United States
Environmental Protection
Agency

Technical Support Document for the Draft Sixth Contaminant Candidate List (CCL 6)

-

Chemical Contaminants

Office of Water (4607M)
EPA 815-R-26-004
February 2026
www.epa.gov/safewater

Table of Contents

Equations, Figures, and Tables

List of Abbreviations and Acronyms

ADAF	Age-Dependent Adjustment Factor
ASTM	American Society for Testing and Materials
ATSDR	Agency for Toxic Substances and Disease Registry
CalEPA	California Environmental Protection Agency
CASRN	Chemical Abstracts Services Registry Number
CCL	Contaminant Candidate List
CCL 1	First Contaminant Candidate List
CCL 2	Second Contaminant Candidate List
CCL 3	Third Contaminant Candidate List
CCL 4	Fourth Contaminant Candidate List
CCL 5	Fifth Contaminant Candidate List
CCL 6	Sixth Contaminant Candidate List
CDC	Centers for Disease Control and Prevention
CDPR	California Department of Pesticide Regulation
CDR	Chemical Data Reporting
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CIS	Contaminant Information Sheets
CSF	Cancer Slope Factor
CWS	Community Water System
CWSS	Community Water System Survey
DBP	Disinfection Byproduct
DBP ICR	Disinfection By-product Information Collection Rule
DOE	Department of Energy
DSSTox	Distributed Structure-Searchable Toxicity Public Database Network
DTXSID	Distributed Structure-Searchable Toxicity Substance Identifier
ECHA	European Chemicals Agency
EDSP	Endocrine Disruptor Screening Program
EDWC	Estimated drinking water concentration
EPA	U.S. Environmental Protection Agency
FDA	Food and Drug Administration
FGD	Flue Gas Desulfurization
FFDCA	Federal Food, Drug, and Cosmetic Act
fHQ	Final Hazard Quotient
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FR	Federal Register
HBSL	Health-Based Screening Levels
HHBP	Human Health Benchmarks for Pesticides
HHB-Rx	Human Health Benchmarks for Pharmaceuticals
HSL	Health Screening Level
HRL	Health Reference Level
HSDB	Hazardous Substances Data Bank
IARC	International Agency for Research on Cancer
ICR	Information Collection Request
IRIS	Integrated Risk Information System
KH	Henry’s Law Coefficient
Koc	Organic Carbon-Water Partition Coefficient
Kow	Octanol-Water Partition Coefficient
LD_{50}	Median Lethal Doses

LOD	Limit of Detection
LOAEL	Lowest Observed Adverse Effect Level
MADL	Maximum Allowable Dose Level
MCMC	Markov Chain Monte Carlo
MRDD	Maximum Recommended Daily Dose
MCL	Maximum Contaminant Level
MCLG	Maximum Contaminant Level Goal
mg/kg/day	Milligrams per Kilogram per Day
mg/L	Milligrams per Liter
MRL	Minimal Risk Level
NAWQA	National Water Quality Assessment
NDWAC	National Drinking Water Advisory Council
NEWTS	National Energy Water Treatment and Speciation Database
ng/mL	Nanograms per milliliter
NHANES	National Health and Nutrition Examination Survey
NIH	National Institutes of Health
NIRS	National Inorganics and Radionuclides Survey
NOAEL	No Observed Adverse Effect Level
NRC	National Research Council
NPDWR	National Primary Drinking Water Regulation
NTP	National Toxicology Program
NWIS	National Water Information System
OEHHA	Office of Environmental Health Hazard Assessment
OPERA	Open structure–activity/property Relationship App
OPP	Office of Pesticide Programs
ORD	Office of Research and Development
PAD	Population-Adjusted Dose
PCCL 5	Preliminary Contaminant Candidate List 5
PCCL 6	Preliminary Contaminant Candidate List 6
PDP	Pesticide Data Program
PECO	Population, Exposure, Control, and Outcome
PFAS	Per- and Polyfluoroalkyl Substances
PPLS	Pesticide Product and Label System
PPRTVs	Provisional Peer-Reviewed Toxicity Values
PWS	Public Water System
QA/QC	Quality Assurance/Quality Control
QSAR	Qualitative Structure-Activity Relationship
RD 1	Regulatory Determination 1
RD 2	Regulatory Determination 2
RD 4	Regulatory Determination 4
RD 5	Regulatory Determination 5
RfD	References Dose
RSC	Relative Source Contribution
RSR	Rapid Systematic Review
SAB	Science Advisory Board
SDWA	Safe Drinking Water Act
STEWARDS	Sustaining the Earth’s Watersheds Agricultural Research Database System
sHQ	Screening Hazard Quotient
SURF	Surface Water Database – California Department of Pesticide Regulation
SYR 3	Third Six-Year Review

TEST	Toxicity Estimation Software Tool
ToxRefDB	Toxicity Reference Database
TRI	Toxics Release Inventory
TSCA	Toxic Substances Control Act
UCMR 1	First Unregulated Contaminant Monitoring Rule
UCMR 2	Second Unregulated Contaminant Monitoring Rule
UCMR 3	Third Unregulated Contaminant Monitoring Rule
UCMR 4	Fourth Unregulated Contaminant Monitoring Rule
UCM-State	Unregulated Contaminant Monitoring-State
UF	Uncertainty Factor
USDA	U.S. Department of Agriculture
USGS	U.S. Geological Survey
UV	Ultraviolet
WHO	World Health Organization

Chapter 1 Introduction

1.1 Background

Section 1412(b)(1)(B)(i) of the Safe Drinking Water Act (SDWA), as amended in 1996, requires the U.S. Environmental Protection Agency (EPA or "the Agency") to publish every five years a list of drinking water contaminants that at the time of publication:

- Are not subject to any proposed or promulgated National Primary Drinking Water Regulation (NPDWR).
- Are known or anticipated to occur in public water systems (PWSs)
- May require regulation under SDWA.

This list is known as the Contaminant Candidate List (CCL).

SDWA directs the Agency to consider health effects and occurrence information for unregulated contaminants to identify those contaminants that present the greatest public health concern related to exposure from drinking water. In identifying these contaminants, SDWA requires that, when developing the CCL, the EPA considers the National Contaminant Occurrence Database established under Section 1445(g) of SDWA and consults the scientific community including the Science Advisory Board (SAB). The EPA must consider substances identified in Section 101(14) of the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) and substances registered as pesticides under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) as well as other relevant data sources.

As required by SDWA, the Agency also considers adverse health effects, after exposure to a contaminant, that may pose a greater risk to life stages and other sensitive groups that represent a meaningful portion of the population. Sensitive groups or life stages include infants, children, pregnant women, the elderly, and individuals with a history of serious illness.

In an action separate from the CCL, SDWA Section 1412(b)(1)(B)(ii) directs the EPA to make regulatory determinations on at least five of the contaminants from the CCL every five years. Section 1412(b)(1)(A) of SDWA specifies that the EPA shall regulate a contaminant if the EPA Administrator determines the following:

- The contaminant may have an adverse effect on the health of persons;
- The contaminant is known to occur or there is a substantial likelihood that the contaminant will occur in PWSs with a frequency and at levels of public health concern; and
- In the sole judgment of the Administrator, regulation of such contaminant presents meaningful opportunity for health risk reduction for persons served by PWSs.

The CCL itself does not pose a burden or place requirements on the states or PWSs. Rather, the CCL serves as a first level of evaluation for unregulated drinking water contaminants that may need further investigation of potential health effects and the levels at which they are found in drinking water, including prioritized consideration for research and data collection efforts under the Unregulated Contaminant Monitoring Rule (UCMR). Only after additional data and information are collected are contaminants from the CCL considered for regulatory determination and rulemaking under SDWA.

Prior to CCL 6, the EPA had completed five previous cycles of CCLs since 1996 that are briefly described as follows:

Consistent with Executive Order 14303 “Restoring Gold Standard Science,” and guidance from the Office of Science and Technology Policy, the CCL 6 has adhered to the proactive approach outlined therein to ensure that science generated and used throughout the process is transparent, withstands scrutiny, fosters cross disciplinary collaboration, and remains free from bias or undue influence ((Executive Office of the President, 2025); (USEPA, 2026a)).

- The EPA published the First Contaminant Candidate List (CCL 1) on March 2, 1998 (63 FR 10274, (USEPA, 1998). The CCL 1 was developed based on recommendations by the National Drinking Water Advisory Council (NDWAC) and reviewed by technical experts. It contained 50 chemicals and 10 microbial contaminants/groups.

- The EPA published the Second Contaminant Candidate List (CCL 2) on February 24, 2005 (70 FR 9071, (USEPA, 2005b)). The EPA carried forward the 51 chemical and microbial contaminants from the CCL 1 that did not have regulatory determinations to the CCL 2.

- The EPA published the Third Contaminant Candidate List (CCL 3) on October 8, 2009 (74 FR 51850, (USEPA, 2009f)). In developing the CCL 3, the EPA implemented an improved, stepwise process that built on the previous CCL process and was based on expert input and recommendations from the National Academy of Sciences' National Research Council (NRC), NDWAC, and SAB. It contained 104 chemicals or chemical groups and 12 microbial contaminants.

- The EPA published the Fourth Contaminant Candidate List (CCL 4) on November 17, 2016 (81 FR 81099, USEPA, 2016a (USEPA, 2016a)). The EPA carried forward the chemical and microbial contaminants from the CCL 3 that did not have regulatory determinations to the CCL 4. The Final CCL 4 contained 97 chemicals or chemical groups and 12 microbial contaminants. All contaminants listed on the Final CCL 4 were carried forward from the CCL 3, along with two additional public nominations.

- The EPA published the Fifth Contaminant Candidate List (CCL 5) on November 14, 2022 (87 FR 68060 (USEPA, 2022a)). For CCL 5, the EPA established and applied a modified data-driven screening points system to identify and prioritize a subset of chemicals with the greatest potential for public health concern. The Final CCL 5 included 66 chemicals, three chemical groups (cyanotoxins, disinfection byproducts (DBPs), and per- and polyfluoroalkyl substances (PFAS)), and 12 microbial contaminants.

1.2 Overview of the CCL 6 Development Process

The methodology for developing the Sixth Contaminant Candidate List (CCL 6) is based on the three-step framework used previously for the CCLs (CCL 3, CCL 4, and CCL 5). The CCL 6 process closely mirrored the updates implemented during the most recent CCL 5 cycle (USEPA, 2022a) that have allowed the EPA to consider a larger number of contaminants, enhance transparency in the data being evaluated, and improve efficiency of transferring information compiled for CCL to other SDWA processes such as Regulatory Determination and UCMR activities.

A simplified illustration of the CCL development framework for chemicals (adapted from Exhibit 1 in (USEPA, 2009a) is shown in Figure 1. The CCL framework involves three steps:

- **Step 1** includes compiling the CCL 6 Universe of potential drinking water contaminants. During this step, the EPA identified the main data sources (i.e., those data sources for drinking water contaminants that met the four assessment criteria for usability on the CCL described in Section 2.2). As directed by SDWA, the EPA considered health effects and occurrence information on unregulated contaminants to identify those that present the greatest public health concern related to exposure from drinking water. Chemical contaminant data from data sources that met four assessment factors (relevance, completeness, redundancy, and retrievability) were compiled into a single file, with a uniform format and identifiers for chemical contaminants.

- **Step 2** involves screening the CCL 6 Universe and publicly nominated chemicals to identify a subset of chemicals that have the potential to occur in PWSs and thereby may pose a public health concern. This subset of chemicals is called the Preliminary CCL 6 (PCCL 6). In this step, the EPA applied a screening points system that was related to the potential for a chemical to occur in PWSs and the potential for public health concern. The EPA screened chemicals for inclusion on the PCCL 6 by evaluating the health effects and occurrence information provided in the data sources used to compile the CCL 6 Universe. The screening procedure is designed to balance known information regarding toxicity, exposure, and risk by assigning higher value to data that are indicative of an occurrence in finished drinking water and/or potential to cause health effects.

- **Step 3** encompasses a structured approach for selecting the CCL 6 from the PCCL 6. The EPA summarized all relevant data elements in a standardized document format called Contaminant Information Sheets (CISs) in order to further evaluate each chemical. EPA scientists with a broad range of professional experience and relevant expertise evaluated chemicals for the CCL 6 based on the available health effects and occurrence data. These chemical evaluators used the CISs to assess potential public health risk when comparing metrics across chemicals with diverse types of available data and made recommendations on which of the PCCL 6 chemicals should be listed on the CCL 6. The EPA then reviewed these recommendations and made refinements to the list based on the history of each contaminant under the SDWA regulatory determination process.

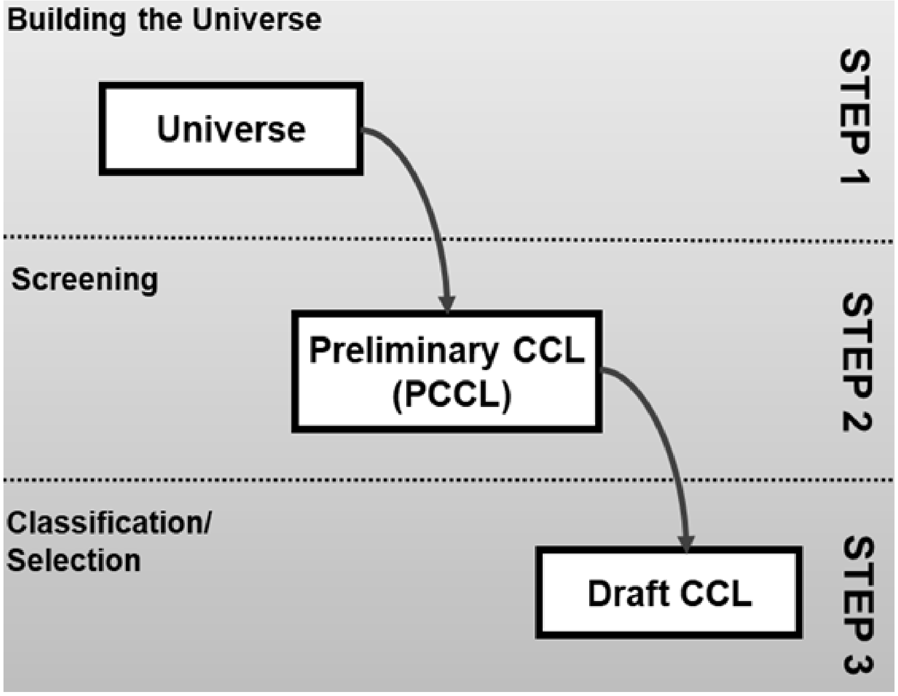


Figure 1. CCL Development Framework

This Technical Support Document describes in detail the process used to develop the draft CCL 6 for chemical contaminants and the updates made in response to expert input and recommendations from the SAB and the public. This document is organized in six chapters:

- Chapter 1 provides background information on the CCL process and an overview of the CCL 6 development process.
- Chapters 2, 3 and 4 describe in detail Steps 1, 2, and 3, respectively.
- Chapter 5 presents the data availability assessment of draft CCL 6 chemicals.
- Chapter 6 describes data management and quality assurance.

