AIRBORNE CHARACTERIZATION AND CHEMICAL PROPERTIES

Studies that met PECO criteria include those that provide data on inhaled Cr(VI) in a variety of physical and chemical forms. Airborne Cr(VI) can exist in different sizes and forms (e.g., particulates, dusts, aerosols, fumes, or mists) that affect respiratory tract deposition. Furthermore, the studies that met PECO criteria include compounds containing Cr(VI) that have different chemical properties. All forms of Cr(VI) meeting PECO criteria will be evaluated for hazard identification, regardless of chemical properties or airborne characteristics. However, the evidence synthesis will consider the possibility that some forms or mixtures (such as Cr[VI] in extremely acidic or alkaline solutions) may have properties that alter the toxicity or introduce uncertainties. In addition, the physical and chemical properties of airborne Cr(VI) will be taken into consideration when evaluating the suitability of studies for dose-response analysis.

Nine studies involving occupational exposure to welding fume were identified in the set of studies meeting PECO criteria. Cr(VI) exposure via welding fume may occur if chromium is a component of the base materials being joined (e.g., stainless steel), is present as a surface coating, or is a component of materials consumed during the welding process, such as metal filler rod. Occupational exposures to Cr(VI) in welding fume are variable due to differences in welding types, practices, and duration of welding tasks (Shaw Environmental, 2006). Further, welding fume components vary by the type of welding and base materials (Shaw Environmental, 2006). Because variability in occupational exposure makes exposure to Cr(VI) difficult to quantify, toxicity data for welding fume will not be considered for dose-response analysis. However, exposures to Cr(VI) are high among stainless steel welders relative to workers performing other types of welding due to the high chromium content of stainless steel compared with other base metals or alloys [e.g., mild steel;

([NIOSH, 2013a](); [Shaw Environmental, 2006]())]. Therefore, studies comparing stainless steel welders to a less exposed reference group may be evaluated for noncancer hazard identification.

5.2. TOXICOKINETICS

Information on the toxicokinetics of Cr(VI) is provided elsewhere in this document (see Sections 3.1 and 6.4). Of the PBPK models available that met PECO criteria, evaluations will be limited to those accounting for Cr(VI) reduction in the stomach compartment and interspecies differences in gastric pH and physiology. Models must also include parameterization for mice, rats, and humans. This narrows the evaluation to models that may be suitable for the dose-response assessment. Furthermore, based on the issues related to toxicokinetics outlined in Sections 3.1 and 6.4 and discussions and comments from public meetings ([U.S. EPA, 2014c](), [2013]()), route-to-route extrapolations will not be considered.

5.3. TOXICOGENOMICS

Eighteen studies reporting gene expression data following Cr(VI) exposures were identified during screening as "potentially relevant supplemental material." Nine of these studies were conducted in animals and will be subject to study evaluation using the criteria described in Section 6.3. In addition, for both in vitro and in vivo toxicogenomic studies, the conduct of the expression data generation and reporting will be evaluated using publicly available criteria based on standard practices in the field ([Bourdon-Lacombe et al., 2015]()); specifically, the Minimum Information About a Microarray Experiment (MIAME) ([Brazma et al., 2001]()) and the Systematic Omics Analysis Review (SOAR) tool ([McConnell et al., 2014]()).

The applicability of the available microarray data to making toxicological inferences will be assessed indirectly based on (1) comparison between the dose-response relationships derived from transcriptomics data and apical outcomes and (2) evaluation of biological plausibility, as well as external and internal consistency of the results of the gene expression analysis. Where appropriate, tools such as BMDExpress 2.20.0148 beta ([Sciome, 2018]()) will be used to examine dose-response relationships for gene expression and to identify pathways enriched with genes that demonstrate significant dose-response trends and to determine the points of departure.

To use toxicogenomic data to inform biological processes associated with the exposure to Cr(VI), the expression data will be analyzed using several complementary approaches. Pathways and upstream regulators relevant to the genes identified as differentially expressed between Cr(VI)-exposed versus unexposed controls will be explored using Ingenuity Pathway Analysis ([Qiagen, 2018]()). Gene sets enriched in Cr(VI)-exposed versus unexposed control animals will be determined by Gene Set Enrichment Analysis [Broad Institute; ([Subramanian et al., 2005]())]. Similarity of gene expression changes induced by Cr(VI) to public expression data corresponding to various human and animal diseases and exposures to xenobiotics will be examined. This will be

done to identify conditions associated with gene expression profiles like those resulting from animal exposure to Cr(VI).

Similarities between gene expression profiles will be examined using Basespace Correlation Engine (Illumina, 2018) and Signature Search Tool [Genevestigator; (Kupershmidt et al., 2010; Hruz et al., 2008)]. Available genomic biomarkers will also be used to detect specific events. For example, the TGx-DDI biomarker for DNA damage classification (Jackson et al., 2017) will be used as an auxiliary tool to detect the presence of DNA damage expression signatures in the analyzed expression data set using the NTP web service (note that limitations due to differences between actual and recommended specimen type/exposure time/species will be considered).

5.4. OUTCOMES CONSIDERED IN THE CR(VI) ASSESSMENT

As previously stated in Section 3.2, the assessment will evaluate evidence for all cancer outcomes, and will evaluate noncancer effects for the following potential target systems: respiratory, GI, hepatic, hematological, immunological, reproductive, and developmental. The systematic review will focus on identifying data from inhalation exposures that are useful for deriving quantitative estimates for lung cancer and nasal effects rather than revisiting the qualitative identification of hazard for these outcomes. Additional details on how studies were screened and sorted are contained in Sections 4.3 and 4.4.

The endpoints that will be the primary focus of the outcome-specific evaluations—grouped by health outcome—are identified in Tables 6 and 7, along with the number of studies that examined these endpoints. Identification of these endpoints will guide the development of endpoint-specific study evaluation criteria (discussed further in Section 6). Table 8 provides an inventory of a selection of categories used when screening studies identified as "potentially relevant supplemental materials." This table is not comprehensive but provides a high-level indication of the relative density of publications in these reference topic areas. A graphical representation of the information in Table 8 for mechanistic studies identified from the "potentially relevant supplemental materials" is provided in EPA's version of Health Assessment Workspace Collaborative (HAWC), a free and open source web-based software application (https://hawcprd.epa.gov/lit/assessment/100500006/references/visualization/).[5]

[5]HAWC: A Modular Web-Based Interface to Facilitate Development of Human Health Assessments of Chemicals. https://hawcproject.org/portal/.

Table 6. Outcomes and associated endpoints to be considered for animal study evaluation

Health outcome and endpoints	Number of references
Gastrointestinal tract (oral)	4
Epithelial effects of small intestine	4
Stomach ulcer	2
Tumors of the GI tract	2
Respiratory tract (inhalation)	7
Nasal	2
General respiratory and pulmonary	5
Tumors of the lung	2
Hepatic (oral)	13
Clinical chemistry changes	10
Histopathological changes	11
Organ-weight changes	7
Hepatic (inhalation)	4
Clinical chemistry changes	3
Histopathological changes	3
Organ-weight changes	3
General (including gross changes, liver disease mortality)	1
Hematological (oral)	9
Clinical chemistry changes	9
Hematological (inhalation)	4
Clinical chemistry changes	4
Immune (oral)	5
Clinical chemistry and functional assays	3
Histopathological changes	2
Organ-weight changes	2
Immune (inhalation)	3

Health outcome and endpoints	Number of references
Clinical chemistry and functional assays	2
Organ-weight changes	3
Reproductive/developmental (oral)	40
Male reproductive	14
Female reproductive	10
Developmental (in utero and postnatal)	20
Reproductive/developmental (inhalation)	3
Male reproductive	3
PBPK modeling (see Section 6.4)	8

Note: Number of references indicates studies examining the outcome and associated endpoints, not the number of observed effects. Some studies are counted in multiple categories.

Table 7. Outcomes and associated endpoints to be considered for human study evaluation

Health outcome and endpoints	Number of references
Lung cancer (inhalation)	10
Other cancer (inhalation)	1
Cancer (oral route of exposure)	7
Cancer in offspring (inhalation)	2
Respiratory noncancer, lung	6
Respiratory noncancer, nasal	11
Asthma	5
Hepatic	8
Hematological	5
Immunological	8
Reproductive and developmental	13
PBPK modeling (see Section 6.4)	7

Note: Number of references indicates studies examining the outcome and associated endpoints, not the number of observed effects. Some studies are counted in multiple categories.

Table 8. Inventory of selected reference topics screened as "potentially relevant supplemental material" to be considered in the assessment

Reference topic	Number of references	
	Animal[a]	Human
In vivo toxicokinetics	48	6
Oral	8	6
Inhalation	3	0
Other	38	0
In vitro/ex vivo toxicokinetics	8	16
Gastric systems	4	6
Red blood cells	4	10
Mechanistic ADME	30	13
Liver	15	3
Gastrointestinal	2	0
Lung	4	6
Red blood cells	1	4
Other	10	0
Biomonitoring and biomarkers[b]	N/A	18
Blood/plasma/red blood cells	N/A	9
Urine	N/A	13
Other	N/A	6
Epidemiology studies related to included studies	N/A	18
Mechanistic studies (total number of studies)		
Cancer (843)	358	334
Electrophilicity (144)	88	42
Genotoxicity (413)	183	172
Altered DNA repair (78)	27	50
Epigenetic alterations (24)	3	14
Oxidative stress (255)	100	105

Reference topic	Number of references	
	Animal[a]	Human
Chronic inflammation (24)	9	9
Immunosuppression (3)	4	2
Receptor-mediated effects (111)	66	104
Immortalization/transformation (38)	16	26
Altered cell proliferation, death, or nutrient supply (224)	58	99
Gastrointestinal (31)	21	15
Respiratory (112)	53	128
Hepatic (59)	85	19
Hematological (11)	8	13
Immune (24)	27	27
Reproductive or developmental (38)	33	4

N/A = not applicable.
[a]Count does not include nonmammalian animal models or acellular systems.
[b]Count does not include epidemiology studies reporting human biomarker data.

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

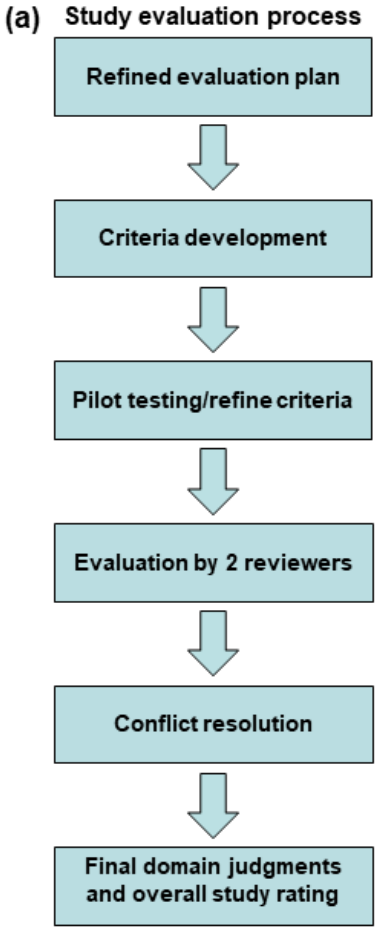
The general approach for evaluating primary health effect studies meeting PECO criteria for all study types is described in Section 6.1; the specifics of applying the approach for evaluating epidemiology and animal toxicology studies are described separately in Sections 6.2 and 6.3, respectively. Different approaches are used for evaluating PBPK models (see Section 6.4) and mechanistic studies (see Sections 6.5 and 9.2).

6.1. STUDY EVALUATION OVERVIEW FOR HEALTH EFFECT STUDIES

Key concerns for the review of epidemiology and animal toxicology studies are potential bias (factors that affect the magnitude or direction of an effect in either direction) and insensitivity (factors that limit the ability of a study to detect a true effect; low sensitivity is a bias towards the null when an effect exists). The study evaluations are aimed at discerning the expected magnitude of any identified limitations (focusing on limitations that could substantively change a result), considering also the expected direction of the bias. 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 2.

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 are developed in consultation with topic-specific technical experts, and existing guidance documents will be used when available, including EPA guidance for carcinogenicity, neurotoxicity, reproductive toxicity, and developmental toxicity (U.S. EPA, 2005a, 2002, 1998a, 1996, 1991). These 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 methodologic 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.



(b) Individual evaluation domains

Animal	Epidemiology
Selection and performance • Allocation • Observational bias/blinding	Participant selection
Confounding/variable control	Confounding
Selective reporting and attrition	Selective reporting
Exposure methods sensitivity • Chemical administration and characterization • Exposure timing, frequency, and duration	Exposure measurement
Outcome measures and results display • Endpoint sensitivity and specificity • Results presentation	Outcome ascertainment Analysis
Reporting quality	Other Sensitivity

Domain judgments

Judgment	Interpretation
Good	Appropriate study conduct relating to the domain and minor deficiencies not expected to influence results.
Adequate	A study that may have some limitations relating to the domain, but they are not likely to be severe or to have a notable impact on results.
Deficient	Identified biases or deficiencies interpreted as likely to have had a notable impact on the results or prevent reliable interpretation of study findings.
Critically Deficient	A serious flaw identified that makes the observed effect(s) uninterpretable. Studies with a critical deficiency will almost always be considered "uninformative" overall.

Overall study rating for an outcome

Rating	Interpretation
High	No notable deficiencies or concerns identified; potential for bias unlikely or minimal; sensitive methodology.
Medium	Possible deficiencies or concerns noted, but resulting bias or lack of sensitivity is unlikely to be of a notable degree.
Low	Deficiencies or concerns were noted, and the potential for substantive bias or inadequate sensitivity could have a significant impact on the study results or their interpretation.
Uninformative	Serious flaw(s) makes study results unusable for hazard identification or dose response.

Figure 2. Overview of Integrated Risk Information System (IRIS) study evaluation process.

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,[6] 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 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 to seek missing information is largely based on the likelihood that such information would affect the overall confidence of the study. Outreach to study authors will be documented and considered unsuccessful if researchers do not respond to an email or phone request within one month of the attempt to contact.

For each outcome in a study,[7] reviewers will reach a consensus judgment of *good*, *adequate*, *deficient*, *not reported*, or *critically deficient* for each evaluation domain (see Sections 6.2 and 6.3 for a description of evaluation domains for epidemiology and experimental animal studies). 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 use for identifying individual hazards. Study limitations specific to the usability of the study for dose-response analysis may be important for later decisions but 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 minor deficiencies that are noted would not be expected to influence the study results.

- *Adequate* indicates a judgment that there may be 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 prevent 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 of unreported items and severity of other limitations identified in the study, it may or may not be worth reaching out to the study authors for this information (see discussion below).

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

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

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

    - An inhalation study of Cr(VI) in which the only control group is intentionally or unintentionally infected with a respiratory virus (confounding/variable control).

    - An oral ingestion study of Cr(VI) in which the chemical compound is not stated, drinking water or gavage administration is not specified, control group exposure and husbandry not specified, and the oral doses are not provided or cannot be verified due to missing information (exposure methods sensitivity, reporting quality).