The companion documents to the Draft CCL 6 Chemical Technical Support Document include the following:

- Technical Support Document for the Draft Sixth Contaminant Candidate List (CCL 6) – Contaminant Information Sheets (CISs), hereafter referred to as the CIS (USEPA, 2026b)
- Technical Support Document for the Draft Sixth Contaminant Candidate List (CCL 6) – Microbial Contaminants (USEPA, 2026c)

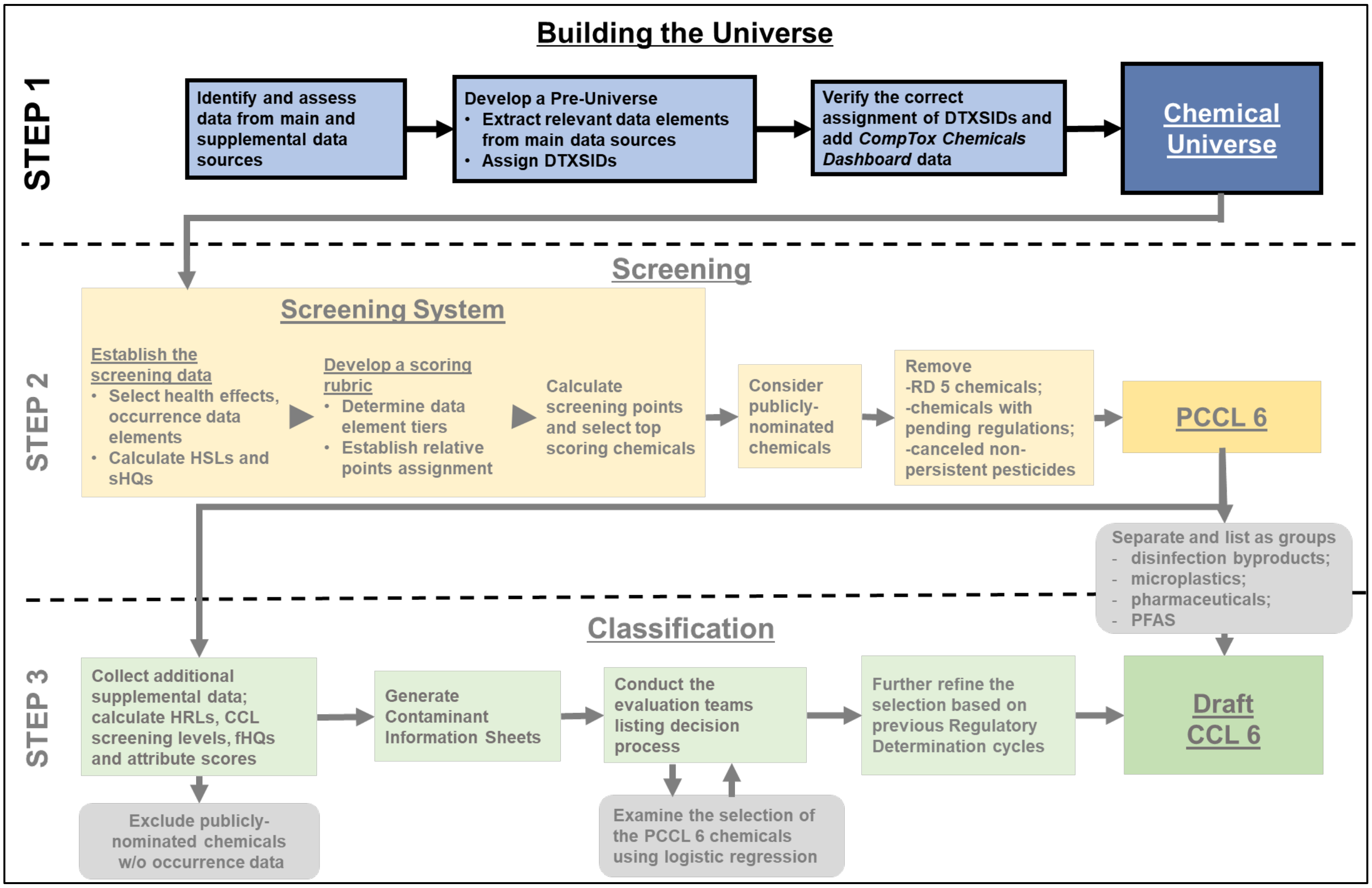
All three technical support documents are accessible via the EPA docket (Docket ID No. EPA-HQ-OW-2022-0946) at https://www.regulations.gov.

Chapter 2 Building the Universe

2.1 Overview

The purpose of Step 1 of the CCL 6 development process is to construct a broad universe of potential drinking water chemical contaminants, as shown in blue in Figure 2. During this step, the EPA compiled main and supplemental data sources, identified approximately 25,000 chemicals from the main data sources to form a Pre-Universe, then added CompTox Chemicals Dashboard data and verified unique chemical identifiers to create the Universe. For CCL 6, the Agency retained all chemical contaminants in the Universe that were identified during the initial Pre-Universe stage of the process, resulting in the largest and most data-rich CCL Universe generated at the time of publication and containing more than 450,000 rows of data.

Figure 2. CCL 6 Development Framework Step 1 - Building the Universe



2.2 Assessing and Identifying Data Sources

To begin the CCL 6 development process, the EPA started by compiling potential health effects and occurrence data sources that could be used to prioritize chemical contaminants for listing on the CCL 6. The EPA gathered data sources identified by the SAB from CCL 5 along with data sources recommended by the CCL 6 EPA workgroup and other Agency subject matter experts. The EPA also conducted a literature search to identify new main data sources using keywords related to water monitoring data and emerging contaminants. Information on how the EPA addressed data sources provided through the public nomination process is described in Section 3.6.

Through this effort, the EPA identified 121 potential new data sources and further assessed their usability for the CCL 6 development process. The EPA accessed each potential data source and evaluated them using the following four assessment factors, according to the process depicted in Figure 3:

- **Relevance:** The source must contain data that either show that the contaminant occurs or has the potential to occur in the environment or the contaminant has known or potential health effects in humans. For example, the EPA collects data on the volume of different chemicals produced in the U.S. under the Chemical Data Reporting (CDR) rule (USEPA, 2020). This information can indicate potential occurrence of chemicals in the environment and therefore would be considered a relevant source of data for CCL 6 development. The EPA's Integrated Risk Information System (IRIS) database would also be considered a relevant source of data because it provides toxicity values such as references doses (RfDs) and cancer slope factors (CSFs) that indicate potential human health effects of chemicals (USEPA-n.d.-a). For example, an RfD serves as an estimate of a daily oral exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious effects during a lifetime. Sources with data on marine water quality or international water quality were not considered relevant to drinking water in the United States and were not used for CCL 6. For example, a dataset from Canada (National Long-term Water Quality Monitoring Data) was evaluated, and it was determined that none of the monitoring sites in the dataset were relevant to U.S. drinking water sources.
- **Completeness:** The minimum completeness requirements are that the data source must either have been peer-reviewed or provide a description of the data, information on how the data were obtained, and information for a person to contact about the data source. The California Department of Pesticide Regulation (CDPR) Surface Water Database (SURF) is an example of a complete source because it provides information on who to contact about the data source as well as a description of the data and how the data were obtained (CDPR, n.d.).
- **Redundancy:** The data source must not duplicate or contain information that is identical to other, more comprehensive data sources. That is, the source should not be identical in terms of what data were collected, the time and place of collection, who collected the data, and how the data were collected and modified. If multiple data sources present identical information, data from the most comprehensive source are used. For example, the EPA's Database of Sources of Environmental Releases of Dioxin-Like Compounds in the United States contains data on chlorinated dibenzo-p-dioxin/dibenzofuran emissions from all known sources in the United States (USEPA, 2000a). However, the same data can also be found in another, more comprehensive source, the EPA's Toxics Release Inventory (TRI) (USEPA-n.d.-b). Therefore, data from the more comprehensive source, TRI, were used for CCL 6 while the other source was considered redundant and was not used. Additionally, although multiple data sources may present values for the same data element, this does not make the data sources redundant. For example, the EPA's IRIS database and California Environmental Protection Agency's (CalEPA) Office of Environmental Health Hazard Assessment's (OEHHA) Chemical Database both provide a reference concentration (RfC) for the chemical 1,4-dioxane (USEPA-n.d.-a) (CalEPA, n.d.). The EPA used both data sources and retained both RfC values because they were developed by different organizations at different times (see the CIS for 1,4-dioxane in (USEPA, 2026b)).

- **Retrievability:** The data must be publicly accessible and formatted for automated retrieval (i.e., data are stored in a tabular format). For example, the Agency for Toxic Substances and Disease Registry (ATSDR) provides Minimal Risk Levels (MRLs) in a tabular format that can be easily copied and pasted into a Microsoft Excel spreadsheet and subsequently added to a data directory to support CCL 6 development; the data are also accessible to the public (CDC, n.d.).

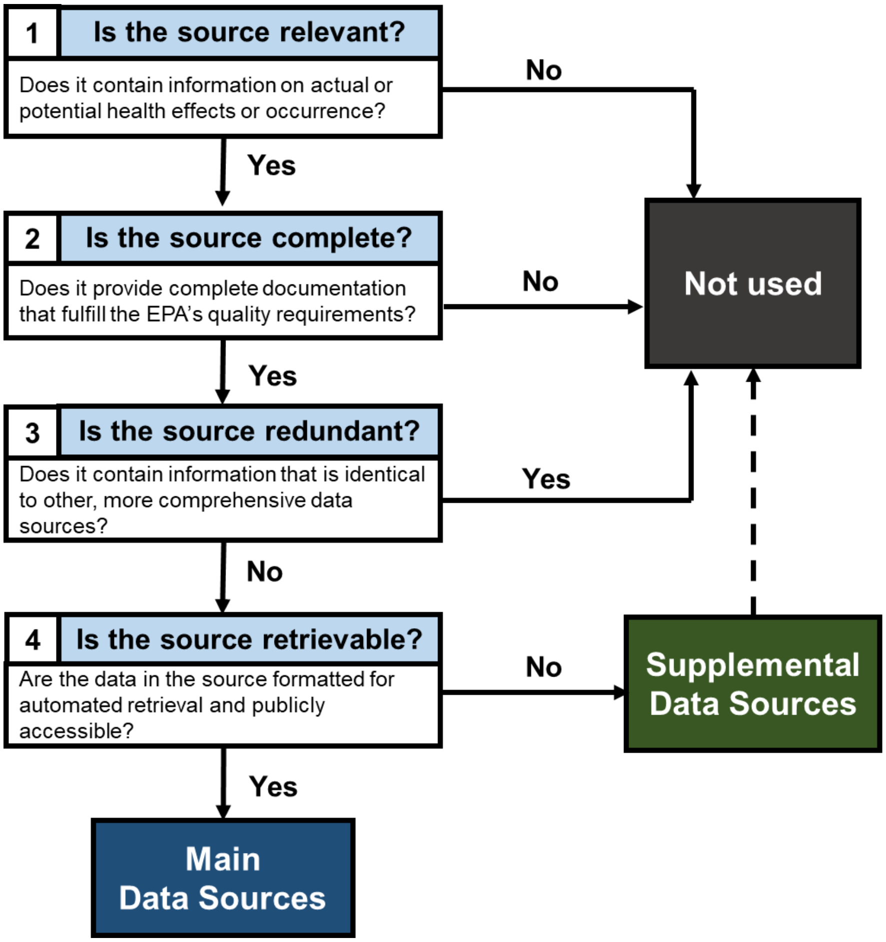


Figure 3. Data Source Assessment Process

These four assessment factors were first used to evaluate data sources under the CCL 3 development process (USEPA, 2009a) based on guidance from NDWAC. NDWAC recommended that data sources should have data and information about actual or potential occurrence of contaminants in drinking water or source water and/or about health effects, provide data that are readily available, and meet the EPA's minimum guidelines for documentation and quality (NDWAC, 2004).

Data sources identified as relevant, complete, not redundant, and retrievable were considered main data sources. If any data sources met all criteria except retrievability, they would have been set aside as supplemental sources.

2.2.1 CCL 6 Main Data Sources

Out of the 121 potential new sources for chemical data that the EPA evaluated for the CCL 6 process, 18 met all four assessment factors required for consideration as main data sources. One existing source from CCL 5, the Hazardous Substances Data Bank (HSDB), now part of PubChem, did not meet retrievability criteria because it lacked a format for automated retrieval but was still used as a main data source (HHS, n.d.); the HSDB is a data-rich source and the only source of median Lethal Doses (LD_{50}s) for the CCL 6 process. The EPA downloaded data from the 18 new main sources and 21 updated sources from CCL 5, along with using 22 static data sources from CCL 5.

Sixty-one main data sources in total are included in the CCL 6 data directory used to identify chemical contaminants for the Pre-Universe. These included 20 sources of health effects data listed in Table 1 and 41 sources of occurrence data as shown in Table 2. Further descriptions of these sources are provided in (Appendix A – Main Data Source Descriptions), while references for the sources and the data processing performed by the EPA to each one are detailed in Appendix M - Data Management for CCL 6. The EPA included health effects and occurrence data from main data sources through July 2023 for health sources and December 2023 for occurrence sources. Complete references for the main data sources listed in Table 1 and Table 2 are provided in Appendix M.

Table 1. CCL 6 Health Effects Main Data Sources

Data Source	Agency or Author[1]	Static, Updated, or New Since CCL 5
Agency for Toxic Substances and Disease Registry (ATSDR) Minimal Risk Levels (MRLs) and Toxicological Profiles	Centers for Disease Control and Prevention (CDC)	Updated
Cancer Potency Data Bank (CPDB)	National Library of Medicine, U.S. Department of Health and Human Services (HHS)	Static
“Development and Validation of a Computational Model for Androgen Receptor Activity”	Kleinstreuer et al. 2017	Updated[2]
Drinking Water Standards and Health Advisories (DWSHA) Tables, Health Advisory Documents, and Health Effects Support Documents	U.S. Environmental Protection Agency (EPA)	Static
Endocrine Disruptor Screening Program (EDSP) – Estrogen Receptor Bioactivity	U.S. Environmental Protection Agency (EPA)	New
Guidelines for Canadian Drinking Water Quality	Health Canada	Updated
Guidelines for Drinking-Water Quality	World Health Organization (WHO)	Updated
Hazardous Substances Data Bank (HSDB)	National Library of Medicine, U.S. Department of Health and Human Services	Static
Health-Based Screening Levels (HBSLs)	U.S. Geological Survey (USGS)	Static
Human Health-Based Water Guidance Table	Minnesota Department of Health	Updated
Human Health Benchmarks for Pesticides	U.S. Environmental Protection Agency (EPA)	Updated
Integrated Risk Information System (IRIS)	U.S. Environmental Protection Agency (EPA)	Updated
International Agency for Research on Cancer (IARC) Classifications	World Health Organization (WHO)	Updated

Data Source	Agency or Author[1]	Static, Updated, or New Since CCL 5
Agency for Toxic Substances and Disease Registry (ATSDR) Minimal Risk Levels (MRLs) and Toxicological Profiles	Centers for Disease Control and Prevention (CDC)	Updated
Cancer Potency Data Bank (CPDB)	National Library of Medicine, U.S. Department of Health and Human Services (HHS)	Static
"Development and Validation of a Computational Model for Androgen Receptor Activity"	Kleinstreuer et al. 2017	Updated[2]
Maximum Recommended Daily Dose (MRDD) Database	Food and Drug Administration (FDA)	Static
National Recommended Water Quality Criteria – Human Health Criteria	U.S. Environmental Protection Agency (EPA)	Static
National Toxicology Program (NTP) Cancer Classifications	U.S. Department of Health and Human Services	Static
Provisional Peer-Reviewed Toxicity Values (PPRTVs)	U.S. Environmental Protection Agency (EPA)	Updated
Human Health Benchmarks for Pharmaceuticals (HHB-Rx) in Drinking Water	U.S. Environmental Protection Agency (EPA)	Updated[3]
Toxicity Criteria Database	California Environmental Protection Agency (CalEPA) - Office of Environmental Health Hazard Assessment (OEHHA)	Updated
Toxicity Reference Database (ToxRefDB)	U.S. Environmental Protection Agency (EPA)	Updated

[1] References for the data sources listed in this table are provided in Appendix M.

[2] Qualitative data from Kleinstreuer et al. (2017) were used to support CCL 5 development. For CCL 6, the EPA updated this data source by using quantitative data presented in the publication.

[3] This data source is an updated version of the "Screening Levels for Pharmaceuticals" that was used in CCL 5.

Table 2. CCL 6 Occurrence Main Data Sources

Data Source	Agency or Author[1]	Static, Updated, or New Since CCL 5
ATSDR Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) Substance Priority List	Centers for Disease Control and Prevention (CDC)	Updated
Chemical Data Reporting (CDR) Results	U.S. Environmental Protection Agency (EPA)	Updated
"Concentrations of prioritized pharmaceuticals in effluents from 50 large wastewater treatment plants in the US and implications for risk estimation"	Kostich et al. 2014	Static
Disinfection By-product Information Collection Rule (DBP ICR)	U.S. Environmental Protection Agency (EPA)	Static
"Evaluating the extent of pharmaceuticals in surface waters of the United States using a National-scale Rivers and Streams Assessment survey"	Batt et al. 2016	Static
"Expanded target-chemical analysis reveals extensive mixed-organic-contaminant exposure in U.S. streams"	Bradley et al. 2017	Static

Data Source	Agency or Author[1]	Static, Updated, or New Since CCL 5
Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) registered pesticides and pesticide ingredients	U.S. Environmental Protection Agency (EPA)	Updated
"Legacy and emerging perfluoroalkyl substances are important emerging water contaminants in the Cape Fear River Watershed of North Carolina"	Sun et al. 2016	Static
"Multimedia screening of contaminants of emerging concern (CECS) in coastal urban watersheds in southern California (USA)."	Maruya et al. 2016	New
National Aquatic Resource Surveys, 2017 National Lakes Assessment	U.S. Environmental Protection Agency (EPA)	New
National Aquatic Resource Surveys, Rivers and Streams 2018-2019 Algal Toxin	U.S. Environmental Protection Agency (EPA)	New
National Health and Nutrition Examination Survey (NHANES)	Centers for Disease Control and Prevention (CDC)	Updated
National Inorganics and Radionuclides Survey (NIRS)	U.S. Environmental Protection Agency (EPA)	Static
National Water Information System (NWIS)	U.S. Geological Survey (USGS)	Updated
National Water-Quality Assessment (NAWQA)	U.S. Geological Survey (USGS)	Updated
"Nationwide reconnaissance of contaminants of emerging concern in source and treated drinking waters of the United States"	Glassmeyer et al. 2017	Static
"Nationwide reconnaissance of contaminants of emerging concern in source and treated drinking waters of the United States: Pharmaceuticals"	Furlong et al. 2017	Static
National Energy Water Treatment and Speciation Database (NEWTS) Coal Mine Drainage Dataset from Cravotta Brady (2015)	U.S. Department of Energy (DOE)	New
NEWTS Geothermal Technologies Office Rare Earth Element Assessment Report Database	U.S. Department of Energy (DOE)	New
NEWTS Ash Database	U.S. Environmental Protection Agency (EPA)	New
NEWTS Flue Gas Desulfurization (FGD) Effluent Database	U.S. Environmental Protection Agency (EPA)	New
NEWTS Leachate Dataset	U.S. Environmental Protection Agency (EPA)	New
NEWTS Brackish Water Database	U.S. Geological Survey (USGS)	New
NEWTS National Produced Waters Geochemical Database v2.3	U.S. Geological Survey (USGS)	New
"Occurrence of antibiotics, estrogenic hormones, and ultraviolet (UV) filters in water, sediment, and oyster tissue from the Chesapeake Bay."	He et al. 2019	New

Data Source	Agency or Author[1]	Static, Updated, or New Since CCL 5
Pesticide Data Program (PDP)	U.S. Department of Agriculture (USDA)	Static
Pesticide Use Estimates	U.S. Geological Survey (USGS)	Updated
“Pharmaceuticals, hormones, pesticides, and other bioactive contaminants in water, sediment, and tissue from Rocky Mountain National Park”	Battaglin et al. 2018	New
“Pharmaceutical manufacturing facility discharges can substantially increase the pharmaceutical load to U.S. wastewaters”	Scott et al. 2018	Static
“Pharmaceuticals, perfluorosurfactants, and other organic wastewater compounds in public drinking water wells in a shallow sand and gravel aquifer.”	Schaider et al. 2014	New
“Predicting variability of aquatic concentrations of human pharmaceuticals”	Kostich et al. 2010	Static
“Reconnaissance of mixed organic and inorganic chemicals in private and public supply tapwaters at selected residential and workplace sites in the United States”	Bradley et al. 2018	Static
“Simultaneous Determination of Selected Trace Contaminants in Drinking Water Using Solid-Phase Extraction-High Performance Liquid Chromatography-Tandem Mass Spectrometry”	Zhang et al. 2019	New
Surface Water Database (SURF)	California Department of Pesticide Regulation (CDPR)	Updated
“Suspect screening and non-targeted analysis of drinking water using point-of-use filters”	Newton et al. 2018	Static
Sustaining the Earth’s Watersheds - Agricultural Research Database System (STEWARDS)	U.S. Department of Agriculture – Agricultural Research Service (USDA-ARS)	New
Toxics Release Inventory (TRI)	U.S. Environmental Protection Agency (EPA)	Updated
Unregulated Contaminant Monitoring Rule (UCMR) Cycles 1-4	U.S. Environmental Protection Agency (EPA)	Static
Unregulated Contaminant Monitoring Rule (UCMR) Cycle 5	U.S. Environmental Protection Agency (EPA)	New
Unregulated Contaminant Monitoring-State (UCM-State) Rounds 1 and 2	U.S. Environmental Protection Agency (EPA)	Static
“Urban Stormwater: An Overlooked Pathway of Extensive Mixed Contaminants to Surface and Groundwaters in the United States”	Masoner et al. 2019	New

[1] References for the data sources listed in this table are provided in Appendix M.

2.2.2 CCL 6 Supplemental Data Sources

Depending on the data availability for the PCCL chemicals identified from main sources during the Step 1 stage of the CCL development process, it is often necessary to gather and extract additional data during Step 3 (Classification) of the process to further evaluate chemicals for listing on the CCL 6. The data sources set aside as supplemental during CCL 5 (because they met the relevance, completeness, and redundancy assessment factors but were not retrievable because of their format) continued to be considered for use in CCL 6. The EPA conducted literature searches to identify further supplemental data on occurrence and health effects to aid in evaluating chemicals of interest (see Section 4.2). Although supplemental sources could not be efficiently or effectively incorporated into the Step 2 screening process because they did not meet retrievability criteria (see Chapter 3), they often provided important additional data and chemical descriptions that supported CCL 6 listing recommendations. See Appendix B for a complete list of supplemental data sources.

For health effects data, supplemental data sources were often closely related to a main data source. For example, the EPA's IRIS program provides an easily accessible and downloadable online database (Integrated Risk Information System (IRIS) database) that contains toxicity values for several hundred chemicals. The IRIS database met the four assessment factors to be a main data source for CCL 6. Although the online IRIS database fulfilled data needs for screening purposes, background information related to developing toxicity values for individual chemicals of potential importance for the classification process of CCL 6 were not available in this database and had to be manually extracted from IRIS assessments. Therefore, for PCCL 6 chemicals, the EPA also downloaded IRIS Chemical Assessment Summaries and Toxicological Reviews as supplemental data sources. Other supplemental health effects data sources are discussed further in Section 4.2.2 and Section 4.3.1 as well as Appendix G.

Supplemental occurrence data sources could also be used to fill data gaps during the Step 3 process as needed. Data sources identified during CCL 5 continued to be used to supplement the occurrence data from main data sources (see Appendix B). This includes non-national water databases that report contaminant monitoring data in finished water or ambient water.

2.3 Developing a Pre-Universe

2.3.1 Overview

The Pre-Universe is a list of chemical contaminants identified through health and occurrence data extracted from the main data sources. For CCL 6, the EPA developed the Pre-Universe in three steps: extracting chemicals and relevant data elements, matching unique identifiers to each chemical, and transforming the extracted data into a simple data format. Each step is described under subsections 2.3.2 through 2.3.4.

Generally, the Pre-Universe development involved "pre-processing", which refers to actions the Agency took to identify chemicals from each main data source and transform data of various types, formats, and structures into a uniform and understandable format. Each data source used for CCL 6 has a unique data formatting type and requires specific steps to properly extract relevant data elements and create data entries. Additional information on the pre-processing of main data sources is provided in Appendix M. For CCL 6, data elements are defined as values or descriptors with unique meanings that characterize either the toxicological or occurrence information associated with chemical contaminants. Two examples of CCL 6 data elements are RfDs and finished water detection rates for nationally representative monitoring programs. Data entries are defined as singular data elements relating to a specific chemical. An example of a data entry is the 89% finished water detection rate in public water systems for manganese from UCMR 4, a nationally representative monitoring program.

The EPA extracted approximately 132,000 rows of data from the 61 main data sources used to generate a Pre-Universe file. Included among those data rows were 49 unique data elements described in Table 4 of Section 3.2, as well as Section M.5 of Appendix M. Later in Step 2 (Screening) and Step 3 (Classification), the EPA extracted additional finished water and ambient water occurrence data elements for the PCCL chemicals.

2.3.2 Extracting Relevant Data Elements for Developing the Pre-Universe

Several relevant data elements were extracted for the development of the Pre-Universe. The categories of data elements include dose-response data, categorical toxicity data (e.g., cancer classifications), modeled bioactivity data, finished drinking water data, ambient water data, environmental release data, and chemical production data. Each data element type may contain several relevant data elements. For example, dose-response data include data elements such as No Observed Adverse Effect Levels (NOAELs), Lowest Observed Adverse Effect Levels (LOAELs), RfDs, and LD_{50}s. Similarly, finished drinking water data include relevant data elements, such as maximum concentration and percentage of sites or number of samples with detections.

Data sources may provide one or multiple data elements relevant to the CCL 6 development process. For example, national finished drinking water monitoring programs, such as the Unregulated Contaminant Monitoring Rule (UCMR), provide both maximum concentrations and percent detection data.

During the development of the prior CCL 5, the EPA recognized the potential risk for exposure even when a chemical is reported as a non-detection; therefore, the EPA adopted a more health-protective approach to handle non-detections in ambient and finished water data in the screening stage. For the CCL 6 cycle, the EPA substituted maximum concentration values for chemicals with non-detections in two ways. First, if the data source provided a single reporting or detection limit, half the value of that detection limit was substituted for the maximum concentration. For example, nationally representative finished water monitoring data from the First Unregulated Contaminant Monitoring Rule (UCMR 1) for diazinon reported zero detects and a method reporting limit of 0.5 µg/L. Therefore, the reported UCMR 1 maximum concentration for diazinon was changed from zero to 0.25 µg/L. Second, if a data source provided a detection limit range, the EPA used half of the highest level of detection in the range. For example, finished water monitoring data for propoxur provided by the U.S. Department of Agriculture Pesticide Data Program (USDA PDP) reported zero detections and a limit of detection (LOD) range of 6×10^{-6} µg/L – 4.13×10^{-4} µg/L. Therefore, the EPA used 2.065×10^{-4} µg/L as the maximum concentration. If no reporting or detection limits were available, maximum concentration values for non-detections were simply reported as "NA." Further details on how non-detections were handled for a specific data source are included in Chapter 6 and Appendix M.

Appendix M describes specifics about the pre-processing required to extract data elements from CCL 6 main data sources used to develop the Pre-Universe file, including how to access the source data on the internet, when the data were accessed, and any manipulation or calculations performed on the raw data. The EPA documented the exact process used to manipulate and extract data in the form of R Markdown files (Allaire, et al., 2020); (R Core Team, 2025), which include code and relevant notes.

2.3.3 Assigning Unique Contaminant Identifiers

The data directory for CCL 6 development must be able to accurately identify and map thousands of chemicals with their respective health and occurrence information across all main and supplemental sources. To accomplish this, the EPA relied on Distributed Structure-Searchable Toxicity Substance

Identifiers (DTXSIDs) and Chemical Abstracts Service Registry Numbers (CASRNs) along with any other identifiers provided by the data source authors.