    - A reproductive study of Cr(VI) in which rodents were administered high doses (known to induce severe toxicity and death), and the numbers of dams in the results are less than the sample sizes stated in the methods, with no documentation of animal deaths (reporting or attrition)

Once the evaluation domains have been rated, the identified strengths and limitations will be considered as a whole to reach a study confidence classification of *high*, *medium*, or *low* confidence, or *uninformative* for a specific health outcome. This classification is based on the reviewer judgments across the evaluation domains and includes consideration of the likely impact the noted deficiencies in bias and sensitivity, or inadequate reporting have on the results. 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 19 and 20), and are generally not used as the primary sources of information for hazard identification or to derive toxicity values unless they are the only studies available. Studies rated as *low* confidence only because of sensitivity concerns about bias towards the

null will be asterisked or otherwise noted because these studies may 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 recorded in the EPA's version of HAWC. Final study evaluations housed in HAWC will be made available when the draft is publicly released. The study confidence classifications and their rationales will be carried forward and considered as part of evidence synthesis (see Section 9) to aid in the interpretation of results across studies.

6.2. EPIDEMIOLOGY STUDY EVALUATION

Evaluation of epidemiology studies of health effects to assess risk of bias and study sensitivity will be conducted for the following domains: exposure measurement, outcome ascertainment, participant selection, potential confounding, analysis, study sensitivity, and selective reporting. Bias can result in false positives and negatives, while study sensitivity is typically concerned with identifying the latter.

The principles and framework used for evaluating epidemiology studies are based on the Cochrane Risk of Bias in Nonrandomized Studies of Interventions [ROBINS-I; (Sterne et al., 2016)] but modified to address environmental and occupational exposures. The underlying philosophy of ROBINS-I is to describe attributes of an "ideal" study with respect to each of the evaluation domains (e.g., exposure measurement, outcome classification, etc.). Core and prompting questions are used to collect information to guide the evaluation of each domain.

Core and prompting questions for each domain, as well as additional considerations that apply to most outcomes, are presented in Table 9. Core questions represent key concepts while the prompting questions help the reviewer focus on relevant details under each key domain. Exposure- and outcome-specific criteria to use during evaluation of the 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 10.

Exposures to Cr(VI) by the inhalation and oral routes may be assessed based on administered dose or concentration, biomonitoring data (e.g., urine, blood, or other specimens), environmental or occupational-setting measures (e.g., air, water, dust levels), or job title or

residence. Air concentration measurements are preferred to biomarker measurements for the assessment of human exposure by inhalation in epidemiology studies. Studies in which human exposure is quantified by measurements of total chromium in urine, blood, plasma, or erythrocytes will be considered for determination of hazard if conducted in workers with known occupational exposure to Cr(VI). Air concentrations of Cr(VI) are correlated with measurements of total chromium in these biological matrices (Kuo et al., 1997; Miksche and Lewalter, 1997). However, uncertainty in biomarker measurements arises from reduction of Cr(VI) to Cr(III) throughout the body (NIOSH, 2013a). The rate at which Cr(VI) is reduced to Cr(III) following exposure varies by individual, further contributing to uncertainty in biomarker measurements.

When available, existing outcome-specific standard protocols for research studies will be consulted in developing outcome-specific criteria for evaluating epidemiology studies. For example, guidelines published by the American Thoracic Society for collecting spirometry measurements will inform evaluations of epidemiology studies of pulmonary function (Culver et al., 2017; Miller et al., 2005; ATS, 1995, 1987). Likewise, EPA will refer to World Health Organization (WHO) protocols when evaluating epidemiologic studies of semen parameters to assess toxicity to the male reproductive system (WHO, 2010, 1999).

Table 9. Questions to guide the development of criteria for each domain in epidemiology studies

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

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
<u>Outcome ascertainment</u> 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*.

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
Participant selection Is there evidence that selection into or out of the study (or analysis sample) was jointly related to exposure and to outcome?	For longitudinal cohort: • Did participants volunteer for the cohort based on knowledge of exposure and/or preclinical disease symptoms? Was entry into the cohort or continuation in the cohort related to exposure and outcome? For occupational cohort: • Did entry into the cohort begin with the start of the exposure? • Was follow-up or outcome assessment incomplete, and if so, was follow-up related to both exposure and outcome status? • Could exposure produce symptoms that would result in a change in work assignment/work status ("healthy worker survivor effect")? For case-control study: • Were controls representative of population and time periods from which cases were drawn? • Are hospital controls selected from a group whose reason for admission is independent of exposure? • Could recruitment strategies, eligibility criteria, or participation rates result in differential participation relating to both disease and exposure? For population-based survey: • Was recruitment based on advertisement to people with knowledge of exposure, outcome, and hypothesis?	Were differences in participant enrollment and follow-up evaluated to assess bias? If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)? 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	These considerations may require customization to the outcome. This could include determining what study designs effectively allow analyses of associations appropriate to the outcome measures (e.g., design to capture incident vs. prevalent cases, design to capture early pregnancy loss). *Good* • Minimal concern for selection bias based on description of recruitment process (e.g., selection of comparison population, population-based random sample selection, recruitment from sampling frame including current and previous employees). • Exclusion and inclusion criteria 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 and/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

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
		differential selection is likely?	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)? • Accurate information on potential confounders and statistical adjustment procedures? • Lack of association between confounder and outcome, or confounder and exposure in the study? • Information from other sources? Is the assessment of confounders based on a thoughtful review of published literature, potential relationships (e.g., as can be gained through directed acyclic graphing), and minimizing potential overcontrol (e.g., inclusion of a variable on the pathway between exposure and outcome)?	If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations require customization to the exposure and outcome, but this may be limited to identifying key covariates. *Good* • Conveys strategy for identifying key confounders. This may include: a priori biological considerations, published literature, causal diagrams, or statistical analyses; with recognition that not all "risk factors" are confounders. • Inclusion of potential confounders in statistical models not based solely on statistical significance criteria (e.g., $p < 0.05$ from stepwise regression). • Does not include variables in the models that are likely to be influential colliders or intermediates on the causal pathway. • Key confounders are evaluated appropriately and considered to be unlikely sources of substantial confounding. This often will include: o Presenting the distribution of potential confounders by levels of the exposure of interest and/or the outcomes of interest (with amount of missing data noted); o Considering that potential confounders were rare among the study population or were expected to be poorly correlated with exposure of interest; o Considering the most relevant functional forms of potential confounders; and o Examining 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

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
			(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 the outcomes and exposure levels) are not presented; or • Strategy of evaluating confounding is unclear or is not recommended (e.g., only based on statistical significance criteria or stepwise regression [forward or backward elimination]). *Critically deficient* • Includes variables in the models that are colliders and/or intermediates in the causal pathway, indicating that substantial bias is likely from this adjustment; or • Confounding is likely present and not accounted for, indicating that all 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	• Are missing outcome, exposure, and covariate data recognized, and if necessary, accounted for in the analysis? • Does the analysis appropriately consider variable distributions and modeling assumptions?	If there is a concern about the potential for bias, what is the predicted direction or distortion of the	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.

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
familiarity with the data and 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)?	bias on the effect estimate (if there is enough information)?	• Quantitative results presented (effect estimates and confidence limits or variability in estimates; i.e., not presented only as a *p*-value or "significant"/"not significant"). • 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 limit of detection (and percentage below the limit of detection), and decision to use log transformation. • Includes analyses that address robustness of findings, e.g., examination of exposure-response (explicit consideration of nonlinear possibilities, quadratic, spline, or threshold/ceiling effects included, when feasible); relevant sensitivity analyses; effect modification examined based only on a priori rationale with sufficient numbers. • No deficiencies in analysis evident. Discussion of some details 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).

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
			• Effect estimate and *p*-value presented, without standard error or confidence interval. • Results presented as statistically "significant"/"not significant." *Critically deficient* • Results of analyses of effect modification examined without clear a priori rationale and without providing main/principal effects (e.g., presentation only of statistically significant interactions that were not hypothesis driven). Analysis methods are not appropriate for design or data of the study.
Selective reporting Is there reason to be concerned about selective reporting?	• Were results provided for all the primary analyses described in the methods section? • Is there appropriate justification for restricting the amount and type of results that are shown? • Are only statistically significant results presented?	If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?	These considerations generally do not require customization and 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.

Domain and core question	Prompting questions	Follow-up questions	Considerations that apply to most exposures and outcomes
Sensitivity Is there a concern that sensitivity of the study is not adequate to detect an effect?	• Is the exposure range adequate? • 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 primary hypotheses in study. • 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 for outcome development (i.e., adequate follow-up interval). • The study was adequately powered to observe an effect. • No other concerns raised regarding study sensitivity. *Deficient* • Concerns were raised about the issues described for *good* that are expected to notably decrease the sensitivity of the study to detect associations for the outcome.

Table 10. Information relevant to evaluation domains for epidemiology studies

Domain	Types of information that may need to be collected or are important for evaluating the domain
Exposure measurement	Source(s) of exposure (e.g., consumer products, occupational, an industrial accident) and source(s) of exposure data, blinding to outcome, level of detail for job history data, when measurements were taken, type of biomarker(s), assay information, reliability data from 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; degree of exposure to the confounder in the population.
Analysis	Extent (and if applicable, treatment) of missing data for exposure, outcome, and confounders; approach to modeling; classification of exposure and outcome variables (continuous vs. categorical); testing of assumptions; sample size for specific analyses; and relevant sensitivity analyses.
Sensitivity	What are the ages of participants (e.g., not too young in studies of pubertal development)? What is the length of follow-up (for outcomes with long latency periods)? Choice of referent group, the exposure range, and the level of exposure contrast between groups (i.e., the extent to which the "unexposed group" is truly unexposed, and the prevalence of exposure in the group designated as "exposed").
Selective reporting	Are results presented with adequate detail for all the endpoints and exposure measures reported in the methods section, and are they relevant to the PECO? Are results presented for the full sample as well as for specified subgroups? Were stratified analyses (effect modification) motivated by a specific hypothesis?

6.3. EXPERIMENTAL ANIMAL STUDY EVALUATION

The evaluation of experimental animal studies applies principles similar to those described above for evaluating epidemiology studies. The evaluation process focuses on assessing aspects of the study design and conduct through three broad types of evaluations: reporting quality, risk of bias, and study sensitivity. A set of domains with accompanying core questions 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). Assessment-specific considerations are documented and made publicly available with the assessment.

Each domain receives a consensus judgment of *good*, *adequate*, *deficient*, *not reported*, or *critically deficient* (as described in Section 6.1) accompanied by a rationale for the judgment. Once all domains are rated, an overall confidence classification of *high*, *medium*, or *low* confidence or *uninformative* is assigned (as described in Section 6.1). The rationale for the classification, including a brief description of any identified strengths and/or limitations from the domains and their potential impact on the overall confidence determination, should be documented clearly and consistently. This rationale should, to the extent possible, reflect an interpretation of the potential influence on the results (including the direction and/or magnitude of influence).

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

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

Evaluation 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.

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.
				  ○ (Interpreted as *adequate*) The potential concern for bias was mitigated based on use of automated/computer driven systems, standard laboratory kits, relatively simple, objective measures (e.g., body or tissue weight), or screening-level evaluations of histopathology.
				  ○ (Interpreted as *deficient*) The potential impact on the results is major (e.g., outcome measures are highly subjective).
				• *Critically deficient*: Strong evidence for observational bias that impacted the results.

Evaluation concern		Domain—core question	Prompting questions	General considerations
Risk of bias (continued)	**Confounding/variable control**	**Confounding** Are variables with the potential to confound or modify results controlled for and consistent across all experimental groups?	For each study: • Are there differences across the treatment groups (e.g., co-exposures, vehicle, diet, palatability, husbandry, health status, etc.) that could bias the results? • If differences are identified, to what extent are they expected to impact the results?	These considerations may need to be refined by assessment teams, as the specific variables of concern can vary by experiment or chemical. A judgment and rationale for this domain should be given for each cohort or experiment in the study, noting when the potential for confounding is restricted to specific endpoints/outcomes. • *Good*: Outside of the exposure of interest, variables that are likely to confound or modify results appear to be controlled 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.

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 and/or attrition 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 is available from supplemental material. If results omissions or animal attrition are identified, the authors provide an explanation, and these are not expected to impact the interpretation of the results. • *Adequate*: Quantitative or qualitative results are reported for most prespecified outcomes (explicitly stated or inferred), exposure groups and evaluation time points. Omissions and/or attrition are not explained but are not expected to significantly impact the interpretation of the results. • *Deficient*: Quantitative or qualitative results are missing for many prespecified outcomes (explicitly stated or inferred), exposure groups and evaluation time points and/or high animal attrition; omissions and/or attrition are not explained and may significantly impact the interpretation of the results. • *Critically deficient*: Extensive results omission and/or animal attrition are identified and prevents comparisons of results across treatment groups.

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity	**Exposure methods sensitivity**	**Chemical administration and characterization** Did the study adequately characterize exposure to the chemical of interest and the exposure administration methods? *Note: Consideration of the appropriateness of the route of exposure is not evaluated at the individual study level. Relevance and utility of the routes of exposure are considered in the PECO criteria for study inclusion and during evidence synthesis.*	For each study: • Does the study report the source and purity and/or composition (e.g., identity and percent distribution of different isomers) of the chemical? If not, can the purity and/or composition be obtained from the supplier (e.g., as reported on the website)? • Was independent analytical verification of the test article purity and composition performed? • Did the authors take steps to ensure the reported exposure levels were accurate? • Are there concerns about the methods used to administer the chemical (e.g., inhalation chamber type, gavage volume, etc.)? For inhalation studies: • Were target concentrations confirmed using reliable analytical measurements in chamber air?	It is essential that these considerations are considered, and potentially refined, by assessment teams because 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. For inhalation studies, chemical concentrations in the exposure chambers are verified using reliable analytical methods. • *Adequate*: Some uncertainties in the chemical administration and characterization are identified but these are expected to have minimal impact on interpretation of the results (e.g., source and vendor-reported purity are presented, but not independently verified; purity of the test article is suboptimal but not concerning; for inhalation studies, actual exposure concentrations are missing or verified with less reliable methods).

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity (continued)	**Exposure methods sensitivity (continued)**	**Chemical administration and characterization** (continued)	For oral studies: • If necessary based on consideration of chemical specific-knowledge (e.g., instability in solution; volatility) and/or exposure design (e.g., the frequency and duration of exposure), were chemical concentrations in the dosing solutions or diet analytically confirmed?	• *Deficient*: Uncertainties in the exposure characterization are identified and expected to substantially impact the results (e.g., source of the test article is not reported; levels of impurities are substantial or concerning; deficient administration methods, such as use of static inhalation chambers or a gavage volume considered too large for the species and/or life stage at exposure). • *Critically deficient*: Uncertainties in the exposure characterization are identified, and there is reasonable certainty that the results are largely attributable to factors other than exposure to the chemical of interest (e.g., identified impurities are expected to be a primary driver of the results).
		Exposure timing, frequency, and duration Was the timing, frequency, and duration of exposure sensitive for the endpoint(s)/outcome(s) of interest?	For each endpoint/outcome or grouping of endpoints/outcomes in a study: • Does the exposure period include the critical window of sensitivity? • Was the duration and frequency of exposure sensitive for detecting the endpoint of interest?	Considerations for this domain are highly variable depending on the endpoint(s)/outcome(s) of interest and must be refined by assessment teams. A judgment and rationale for this domain should be given for each endpoint/outcome or group of endpoints/outcomes investigated in the study. • *Good*: The duration and frequency of the exposure was sensitive, and the exposure included the critical window of sensitivity (if known). • *Adequate*: The duration and frequency of the exposure was sensitive, and the exposure covered most of the critical window of sensitivity (if known). • *Deficient*: The duration and/or frequency of the exposure is not sensitive and did not include the majority of the critical window of sensitivity (if known). These limitations are expected to bias the results towards the null. • *Critically deficient*: The exposure design was not sensitive and is expected to strongly bias the results towards the null. The rationale should indicate the specific concern(s).