The EPA's Distributed Structure-Searchable Toxicity (DSSTox) Public Database Network (https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox-database) is a curated compilation of chemical names and structures with a unique identifier system called the DSSTox substance identifier (DTXSID), which the EPA used to help identify chemicals and compile chemical-specific data for CCL 6. There are benefits of using DTXSIDs as the identifier system during the CCL 6 Pre-Universe and Universe development. First, DTXSIDs are curated by the EPA to ensure that each DTXSID refers to one unique chemical or chemical group (Williams, et al., 2017). Second, the EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard) publishes mapping files that match DTXSIDs to other chemical identifiers, including chemical names, CASRNs, International Chemical Identifier strings, and International Chemical Identifier keys. These mapping files allowed the EPA to efficiently and accurately compile data provided by multiple data sources that used different chemical identifiers. Verification of DTXSIDs occurred while building the Pre-Universe file as discussed later in Section 2.4.2.

2.3.4 Saving Extracted Metrics in a Simple Data Format

A uniform "simple" data format was chosen so that all main and supplemental data could be easily combined and used in later steps of the CCL 6 development process. This simple format includes seven critical pieces of information about each data entry required for the second step of the CCL 6 process:

- Chemical name or identifier as reported in the data source
- CASRN
- DTXSID
- Value of the data element extracted from the data source
- Units of the data element
- Name of the data source from which the data element was extracted
- Type of data element extracted

Further information about the simple data format can be found in Appendix M.

2.4 Enhancing the Universe

2.4.1 Overview

The EPA used the Pre-Universe as the foundation to prepare a finalized Universe of chemicals and their associated data elements for efficient and effective use during Steps 2 and 3 of the CCL 6 development process (described later in Chapter 3 and Chapter 4). During this enhancement substep of Universe development, the Agency verified DTXSIDs of chemicals identified in the Pre-Universe, added relevant supplemental data collected for Pre-Universe chemicals from the EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard), and generated a file to present data elements from different data sources in a uniform format. This Universe file was used to screen chemicals for inclusion on the PCCL 6 and classify chemicals for inclusion on the CCL 6.

The number of chemicals included in the CCL 6 Pre-Universe and CCL 6 Universe remained the same. However, the amount of data associated with the chemicals improved significantly, demonstrated by the increase in data file rows from 132,000 in the Pre-Universe file to more than 450,000 rows in the Universe file.

2.4.2 Verifying DTXSID Assignments

Based on experience with the CCL 5 process, the Agency made the decision during the early stages of CCL 6 development to pull both the names and (whenever provided by the original data sources) CASRNs for chemicals to populate the updated Simple Data Format (see Section 2.3.4); this decision allowed the EPA to more effectively use a batch search function within the CompTox Chemicals Dashboard that could rely on two fields of information to more accurately map chemicals entries provided by the main data sources to DTXSID assignments, and also to verify any pre-existing DTXSID assignments from any of the static CCL 5 sources that were brought forward when constructing the Pre-Universe file. With this new approach, nearly all entries from the Pre-Universe files were able to be correctly mapped, condensing the 130,000 rows in the Pre-Universe file down to approximately 25,000 unique chemicals. Any entry that could not be assigned a DTXSID using this process was temporarily designated "NA" for its unique identifier key field in the Pre-Universe file. Data entries with "NA" were subsequently reviewed on an individual basis for follow-up curation. If no DTXSID could be assigned after this additional manual step, those entries were removed from the finalized Pre-Universe file.

Out of the 130,000 rows of data in the Pre-Universe file, approximately 750 entries required manual curation by curators from the CompTox staff in cases that data entries couldn't be identified using the batch search function. The EPA identified 57 data entries that could not be matched to an existing DTXSID in the CompTox Chemicals Dashboard. For these entries, the CompTox curators noted the reason for these omissions, including ambiguous or unregistered names, classes of chemicals, radiation, and biological entries like spores and moths. The EPA ultimately excluded these 57 entries.

The EPA further cleaned up contaminant identifiers matched during the CCL 6 Pre-Universe development by grouping DTXSIDs for chemicals that would dissociate to the same compound in water (e.g., the EPA assigned the same DTXSIDs to lithium and lithium salts because they all form the lithium ion in water) and editing incorrectly matched DTXSIDs during Pre-Universe development. The EPA corrected DTXSIDs manually when evidence suggested that certain chemicals should be grouped or distinguished from one another. The EPA performed an extensive quality assurance/quality control (QA/QC) review of DTXSID assignments throughout the CCL 6 development process to flag and correct any mismatched DTXSIDs (see Section 6.2.4 QA/QC for DTXSID Assignments).

The EPA's analysis showed that several chemicals with different DTXSIDs should be grouped under a single DTXSID. For example, many studies related to the oral toxicity of lithium report lithium chloride salt (DTXSID2025509) as the compound tested in the study because this salt was used to generate the lithium solution dosed to the animals in the experiment. In contrast, monitoring studies measuring lithium in drinking water or ambient water frequently report the resulting concentrations simply as "lithium" (DTXSID5036761). Lithium can also be matched to a DTXSID describing "lithium ions" (DTXSID10169612). Due to the level of detailed review that would be required to determine any differences in toxicity between various lithium salts and the speciation of lithium expected in drinking water, for the CCL 6 Universe, the EPA considered all data relevant to lithium and lithium salts as one group and therefore grouped them under a single DTXSID (DTXSID5036761).

A similar example of grouping contaminants under a single DTXSID in the Universe is entries describing "1-butanol" grouped with entries describing "1-butanol, sodium salt," entries describing "dalapon" grouped with entries describing "dalapon sodium," and entries describing "potassium bromate" and "sodium bromate" grouped with entries describing "bromate ion." Although this type of correction may apply to many chemicals in the Universe, it was not feasible for the EPA to identify all instances, so efforts focused on identifying chemicals with ionized and/or salt forms (e.g., inorganic ions) as was done for CCL 5.

2.4.3 Additional Data Accessed via the CompTox Chemicals Dashboard

The EPA downloaded and appended supplemental data from the CompTox Chemicals Dashboard to broaden the available data for chemicals identified during Pre-Universe development. The CompTox Chemicals Dashboard provides easy access to results from qualitative structure-activity relationship (QSAR) and to ExpoCast models that the EPA and others developed to predict toxicity endpoints, physical properties, and exposure and environmental fate parameters for chemicals based on their structures. QSAR models are useful and valid only within their applicability domain; that is, if the types of chemicals tested were not included in the training dataset for the model, the model could produce unrealistic predictions.

Data from the CompTox Chemicals Dashboard was the only supplemental source the EPA relied on for additional data elements used for screening (Chapter 3), although only select data elements were used during this step. Although the CompTox Chemicals Dashboard technically meets the four assessment criteria for inclusion under the CCL protocol as a main data source, its primary function is to assist the Agency with mapping and assigning the unique contaminant keys (i.e., DTXSIDs) contained within in its database to the chemicals identified from the main sources of health effects and occurrence data. If the EPA were to treat the CompTox Chemicals Dashboard as a traditional main data source, the total count for chemicals in the Universe would expand from approximately 25,000 chemicals to over one million chemicals, most of those with minimal data to contribute towards assigning points for prioritization of chemicals to create the PCCL under Step 2 (Screening) of the process (see Section 3.4).

As described in Section 2.2.2, the EPA downloaded supplemental data from other sources for use during the classification step (Chapter 4). These supplemental data, including all data downloaded from the CompTox Chemicals Dashboard, were provided to chemical evaluators on CISs, as further described in Chapter 4. Pre-processing specifics related to downloading, manipulating, and extracting CompTox Chemicals Dashboard data elements can be found in Appendix M.

2.4.4 Creating a Uniform Universe File

The EPA followed several steps to ensure data elements from the multiple sources used for building the Universe file were converted to the same units and reported in the same format. For example, all concentrations in the Universe file referred to as benchmarks were converted to mg/L and all units of dose for oral toxicity values were converted to mg/kg/day or $(\text{mg/kg/day})^{-1}$. To calculate distributions and compare the relative magnitude of data entries, all entries were also converted to a single numeric form. For example, the EPA temporarily modified production data, which can be reported as a range of pounds produced, to a single value (see Section 3.3.2 for additional details).

The EPA also converted categorical cancer classifications to a numeric scheme (1, 2, or 3) according to the same methodology used for CCL 3, CCL 4 and CCL 5 (USEPA, 2009b); (USEPA, 2022b). In CCL 3, cancer classifications were distributed into numerical categories 1, 2, or 3 according to the designations provided in Table 3. The EPA included both the original cancer classifications as designated by the source in the Universe file along with an additional element for the corresponding numerical categories of each cancer classification entry. In this way, cancer classifications from different sources could be compared while maintaining the cancer descriptors as written in the original data sources. The numeric category equivalents for cancer classifications are listed in Table 3. If the cancer classification for a chemical was available from a data source compiled while building the Universe file but was not included in Table 3, the EPA retained the cancer classification from the source but created no new numeric data entry. For example, if a chemical has a cancer classification of "Not likely to be carcinogenic (NL)," which was not associated with a numerical category as defined in CCL 3 (USEPA, 2009b), no numeric entry was assigned. The numeric entries were used for screening (see Chapter 3); however, the EPA reverted back to the original cancer classification entries for the classification step of the CCL 6 process (see Chapter 4).

Table 3. Cancer Classification Numeric Conversions

EPA	International Agency for Research on Cancer (IARC)	National Toxicology Program (NTP)	Numeric Classification
A, H, CA or Ca	1	CE or P in 2 species or 2 sexes	1
B1, B2, Li, L	2A	Combinations of CE, SE, EE and NE or combinations of P, E, and N	2
C, S, SU, Su	2B	Combinations of SE, EE, and NE or combinations of E and N	3

Source: (USEPA, 2009b)

EPA: A = Human carcinogen; H/CA/Ca = Carcinogenic to humans; B1 = Probable human carcinogen; B2 = Limited evidence in animals and inadequate or no evidence in humans; L/Li = Likely to be carcinogenic to humans; C = Possible human carcinogen; S/SU/Su = Suggestive evidence for carcinogenicity

IARC: 1 = Carcinogenic to humans; 2A = Probably carcinogenic to humans; 2B = Possibly carcinogenic to humans

NTP: CE/P = Clear evidence of carcinogenicity; SE = Some evidence of carcinogenicity; EE/E = Equivocal evidence of carcinogenicity; NE/N = No evidence of carcinogenicity

These modifications allowed the Agency to compile and compare data from multiple sources for use during Steps 2 and 3 of the CCL 6 process (see Chapter 3 and Chapter 4). This was especially important for Step 2 (Screening) of the CCL 6 process, which requires a uniform and comprehensive set of data elements to accurately score the approximately 25,000 Universe chemicals for screening down to a PCCL.

Chapter 3 Screening Universe Chemicals to Select the PCCL

3.1 Overview

The purpose of Step 2 of the CCL 6 development process was to screen Universe chemicals for inclusion on the PCCL 6 for further evaluation and potential listing on the draft CCL 6. The PCCL 6 is comprised of the top scoring Universe chemicals and any publicly nominated chemicals not already part of the top scoring Universe chemicals. Certain top scoring chemicals and publicly nominated chemicals were not included on the PCCL 6 because they 1) had ongoing Agency actions under the Regulatory Determination 5 (RD 5) process or other ongoing actions under SDWA, or 2) were canceled and non-persistent pesticides that had no reported alternative uses and that the Agency determined did not warrant further evaluation, as described further in Section 3.7.

In this step, the EPA developed screening scores for Universe chemicals based on the health effects and occurrence data compiled in Step 1 (Building the Universe). To screen chemicals for the PCCL 6, the EPA maintained a similar screening framework as was used in CCL 5, with minor modifications (USEPA, 2022b). Like CCL 5, the CCL 6 screening process requires little to no manual review of data and considers chemicals that are relatively data-poor and data-rich in terms of relevant health effects and drinking water occurrence data. Development of the CCL 6 screening system included the following actions, described in detail in this chapter:

1. Determining the data elements to be used for screening.
2. Determining health screening levels and calculating screening hazard quotients.
3. Establishing a scoring rubric for the relative point assignment across health effects and occurrence data elements.
4. Assigning points to the data elements available for each chemical and calculating a screening score.
5. Selecting chemicals based on screening scores for inclusion on the PCCL 6.

The CCL 6 screening process relied on a transparent and reproducible scoring rubric and a point-based screening system implemented using the R programming language (R Core Team, 2025). The EPA assigned points based on the data elements available for each chemical and the relative toxicity or occurrence indicated by each value. The R script developed and used in the CCL 6 screening process requires only the Universe file as an input and writes an output file containing point assignments for data elements and the screening score (i.e., the sum of a chemical's screening points assigned for each available data element) for each chemical. The EPA used the screening score to identify chemicals most relevant to drinking water exposure that have the potential to cause the greatest health concern.[1] The points assignment and screening processes are further described in Section 3.2 and Section 3.3.