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, and/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, etc.). • Decreased specificity or sensitivity of the response due to the timing of endpoint evaluation, as compared to exposure (e.g., short-acting depressant or irritant effects of chemicals; insensitivity due to prolonged period of nonexposure prior to testing).

Evaluation concern		Domain—core question	Prompting questions	General considerations
Sensitivity (continued)	**Outcome measures and results display (continued)**	**Results presentation** Are the results presented in a way that makes the data usable and transparent?	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).

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 towards the null) for additional consideration during evidence synthesis. If the study is otherwise well conducted and an effect is observed, the confidence may be increased.*	For each endpoint/outcome or grouping of endpoints/outcomes in a study: • Were concerns (i.e., limitations or uncertainties) related to the reporting quality, risk of bias, or sensitivity identified? • If yes, what is their expected impact on the overall interpretation of the reliability and validity of the study results, including (when possible) interpretations of impacts on the magnitude or direction of the reported effects?	The overall confidence rating considers the likely impact of the noted concerns (i.e., limitations or uncertainties) in reporting, bias, and sensitivity on the results. A confidence rating and rationale should be given for each endpoint/outcome or group of endpoints/outcomes investigated in the study. Confidence ratings are described above (see Section 6.1).

OECD = Organisation for Economic Co-operation 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; U.S. EPA, 2002).

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

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

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 alternate quantitative approach. Data for PBPK models may come from studies across various species and may be in vitro or in vivo in design. Because Cr(VI) can be reduced to Cr(III) extracellularly by biological fluids (e.g., gastric juices) of humans and rodents (De Flora et al., 1997), ex vivo studies and models are also available. The relationship between ex vivo and whole-body toxicokinetic models of Cr(VI) for the oral route of exposure is presented below in Figure 3.

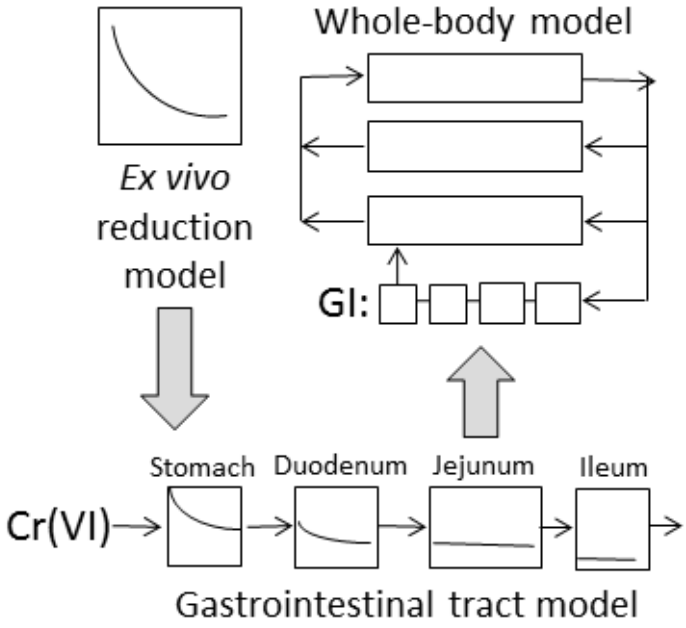


Figure 3. Relationship between ex vivo reduction models, in vivo gastric models, and whole-body physiologically based pharmacokinetic (PBPK) models.

In the trivalent state, chromium is poorly absorbed by cells and has not been shown to induce the same effects as Cr(VI) (Collins et al., 2010). Thus, extracellular reduction is a pathway for detoxification because it decreases the systemic uptake and distribution of Cr(VI) and the exposure of epithelial cells to Cr(VI). Following oral ingestion, most extracellular reduction and

detoxification will occur in the stomach prior to systemic absorption due to the acidity of gastric juice, and the length of time ingested water and food are stored in the stomach. However, this mechanism is less important following inhalation exposure, because the thin layer of respiratory tract lining fluid is less acidic and less effective at reducing Cr(VI) (Krawic et al., 2017; Ng et al., 2004). Deposition in the lung is not uniform, and particulates may locally accumulate at high quantities in susceptible areas such as airway bifurcation sites (Balashazy et al., 2003). This is supported by studies showing high chromium deposition at these sites in the lungs of chromate workers, and a correlation between lung chromium burden and lung cancer (Kondo et al., 2003; Ishikawa et al., 1994a, b).

Because extracellular gastric reduction kinetics are expected to significantly impact dosimetry, the scope of the PBPK model evaluations for this assessment will be limited to models accounting for Cr(VI) reduction in the stomach compartment and interspecies differences in gastric pH and physiology (mice, rats, and humans). For the inhalation route of exposure, the regional deposited dose ratio (RDDR) for the respiratory tract region of interest, estimated by airway particle deposition modeling, will be used to account for species differences (U.S. EPA, 1994). Route-to-route extrapolation will not be considered.

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

PBPK models were identified from the literature search, screening, and survey processes (see Table 12). The two models listed in the bottom two rows of Table 12 [referenced by Schlosser and Sasso (2014), Sasso and Schlosser (2015), Kirman et al. (2017), and Kirman et al. (2016)] will be evaluated for this assessment because they are the only models incorporating the effects of gastric pH and physiology on Cr(VI) toxicokinetics of mice, rats, and humans.

Parameters and codes from the earlier models may still undergo limited evaluations due to the shared lineage in derivation. Scientific or technical errors in earlier models may propagate to the later versions. For example, Kirman et al. (2017) and Kirman et al. (2016) supersedes Kirman et al. (2012) and Kirman et al. (2013), and both sets of models use many of the same data sets, codes, and parameters as the O'Flaherty models. The Sasso and Schlosser (2015) model uses codes and parameters from the Kirman et al. (2012) and Kirman et al. (2013) models. PBPK parameters that originated from the O'Flaherty models may need to be evaluated if they are used in the later Kirman et al. (2017), Kirman et al. (2016), and Sasso and Schlosser (2015) models.

Table 12. Physiologically based pharmacokinetic models for Cr(VI)

Reference	Species	Notes
O'Flaherty (1996) O'Flaherty (1993) O'Flaherty et al. (2001) O'Flaherty and Radike (1991)	Rat	Compartments include kidney, liver, bone, GI tract, two lung pools (for inhalation only), plasma, red blood cells, and lumped compartments for remaining tissues (rapidly and slowly perfused). A single lumped compartment represents the GI tract, and reduction kinetics do not include pH-reduction relationships. This model is not readily extendable to the mouse.
O'Flaherty et al. (2001)	Human	Calibrated to data from exposure via intravenous injection, gavage, inhalation (intratracheal), and drinking water (all data are from studies dated 1985 and earlier). Background Cr(III) exposure is simulated in the model and contributes to predicted total chromium concentrations.
Kirman et al. (2012)	Rat, mouse	Compartments include kidney, liver, bone, GI tract, plasma, red blood cells, and a lumped compartment for remaining tissues. A multicompartment model represents the GI tract (oral cavity, stomach, duodenum, jejunum, ileum, large intestine), with reduction kinetics based on the model by Proctor et al. (2012).
Kirman et al. (2013)	Human	Incorporates toxicokinetic data from experiments designed by the study authors and data from other studies. Only data for drinking water and dietary routes of exposure are incorporated. Total concentrations in control groups are subtracted from exposure groups to account for background Cr(III) exposure.
Schlosser and Sasso (2014); Sasso and Schlosser (2015)	Rat, mouse, human	Simulates Cr(VI) reduction kinetics and transit in the stomach. Incorporates toxicokinetic model of the stomach lumen by Kirman et al. (2012) and Kirman et al. (2013), but with a revised model for Cr(VI) reduction based on reanalysis of ex vivo data to improve model/data fit.
Kirman et al. (2017) and Kirman et al. (2016)	Rat, mouse, human	Same structure as Kirman et al. (2012) and Kirman et al. (2013), but incorporates a revised model for Cr(VI) reduction based on additional human gastric juice data. This model supersedes earlier models by the same investigators.

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

EPA will undertake model evaluation in accordance with criteria outlined by U.S. EPA (2018b). Judgments on the suitability of a model are separated into two categories: scientific and technical (see Table 13). The scientific criteria focus on whether the biology, chemistry, and other information available for chemical modes of action (MOAs) are justified (i.e., preferably with citations to support use) and represented by the model structure and equations. The scientific criteria are judged based on information presented in the publication or report that describes the model and does not require evaluation of the computer code. Preliminary technical criteria include

availability of the computer code and completeness of parameter listing and documentation. Studies that meet the preliminary scientific and technical criteria are then subjected to an in-depth technical evaluation, which includes a thorough review and testing of the computational code. The in-depth technical and scientific analyses focus on the accurate implementation of the conceptual model in the computational code, use of scientifically supported and biologically consistent parameters in the model, and reproducibility of model results reported in journal publications and other documents. This approach stresses (1) clarity in the documentation of model purpose, structure, and biological characterization; (2) validation of mathematical descriptions, parameter values, and computer implementation; and (3) evaluation of each plausible dose metric. The in-depth analysis is used to evaluate the potential value and cost of developing a new model or substantially revising an existing one. Further details of the initial and in-depth evaluation criteria can be found in the *Umbrella Quality Assurance Project Plan (QAPP) for PBPK Models* (U.S. EPA, 2018b).

6.5. MECHANISTIC STUDY EVALUATION

Sections 9 and 10 outline an approach for considering information from mechanistic studies (including in vitro, in vivo, ex vivo, and in silico studies) where the specific analytical approach is targeted to the assessment needs depending on the extent and nature of the human and animal evidence. In this way, the mechanistic synthesis might range from a high-level summary of potential mechanisms of action to specific, focused questions needed to fill data gaps identified from the human and animal syntheses and integration (e.g., shape of the dose-response curve, applicability of the animal evidence to humans, identifying susceptible populations). Individual study-level evaluation of mechanistic endpoints will typically be pursued only when the interpretation of studies is likely to significantly affect hazard conclusions or assumptions about dose-response analysis, and the issues that need resolution have not been sufficiently addressed in previous assessments or reviews published in peer-reviewed journals. Toxicogenomic studies will be evaluated using the criteria identified in the refined evaluation plan (see Section 5). If other mechanistic endpoints require study-level evaluation using endpoint-specific criteria that have not been predefined, these criteria will be described in the updated protocol released with the draft assessment.

Table 13. Criteria for evaluating physiologically based pharmacokinetic (PBPK) models

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

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

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

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

$BW^{3/4}$ = body weight raised to the ¾ power.

Assessing potential bias in in vitro studies is an active area of method development in the field of systematic review. Historically, most tools used to assess these studies have focused on reporting quality; tools to assess risk of bias (internal validity) of mechanistic evidence are not well established (NASEM, 2018; NTP, 2015), although current trends are to expand the assessment to include methodological quality with consideration of potential bias (IRIS, 2015). One of the more recently developed tools that has undergone user testing and refinement is the *Science in Risk Assessment and Policy* (SciRAP) approach for the evaluation of reliability for in vitro studies (Beronius et al., 2018; Molander et al., 2015; Beronius et al., 2014; Agerstrand et al., 2011; U.S. EPA, 2002). The IRIS Program is in the pilot phase of testing approaches for arriving at study level judgments for in vitro studies. Currently, two methods for evaluating in vitro mechanistic studies are being considered for use in IRIS assessments: (1) a tool used for conducting assessments under the Toxic Substances Control Act (TSCA), which uses a numerical scoring approach to rate studies (U.S. EPA, 2018a) and (2) the SciRAP tool (Beronius et al., 2018), which separately presents domain judgments for reporting quality, methodological quality, and relevance. A comparison of study level judgments based on use of both tools should assist in refining an approach for routine use in IRIS assessments. The IRIS Program is aware of other tools being developed (NASEM, 2018) and will monitor developments through its engagements with the systematic review community. The tool(s) and/or criteria used for testing specific questions that arise during the evaluation of mechanistic events in the chromium assessment will be described in the updated protocol released with the draft assessment.

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, as well as the results of the evaluation of sources of bias and sensitivity. The hazard evaluations will be organized around organ systems (e.g., respiratory, hepatic 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 14 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 14. 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 inter-related outcomes? If so, consider whether some outcomes are more useful and/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*, *uninformative*) 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?

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 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, or 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 listed in Appendix B. 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 PECO refinement 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 EPA HAWC in Excel format when the draft is publicly released.[8] 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 discussion or 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

[8]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.

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

For animal data already extracted to evidence tables released in 2014 (U.S. EPA, 2014b) and contained in Microsoft Word, data extraction procedures were followed according to data type.

- **Dichotomous data**: Following the 2014 public meetings, these data were revised to correct errors identified by public commenters, EPA staff, and contractors. Additional dichotomous data sets were extracted during this revision process. The revised dichotomous data tables will be imported into HAWC from Microsoft Word.

- **Continuous data**: Because the evidence tables released in 2014 expressed continuous data only as a percent control response, the values in those tables do not contain enough information for quality revisions or HAWC importation. As a result, the raw data (means and standard deviations or standard errors) will be re-extracted from the publications and entered into HAWC. For chronic studies that collected data at multiple sampling times (e.g., 4 days, 22 days, 3 months, 6 months, or 12 months), data extraction will be performed for the chronic sampling time only. On a case-by-case basis, data extraction at earlier sampling times will be performed (e.g., to illustrate the dynamic behavior of some endpoints).

- **Qualitative results**: Results that were only presented qualitatively by the study authors and extracted for the evidence tables released in 2014 were imported into appropriate HAWC text fields.

- ***Uninformative* and *low*-confidence studies**: Data and results from studies determined to be *uninformative* or *low* confidence by study evaluation (further described in Section 6) will generally not be imported into HAWC. HAWC entries for these studies may be limited to basic information about the references. Additional study information or data may be available in HAWC for these studies on a case-by-case basis (e.g., if HAWC data extraction occurred before final study evaluation).

For human data already extracted to evidence tables released in 2014 (U.S. EPA, 2014c) and contained in Microsoft Word, data extraction procedures will depend on the quality of the study and the study design. In general, study summary information will be imported into appropriate HAWC text fields for all studies that are evaluated in HAWC. If sufficient *medium*- and *high*-confidence studies are not available following study evaluation, data from *low*-confidence studies will be extracted. Studies will undergo a more thorough data extraction than was performed in 2014 (see Appendix B).

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 or not meta-analysis is feasible for an assessment (see Section 9.1). Many such data transformations can be performed automatically in HAWC. The following summary of effect size metrics by data type outlines issues in selecting the most appropriate common metric for a collection of related endpoints (Vesterinen et al., 2014).