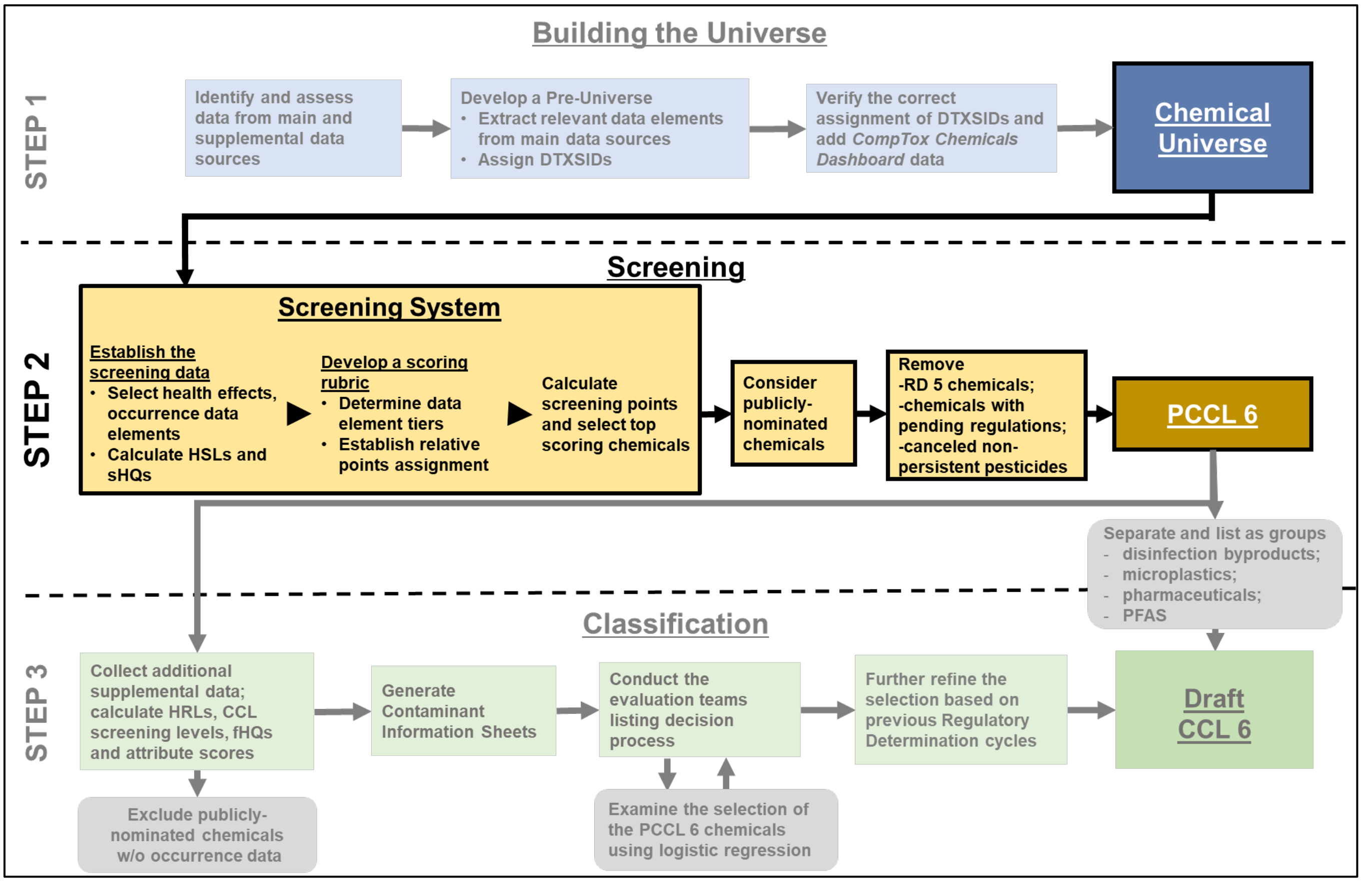
---

[1] Although screening scores were used to prioritize chemicals for inclusion on the PCCL 6, these scores do not reflect the EPA's regulatory priorities for chemicals. The screening points system was designed to reflect the likelihood of a chemical being listed on the CCL 6, but the screening score itself did not influence the recommendations of the chemical evaluators. As discussed in Chapter 4, the evaluation teams were not provided with screening scores to use while assessing chemicals for the CCL 6.

The EPA applied a points-based screening system across the approximately 25,000 CCL 6 Universe chemicals to determine which ones warranted further consideration under the time- and resource-intensive classification process (see Chapter 4). Section 3.5 discusses the use of the CCL 6 screening system for this purpose. Figure 4 illustrates the screening process, indicated by yellow boxes.

The EPA also evaluated publicly nominated chemicals for inclusion on the PCCL 6, as discussed in Section 3.6. Finally, the EPA excluded chemicals from the PCCL 6 that did not warrant further evaluation, as discussed in Section 3.7. Section 3.8 contains a summary of the PCCL 6.

Figure 4. Development Framework Step 2 – Screening



3.2 Establishing the Screening Data

3.2.1 Incorporating Universe Data Elements

The CCL 6 screening process is the same as the CCL 5 screening process, with minor modifications (USEPA, 2022b). The EPA designed the CCL 5 screening process to systematically consider the health effects and occurrence data from the Universe file and advance chemicals for further evaluation using consistent and transparent methods. During the CCL 6 Universe development process, the EPA compiled 71 different data elements to consider for point assignment or as additional information for individual chemicals. Of these data elements, the EPA assigned points to 20 data elements related to health effects and 13 data elements related to occurrence. The data elements used for point assignment are listed in Table 4. The remaining data elements not assigned points are included in Section M.5 of Appendix M.

Many of the data elements assigned points in CCL 6 are the same as were used in the CCL 5 screening and classification processes. These include health effects information such as categories of cancer classifications and toxicity values (e.g., RfD, NOAEL, LOAEL, and LD_{50}), and occurrence information such as measures of concentration and frequency of detections in finished water, CDR production volume and TRI chemical release data, and others. In this step, the EPA interprets broadly the SDWA criterion that contaminants are known or anticipated to occur in PWSs. The EPA considered not only PWS monitoring data but also data on concentrations in ambient surface and ground waters, releases to the environment (e.g., TRI), and production. Although such data may not establish conclusively that contaminants are known to occur in PWSs, the EPA considers these data sufficient to anticipate that contaminants may occur in PWSs and were therefore included in the points assignment.

There are also new data elements related to health endpoints that the EPA included in the CCL 6 screening process that were not used in previous CCL cycles. For example, the EPA created a new data element for and assigned health effects screening points to estrogen receptor-active chemicals analyzed under the EPA's Endocrine Disruptor Screening Program (EDSP) ((USEPA, 2023b); https://www.epa.gov/endocrine-disruption) in an effort to address a public nomination that requested increased focus on estrogenic compounds (see Section 3.6). There were no new occurrence data elements that received screening points; however, some of the data sources used for the existing occurrence data elements that received screening points (e.g., national finished water detection rates) were updated between CCL 5 and CCL 6. For example, in CCL 6 the data from the National Water Information System (NWIS) and National Water Quality Assessment (NAWQA) sources were separated into finished and ambient water. The addition of more finished water sources allowed for the calculation of the screening hazard quotient (Section 3.2.2) for more contaminants.

The EPA also revised existing data elements from CCL 5 to better represent potential risk to the general population. For example, under CCL 5, contaminants could receive multiple sets of points from four different data elements related to neurotoxicity ("Developmental neurotoxins," "Developmental neurotoxins *in vivo,*" "human neurotoxicants," and "mined literature for neurotoxins") (USEPA, 2022b). For CCL 6, the EPA revised the screening system to combine these four lists of contaminants and only award a single set of points regardless of whether the contaminant of interest was on more than one of those lists. This was done to avoid overemphasis of one type of adverse health effect (e.g., neurotoxicity) over other types of effects or contaminants that cause toxicity through other modes of action.

Table 4. Data Elements Assigned Points in the CCL 6 Screening System

Data Element	Description
Health Effects	
Acute benchmark	Health-based concentration in water not expected to result in adverse health effects due to acute (short-term) exposure -- e.g., 10-day Health Advisories, acute or short-term guidance values from the Minnesota Department of Health, acute Human Health Benchmark for Pesticides (HHBPs)
Acute reference dose (RfD)	A toxicity value derived from a study with an acute exposure duration that estimates the dose below which no adverse non-cancer effects are expected to occur -- e.g., acute-duration Minimum Risk Levels (MRLs), acute population-adjusted doses from HHBPs

Data Element	Description
Cancer slope factor (CSF)	A toxicity value that estimates the increased cancer risk per unit of oral dose. This value may also be called an oral slope factor or cancer quantification value (q_1^*) in assessments.
Chronic benchmark	Health-based concentration in water not expected to result in adverse health effects due to chronic exposure -- e.g., Lifetime Health Advisories, 10^{-6} cancer risk concentrations, chronic HHBPs, drinking water guidelines from the World Health Organization (WHO) and Health Canada
Chronic Lowest Observed Adverse Effect Level (LOAEL)	Lowest exposure level that causes adverse effects from a study with a chronic exposure duration, a two-generation study, or a developmental toxicity study
Chronic No Observed Adverse Effect Level (NOAEL)	Exposure level that causes no adverse effects from a study with a chronic exposure duration, a two-generation study, or a developmental toxicity study
Androgen Receptor Bioactivity	Results from Kleinstreuer et al. (2017) modeling of quantitative *in vitro* androgen receptor bioactivity
Estrogen Receptor Bioactivity	Results from the EPA's EDSP modeling of quantitative *in vitro* Estrogen Receptor Bioactivity
Neurotoxins/neurotoxicants[1]	A list of contaminants that cause neurotoxicity, including developmental neurotoxicity, as identified by Grandjean and Landrigan (2006), Mundy et al. (2015), Aschner et al. (2017), and automated literature mining of PubMed (all downloaded from the EPA's CompTox Chemicals Dashboard)
Median lethal dose (LD_{50})	The lethal dose of a contaminant for 50% of the tested animals in a study after a specified exposure duration
Maximum Recommended Daily Dose (MRDD)	Maximum Recommended Daily Dose for FDA-approved pharmaceuticals
Numeric cancer classification	Numeric equivalent of cancer classification according to CCL 3 health effect categories (see Section 2.4.4 for numerical conversions)
PubMed articles	Number of articles returned from a PubMed search for the contaminant of interest (downloaded from the EPA's CompTox Chemicals Dashboard)
Reference Dose (RfD)	A toxicity value derived from a study with a chronic exposure duration or with a developmental or two-generation study design that estimates the dose below which no adverse non-cancer effects are expected to occur – e.g., chronic MRLs and chronic population-adjusted doses from the HHBPs
Subchronic benchmark	Health-based concentration in water not expected to result in adverse health effects due to subchronic exposure -- e.g., Minnesota Department of Health subchronic guidance values
Subchronic LOAEL	Lowest exposure level that causes adverse effects from a study with a subchronic exposure duration
Subchronic NOAEL	Exposure level that causes no adverse effects from a study with a subchronic exposure duration
Subchronic RfD	A toxicity value derived from a study with a subchronic exposure duration that estimates the dose below which no adverse non-cancer effects are expected to occur -- e.g., intermediate-duration MRLs
Median toxic dose (TD_{50})	Dose of a contaminant associated with 50% of animals in the study developing tumors, compiled by the Cancer Potency Data Bank
ToxCast assay percent active	Percent of ToxCast *in vitro* assays for which a contaminant tested as bioactive (downloaded from the EPA's CompTox Chemicals Dashboard)
Occurrence	
Biodegradation half-life – OPERA model	The predicted biodegradation half-life in days, according to the Open structure-activity/property Relationship App (OPERA) model (downloaded from the EPA's CompTox Chemicals Dashboard)
Blood concentrations	90th percentile concentration in human blood, according to NHANES biomonitoring data

Data Element	Description
National ambient water detection rates	Detection rates in ambient water from nationally representative surveys – e.g., USGS Water Quality Portal National Water-Quality Assessment (NAWQA)
National finished water detection rates	Detection rates in finished water from nationally representative monitoring programs – UCMR 1-5, UCM-State Rounds 1 and 2, and NIRS
Non-national ambient water detection rates	Detection rates in ambient water from non-nationally representative studies – e.g., Batt et al. (2016) and Bradley et al. (2017) and others
Non-national finished water detection rates	Detection rates in finished water from non-nationally representative studies – e.g., Bradley et al. (2018) and Furlong et al. (2017)
Pesticide application	Pesticide application rate in kilograms per year (USGS Pesticide Use Estimates)
Presence on FIFRA or CERCLA lists	The contaminant is included on lists from FIFRA or CERCLA (points assigned separately for each applicable list)
Production volume	Total chemical production volume in pounds per year from the EPA's Chemical Data Reporting
Release quantity	Environmental release data from the Toxics Release Inventory in total pounds released per year
Screening hazard quotient	The ratio of the maximum concentration in finished water[2] to the minimum Health Screening Level (see Section 3.2.2)
Serum concentration	90th percentile concentration in human serum, according to NHANES biomonitoring data
Urine concentrations	90th percentile concentration in human urine, according to NHANES biomonitoring data

[1] This data element is a combination of the lists captured under the "developmental neurotoxins," "developmental neurotoxins *in vivo,*" "human neurotoxicants," and "mined literature for neurotoxins" data elements established under CCL 5 (USEPA, 2022b).

[2] EPA's method for assigning maximum concentration values to non-detected chemicals in the screening step of CCL 6 is described in Chapter 2 and Appendix M.

Some data elements in the Universe file were not assigned points for CCL 6 screening purposes. In general, the EPA did not assign points to data elements if they met one or more of the following exclusion criteria:

- Data element was not available for many chemicals.
- Data element was not considered highly relevant to hazards associated with lifetime drinking water consumption.
- Data element required chemical-specific data manipulation (e.g., unit conversions requiring chemical-specific molecular weight) and/or was not comparable to others in the Universe.
- Another data element extracted from the same data source and describing the same data was assigned points.
- Data element was not relevant to unregulated chemicals.

Section M.5 of Appendix M lists the data elements in the Universe file that were not assigned points because the data element met one or more of these exclusion criteria. Examples of data elements meeting these exclusion criteria are detailed below.

California EPA's Maximum Allowable Dose Level (MADL) exposure values, which are designed to reflect a "No Observable Effect Level" related to reproductive toxicity, meet several of these exclusion criteria. MADLs were not assigned points because they often represent a total exposure level for multiple routes of exposure (oral, dermal, intravenous, etc.) that are not necessarily considered highly relevant to hazards associated with drinking water. They are also reported in units of µg/day and subsequently cannot be directly compared to standard toxicity values like oral RfDs derived by the EPA (reported in units of mg/kg/day). However, the EPA did include MADLs as a supplementary source of health effects information on the Chemical Information Sheets (CISs; see Chapter 4), particularly because they describe potential hazards for sensitive populations (reproductive and developmental life stages).

Physical and chemical properties estimated by the EPA QSAR models Toxicity Estimation Software Tool (TEST) and the Open structure-activity/property Relationship App (OPERA), as well as toxicity values based on inhalation data, were also not considered for point assignments. Although these data provide context to occurrence or health effects information, they are not considered directly relevant to potential hazards due to drinking water exposure. More specifically, the EPA prioritized toxicity values based on oral data rather than inhalation data based on the assumption that the oral consumption of drinking water is the primary exposure source of chemicals that may occur in drinking water. Additionally, some predictions, for example the oral rat LD_{50} provided by the TEST model, are in units that would require chemical-specific manipulation (i.e., molar mass conversion to mg/kg from mol/kg for each Universe chemical). LD_{50} values from the TEST model are not readily comparable to LD_{50} values from other data sources and were therefore not included for point assignment. Although data elements meeting the exclusion criteria described above were not assigned points in the CCL 6 screening system, these data elements could potentially be informative and, along with MADLs, were provided to chemical evaluators during the classification process (see Chapter 4). These data may provide important context for evaluators, such as whether the contaminant under consideration is a potent toxicant via the inhalation route of exposure rather than the oral route of exposure.

For certain data elements, points were not assigned because the EPA decided to assign points to another equivalent data element or another data element describing similar data. In the case of cancer classifications, the EPA assigned points to the numeric rather than the original cancer classification data element because the numeric cancer classification data element incorporates all of the same data in a standardized way that is comparable across sources (see Section 2.4.4). In this way, the EPA prevented chemicals from receiving multiple sets of points for the same information.

For occurrence monitoring data in finished and ambient waters, the EPA assigned points to detection rates but not maximum concentrations. Maximum concentration and corresponding detection rate describe different aspects of occurrence monitoring data. Detection rates are more relevant to identifying the frequency of contaminant exposure through drinking water. Maximum concentrations in finished water are used to derive screening hazard quotients (sHQs, see Section 3.2.2), which were also assigned points; therefore, maximum concentrations in finished water are not assigned points directly but are embedded in the points assignment for a chemical's sHQ.

3.2.2 Calculating Screening Hazard Quotients

During the CCL 3 process, the EPA determined that one of the important measures for screening chemicals was a comparison between the Potency and Magnitude of a chemical. In CCL 3, the EPA addressed this during the classification step by calculating the "HRL/concentration ratio." This ratio is a comparison between a health reference level (HRL), which is a concentration of a chemical in drinking water at or below which is not expected to result in adverse health outcomes over a lifetime of exposure, and the 90th percentile concentration of the chemical in ambient or finished water (USEPA, 2009c).

As done in CCL 5, for CCL 6 chemicals that had the necessary health effects and occurrence information, the EPA calculated an sHQ. The sHQ represents the chemical-specific ratio of the drinking water concentration to the screening level at which no adverse health effects are expected, as further described in this section.

To calculate the sHQ, the EPA derived an element called the health screening level (HSL). The comparison of the HSL against the drinking water occurrence data for each chemical helps to inform whether a chemical has the potential to occur in finished drinking water at concentrations relevant to adverse health effects. A CCL 6 HSL is a calculated concentration of a chemical in drinking water derived from chronic toxicity values identified from main data sources. It is important to note that these HSLs are different metrics than the CCL 6 health concentrations (HRLs and CCL screening levels) introduced in Chapter 4. HSLs were used in CCL 6 for initial coarse screening purposes only and were replaced by HRLs and CCL screening levels, which relied on fewer assumptions and underwent manual review and expert discussion during their derivation, for classification.

HSLs were calculated according to the equations in Table 5, assuming a bodyweight-adjusted drinking water intake (DWI) of 33.8 mL/kg-day and 20% relative source contribution (RSC) (USEPA, 2019a); (USEPA, 2000b). At this stage in the CCL 6 process, the EPA did not differentiate between carcinogens that have a linear or mutagenic mode of action; the same equation was used for all CSFs. When toxicity values such as NOAELs and LOAELs were available, the same default composite uncertainty factors (UFs) were applied as were used in CCL 5 (1,000x for NOAELs and 3,000x for LOAELs). If multiple types of toxicity values were available for a chemical, the EPA calculated corresponding HSLs using each type of toxicity value and the most health-protective HSL was used to compare against finished water concentrations.

Table 5. Formulas for Calculating Health Screening Levels (HSLs)

Health Data Element	Default UF	Equation for HSL
Benchmark	NA	Use benchmark as derived by source as HSL
RfD	NA	$HSL = \frac{RfD}{DWI} * RSC$
CSF	NA	$HSL = \frac{\left(\frac{1x10^{-6}}{\mathrm{CSF}}\right)}{DWI}$
NOAEL	1,000	$HSL = \frac{\left(\frac{NOAEL}{1000}\right)}{DWI} * RSC$
LOAEL	3,000	$HSL = \frac{\left(\frac{LOAEL}{3000}\right)}{DWI} * RSC$

After identifying the most health-protective HSL, the EPA calculated the sHQ for a chemical by dividing the maximum finished water concentration by the HSL (Equation 1). The EPA only chose maximum concentrations of a chemical in finished water for use in the calculation of sHQs to focus on chemicals most relevant to drinking water exposure and having the potential for the greatest public health concern.

Equation 1. Formula for Calculating Screening Hazard Quotients

$$sHQ = \frac{\max finished\ water\ concentration}{HSL}$$

If maximum finished water concentration values were available from multiple data sources for a chemical, the overall highest concentration of the maximum finished water concentrations (the most health-protective) was chosen. sHQs were calculated for 350 of the Universe chemicals. The logarithmic distribution of sHQs calculated for the screening step of CCL 6 is shown in Figure 5. Similarly to the difference between HSLs and health concentrations, it should be noted that the sHQ differs from the final hazard quotient (fHQ) calculated in the classification step of the CCL 6 process (see Chapter 4).

The EPA incorporated the sHQ as a data element in the Universe file and assigned points in the same way as other CCL 6 data elements. The process for distributing and applying screening points to each type of data element is described in Section 3.3.

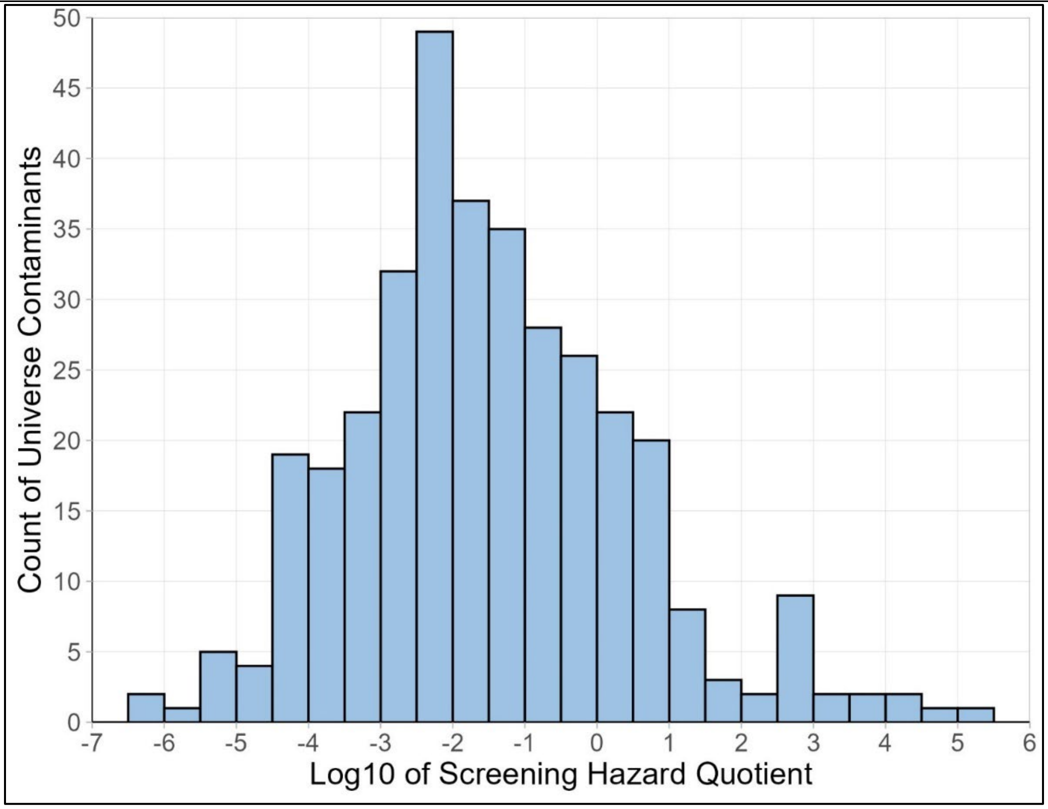


Figure 5. Empirical Histogram of Log Transformed Screening Hazard Quotients Calculated for the Screening Step

3.3 Developing a Scoring Rubric

3.3.1 Determining Screening Tiers

The EPA categorized the data elements selected for screening into one of two groups: data elements related to occurrence or data elements related to health effects. These two groups of data elements were further categorized into five tiers each, with Tier 1 containing data elements most relevant to understanding potential drinking water risk and Tier 5 containing data elements less directly describing potential drinking water risk (Table 6).

For example, as shown in Table 6, the highest tier of health effects data elements (Health Tier 1) includes RfD, CSF, and chronic benchmark. These data elements are generally available for chemicals that have a health assessment conducted by the EPA or another health agency and are directly related to potential lifetime drinking water risks because they describe health effects resulting from chronic oral exposures to chemical contaminants. The highest tier of occurrence data elements (Occurrence Tier 1) is the sHQ (see Section 3.2.2), which is the ratio of the maximum concentration of the chemical in finished drinking water to the lowest health screening level for a chemical. The maximum concentration of a chemical in finished water is the occurrence data element most applicable to potential hazards through drinking water. The lowest HSL is the most health-protective value indicating potential toxicity due to chronic oral exposure. Chemicals with higher sHQs have the greatest potential to be of public health concern in terms of exposure via finished water.

The lowest occurrence tier (Occurrence Tier 5) includes information like chemical release quantity, estimated pesticide application rate, and chemical production volume. These data are useful predictors of potential occurrence in finished water but are not as directly relevant as detection rates of a chemical in finished water or ambient water to inform listing recommendations. Similarly, the lowest health tier (Health Tier 5) includes modeled or *in vitro* bioactivity data (e.g., androgen receptor bioactivity) and LD_{50}s. These data elements may give an indication of relative toxicity or the potential to cause toxicity but do not provide the information needed to derive chronic toxicity values such as RfDs or CSFs, which are necessary for assessing lifetime drinking water risk.

Table 6. Health and Occurrence Tiers for Points Assignments

Health Tiers	Data Elements
Tier 1	Reference dose, cancer slope factor, chronic benchmark
Tier 2	Chronic LOAEL, chronic NOAEL
Tier 3	Numeric cancer classification, subchronic benchmark, subchronic reference dose
Tier 4	Acute benchmark, acute reference dose, subchronic LOAEL, subchronic NOAEL, MRDD, neurotoxins/neurotoxicants,
Tier 5	TD_{50}, LD_{50}, ToxCast assay percent active, number of PubMed articles, androgen receptor bioactivity, estrogen receptor bioactivity
Occurrence Tiers	**Data Elements**
Tier 1	Screening hazard quotient (sHQ)
Tier 2	Nationally representative monitoring program and survey, finished water detection rates
Tier 3	Nationally representative monitoring program, ambient water detection rates Non-nationally representative study, finished water detection rates
Tier 4	Non-nationally representative study, ambient water detection rates
Tier 5	Chemical Release Quantity, Estimated Pesticide Application Rate, Chemical Production Volume, Presence of CERCLA or FIFRA lists, NHANES blood, urine, and serum concentrations, Biodegradation half-life

Altogether, a chemical can receive screening points for each data element in every tier and screening points for multiple data elements within a tier. For example, a chemical may have estimated pesticide application rate data, chemical release quantity data, ambient water detection rates from a non-nationally representative study, finished water detection rates from a non-nationally representative study, and finished water detection rates from a nationally representative monitoring program or survey. In this case, screening points are assigned to each of these data elements. Lower tiers have fewer points associated with them because they are considered less relevant to hazards associated with chemical exposures via drinking water. The point assignments for each tier of data, along with the categories within them, were designed to allow consideration of chemicals with ample data as well as chemicals with data indicating concern but limited overall data availability for listing on the CCL. The detailed process for determining screening point assignments is described in the next section.

3.3.2 Determining Relative Point Assignments Within Each Screening Tier

The EPA analyzed the chemical-specific data for each data element and plotted distributions to ensure the data contained no obvious irregularities. The EPA calculated summary statistics (minimum, median, maximum) and quantiles (20th, 40th, 60th, and 80th percentiles, etc.) for data elements when possible. For most data elements, these quantiles were used to establish screening point categories for each health and occurrence tier (see Table 7 and Table 8 at the end of this section). Values of zero were excluded from summary statistics and quantiles for screening point categories. The EPA retained all data associated with chemicals that are regulated with NPDWRs when establishing the relative point assignments for each data element. Although points were assigned to regulated chemicals, they were not considered further in the CCL 6 process. An example of the distribution of CSFs with the calculated quantiles represented with red dotted lines is provided in Figure 6. Point assignments for categorical data elements could not be established based on distribution of values; these data elements include cancer classifications, NHANES biomonitoring detections in blood, serum and urine, presence of a chemical on the CERCLA or FIFRA list, and a few other data elements.

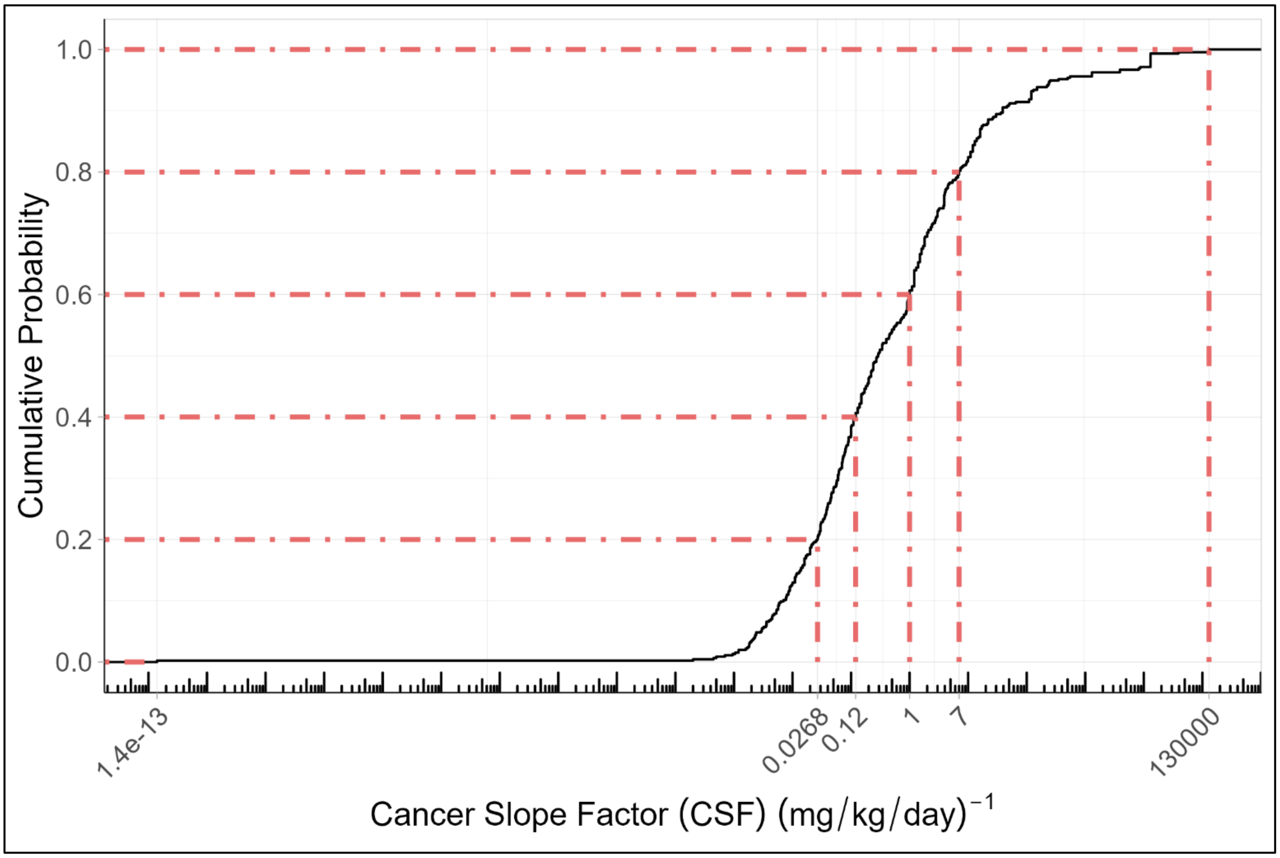


Figure 6. Empirical Distribution of Cancer Slope Factors in the CCL 6 Universe

Relative point assignments for data elements that were not established based on quantiles and the distribution of values or required data manipulation steps are detailed below. For the sHQ and detection rate data elements, the summary statistics were deemed similar to the CCL 5 summary statistics and thus the points assignments did not change. For the chemical production volume, biodegradation half-life, and PubMed articles data elements, the EPA followed the same procedures for establishing point assignments as were done for CCL 5 (USEPA, 2022b).