Common metrics for continuous outcomes include:

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

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

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

Common metrics for event or incidence data include:

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

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

An additional approach for epidemiology studies 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 Cr(VI)/kg-day. When 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 if available. When possible, time-weighted average daily doses will be calculated from the start of the bioassay through the time of data collection. Otherwise, EPA defaults will be used (U.S. EPA, 1988), addressing age and study duration as relevant for the species/strain and sex of the animal of interest. Exposure levels in inhalation studies will be expressed in units of mg/m^3. Assumptions used in performing dose conversions will be documented.

Exposure levels will be converted to Cr(VI) equivalents depending on the chemical compound. For example, doses of test material administered as sodium dichromate dihydrate ($Cr_2H_4Na_2O_9$) were expressed as Cr(VI) using a molecular weight conversion of approximately 0.3490 g Cr(VI) per g $Cr_2H_4Na_2O_9$.

9. SYNTHESIS WITHIN LINES OF EVIDENCE

For the purposes of this assessment, evidence synthesis and integration are considered distinct, but related processes. 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 health hazard or smaller subset of related outcomes, EPA will separately synthesize the available human and animal health effect evidence. Mechanistic evidence is also considered, although the specific analytical approach is targeted to the assessment needs depending on the extent and nature of the human and animal evidence (see Sections 9.2 and 10). Each synthesis will be written to provide a summary discussion of the available evidence that may suggest causation 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 [(U.S. EPA, 2005a, 1994); see Table 15]. Importantly, the evidence synthesis process explicitly considers and incorporates the conclusions from the individual study evaluations (see Section 6).

Data permitting, the syntheses will also discuss analyses relating to potential susceptible populations[9]. 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 16). 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.

[9]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, as well as social and behavioral determinants that can modify the effect of a specific exposure. However, certain factors resulting in higher exposures to specific groups (e.g., proximity, occupation, housing) may not be analyzed to describe potential susceptibility among specific populations or subgroups.

Table 15. Information most relevant to describing primary considerations informing causality during evidence syntheses

Consideration	Description and synthesis methods
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)] based on analyses of potentially important explanatory factors such as: • Confidence in studies' results, including study sensitivity (e.g., some study results that appear to be inconsistent 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 epidemiology studies). • Specificity and sensitivity of the endpoint for evaluating the health effect in question (e.g., functional measures can be more sensitive than organ weights). • Populations or species, including consideration of potential susceptible groups or differences across 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. 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.
Strength (effect magnitude) and precision	Description: Examines the effect magnitude or relative risk, based on what is known about the assessed endpoint(s), and considers the precision of the reported results based on analyses of variability (e.g., confidence intervals; standard error). This 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.

Consideration	Description and synthesis methods
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 as well as the sensitivity and specificity of the measures used.
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 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 alternate or previously unidentified mechanisms of toxicity.

Consideration	Description and synthesis methods
Natural experiments	Description: Specific to epidemiology 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 lead in gasoline). Application: Compared to other observational designs, natural experiments have the benefit of dividing people into exposed and unexposed groups without them influencing their own exposure status. During synthesis, associations in *medium*- and *high*-confidence natural experiments can substantially reduce concerns about residual confounding.

Table 16. 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	Pre-existing conditions or disease such as psychosocial stress, 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). In this way, the mechanistic synthesis might range from a high-level summary of potential mechanisms of action to specific, focused questions needed to fill data gaps 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).

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, then a careful examination of risk of bias and sensitivity with potential impacts on the evidence synthesis conclusions will be included in the narrative.

As previously described, these syntheses will articulate the strengths and the weaknesses of the available evidence organized around the considerations described in Table 15, 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 an 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. 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

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 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. 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 a single, 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 difficult 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 will depend 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.

The synthesis of mechanistic information is used to inform 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 will focus 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.

To identify the focused set(s) of studies for use in analyses of critical mechanistic questions, the synthesis will apply a phased approach that progressively focuses the scope of the mechanistic information to be considered. This stepwise focusing, which began 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. Cr(VI) mechanistic information will be collected and inventoried (i.e., capturing details relating to exposure characteristics, model system, and assays tested to allow for sorting and retrieval to address critical mechanistic questions) for all health outcomes meeting PECO criteria, including cancer and effects on the GI, respiratory, reproductive, developmental, immune, and hematological systems. Other mechanistic information (e.g., relevant to non-PECO health outcomes) will be reviewed and sorted to facilitate later decisions, including identification of areas of research unexamined in the available human or animal health effect studies.

For cancer, it is acknowledged that the issue of whether Cr(VI) causes cancer by the oral route of exposure via a mutagenic mode of action is critical to address (see Section 2.3); therefore, a specific and thorough analysis integrating the evidence for potential mechanisms of cancer relevant to the oral route of exposure will be conducted. Given the focus of the lung cancer assessment on dose-response analysis, the mechanistic information relevant to cancer via the inhalation route will be investigated to identify and synthesize those studies that could influence the dose-response assessment for lung cancer, if available. It is not anticipated that other mechanistic analyses relevant to cancer will be conducted in the assessment; however, if other cancer types are identified that require a focused mechanistic analysis, these will be documented in the updated protocol released with the draft assessment. To facilitate the two primary mechanistic evaluations for

cancer, an inventory of the available mechanistic studies was developed. As shown in Table 8, mechanistic studies investigating genotoxicity, oxidative stress, alterations in cell proliferation and cell death, electrophilicity, receptor-mediated effects, altered DNA repair, immortalization, chronic inflammation, and epigenetic alterations have been identified in the mechanistic studies database relevant to cancer. Mechanistic events relevant to these characteristics will be investigated, and any areas lacking evidence will be identified. The identification of mechanistic evidence that may indicate potentially relevant susceptible subpopulations or life stages will be particularly important.

The information collected as described above (e.g., in sortable inventories) will be used to identify studies available for consideration in addressing the specific gaps in understanding identified as critical to address from the evaluation and synthesis of the human and animal lines of evidence, 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 considered relevant for a particular question), with the rationale for any focusing transparently documented. 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 will first prioritize studies based on their toxicological relevance to answering the specific question (e.g., model system, specificity of the assay for the effect of interest). For example, mechanistic information from in vivo studies will be analyzed first, with primary consideration given to endpoint-relevant routes. Analysis of ex vivo and in vitro studies will then be prioritized by those most informative to evaluating the mechanistic events indicated by the in vivo data, including studies conducted under conditions most relevant to human exposures and in model systems best replicating in vivo human biology. The path for focusing the mechanistic database will be documented in the updated protocol released with the draft assessment.

More rigorous analyses will be 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 study design aspects in the available human and animal health effect

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).

For the more rigorous mechanistic analyses, the review will be facilitated using 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 key characteristics of carcinogens have been used to organize the large mechanistic database relevant to cancer for Cr(VI) exposure (see Table 8) and will serve to organize the mechanistic analysis and help identify key events that will be evaluated using the MOA analysis framework described in EPA's cancer guidelines (U.S. EPA, 2005a). Similar approaches (e.g., identification of key characteristics or mechanistic events anchored to a specific health effect) will be used to organize mechanistic databases for noncancer health effects. The mechanistic analyses will inform the evidence integration across lines of evidence, as well as the dose-response analyses, that are described in Sections 11 and 12.

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 4):

- First, judgments regarding the strength of the evidence from the available human and animal studies are made in parallel. These judgments incorporate mechanistic evidence (or MOA understanding) in exposed humans and animals, respectively, 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.

- Second, the animal and human evidence judgments are combined to draw an overall conclusion(s) that incorporates inferences drawn based on information on the human relevance of the animal evidence (i.e., based on default assumptions or empirical evidence), coherence across the human and animal evidence streams, and susceptibility.

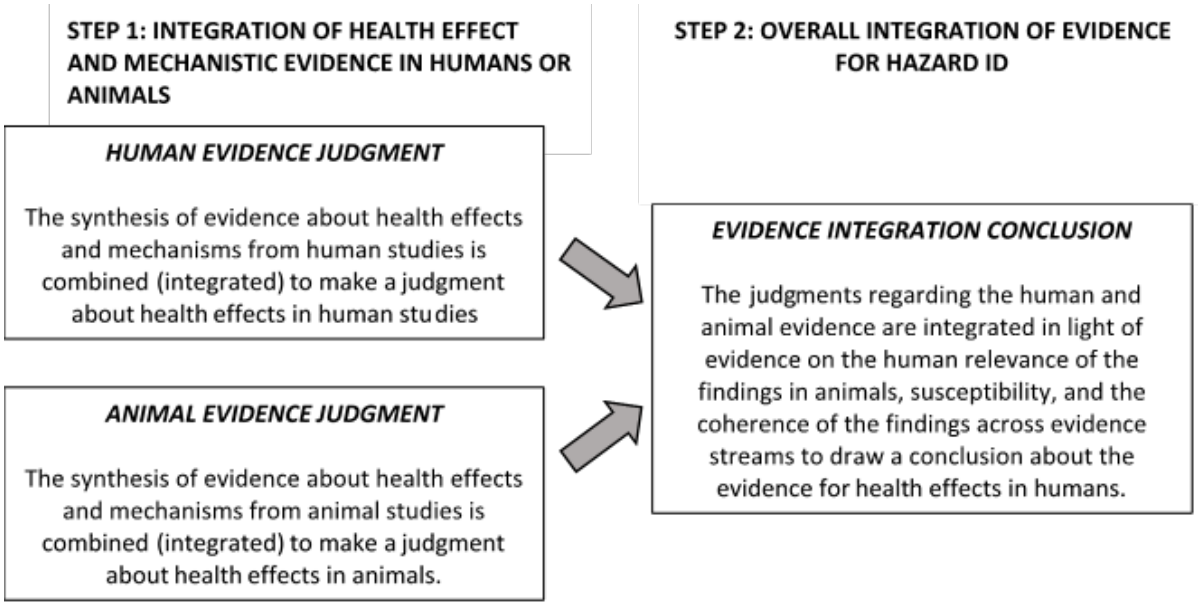


Figure 4. Process for evidence integration.

The decision points within the structured two-step evidence integration process will be summarized in an evidence profile table for each hazard (see Table 17 for a template) 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 level of confidence conclusion based on considering the body of evidence. As described in Section 9, the human, animal, and mechanistic syntheses serve as inputs providing a foundation for the evidence integration decisions; thus, the major conclusions from these syntheses will be summarized in the evidence profile table (see Table 17) 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 informed each judgment.

10.1. INTEGRATION WITHIN THE HUMAN AND ANIMAL EVIDENCE

As summarized above, before 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 15 (see Section 9) are evaluated in the context of how they impact the strength of evidence (see Table 18), and the judgments are reached using the structured frameworks explained in Tables 19 and 20 (for human and animal evidence, respectively). These judgments are summarized in tabular format using the template in Table 17 to transparently convey expert judgments made throughout the evidence synthesis and integration processes. The evidence profile table allows for consistent documentation of the supporting rationale for each decision.

Table 17. Evidence profile table template

Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of findings	Human and animal evidence judgments	Inference across lines of evidence	Overall evidence integration conclusion
[Health effect or outcome grouping]						
Evidence from human studies [route]					• Human relevance of findings in animals • Coherence across lines of evidence (i.e., for both health effect-specific and mechanistic data) • Other inferences ○ Information on susceptibility ○ MOA analysis inferences (e.g., cross-species inferences of toxicokinetics, or quantitative implications) ○ Relevant information from other sources (e.g., read across; other potentially related health hazards)	Describe conclusion(s) and primary basis for the integration of all available evidence (across human, animal, and mechanistic): • ⊕⊕⊕ Evidence demonstrates • ⊕⊕⊙ Evidence indicates • ⊕⊙⊙ Evidence suggests • ⊙⊙⊙ Evidence inadequate • – – – Strong evidence supports no effect
• References • Study confidence (based on evaluation of risk of bias and sensitivity) • Study design description	• Consistency or replication • Dose-response gradient • Coherence of observed effects (apical studies) • Effect size (magnitude, severity) • Mechanistic evidence providing plausibility • *Medium*- or *high*-confidence studies[a]	• Unexplained inconsistency • Imprecision • *Low*-confidence studies[a] or other concerns about methods or design across studies • Other (e.g., single/few studies) • Evidence demonstrating implausibility (e.g., mechanistic)	• Results information (general endpoints affected/unaffected) across studies • Human mechanistic evidence informing biological plausibility: discuss how data influenced the human evidence judgment (e.g., evidence of precursors in exposed humans) Could be multiple rows (e.g., grouped by study confidence or population) if this informs heterogeneity of results	Describe the strength of the evidence from human studies, and primary basis for judgment: • ⊕⊕⊕ *Robust* • ⊕⊕⊙ *Moderate* • ⊕⊙⊙ *Slight* • ⊙⊙⊙ *Indeterminate* • – – – *Compelling evidence of no effect*		

Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of findings	Human and animal evidence judgments	Inference across lines of evidence	Overall evidence integration conclusion
Evidence for an effect in animals [route]						
• References • Study confidence (based on evaluation of risk of bias and sensitivity) • Study design description	• Consistency or replication • Dose-response gradient • Coherence of observed effects (apical studies) • Effect size (magnitude, severity) • Mechanistic evidence providing plausibility • *Medium*- or *high*-confidence studies[a]	• Unexplained inconsistency • Imprecision • *Low*-confidence studies[a] or other concerns about methods or design across studies • Other (e.g., single/few studies) • Evidence demonstrating implausibility (e.g., mechanistic)	• Results information (general endpoints affected/unaffected) across studies • Animal mechanistic evidence informing biological plausibility: discuss how mechanistic data influenced the animal evidence judgment (e.g., evidence of coherent molecular changes in animal studies) Could be multiple rows (e.g., grouped by study confidence or population) if this informs heterogeneity of results	Describe the strength of the evidence from animal studies, and primary basis for judgment: • ⊕⊕⊕ *Robust* • ⊕⊕⊙ *Moderate* • ⊕⊙⊙ *Slight* • ⊙⊙⊙ *Indeterminate* • – – – *Compelling evidence of no effect*		

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

Table 18. 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 19 and 20 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 17.		
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, 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.

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 also may not be a monotonic dose-response (monotonicity should not necessarily be expected, e.g., different outcomes may be expected at low vs. high doses 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 epidemiology studies because of their observational nature).	• A lack of dose-response when expected based on biological understanding and having a wide-range of doses/exposures evaluated in the evidence base can decrease strength. • In experimental studies, strength may be decreased when effects resolve under certain experimental conditions (e.g., rapid reversibility after removal of exposure). However, many reversible effects are of high concern. Deciding between these situations is informed by factors such as the toxicokinetics of the chemical and the conditions of exposure [see U.S. EPA (1998a)], 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 toxicology 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 or decreased.
Coherence	• Biologically related findings within an organ system, or across populations (e.g., sex) increase strength, particularly when a temporal- or dose-dependent progression of related effects is observed within or across studies, or when related findings of increasing severity are observed with increasing exposure.	• An observed lack of expected coherent changes (e.g., well-established biological relationships) 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.

Consideration	Increased evidence strength (of the human or animal evidence)	Decreased evidence strength (of the human or animal evidence)
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).

For human and animal evidence, the analyses of each consideration in Table 18 will be used to develop a strength-of-evidence judgment. Tables 19 and 20 provide the judgments for each category and the criteria that will guide how to apply the judgments. Briefly, the terms *robust* and *moderate* are standardized characterizations for judgments on the extent of support provided by human or animal studies that the health effect(s) result from chemical exposure. 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. These terms are applied to human and animal evidence separately and are differentiated by the quantity and quality of information available to rule out alternative explanations for the results. 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 stronger judgments from being drawn. *Indeterminate* reflects evidence stream judgments when no studies are available, or situations in which the evidence is inconsistent and/or primarily of *low* confidence. *Compelling evidence of no effect* represents a situation in which extensive evidence across a range of populations and exposures has identified no effects/associations. This 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. 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 and is not a component of the strength-of-evidence framework for human or animal studies. If potential publication bias is evaluated for an outcome, it may inform the level of certainty regarding the completeness of the assessment database for that outcome.

Table 19. 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*.

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 20. 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*.
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*.

Within-stream strength-of-evidence judgment	Description
Slight (⊕⊙⊙) …evidence in animals *(signal of effect with large amount of uncertainty)*	Scenarios in which there is a signal of a possible effect, but the evidence is conflicting or weak. Most commonly, this includes situations 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*.
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. Based on the integration across lines of evidence, this stage culminates in an evidence integration narrative as described at the beginning of this chapter 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 the 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 21 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.

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 dose or exposure 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's cancer guidelines (U.S. EPA, 2005a). If a systematic review of more than one cancer type was conducted, then 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 streams, 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; dose) and susceptibility (e.g., genetics; life stage), as the data allow (Farland, 2005; U.S. EPA, 2005a, b).

Table 21. 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 outstanding questions or limitations may 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.

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**	An evidence base that suggests that (chemical) exposure may cause (health effect) in humans, but there are very few studies that have contributed to the evaluation, the evidence is very weak or conflicting, and/or the methodological conduct of the studies is poor. • This conclusion level is used if there is *slight* human evidence and *indeterminate* to *slight* animal evidence. • This conclusion level is also used with *slight* animal evidence and *indeterminate* to *slight* human evidence. • 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, outstanding issues regarding the *moderate* evidence have 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).
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.

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

[a]Evidence integration 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 only allow for evaluations at the category of health effects examined; this will more frequently be the case for conclusion levels of "evidence suggests" and "evidence is 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 (AOPs) and of the human implications of new toxicity testing methods (e.g., from high-throughput screening, from short-term in vivo testing of alternative species, or from new in vitro testing) will continue to increase. This may make possible the development of hazard conclusions when there are mechanistic or other relevant data that can be interpreted with a similar level of confidence to positive animal results in the absence of conventional studies in humans or in animals.

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

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) identification of biologically based benchmark response (BMR) levels. The pool of outcomes and study-specific endpoints will be discussed to identify which categories of effects and study designs are considered the strongest and most appropriate for quantitative assessment of a given health effect. Health effects 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. The existence and validity of PBPK models or toxicokinetic information that may allow the estimation of internal dose 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 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 and/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, 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: SELECTING STUDIES AND QUANTITATIVE ANALYSIS

The previous sections of this protocol describe how systematic review principles are applied to evaluate study quality (potential bias and sensitivity) and reach evidence synthesis and integration conclusions on health outcomes (or hazard identification) associated with exposure to the chemical of interest. Selection of specific data sets for dose-response assessment and performance of the dose-response assessment is conducted after hazard identification is complete and involves database and chemical-specific biological judgments. A number of EPA guidance and support documents detail data requirements and other considerations for dose-response modeling, especially EPA's *Benchmark Dose Technical Guidance* (U.S. EPA, 2012b), EPA's *Review of the Reference Dose and Reference Concentration Processes* (U.S. EPA, 2005a, 2002), *Guidelines for Carcinogen Risk Assessment* (U.S. EPA, 2005a), and *Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens* (U.S. EPA, 2005b). This section of the protocol provides an overview of considerations for conducting the dose-response assessment, particularly statistical considerations specific to dose-response analysis that support quantitative risk assessment. Importantly, these considerations do not supersede existing EPA guidance.

For IRIS assessments, dose-response assessments are typically performed for both noncancer and cancer hazards, and for both oral and inhalation routes of exposure following chronic exposure[10] to the chemical of interest, if supported by existing data. For noncancer hazards, an oral reference dose (RfD) and/or an inhalation reference concentration (RfC) are usually derived. An RfD or an RfC 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). These health effects may also include cancer effects [e.g., in a case 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]. Reference values are not predictive risk values; that is, they provide no information about risks at higher or lower exposure levels.

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) and/or an inhalation unit risk (IUR) are used to estimate human cancer risks. An OSF is a plausible upper-bound lifetime cancer risk from chronic ingestion

[10]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).

of a chemical per unit of mass consumed per unit body weight, per day (mg/kg-day). An IUR is a plausible upper-bound lifetime cancer risk from chronic inhalation of a chemical per unit of air concentration (expressed as ppm or $\mu g/m^3$). In contrast with reference values (RfVs), an OSF or IUR 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 Summary") of this assessment, IRIS will conduct the assessment with a goal of developing an RfD and RfC for the noncancer effects of Cr(VI) and quantitative cancer assessments for inhaled and ingested Cr(VI) consistent with the available mechanistic evidence.

The derivation of cancer risk estimates may also depend on the nature of the hazard conclusion. Specifically, EPA generally conducts dose-response assessments and derives cancer values for chemicals that are classified as *carcinogenic* or *likely to be carcinogenic* to humans. When there is *suggestive evidence* of carcinogenicity to humans, EPA generally would not conduct a dose-response assessment and derive a cancer value. Similarly, for noncancer outcomes, EPA will make decisions on whether to conduct dose-response assessments based on the strength of the evidence of a hazard. However, when the evidence includes a well-conducted study, quantitative analyses may be useful for some purposes, for example, providing a sense of the magnitude and uncertainty of potential risks, ranking potential hazards, or setting research priorities (U.S. EPA, 2005a).

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 22. 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.

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/lowest-observed-adverse-effect level [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 22. Attributes used to evaluate studies for derivation of toxicity values

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

Study attributes	Considerations	
	Human studies	**Animal studies**
Measurement of exposure	Studies that can reliably distinguish between levels of exposure in a time window considered most relevant for development of a causal effect are preferred. Exposure assessment methods that provide measurements at the level of the individual and that reduce measurement error are preferred. Measurements of exposure should not be influenced by knowledge of health outcome status.	Studies providing actual measurements of exposure (e.g., analytical inhalation concentrations vs. target concentrations) are preferred. Relevant internal dose measures may facilitate extrapolation to humans, as would availability of a suitable animal PBPK model in conjunction with an animal study reported in terms of administered exposure.
Measurement of health outcome(s)	Studies that can reliably distinguish the presence or absence (or degree of severity) of the outcome are preferred. Outcome ascertainment methods using generally accepted or standardized approaches are preferred.	
	Studies with individual data are preferred in general. Examples include: to characterize experimental variability more realistically, to characterize overall incidence of individuals affected by related outcomes (e.g., phthalate syndrome).	
	Among several relevant health outcomes, preference is generally given to those with greater biological significance.	
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).

Among the studies that support the 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, 2012b). These attributes support a more complete characterization of the shape of the exposure-response curve and decrease the uncertainty in the associated exposure-response metric (e.g., IUR or RfC) by reducing statistical uncertainty in the point of departure and minimizing the need for low-dose extrapolation. 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, 2012b).

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, 2012b; 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. This modeling yields a POD, an exposure level ideally near the lower end of the range of observation, without significant extrapolation to lower exposure levels. 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, if there are multiple tumor sites 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 MOA and quantitatively support model parameters that represent rates and other quantities associated with the key precursor events of the MOA. Toxicodynamic modeling is potentially the most comprehensive way to account for the biological processes involved in a response. Such models seek to reflect the sequence of key precursor events that lead to a response. Toxicodynamic models can contribute to dose-response assessment by revealing and describing nonlinear relationships between internal dose and response. Such models may provide a useful approach for analysis in the range of observation, provided the purpose of the assessment justifies the effort involved.

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. The 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, 2012b)]. 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. The POD is used as the starting point for subsequent extrapolations and analyses. For linear extrapolation of cancer risk, the POD is used to calculate an OSF or IUR, and for nonlinear extrapolation, the POD is used in calculating an RfD or RfC.

The response level at which the POD is calculated is guided by the severity of the endpoint. If linear extrapolation is used, selection of a response level corresponding to the point of departure is not highly influential, so standard values near the low end of the observable range are generally used (for example, 10% extra risk for cancer bioassay data, 1% for epidemiologic data, lower for rare cancers). Nonlinear approaches 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 life span. 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). Inhalation exposures are scaled using dosimetry models that apply species-specific physiologic and anatomic factors and consider whether the effect occurs at the site of first contact or after systemic circulation (U.S. EPA, 2012a; 1994, §3).

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

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

11.2.2. Extrapolation: Slope Factors and Unit Risks

An OSF or IUR facilitates estimation of human cancer risks when low-dose linear extrapolation for cancer effects is supported, particularly for chemicals with direct mutagenic activity or those for which the data indicate a linear component below the POD. Low-dose linear extrapolation is also used as a default when the data are insufficient to establish the MOA (U.S. EPA, 2005a). If data are sufficient to ascertain one or more MOAs 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 or unit risks, 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, then 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 and IUR for Cr(VI) conducted as part of the current assessment will be performed consistent with EPA guidance. For the oral assessment, both linear and nonlinear approaches will be presented for Cr(VI) carcinogenicity (U.S. EPA, 2005a) to provide insights into uncertainties related to model choice and mechanisms.

11.2.3. Extrapolation: Reference Values

Reference value derivation is EPA's most frequently used type of nonlinear extrapolation method and 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). Extrapolation between routes of exposure will not be performed for Cr(VI) (see Sections 3.1 and 6.4). Increasingly, data-based adjustments (U.S. EPA, 2014a) 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 that may range from detailed chemical-specific to default approaches (U.S. EPA, 2014a, 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, 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, 2014a, 2002).[11, 12] 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; 1998a, §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 no evidence is 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, 1998a, 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, 1998a, 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, 1998a, 1996, 1994, 1991). The size of the factor depends on the nature of the database deficiency. For example, the EPA typically follows the recommendation that a factor of 10 be applied if both a prenatal toxicity study and a two-generation reproduction study are missing and a factor of $10^{1/2}$ (i.e., 3) if either one or the other is missing (U.S. EPA, 2002, §4.4.5).

The POD for 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 and RfC for the noncancer effects of Cr(VI) will be conducted consistent with EPA guidance summarized above.

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

[12]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.

12. PROTOCOL HISTORY

1 Release date: March 15, 2019

REFERENCES

Agerstrand, M; Kuester, A; Bachmann, J; Breitholtz, M; Ebert, I; Rechenberg, B; Ruden, C. (2011). Reporting and evaluation criteria as means towards a transparent use of ecotoxicity data for environmental risk assessment of pharmaceuticals. Environ Pollut 159: 2487-2492. http://dx.doi.org/10.1016/j.envpol.2011.06.023.

Anger, G; Halstenberg, J; Hochgeschwender, K; Scherhag, C; Korallus, U; Knopf, H; Schmidt, P; Ohlinger, M. (2005). Chromium compounds. In B Elvers (Ed.), Ullmann's Encyclopedia of Industrial Chemistry (7th ed., pp. 1-36). Hoboken, NJ: WileyVCH Verlag GmbH & Co. http://dx.doi.org/10.1002/14356007.a07_067.

ATS (American Thoracic Society). (1987). Standardization of spirometry--1987 update. Am J Respir Crit Care Med 136: 1285-1298.

ATS (American Thoracic Society). (1995). Standardization of spirometry, 1994 update. American Thoracic Society. Am J Respir Crit Care Med 152: 1107-1136. http://dx.doi.org/10.1164/ajrccm.152.3.7663792.

ATSDR (Agency for Toxic Substances and Disease Registry). (2012). Toxicological profile for chromium. Atlanta, GA: US Department of Health and Human Services, Public Health Service. http://www.atsdr.cdc.gov/ToxProfiles/tp.asp?id=62&tid=17.

Bahadori, T; Thayer, K. (2018). National Academy of Sciences committee to review advances made to the IRIS process. Presentation presented at National Academy of Sciences Review of Advances Made to the IRIS Process: A Workshop, February 1-2, 2018, Washington, D.C.

Balashazy, I; Hofmann, W; Heistracher, T. (2003). Local particle deposition patterns may play a key role in the development of lung cancer. J Appl Physiol (1985) 94: 1719-1725. http://dx.doi.org/10.1152/japplphysiol.00527.2002.

Barnhart, J. (1997). Chromium chemistry and implications for environmental fate and toxicity. Journal of Soil Contamination 6: 561-568.

Bedinger, GM. (2015). 2014 Minerals yearbook. Arsenic [advance release]. Reston, VA: U.S. Geological Survey.

Beronius, A; Molander, L; Rudén, C; Hanberg, A. (2014). Facilitating the use of non-standard in vivo studies in health risk assessment of chemicals: a proposal to improve evaluation criteria and reporting. J Appl Toxicol 34: 607-617. http://dx.doi.org/10.1002/jat.2991.

Beronius, A; Molander, L; Zilliacus, J; Rudén, C; Hanberg, A. (2018). Testing and refining the Science in Risk Assessment and Policy (SciRAP) web-based platform for evaluating the reliability and relevance of in vivo toxicity studies. J Appl Toxicol 38: 1460-1470. http://dx.doi.org/10.1002/jat.3648.

Bourdon-Lacombe, JA; Moffat, ID; Deveau, M; Husain, M; Auerbach, S; Krewski, D; Thomas, RS; Bushel, PR; Williams, A; Yauk, CL. (2015). Technical guide for applications of gene expression profiling in human health risk assessment of environmental chemicals. Regul Toxicol Pharmacol 72: 292-309. http://dx.doi.org/10.1016/j.yrtph.2015.04.010.

Brazma, A; Hingamp, P; Quackenbush, J; Sherlock, G; Spellman, P; Stoeckert, C; Aach, J; Ansorge, W; Ball, CA; Causton, HC; Gaasterland, T; Glenisson, P; Holstege, FC; Kim, IF; Markowitz, V; Matese, JC; Parkinson, H; Robinson, A; Sarkans, U; Schulze-Kremer, S; Stewart, J; Taylor, R; Vilo, J; Vingron, M. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29: 365-371. http://dx.doi.org/10.1038/ng1201-365.

CA State Water Board (California State Water Quality Control Board). (2017). Frequently asked questions about hexavalent chromium in public water systems [Fact Sheet]. Sacramento, CA. https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/documents/chromium6/chrome_6_faqs.pdf.

CalEPA (California Environmental Protection Agency). (2003). Ambient air monitoring for hexavalent chromium and metals in Barrio Logan: May 2001 through May 2002. (Report: 2003-10-14). Sacramento, CA.

CalEPA (California Environmental Protection Agency). (2004). Barrio Logan report: A compilation of air quality studies in Barrio Logan. November 2004. (Report: 2005-11-00). Sacramento, CA.

CalEPA (California Environmental Protection Agency). (2011). Public health goal for Hexavalent Chromium (Cr VI) in drinking water. Sacramento, CA: Pesticide and Environmental Toxicology Branch, Office of Environmental Health Hazard Assessment. http://oehha.ca.gov/media/downloads/water/public-health-goal/cr6phg072911.pdf.