The EPA evaluated the distribution of calculated sHQs for point assignments. The distribution of sHQ values is highly positively skewed. Generally, an sHQ equal to 1 indicates the finished water concentration is equal to the HSL and, therefore, the concentration in finished water has reached the threshold at which adverse effects resulting from exposure may be expected to occur. An sHQ greater than 1 indicates the finished water concentration exceeds the HSL and, therefore, the chemical may pose a hazard for public health. However, an sHQ less than 1 does not necessarily indicate that there is no risk to the exposed population.

For CCL 6, the EPA used the points assignment categories for the sHQ that were developed during CCL 5. Instead of limiting sHQ point assignments to chemicals with sHQs of 1 or higher, the EPA assigned points to sHQ values that were equal to or exceeded the CCL 5 median (0.01). The EPA divided sHQ values equal to or greater than 0.01 into five categories based on orders of magnitude (i.e., powers of ten). These five categories are 0.01-0.1, 0.1-1, 1-10, 10-100 and >100, where lower points are allocated to the lowest category and higher points to the highest category (see Table 7). For sHQ values that fall on a category boundary, points are assigned according to the higher category. For example, if a chemical has a sHQ value of 0.1, which is the upper bound of Category 1 and the lower bound of Category 2, screening points are assigned to the sHQ value according to Category 2. For all points assignments, the EPA used this protocol if a data element value fell on a category boundary.

The EPA evaluated the distributions of detection rates in ambient and finished water for point assignments. The distributions are highly skewed to the right – some of this skew may be due to some naturally occurring inorganic elements detected in nearly all samples (see Figure 7). To avoid overemphasizing point assignments to inorganic ions with high detection rates, the EPA developed points categories based on selected percent detection rate values rather than calculated quantiles. These point categories are >0-2.5%, 2.5-5%, 5-7.5%, 7.5%-10%, and >10%, where lower points are allocated to lower detection rates and higher points to higher detection rates (see Table 8).

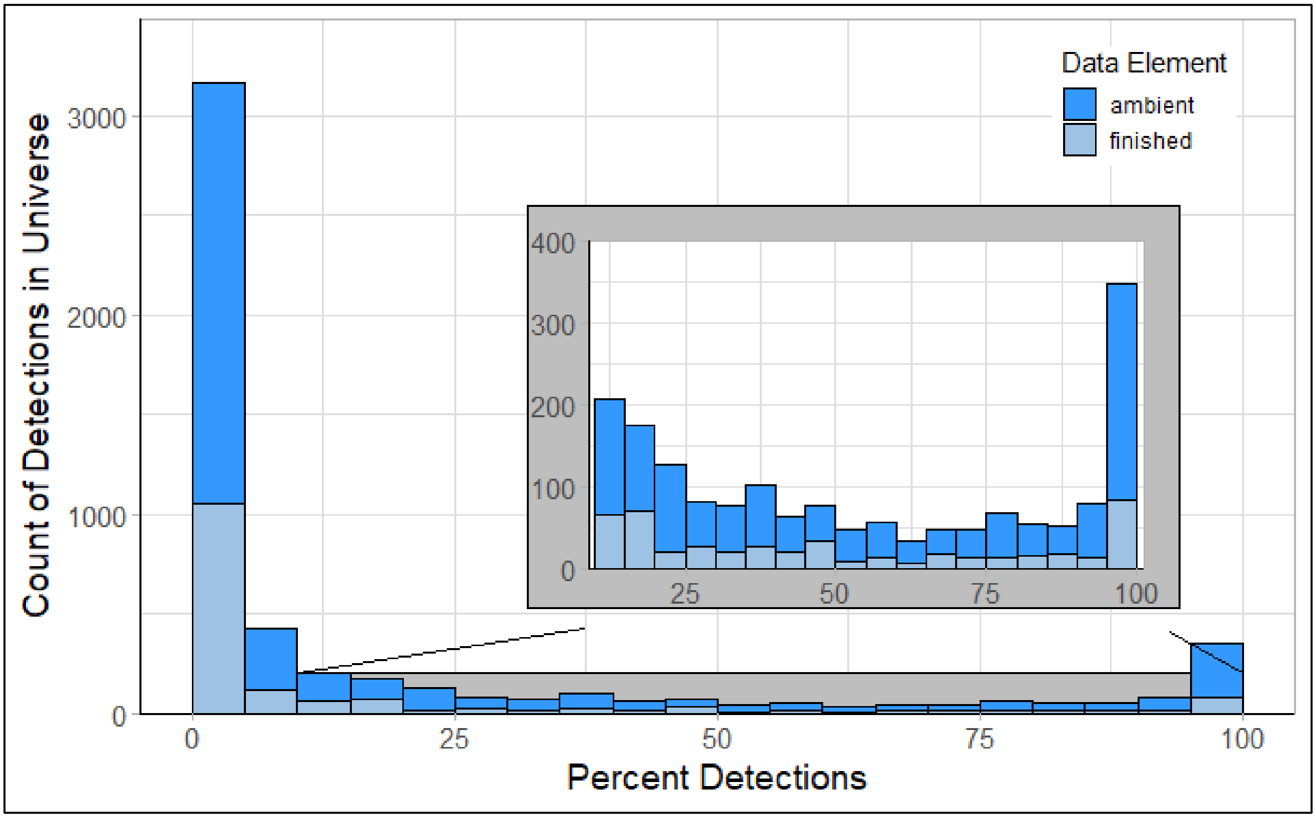


Figure 7. Ambient and Finished Water Detection Rates for the CCL 6 Universe Chemicals

The EPA evaluated the distribution of chemical production volume data from the 2020 Chemical Data Reporting (CDR) rule under the Toxic Substance Control Act (TSCA) for point assignments. Production volume data required special data processing steps to be incorporated into the CCL 6 screening rubric because some of these data are reported as categories (i.e., a range of values such as 100,000 – 500,000 lbs.) or inequalities (e.g., < 25,000 lbs.) of production volume while others are reported as a numeric sum or singular value. For the screening step only, the EPA converted chemical production values to single numerical values so that distributions and quantiles could be calculated, as described in this section. These quantiles were subsequently used to establish relative point assignments for the production volume data element. For the classification step (Chapter 4), chemical production volume data values were not modified and used as originally reported by CDR.

The EPA's method of converting production volumes to single numerical values was driven by how the data were originally reported by CDR. The EPA analyzed the variations of production volume categories and determined that using the minimum value for the category ranges and temporarily substituting ½ of the lowest production volume were the appropriate approaches. For all production volumes expressed as a range, the minimum value of the range was substituted for calculating point assignments. Additionally, values of "<1,000,000" were substituted with a value of 500,000 (i.e., ½ of 1,000,000 lbs.) for point assignments. See Table 8 for points categories and point assignments for the chemical production volume data element.

The EPA analyzed the distribution of predicted biodegradation half-lives in the OPERA model from the CompTox Chemistry Dashboard (see Appendix M for additional information) for point assignments to incorporate physio-chemical considerations into the screening system. The EPA established one point category for this data element (Table 8). For chemicals with biodegradation half-life prediction values shorter than 3.5 days, or below the 20th percentile, the EPA assigned negative screening points. This reflects the reduced likelihood that chemicals with relatively short half-lives occur with similar durations and at similar levels in finished water as chemicals considered to be persistent in the environment.

The EPA analyzed the distribution of the number of PubMed articles data element provided by the CompTox Chemistry Dashboard (see Appendix M for additional information) for point assignments. This data element represents the number of PubMed records associated with a given chemical structure. The value gives a sense of the amount of literature available that may not be "retrievable" for the CCL 6 Universe. The EPA established two points categories for this data element (50th-90th percentile and greater than or equal to the 90th percentile) where lower points are allocated to the lower category and higher points to the higher category. See Table 7 for points categories and point assignments for the number of PubMed articles data element.

For a specific chemical, the number of points assigned to each individual data element depends on the relative toxicity or relative occurrence indicated by the data element compared to values of that data element available for all other chemicals in the Universe. For example, a chemical with a CSF between the 80th percentile and the maximum (most toxic) CSF for all available chemicals would have the highest indication of potential potency and therefore be in the highest point category (Category 5) for the CSF data element. It is important to note that many of the health effects data elements have an inverse relationship between the toxicity value and the expected toxicity (e.g., chemicals with lower RfDs are considered more potent toxicants). In these cases, the upper bound of each point category corresponds with the lowest value in that category.

Table 7 and Table 8 present the upper bound and lower bound values for the points categories (Category 1 through Category 5) of relative potency and prevalence for each data element included in health effect and occurrence tiers, respectively.

Table 7. Point Assignments for Health-Related Data Elements

Data Element		Category 1		Category 2		Category 3		Category 4		Category 5	
		lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
Health Effects Tier 1	Points Assigned [1]	200		300		400		500		600	
Reference Doses	values - mg/kg/day	3000	0.1	0.1	0.03	0.03	0.008	0.008	0.001	0.001	0.0000000007
Cancer Slope Factors	values - (mg/kg/day)$^{-1}$	1.4E-13	0.0268	0.0268	0.12	0.12	1	1	7	7	130,000
Chronic Benchmarks	values - mg/L	200	0.2	0.2	0.03	0.03	0.004197	0.004197	0.000409	0.000409	0.000000000005
Health Effects Tier 2	Points Assigned	150		250		350		450		550	
Chronic NOAELs	values - mg/kg/day	1,200	250	250	69	69	20	20	4.65	4.65	0.00000714
Chronic LOAELs	values - mg/kg/day	14,792	400	400	126	126	50	50	14.5	14.5	0.0000714
Health Effects Tier 3	Points Assigned	100		200		300		400		500	
Numeric Cancer Classifications	See Table 3	NA		NA		3		2		1	
Subchronic RfDs	values - mg/kg/day	3000	0.48	0.48	0.06	0.06	0.008	0.008	0.001	0.001	0.00000002
Subchronic Benchmarks	values - mg/L	50	0.3	0.3	0.074	0.074	0.0116	0.0116	0.00034	0.00034	0.0000002
Health Effects Tier 4	Points Assigned	50		100		150		200		250	
Acute Benchmarks	values - mg/L	100	3	3	0.8	0.8	0.17	0.17	0.01	0.01	0.0000001

Data Element		Category 1		Category 2		Category 3		Category 4		Category 5	
		lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
Acute RfDs	values - mg/kg/day	7	0.5	0.5	0.15	0.15	0.05	0.05	0.006	0.006	0.0000002
Subchronic LOAELs	values - mg/kg/day	20,000	790.48	790.48	300	300	119.2	119.2	30	30	0.005
Subchronic NOAELs	values - mg/kg/day	20,000	600	600	202	202	58.63	58.63	11.3	11.3	0.0024
MRDDs	values - mg/kg/day	999	25	25	6.67	6.67	2.032	2.032	0.333	0.333	0.00001
Neurotoxins/Neurotoxicants	presence on list	Yes		NA		NA		NA		NA	
Health Effects Tier 5	Points Assigned	10		30		50		70		90	
TD_{50}s	values -- mg/kg/day	111,000,000	1568	1568	368	368	101	101	20.42	20.42	0.0000121
LD_{50}s	values -- mg/kg	4,386,000	4100	4100	1688	1688	604	604	140	140	0.0003
ToxCast Assay Percent Active	values -- percent	>0	1	1	3	3	9	9	22	22	100
Estrogen Receptor Bioactivity	AUC for bioactivity	0.01-<0.1		NA		≥0.1		NA		NA	
Androgen Receptor Bioactivity	AUC for bioactivity	0.001-<0.1		NA		0.1-1		NA		NA	
PubMed Articles	number of articles	50th-90th percentiles (81-3494 articles)		90th percentile (>3494 articles)		NA		NA		NA	

---

[1] If a data element value falls on a category boundary, screening points are assigned according to the higher category.

Table 8. Point Assignments for Occurrence-Related Data Elements

Data Element		Category 1		Category 2		Category 3		Category 4		Category 5	
		lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
Occurrence Tier 1	Points Assigned[1]	750		1000		1250		1500		1750	
Screening Hazard Quotient[2]	No units	0.01	0.1	0.1	1	1	10	10	100	100	1.67E+05
Occurrence Tier 2	Points Assigned	600		800		1000		1200		1400	
Nationally representative monitoring program, finished water detection rates	values - percent	>0%	2.50%	2.50%	5%	5%	7.50%	7.50%	10%	10%	100%
Occurrence Tier 3	Points Assigned	500		700		900		1100		1300	
Nationally representative monitoring program, ambient water detection rates	values - percent	>0%	2.50%	2.50%	5%	5%	7.50%	7.50%	10%	10%	100%
Non-nationally representative study, finished water detection rates	values - percent	>0%	2.50%	2.50%	5%	5%	7.50%	7.50%	10%	10%	100%
Occurrence Tier 4	Points Assigned	300		500		700		900		1100	
Non-nationally representative study, ambient	values - percent	>0%	2.50%	2.50%	5%	5%	7.50%	7.50%	10%	10%	100%

Data Element		Category 1		Category 2		Category 3		Category 4		Category 5	
		lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
water detection rates											
Occurrence Tier 5	Points Assigned	50		100		150		200		250	
Chemical release information	values - lbs/year	>0	10.3046	10.3046	2,111.978	2,111.978	29,767.79	29,767.79	488,768.6	488,768.6	558,853,739.8
Estimated Pesticide Application Rate	values - kg/year	>0	80.54	80.54	7677.14	7677.14	49,516.56	49,516.56	306,439.66	306,439.66	124,827,727
Chemical production information[3]	values- lbs/year	1	500,000	NA	NA	500,000	1,000,000	1,000,000	3,306,932.80	3,306,932.80	200,000,000,000
FIFRA registered pesticide	presence on list	Yes		NA		NA		NA		NA	
CERCLA priority substance	presence on list	Yes		NA		NA		NA		NA	
NHANES biomonitoring detection in blood, serum, and/or urine	Values – ng/mL	NA		NA		Any value detected at or above the 90th percentile		NA		NA	
Biodegradation half-life	Points Assigned	-10									
	values - days	<20th percentile (<3.540462)									

[1] If a data element value falls on a category boundary, screening points are assigned according to the higher category.