CDPH (California Department of Public Health). (2013). Notice of proposed rulemaking: Title 22, California Code of Regulations subject: Hexavalent chromium mcl (DPH-11-005). Available online at (accessed

Collins, BJ; Stout, MD; Levine, KE; Kissling, GE; Melnick, RL; Fennell, TR; Walden, R; Abdo, K; Pritchard, JB; Fernando, RA; Burka, LT; Hooth, MJ. (2010). Exposure to hexavalent chromium resulted in significantly higher tissue chromium burden compared with trivalent chromium following similar oral doses to male F344/N rats and female B6C3F1 mice. Toxicol Sci 118: 368-379. http://dx.doi.org/10.1093/toxsci/kfq263.

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.

Crump, C; Crump, K; Hack, E; Luippold, R; Mundt, K; Liebig, E; Panko, J; Paustenbach, D; Proctor, D. (2003). Dose-response and risk assessment of airborne hexavalent chromium and lung cancer mortality. Risk Anal 23: 1147-1163. http://dx.doi.org/10.1111/j.0272-4332.2003.00388.x.

Culver, BH; Graham, BL; Coates, AL; Wanger, J; Berry, CE; Clarke, PK; Hallstrand, TS; Hankinson, JL; Kaminsky, DA; Macintyre, NR; Mccormack, MC; Rosenfeld, M; Stanojevic, S; Weiner, DJ; Laboratories, ACoPSfPF. (2017). Recommendations for a Standardized Pulmonary Function Report. An Official American Thoracic Society Technical Statement. Am J Respir Crit Care Med 196: 1463-1472. http://dx.doi.org/10.1164/rccm.201710-1981ST.

Cummings, DE; Fendorf, S; Singh, N; Sani, RK; Peyton, BM; Magnuson, TS. (2007). Reduction of Cr(VI) under acidic conditions by the facultative Fe(lll)-reducing bacterium Acidiphilium cryptum. Environ Sci Technol 41: 146-152. http://dx.doi.org/10.1021/es061333k.

De Flora, S; Camoirano, A; Bagnasco, M; Bennicelli, C; Corbett, GE; Kerger, BD. (1997). Estimates of the chromium(VI) reducing capacity in human body compartments as a mechanism for attenuating its potential toxicity and carcinogenicity. Carcinogenesis 18: 531-537.

EFSA (European Food Safety Authority). (2017). Guidance on the use of the weight of evidence approach in scientific assessments. EFSA J 15: 4971. http://dx.doi.org/10.2903 / j.efsa.2017.4971.

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

FDA, US. (2013). Beverages: Bottle water. Code of Federal Regulations: 21 CFR 165.110. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm.

Federal Register. (2010). Federal Register: March 29 2010 National Primary Drinking Water Regulations Announcement of the Results of EPAs Review of Existing Drinking Water Standards and Request for Public Comment and/or Information on Related Issues Notice. (815Z10001). http://nepis.epa.gov/exe/ZyPURL.cgi?Dockey=P1007TNX.txt.

Fendorf, S; Wielinga, BW; Hansel, CM. (2000). Chromium transformations in natural environments: The role of biological and abiological. processes in chromium(VI) reduction. International Geology Review 42: 691-701. http://dx.doi.org/10.1080/00206810009465107.

Fendorf, SE. (1995). Surface reactions of chromium in soils and waters. Geoderma 67: 55-71. http://dx.doi.org/10.1016/0016-7061(94)00062-F.

Gibb, H; Lees, P; Pinsky, P; Rooney, B. (2000a). Lung cancer among workers in chromium chemical production. Am J Ind Med 38: 115-126.

Gibb, HJ; Lees, PS; Pinsky, PF; Rooney, BC. (2000b). Clinical findings of irritation among chromium chemical production workers. Am J Ind Med 38: 127-131.

Gibb, HJ; Lees, PS; Wang, J; Grace O'Leary, K. (2015). Extended followup of a cohort of chromium production workers. Am J Ind Med 58: 905-913. http://dx.doi.org/10.1002/ajim.22479.

Glaser, U; Hochrainer, D; Kloeppel, H; Kuhnen, H. (1985). Low level chromium (VI) inhalation effects on alveolar macrophages and immune functions in Wistar rats. Arch Toxicol 57: 250-256. http://dx.doi.org/10.1007/BF00324787.

Glaser, U; Hochrainer, D; Steinhoff, D. (1990). Investigation of irritating properties of inhaled Cr(VI) with possible influence on its carcinogenic action. In NH Seemayer; W Hadnagy (Eds.), Environmental hygiene II (pp. 239-245). Berlin, Germany: Springer-Verlag.

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.

Hausladen, DM; Fendorf, S. (2017). Hexavalent chromium generation within naturally structured soils and sediments. Environ Sci Technol 51: 2058-2067. http://dx.doi.org/10.1021/acs.est.6b04039.

Health Canada. (2016). Guidelines for Canadian drinking water quality: Guideline technical document: Chromium. In Guidelines for Canadian Drinking Water Quality. Ottawa, Ontario: Minister of Health.

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

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

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

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

Hruz, T; Laule, O; Szabo, G; Wessendorp, F; Bleuler, S; Oertle, L; Widmayer, P; Gruissem, W; Zimmermann, P. (2008). Genevestigator v3: A reference expression database for the meta-analysis of transcriptomes. 2008: 420747. http://dx.doi.org/10.1155/2008/420747.

Huang, L; Yu, CH; Hopke, PK; Lioy, PJ; Buckley, BT; Shin, JY; Fan, ZT. (2014). Measurement of soluble and total hexavalent chromium in the ambient airborne particles in New Jersey. Aerosol Air Qual Res 14: 1939-1949. http://dx.doi.org/10.4209/aaqr.2013.10.0312.

IARC (International Agency for Research on Cancer). (2012). A review of human carcinogens. Part C: Arsenic, metals, fibres, and dusts [IARC Monograph]. Lyon, France: World Health Organization. http://monographs.iarc.fr/ENG/Monographs/vol100C/mono100C.pdf.

ICF Consulting. (2018). DRAGON: An online tool for systematic review. Available online at https://www.icf.com/resources/solutions-and-apps/dragon-online-tool-systematic-review (accessed

Illumina, Inc., . (2018). BaseSpace correlation engine [Computer Program]. San Diego, CA. Retrieved from https://www.illumina.com/products/by-type/informatics-products/basespace-correlation-engine.html

IPCS (International Programme on Chemical Safety). (2013). Inorganic chromium(VI) compounds. (78). Geneva, Switzerland: World Health Organization. http://www.who.int/ipcs/publications/cicad/cicad_78.pdf.

IRIS (Integrated Risk Information System). (2015). Advancing Systematic Review Workshop: Agenda.Advancing Systematic Review Workshop, December 16 - 17, 2015, Arlington, VA.

Ishikawa, Y; Nakagawa, K; Satoh, Y; Kitagawa, T; Sugano, H; Hirano, T; Tsuchiya, E. (1994a). Characteristics of chromate workers' cancers, chromium lung deposition and precancerous bronchial lesions: an autopsy study. Br J Cancer 70: 160-166.

Ishikawa, Y; Nakagawa, K; Satoh, Y; Kitagawa, T; Sugano, H; Hirano, T; Tsuchiya, E. (1994b). Hot spots of chromium accumulation at bifurcations of chromate workers' bronchi. Cancer Res 54: 2342-2346.

Jackson, MA; Yang, L; Lea, I; Rashid, A; Kuo, B; Williams, A; Lyn Yauk, C; Fostel, J. (2017). The TGx-28.65 biomarker online application for analysis of transcriptomics data to identify DNA damage-inducing chemicals in human cell cultures. Environ Mol Mutagen 58: 529-535. http://dx.doi.org/10.1002/em.22114.

Jardine, PM; Mehlhorn, TL; Bailey, WB; Brooks, SC; Fendorf, S; Gentry, RW; Phelps, TJ; Saiers, JE. (2011). Geochemical processes governing the fate and transport of chromium(III) and chromium(VI) in soils. Vadose Zone J 10: 1058-1070. http://dx.doi.org/10.2136/vzj2010.0102.

Kim, JG; Dixon, JB. (2002). Oxidation and fate of chromium in soils. Soil Sci Plant Nutr 48: 483-490. http://dx.doi.org/10.1080/00380768.2002.10409230.

Kirman, CR; Aylward, LL; Suh, M; Harris, MA; Thompson, CM; Haws, LC; Proctor, DM; Lin, SS; Parker, W; Hays, SM. (2013). Physiologically based pharmacokinetic model for humans orally exposed to chromium. Chem Biol Interact 204: 13-27. http://dx.doi.org/10.1016/j.cbi.2013.04.003.

Kirman, CR; Hays, SM; Aylward, LL; Suh, M; Harris, MA; Thompson, CM; Haws, LC; Proctor, DM. (2012). Physiologically based pharmacokinetic model for rats and mice orally exposed to chromium. Chem Biol Interact 200: 45-64. http://dx.doi.org/10.1016/j.cbi.2012.08.016.

Kirman, CR; Suh, M; Hays, SM; Gürleyük, H; Gerads, R; De Flora, S; Parker, W; Lin, S; Haws, LC; Harris, MA; Proctor, DM. (2016). Reduction of hexavalent chromium by fasted and fed human gastric fluid. II. Ex vivo gastric reduction modeling. Toxicol Appl Pharmacol 306: 120-133. http://dx.doi.org/10.1016/j.taap.2016.07.002.

Kirman, CR; Suh, M; Proctor, DM; Hays, SM. (2017). Improved physiologically based pharmacokinetic model for oral exposures to chromium in mice, rats, and humans to address temporal variation and sensitive populations. Toxicol Appl Pharmacol. http://dx.doi.org/10.1016/j.taap.2017.03.023.

Kondo, K; Takahashi, Y; Ishikawa, S; Uchihara, H; Hirose, Y; Yoshizawa, K; Tsuyuguchi, M; Takizawa, H; Miyoshi, T; Sakiyama, S; Monden, Y. (2003). Microscopic analysis of chromium accumulation in the bronchi and lung of chromate workers. Cancer 98: 2420-2429. http://dx.doi.org/10.1002/cncr.11818.

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.

Krawic, A; Luczak, M; Zhitkovich, A. (2017). Variation in extracellular detoxification is a link to different carcinogenicity among chromates in rodent and human lungs. Chem Res Toxicol 30: 1720-1729. http://dx.doi.org/10.1021/acs.chemrestox.7b00172.

Kuo, HW; Lai, JS; Lin, TI. (1997). Nasal septum lesions and lung function in workers exposed to chromic acid in electroplating factories. Int Arch Occup Environ Health 70: 272-276.

Kupershmidt, I; Su, QJ; Grewal, A; Sundaresh, S; Halperin, I; Flynn, J; Shekar, M; Wang, H; Park, J; Cui, W; Wall, GD; Wisotzkey, R; Alag, S; Akhtari, S; Ronaghi, M. (2010). Ontology-based meta-analysis of global collections of high-throughput public data. PLoS ONE 5. http://dx.doi.org/10.1371/journal.pone.0013066.

Lindberg, E; Hedenstierna, G. (1983). Chrome plating: Symptoms, findings in the upper airways, and effects on lung function. Arch Environ Occup Health 38: 367-374.

Losi, ME; Amrhein, C; Frankenberger, WT. (1994). Environmental biochemistry of chromium [Review]. Rev Environ Contam Toxicol 136: 91-121.

Luippold, RS; Mundt, KA; Austin, RP; Liebig, E; Panko, J; Crump, C; Crump, K; Proctor, D. (2003). Lung cancer mortality among chromate production workers. Occup Environ Med 60: 451-457.

MacKenzie, RD; Byerrum, RU; Decker, CF; Hoppert, CA; Langham, RF. (1958). Chronic toxicity studies. II. Hexavalent and trivalent chromium administered in drinking water to rats. AMA Arch Ind Health 18: 232-234.

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, D.C.

Malsch, PA; Proctor, DM; Finley, BL. (1994). Estimation of a chromium inhalation reference concentration using the benchmark dose method: a case study [Review]. Regul Toxicol Pharmacol 20: 58-82. http://dx.doi.org/10.1006/rtph.1994.1036.

Mancuso, TF. (1975). Consideration of chromium as an industrial carcinogen. In International conference on heavy metals in the environment: v III, health; October. Toronto, ON, Canada: Institute for Environmental Studies.

Mancuso, TF. (1997). Chromium as an industrial carcinogen: Part I. Am J Ind Med 31: 129-139. http://dx.doi.org/10.1002/(SICI)1097-0274(199702)31:2<129::AID-AJIM1>3.0.CO;2-V.

McClain, CN; Fendorf, S; Webb, SM; Maher, K. (2017). Quantifying Cr(VI) production and export from serpentine soil of the California coast range. Environ Sci Technol 51: 141-149. http://dx.doi.org/10.1021/acs.est.6b03484.

McConnell, ER; Bell, SM; Cote, I; Wang, RL; Perkins, EJ; Garcia-Reyero, N; Gong, P; Burgoon, LD. (2014). Systematic Omics Analysis Review (SOAR) tool to support risk assessment. PLoS ONE 9: e110379. http://dx.doi.org/10.1371/journal.pone.0110379.

Miksche, LW; Lewalter, J. (1997). Health surveillance and biological effect monitoring for chromium-exposed workers [Review]. Regul Toxicol Pharmacol 26: S94-S99. http://dx.doi.org/10.1006/rtph.1997.1146.

Miller, MR; Hankinson, J; Brusasco, V; Burgos, F; Casaburi, R; Coates, A; Crapo, R; Enright, P; van Der Grinten, CP; Gustafsson, P; Jensen, R; Johnson, DC; Macintyre, N; Mckay, R; Navajas, D; Pedersen, OF; Pellegrino, R; Viegi, G; Wanger, J; Force, AET. (2005). Standardisation of spirometry. Eur Respir J 26: 319-338. http://dx.doi.org/10.1183/09031936.05.00034805.

Molander, L; Ågerstrand, M; Beronius, A; Hanberg, A; Rudén, C. (2015). Science in Risk Assessment and Policy (SciRAP): An online resource for evaluating and reporting in vivo (eco) toxicity studies. Hum Ecol Risk Assess 21: 753-762.

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.

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

Ng, AW; Bidani, A; Heming, TA. (2004). Innate host defense of the lung: effects of lung-lining fluid pH [Review]. Lung 182: 297-317. http://dx.doi.org/10.1007/s00408-004-2511-6.

NIOSH (National Institute for Occupational Safety and Health). (2013a). Criteria for a recommended standard: Occupational exposure to hexavalent chromium. (DHHS (NIOSH) Publication

Number 2013-128). Centers for Disease Control and Prevention. https://www.cdc.gov/niosh/docs/2013-128/.

NIOSH (National Institute for Occupational Safety and Health). (2013b). Occupational exposure to hexavalent chromium. (DHHS (NIOSH) Publication No. 2013128). Department of Health and Human Services, Centers for Disease Control and Prevention. http://www.cdc.gov/niosh/docs/2013-128/pdfs/2013_128.pdf.

NJ DEP (New Jersey Department of Environmental Protection). (2009). Derivation of ingestion-based soil remediation criterion for Cr+6 based on the NTP chronic bioassay data for Sodium Dichromate Dihydrate. http://www.state.nj.us/dep/dsr/chromium/soil-cleanup-derivation.pdf.

NRC (National Research Council). (2014). Review of EPA's Integrated Risk Information System (IRIS) process. Washington, DC: The National Academies Press. http://www.nap.edu/catalog.php?record_id=18764.

NTP (National Toxicology Program). (2008). Toxicology and carcinogenesis studies of sodium dichromate dihydrate (Cas No. 7789-12-0) in F344/N rats and B6C3F1 mice (drinking water studies) (pp. 1-192). (NTP TR 546; NIH Publication No. 08-5887).

NTP (National Toxicology Program). (2011). Report on carcinogens: Twelfth edition (12th ed.). Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health. http://ntp.niehs.nih.gov/ntp/roc/twelfth/roc12.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.

O'Flaherty, EJ. (1993). A pharmacokinetic model for chromium. Toxicol Lett 68: 145-158. http://dx.doi.org/10.1016/0378-4274(93)90127-J.

O'Flaherty, EJ. (1996). A physiologically based model of chromium kinetics in the rat. Toxicol Appl Pharmacol 138: 54-64. http://dx.doi.org/10.1006/taap.1996.0097.

O'Flaherty, EJ; Kerger, BD; Hays, SM; Paustenbach, DJ. (2001). A physiologically based model for the ingestion of chromium(III) and chromium(VI) by humans. Toxicol Sci 60: 196-213.

O'Flaherty, EJ; Radike, MJ. (1991). Pharmacokinetic modeling of trivalent and hexavalent chromium based on ingestion and inhalation of soluble chromium compounds. (AL-TR-1991-0162). Wright-Patterson AFB, Ohio: Armstrong Laboratory. https://ntrl.ntis.gov/NTRL/dashboard/searchResults.xhtml?searchQuery=ADA256238.

OAQPS (EPA Office of Air Quality Planning and Standards). (2012). Residual risk assessment for the chromic acid anodizing, decorative chromium electroplating, and hard chromium electroplating source categories. Research Triangle Park, NC: U.S. Environmental Protection Agency.

OEHHA (California Office of Environmental Health Hazard Assessment). (2008). TSD for noncancer RELs - Appendix D.3 Chronic RELs and toxicity summaries using the previous version of the Hot Spots Risk Assessment guidelines (OEHHA 1999). Sacramento, CA: California Environmental Protection Agency. http://oehha.ca.gov/media/downloads/crnr/appendixd3final.pdf.

Oregon DEQ (Oregon Department of Environmental Quality). (2016). Portland Metals air monitoring data. Portland, OR.

OSHA (Occupational Safety & Health Administration). (2006). Occupational exposure to hexavalent chromium. Final rule. Fed Reg 71: 10099-10385.

Oze, C; Bird, DK; Fendorf, S. (2007). Genesis of hexavalent chromium from natural sources in soil and groundwater. Proc Natl Acad Sci USA 104: 6544-6549. http://dx.doi.org/10.1073/pnas.0701085104.

Oze, C; Fendorf, S; Bird, DK; Coleman, RG. (2004). Chrominium geochemistry of serpentine soils. International Geology Review 46: 97-126. http://dx.doi.org/10.2747/0020-6814.46.2.97.

Park, RM; Bena, JF; Stayner, LT; Smith, RJ; Gibb, HJ; Lees, PS. (2004). Hexavalent chromium and lung cancer in the chromate industry: a quantitative risk assessment. Risk Anal 24: 1099-1108. http://dx.doi.org/10.1111/j.0272-4332.2004.00512.x.

Proctor, DM; Suh, M; Aylward, LL; Kirman, CR; Harris, MA; Thompson, CM; Gürleyük, H; Gerads, R; Haws, LC; Hays, SM. (2012). Hexavalent chromium reduction kinetics in rodent stomach contents. Chemosphere 89: 487-493. http://dx.doi.org/10.1016/j.chemosphere.2012.04.065.

Proctor, DM; Suh, M; Mittal, L, iz; Hirsch, S; Salgado, RV; Bartlett, C; Van Landingham, C; Rohr, A; Crump, K. (2016). Inhalation cancer risk assessment of hexavalent chromium based on updated mortality for Painesville chromate production workers. J Expo Sci Environ Epidemiol 26: 224-231. http://dx.doi.org/10.1038/jes.2015.77.

Qiagen (Qiagen Bioinformatics). (2018). Ingenuity Pathway Analysis [Computer Program]. Redwoord City, CA. Retrieved from https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/

RIVM (National Institute for Public Health and the Environment (Netherlands)). (2001). Re-evaluation of human-toxicological maximum permissible risk levels (pp. 58-61). (RIVM report 711701 025). Bilthoven, the Netherlands: Rijksinstituut voor Volksgezondheid en Milieu RIVM. http://www.rivm.nl/bibliotheek/rapporten/711701025.html.

Sasso, AF; Schlosser, PM. (2015). An evaluation of in vivo models for toxicokinetics of hexavalent chromium in the stomach. Toxicol Appl Pharmacol 287: 293-298. http://dx.doi.org/10.1016/j.taap.2015.06.016.

Schlosser, PM; Sasso, AF. (2014). A revised model of ex-vivo reduction of hexavalent chromium in human and rodent gastric juices. Toxicol Appl Pharmacol 280: 352-361. http://dx.doi.org/10.1016/j.taap.2014.08.010.

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.

Sciome (Sciome, LLC.). (2018). BMDExpress 2 software (Version 44) [Computer Program]. Research Triangle Park, NC: National Toxicology Program. Retrieved from https://www.sciome.com/bmdexpress/

Shaw Environmental (Shaw Environmental, Inc). (2006). Final report. Industry profile, exposure profile, technological feasibility evaluation, and environmental impact for industries affected by a revised OSHA standard for hexavalent chromium (General industry, construction, and maritime sectors). Washington, DC: U.S. Department of Labor, Occupational Safety and Health Administration.

Smith, MT; Guyton, KZ; Gibbons, CF; Fritz, JM; Portier, CJ; Rusyn, I; DeMarini, DM; Caldwell, JC; Kavlock, RJ; Lambert, P; Hecht, SS; Bucher, JR; Stewart, BW; Baan, R; Cogliano, VJ; Straif, K.

(2016). Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis [Review]. Environ Health Perspect 124: 713-721. http://dx.doi.org/10.1289/ehp.1509912.

Sterne, J; Higgins, J; Reeves, B. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions, Version 7 March 2016. Br Med J 355: i4919.

Subramanian, A; Tamayo, P; Mootha, VK; Mukherjee, S; Ebert, BL; Gillette, MA; Paulovich, A; Pomeroy, SL; Golub, TR; Lander, ES; Mesirov, JP. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102: 15545-15550. http://dx.doi.org/10.1073/pnas.0506580102.

SWRCB. Title 22, Division 4, Chapter 14, Article 3: Amendments, (2014).

TCEQ (Texas Commission on Environmental Quality). (2014). Development support document for hexavalent chromium (Particulate compounds). Austin, TX. https://www.tceq.texas.gov/assets/public/implementation/tox/dsd/final/august2014/hexavalent_chromium.pdf.

TCEQ (Texas Commission on Environmental Quality). (2016). Development support document for hexavalent chromium oral reference dose. Austin, TX. https://www.tceq.texas.gov/assets/public/implementation/tox/dsd/final/chromium_ord.pdf.

U.S. EPA (U.S. Environmental Protection Agency). (1988). Recommendations for and documentation of biological values for use in risk assessment (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). (1998a). 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). (1998b). Toxicological review of hexavalent chromium [EPA Report] (pp. 80). (630R98009). Washington, DC. http://www.epa.gov/ncea/iris/toxreviews/0144-tr.pdf.

U.S. EPA (U.S. Environmental Protection Agency). (2002). A review of the reference dose and reference concentration processes (pp. 1-192). (EPA/630/P-02/002F). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. http://www.epa.gov/osa/review-reference-dose-and-reference-concentration-processes.

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] (pp. 1-166). (EPA/630/P-03/001F). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. http://www2.epa.gov/osa/guidelines-carcinogen-risk-assessment.

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). (2008a). Cr (VI) quantitative risk assessment Q*1 based on mouse carcinogenicity studies. Washington, DC: Health Effects Division, Office of Pesticide Programs. http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPP-2003-0250-0088.

U.S. EPA (U.S. Environmental Protection Agency). (2008b). Evaluation of the carcinogenic potential of inorganic hexavalent chromium (Cr(VI)). Washington, DC: Health Effects Division, Office of Pesticide Programs. http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPP-2003-0250-0089.

U.S. EPA (U.S. Environmental Protection Agency). (2008c). Reregistration eligibility decision for chromated arsenicals. (EPA 739-R-08-006). Office of Prevention, Pesticides And Toxic Substances; U.S. Environmental Protection Agency.

U.S. EPA (U.S. Environmental Protection Agency). (2010). Toxicological review of hexavalent chromium (external review draft). (EPA/635/R-10/004A). Washington, DC. http://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=221433.

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/R11/0001). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum, Office of the Science Advisor. https://www.epa.gov/risk/recommended-use-body-weight-34-default-method-derivation-oral-reference-dose.

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/0106tr.pdf.

U.S. EPA (U.S. Environmental Protection Agency). (2012a). Advances in inhalation gas dosimetry for derivation of a reference concentration (RfC) and use in risk assessment (pp. 1-140).

(EPA/600/R-12/044). Washington, DC. https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=244650&CFID=50524762&CFTOKEN=17139189.

U.S. EPA (U.S. Environmental Protection Agency). (2012b). Benchmark dose technical guidance. (EPA/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). (2013). Scientific workshop: Factors affecting the reduction and absorption of hexavalent chromium in the gastrointestinal (GI) tract: Potential impact on evaluating the carcinogenicity of ingested hexavalent chromium. Paper presented at Hexavalent Chromium Workshop, September 19 & 24, 2018, Crystal City, VA.

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

U.S. EPA (U.S. Environmental Protection Agency). (2014b). Preliminary materials for the Integrated Risk Information System (IRIS) toxicological review of hexavalent chromium part 1: Experimental animal studies [CASRN) 18540-29-9] [EPA Report]. (EPA/635/R-14/094). Research Triangle Park, NC. https://www.epa.gov/iris/iris-bimonthly-public-meeting-jun-2014.

U.S. EPA (U.S. Environmental Protection Agency). (2014c). Preliminary materials for the Integrated Risk Information System (IRIS) toxicological review of hexavalent chromium part 2: Human, toxicokinetic, and mechanistic Studies [CASRN) 18540-29-9] [EPA Report]. (EPA/635/R-14/218). Research Triangle Park, NC. https://www.epa.gov/iris/iris-bimonthly-public-meeting-oct-2014.

U.S. EPA (U.S. Environmental Protection Agency). (2014d). The third Unregulated Contaminant Monitoring Rule (UCMR 3) occurrence data. Retrieved from http://water.epa.gov/lawsregs/rulesregs/sdwa/ucmr/data.cfm#ucmr2013

U.S. EPA (U.S. Environmental Protection Agency). (2016). US EPAs report on the environment (ROE). Available online at https://cfpub.epa.gov/roe/index.cfm (accessed

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: Office of Chemical Safety and Pollution Prevention. https://www.epa.gov/sites/production/files/2017-06/documents/tsca_ra_guidance_final.pdf.

U.S. EPA (U.S. Environmental Protection Agency). (2018a). Application of systematic review in TSCA risk evaluations. (740-P1-8001). Washington, D.C.: U.S. Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention. https://www.epa.gov/sites/production/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf.

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

USGS (U.S. Geological Survey). (1995). Chromium life cycle study. (IC 9411). Washington, D.C.: U.S. Department of the Interior, Bureau of Mines.

Varghese, A; Cawley, M; Hong, T. (2017). Supervised clustering for automated document classification and prioritization: A case study using toxicological abstracts. Environ Syst Decis. http://dx.doi.org/10.1007/s10669-017-9670-5.

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.

WHO (World Health Organization). (1999). WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction. In WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction (4th ed.). Cambridge, UK: Cambridge University Press.

WHO (World Health Organization). (2000). Air quality guidelines for Europe (2nd ed.). Copenhagen, Denmark: World Health Organization, Regional Office for Europe. http://www.euro.who.int/en/health-topics/environment-and-health/air-quality/publications/pre2009/air-quality-guidelines-for-europe.

WHO (World Health Organization). (2003). Chromium in drinking water. (WHO/SDE/WSH/03.04/04). Geneva, Switzerland. http://www.who.int/water_sanitation_health/dwq/chemicals/chromium.pdf.

WHO (World Health Organization). (2010). WHO laboratory manual for the examination and processing of human semen [WHO EHC]. In WHO laboratory manual for the examination and processing of human semen (5th ed.). Geneva, Switzerland.

APPENDICES

APPENDIX A. ELECTRONIC DATABASE SEARCH STRATEGIES

Table A-1. Literature search query strings for computerized databases

Database search date	Terms
PubMed (1/29/2013) (7/19/2013) (2/5/2014) (4/1/2015) (4/1/2016) (5/24/2017) (5/24/2018)	*hexavalent chromium OR (hexavalent AND chromium) OR CRVI OR CR VI OR Chromium VI OR "Chromic acid" OR "Calcium chromate" OR "Potassium dichromate" OR "Potassium chromate" OR "Sodium chromate" OR "lead chromate" OR "zinc chromate" OR "strontium chromate" OR "ammonium dichromate" OR 13765-19-0[RN] OR 1333-82-0[RN] OR 7789-00-6[RN] OR 7778-50-9[RN] OR 7775-11-3[RN] OR 7789-12-0[RN] OR 13530-65-9[RN] OR 7738-94-5[rn] OR 18540-29-9[rn] OR 7758-97-6[RN] OR 11119-70-3[rn] OR 11103-86-9[rn] OR 13530-65-9[rn] OR 7788-98-9[rn] OR 77898-09-5[rn] OR 7789-06-2[rn]*
Web of Science (1/29/2013) (7/19/2013) (2/5/2014) (4/1/2015) (4/1/2016) (5/24/2017) (5/24/2018)	*Topic = (hexavalent chromium OR (hexavalent AND chromium) Chromium VI OR CrVI OR Cr VI OR "Chromic acid" OR "Calcium chromate" OR "Chromic trioxide" OR "Potassium dichromate" OR "Potassium chromate" OR "Sodium chromate" OR "Sodium dichromate dehydrate" OR "lead chromate" OR "zinc chromate" OR "strontium chromate" OR "ammonium dichromate" OR "ammonium chromate" OR 13765-19-0 OR 1333-82-0 OR 7789-00-6 OR 7778-50-9 OR 7775-11-3 OR 7789-12-0 OR 13530-65-9 OR 7738-94-5 OR 18540-29-9 OR 7758-97-6 OR 11119-70-3 OR 11103-86-9 OR 13530-65-9 OR 7788-98-9 OR 77898-09-5 OR 7789-06-2)* AND Research Areas = Toxicology, Biochemistry molecular biology, Public environmental occupational health, Dermatology, Cell biology, Oncology, Life sciences biomedicine other topics, Allergy, Veterinary sciences, Developmental biology, Immunology, Reproductive biology, Pathology, Physiology, Urology nephrology, Hematology, Neurosciences neurology, Respiratory system, Cardiovascular system cardiology, Obstetrics gynecology, Infectious diseases, Gastroenterology hepatology, Microscopy

Database search date	Terms
Web of Science (1/29/2013) (7/19/2013) (2/5/2014) (4/1/2015) (12/1/2017) (5/24/2017) (5/24/2018)	*Topic = (hexavalent chromium OR (hexavalent AND chromium) Chromium VI OR CrVI OR Cr VI OR "Chromic acid" OR "Calcium chromate" OR "Chromic trioxide" OR "Potassium dichromate" OR "Potassium chromate" OR "Sodium chromate" OR "Sodium dichromate dehydrate" OR "lead chromate" OR "zinc chromate" OR "strontium chromate" OR "ammonium dichromate" OR "ammonium chromate" OR 13765-19-0 OR 1333-82-0 OR 7789-00-6 OR 7778-50-9 OR 7775-11-3 OR 7789-12-0 OR 13530-65-9 OR 7738-94-5 OR 18540-29-9 OR 7758-97-6 OR 11119-70-3 OR 11103-86-9 OR 13530-65-9 OR 7788-98-9 OR 77898-09-5 OR 7789-06-2)* AND Research Areas = Chemistry, Environmental sciences ecology, Spectroscopy, Pharmacology pharmacy, Water resources, Genetics heredity, Science technology other topics, Biophysics, Food sciences technology, Endocrinology metabolism, Research experimental medicine, Nutrition dietetics, Zoology, General internal medicine, Construction building technology, Parasitology, Medical laboratory technology, Education educational research, Otorhinolaryngology, Rheumatology, Anatomy morphology, Emergency medicine, Mycology, Sport sciences, Psychiatry AND cancer* OR carcinogen* OR chronic OR subchronic OR genotox* OR inhalation absorption OR oral absorption OR mice OR mouse OR Mutagenicity OR pharmacokinetic OR rat OR rats OR toxic* NOT (fish OR bacteria* OR microorganism* OR plant*) OR tumor*
Toxline (includes TSCATS) (1/29/2013) (7/19/2013) (2/5/2014) (4/1/2015) (4/1/2016) (5/24/2017) (5/24/2018)	18540-29-9 OR 7789-09-5 OR 13765-19-0 OR 1333-82-0 OR 7758-97-6 OR 7789-00-6 OR 7778-50-9 OR 7775-11-3 OR 7789-12-0 OR 7789-06-2 OR 13530-65-9 OR 7788-98-9 OR 7738-94-5 OR 13530-68-2
TSCATS2 (1/29/2013) (7/19/2013) (2/5/2014) (4/1/2015) (4/1/2016) (5/24/2017) (5/24/2018)	18540-29-9

Database search date	Terms
Combined reference set	(duplicates eliminated through electronic screen)

Toxline = Toxicology Literature Online; TSCATS2 = Toxic Substances Control Act Test Submissions 2.0.

Table A-2. Processes used to augment the search of core computerized databases for Cr(VI)

System used	Selected key reference(s) or sources	Date	Additional references identified
Manual search of citations from health assessment documents	ATSDR (Agency for Toxic Substances and Disease Registry). (2012). Toxicological profile for chromium. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service. http://www.atsdr.cdc.gov/ToxProfiles/tp.asp?id=62&tid=17.	1/2013	40 citations added
	U.S. EPA (U.S. Environmental Protection Agency). (2010). Toxicological review of hexavalent chromium (external review draft). (EPA/635/R-10/004A). Washington, DC. http://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=221433.	1/2013	59 citations added
	OSHA (Occupational Safety & Health Administration). (2006). Occupational exposure to hexavalent chromium. Final rule. Fed Reg 71: 10099–10385.	5/2014	3 citations added
	IPCS (International Programme on Chemical Safety). (2013). Inorganic chromium (VI) compounds. (78). Geneva, Switzerland: World Health Organization. http://www.who.int/ipcs/publications/cicad/cicad_78.pdf.	5/2014	5 citations added
	NIOSH (National Institute for Occupational Safety and Health). (2013b). Occupational exposure to hexavalent chromium. (DHHS [NIOSH] Publication No. 2013128). Department of Health and Human Services, Centers for Disease Control and Prevention. http://www.cdc.gov/niosh/docs/2013-128/pdfs/2013_128.pdf.	5/2014	1 citation added
References obtained during the assessment process	Snowball search	1/2013, Ongoing	
Search of online chemical assessment-related websites	Combination of Chemical Abstracts Service registry number (CASRN) and synonyms searched on the following websites: • American Conference of Governmental Industrial Hygienists (ACGIH) (http://www.acgih.org) • American Industrial Hygiene Association Workplace Environmental Exposure Levels (AIHA WEELs) (http://www.tera.org/OARS/WEEL.html) • Agency for Toxic Substances and Disease Registry (ATSDR) (http://www.atsdr.cdc.gov/substances/index.asp) • California Environmental Protection Agency (CalEPA) Office of Environmental Health Hazard Assessment (OEHHA) (http://www.oehha.ca.gov/risk.html) ○ OEHHA Toxicity Criteria Database (http://www.oehha.ca.gov/tcdb/index.asp)		

System used	Selected key reference(s) or sources	Date	Additional references identified
	○ Biomonitoring California-Priority Chemicals (https://biomonitoring.ca.gov/chemicals/priority-chemicals) ○ Biomonitoring California-Designated Chemicals (https://biomonitoring.ca.gov/chemicals/designated-chemicals) ○ Cal/Ecotox Database (https://oehha.ca.gov/ecotoxicology/general-info/calecotox-database) ○ OEHHA fact sheets (http://www.oehha.ca.gov/public_info/facts/index.html) ○ Noncancer health effects table (reference exposure levels [RELs]: http://www.oehha.ca.gov/air/allrels.html) ○ Cancer Potency Factors (see Appendix A and Appendix B; http://www.oehha.ca.gov/air/hot_spots/tsd052909.html) • CalEPA Drinking Water Notification Levels (http://www.swrcb.ca.gov/drinking_water/certlic/drinkingwater/NotificationLevels.shtml) • Chemical Risk Information Platform (CHRIP) (http://www.safe.nite.go.jp/english/db.html) • Consumer Product Safety Commission (CPSC) (http://www.cpsc.gov) • European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) publications (http://www.ecetoc.org/publications) • European Chemicals Agency (ECHA); general site (http://echa.europa.eu/information-on-chemicals) • ECHA info on Registered Substances (http://echa.europa.eu/information-on-chemicals/registered-substances) • ECHA Information from the Existing Substances Regulation (ESR) (http://echa.europa.eu/information-on-chemicals/information-from-existing-substances-regulation) • eChemPortal [participating databases: Aggregated Computational Toxicology Resource (ACToR), AGRITOX, Canadian Categorization Results (CCR), CCR DATA, Canada's Existing Substances Assessment Repository (CESAR), CHRIP, ECHA CHEM, Data Bank of Environmental Properties of Chemicals (EnviChem), European chemical Substances Information System (ESIS), Globally Harmonized System-Japan (GHS-J), High Production Volume Information System (HPVIS), Hazardous Substances Data Bank (HSDB), Hazardous Substances and New Organisms Chemical Classification Information Database (HSNO CCID), INCHEM, Japan CHEmicals Collaborative Knowledge (J-CHECK), JECDB, NICNAS PEC, OECD HPV, OECD SIDS IUCLID, UNEP SIDS, United Kingdom (UK) Coordinated Chemicals Risk Management Programme Publications (CCRMP) Outputs, US EPA IRIS, US EPA Substance Registry Services (SRS)		

System used	Selected key reference(s) or sources	Date	Additional references identified

	(http://www.echemportal.org/echemportal/participant/page.action?pageID=9)]
	• Environment Canada—search entire site (http://www.ec.gc.ca/default.asp?lang=En&n=ECD35C36) if not found below:
	  ○ Toxic substances managed under Canadian Environmental Protection Act (CEPA) (http://www.ec.gc.ca/toxiques-toxics/Default.asp?lang=En&n=98E80CC6-1) search results
	  ○ Final assessments (http://www.ec.gc.ca/lcpe-cepa/default.asp?lang=En&xml=09F567A7-B1EE-1FEE-73DB-8AE6C1EB7658)
	  ○ Draft assessments (http://www.ec.gc.ca/lcpe-cepa/default.asp?lang=En&xml=6892C255-5597-C162-95FC-4B905320F8C9)
	• EPA Chemical Data Access Tool (CDAT) (http://java.epa.gov/oppt_chemical_search/)
	• EPA Acute Exposure Guideline Levels (http://www.epa.gov/oppt/aegl/pubs/chemlist.htm)
	• EPA National Service Center for Environmental Publications (NSCEP) (http://www.epa.gov/ncepihom/)
	• EPA Office of Pesticide Programs (OPP) (http://iaspub.epa.gov/apex/pesticides/f?p=chemicalsearch:1)
	• EPA Science Inventory (http://cfpub.epa.gov/si/)
	• Emergency Response Planning Guidelines (ERPGs) (https://www.aiha.org/get-involved/AIHAGuidelineFoundation/EmergencyResponsePlanningGuidelines/Pages/default.aspx)
	• Food and Drug Administration (FDA) (http://www.fda.gov/)
	• Federal Docket (www.regulations.gov)
	• Health Canada—search entire site (http://www.hc-sc.gc.ca/index-eng.php)
	• Health Canada Drinking Water Documents (http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/index-eng.php#tech_doc)
	• Health Canada First Priority List Assessments (http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/psl1-lsp1/index-eng.php)
	• Health Canada Second Priority List Assessments (http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/psl2-lsp2/index-eng.php)
	• International Agency for Research on Cancer (IARC) Monographs: (https://monographs.iarc.fr/agents-classified-by-the-iarc)
	• IRISTrack/new assessments and reviews (http://cfpub.epa.gov/ncea/iris/search/)		

System used	Selected key reference(s) or sources	Date	Additional references identified

	• Japan Existing Chemical Data Base (JECDB) (http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPageENG.jsp)
	• National Academies Press (NAP)—search site (http://www.nap.edu/)
	• National Cancer Institute (NCI) (http://www.cancer.gov)
	• National Center for Toxicological Research (NCTR) (http://www.fda.gov/AboutFDA/CentersOffices/OC/OfficeofScientificandMedicalPrograms/NCTR/default.htm)
	• National Industrial Chemicals Notification and Assessment Scheme (NICNAS); priority existing chemical (PEC) only covered by eChemPortal (http://www.nicnas.gov.au)
	• National Institute for Environmental Health Sciences (NIEHS) (http://www.niehs.nih.gov/)
	• National Institute for Occupational Safety and Health (NIOSH) (http://www.cdc.gov/niosh/topics/)
	• National Institute for Occupational Safety and Health Technical Information Center (NIOSHTIC) 2 (http://www2a.cdc.gov/nioshtic-2/)
	• National Toxicology Program (NTP)—Report on Carcinogens (RoC), status, results, and management reports
	  ○ RoC (12th–14th): (https://ntp.niehs.nih.gov/pubhealth/roc/index-1.html)
	  ○ NTP site search: (http://ntpsearch.niehs.nih.gov/texis/search/?query=arsenic&pr=ntp_web_entire_site_all&mu=Entire+NTP+Site)
	• Organisation for Economic Cooperation and Development (OECD) high production volume (HPV)/Screening Information Data Set (SIDS)/International Uniform Chemical Information Database (IUCLID) (cross-check with eChem; http://webnet.oecd.org/hpv/ui/Search.aspx)
	• Occupational Safety and Health Administration (OSHA) (http://www.osha.gov/dts/chemicalsampling/toc/toc_chemsamp.html)
	• Registry of Toxic Effects of Chemical Substances (RTECS) (http://www.ccohs.ca/search.html)
	• United Nations Environment Programme (UNEP) SIDS (through 2007; http://www.inchem.org/pages/sids.html)		

.

APPENDIX B. TYPICAL DATA EXTRACTION FIELDS

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

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

Field label	Data extraction elements
Results	Exposure levels (e.g., mean, median, measures of variance as presented in paper, such as standard deviation (SD), standard error of the mean (SEM), 75^{th}/90^{th}/95^{th} percentile, minimum/maximum); range of exposure levels, number of exposed cases
	Statistical findings (e.g., adjusted β, standardized mean difference, adjusted odds ratio, standardized mortality ratio, relative risk, etc.) or description of qualitative results. When possible, convert measures of effect to a common metric with associated 95% confidence intervals (CI). Most often, measures of effect for continuous data are expressed as mean difference, standardized mean difference, and percentage control response. Categorical data are typically expressed as odds ratio, relative risk (RR, also called risk ratio), or β values, depending on what metric is most commonly reported in the included studies and ability to obtain information for effect conversions from the study or through author query.
	Observations on dose-response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, nonmonotonic)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.
ANIMAL	
Funding	Funding source(s)
	Reporting of conflict of interest by authors
Animal model	Sex
	Species
	Strain
	Source of animals
	Age or life stage at start of dosing and at health outcome assessment
	Diet and husbandry information (e.g., diet name/source)
Treatment	Chemical name and CAS number
	Source of chemical
	Purity of chemical
	Dose levels or concentration (as presented and converted to mg/kg BW-day when possible)
	Other dose-related details, such as whether administered dose level was verified by measurement, information on internal dosimetry
	Vehicle used for exposed animals
	Route of administration (e.g., oral, inhalation, dermal, injection)
	Duration and frequency of dosing (e.g., hours, days, weeks when administration was ended, days per week)

Field label	Data extraction elements
Methods	Study design (e.g., single treatment, acute, subchronic [e.g., 90 days in a rodent], chronic, multigenerational, developmental, other)
	Guideline compliance (i.e., use of EPA, OECD, NTP, or another guideline for study design, conducted under GLP guideline conditions, non-GLP but consistent with guideline study, nonguideline peer reviewed publication)
	Number of animals per group (and dams per group in developmental studies)
	Randomization procedure, allocation concealment, blinding during outcome assessment
	Method to control for litter effects in developmental studies
	Use of negative controls and whether controls were untreated, vehicle-treated, or both
	Report on data from positive controls—was expected response observed?
	Endpoint health category (e.g., reproductive)
	Endpoint (e.g., infertility)
	Diagnostic or method to measure endpoint
	Statistical methods
Results	Measures of effect at each dose or concentration level (e.g., mean, median, frequency, and measures of precision or variance) or description of qualitative results. When possible, convert measures of effect to a common metric with associated 95% confidence intervals. Most often, measures of effect for continuous data will be expressed as mean difference, standardized mean difference, and percentage control response. Categorical data will be expressed as relative risk (RR, also called risk ratio).
	NOEL, LOEL, BMD analysis, statistical significance of other dose levels, or other estimates of effect presented in paper. **Note**: The NOEL and LOEL are highly influenced by study design, do not give any quantitative information about the relationship between dose and response, and can be subject to author's interpretation (e.g., a statistically significant effect may not be considered biologically important). Also, a NOEL does not necessarily mean zero response. Ideally, the response rate at specific dose levels is used as the primary measure to characterize the response.
	Observations on dose-response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, nonmonotonic)
	Data on internal concentration, toxicokinetics, or toxicodynamics (when reported)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

BMD = benchmark dose; CAS = Chemical Abstract Service; GLP = good laboratory practice; HPLC-MS/MS = high-performance liquid chromatography–tandem mass spectrometry; LOEL = lowest-observed-effect level; NOEL = no-observed-effect level; NTP = National Toxicology Program; OECD = Organisation for Economic Cooperation and Development.