[2] The EPA assigned maximum concentration values to non-detected chemicals in the screening step of CCL 6. See Chapter 2 and Appendix M for additional information.

[3] Because of the large number of production volume entries at a value of 500,000 lbs (37% of entries), no contaminants received the Category 2 points assignment for the chemical production information data element

If multiple data entries for a single data element existed for a given chemical (e.g., a chemical had two different RfDs or two different non-nationally representative finished water detection rates available from different data sources), the EPA assigned points using the data entry with the value that represents the greater exposure or toxicity. Examples of data elements assigned points include the highest available detection rate of a chemical in finished and/or ambient water or the lowest available RfD for a chemical.

At this stage of the CCL process, the EPA chose these values for each data element for several reasons:

- This is the most conservative and health-protective approach.
- With approximately 25,000 chemicals in the Universe, it is not feasible to conduct a systematic review or assess the relative quality of the information available for each chemical.
- It is prudent to allow for new, albeit potentially less vetted or complete, information to be factored into the screening process.

3.4 Final Point Assignments and Screening Scores

If a chemical had data available for each data element indicating the most severe health effects or the highest occurrence, the maximum possible health effects and occurrence screening points that a chemical would accumulate were 6,100 and 7,850, respectively. Therefore, the highest total combined health effects and occurrence screening points a chemical could be assigned, known as the “screening score,” was 13,950. The maximum screening score that an unregulated chemical in the CCL 6 Universe accumulated was 10,300 points. A histogram of screening scores for all chemicals in the CCL 6 Universe is shown in Figure 8.

The EPA examined final point assignments and screening scores to ensure the Agency considered chemicals of emerging concern in drinking water in addition to well-studied chemicals with more robust human health and drinking water occurrence data. The point system allowed inclusion of a chemical with limited health effects data, but high occurrence, on the PCCL 6. Cobalt, for example, earned only 1,790 of 6,100 possible points for health effects data but was included on the PCCL because it earned a significant number of points (7,300 of 7,850) from occurrence data. Similarly, a chemical with limited or no finished water occurrence data but with health effects information potentially indicating high toxicity could also be included on the PCCL. For example, thiram earned only 600 points from occurrence data but 3,130 points from health effects data.

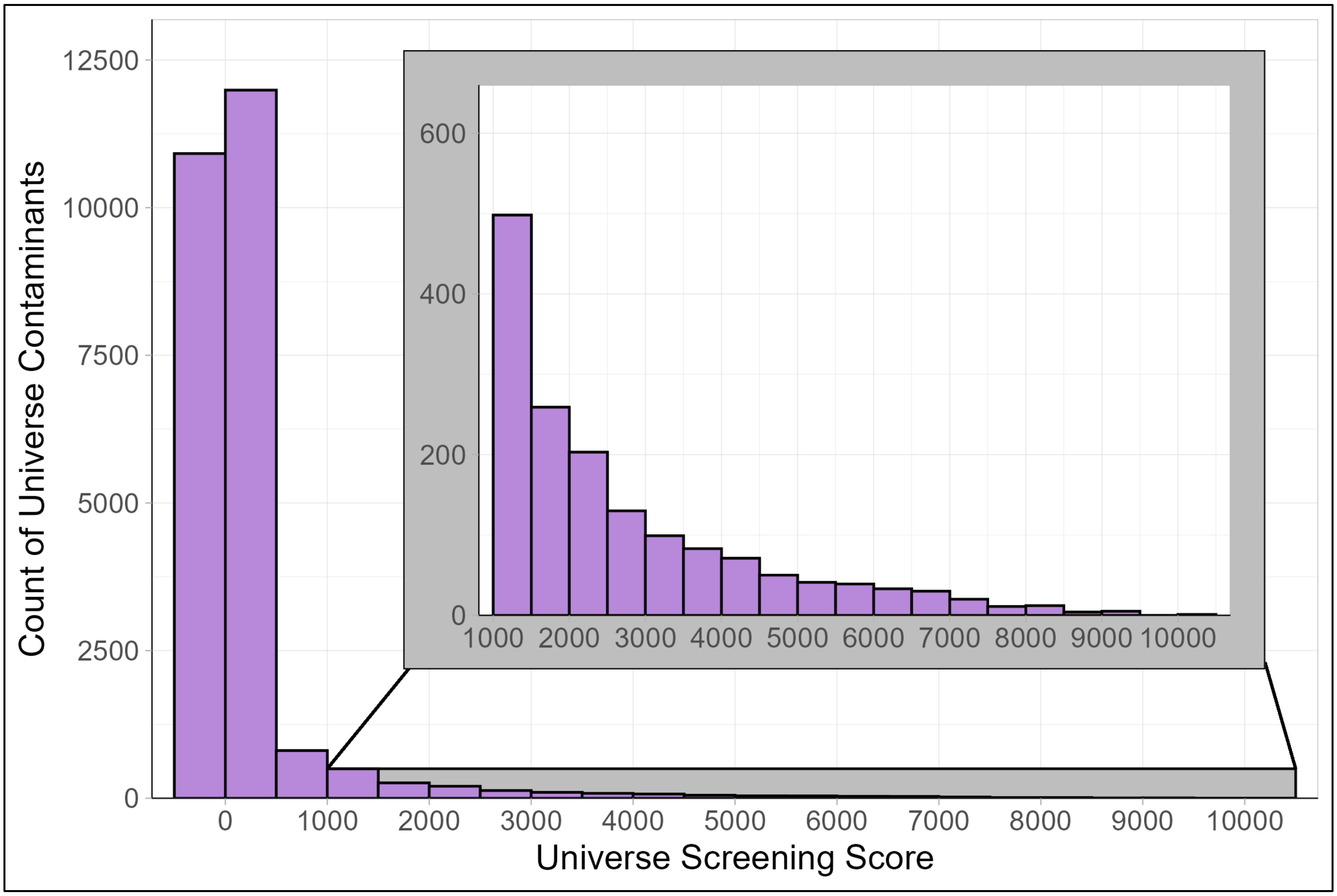


Figure 8. Total Screening Scores for the CCL 6 Universe Chemicals

Figure 9 shows a plot comparing total occurrence score to total health effects score for all chemicals in the Universe. Chemicals with high health effects scores plot in the bottom right quadrant of the diagram (red), chemicals with moderate health effects and occurrence scores plot near the center (purple), and chemicals with high occurrence scores plot in the top left quadrant (blue).

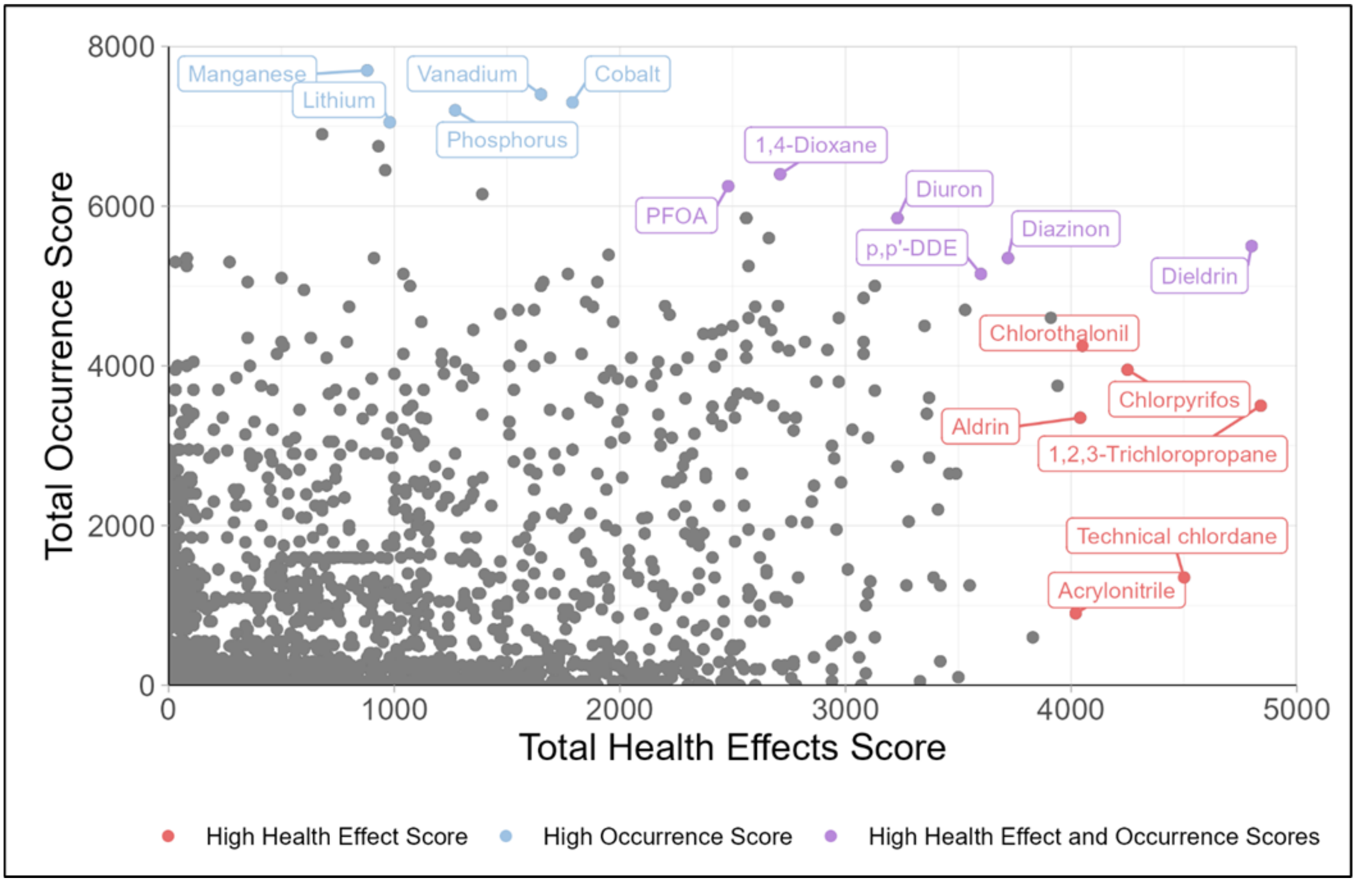


Figure 9. Health Effects and Occurrence Scores for the CCL 6 Universe Chemicals

Although screening scores were used to prioritize chemicals for inclusion on the PCCL, these scores did not influence the recommendations of the chemical evaluators. As discussed in Chapter 4, the CCL 6 Evaluation Teams were not provided with screening scores to use while assessing chemicals for the draft CCL 6. The points system and outcomes are solely a tool for CCL 6 screening purposes and statistical analyses (see Section 4.6) and should not be used or described in other contexts.

3.5 Using the CCL 6 Screening System

The CCL 6 process has identified the most broad and comprehensive universe of health effects and occurrence information to date. The EPA used the screening system to take advantage of this information and identified the top-scoring chemicals, with, theoretically, the potential for the greatest public health concern, for inclusion on the PCCL 6. Narrowing down the CCL 6 Universe to the top scoring chemicals allowed the EPA to consider the resource requirements of compiling additional information, developing CISs and conducting evaluation team reviews during the classification step (see Chapter 4).

When limiting evaluations to the top scoring chemicals, all chemicals scoring at or above 3,720 points were advanced for further consideration for the CCL 6. The highest score accumulated by a chemical in the CCL 6 Universe was 10,300, as mentioned above. When selecting the cutoff for the top scoring chemicals, the EPA increased the size of the PCCL, prior to any additions or removals to it, by 10% compared to CCL 5. The larger PCCL provided more data for use in the logistic regression analysis; for more information on the validation of the size of the PCCL, see Section 4.6.2. A total of 274 top scoring Universe chemicals were elevated for further consideration and potential inclusion on the PCCL 6.

3.6 Consideration of Publicly Nominated Chemicals

3.6.1 Soliciting Public Nominations

On February 17, 2023, the EPA published a request for public nominations of unregulated chemical and microbial contaminants to be considered for possible inclusion on the CCL 6 (88 FR 10316, (USEPA, 2023a)). In accordance with SDWA, which directs the EPA to consider health effects and occurrence information when deciding whether to place contaminants on the CCL, the EPA asked that nominations include responses to the following questions:

1. What is the nominated contaminant's name, CAS Registry Number (CAS RN) or DSSTox substance identifier (DTXSID), and/or common synonym (if applicable)? Note—please do not nominate a contaminant already subject to the National Primary Drinking Water Regulations (NPDWRs) (see the current list at https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations).
2. What are the data you believe support the conclusion that the nominated contaminant is known or anticipated to occur in public water systems? For example, provide information that shows measured occurrence of the contaminant in drinking water, measured occurrence in sources of drinking water that provide water to public drinking water systems, measured occurrence in other water types (*i.e.,* ambient water (rivers, lakes, or streams) groundwater, wastewater, stormwater, or urban runoff) or provide information that shows the contaminant is released in the environment or is manufactured in large quantities and has the potential for contaminating sources of public drinking water. Please provide the source of the information with complete citations for published information (*i.e.,* author(s), title, journal, and date) and/or contact information for the primary investigator. Additionally, please provide original supporting or supplemental information files relevant to the published information (*i.e.,* data tables, data sets, or data files, etc.).
3. What new health effects data are available which you believe supports the conclusion that a contaminant may have an adverse effect on the health of humans? For example, provide information that shows the contaminant may have an adverse health effect on the general population or that the contaminant is potentially harmful to subgroups that comprise a meaningful portion of the population (such as children, pregnant women, the elderly, individuals with a history of serious illness, individuals living in disadvantaged communities with known occurrence of emerging contaminants in their public water system, or others). Please provide the source of this information with complete citations for published information (*i.e.,* author(s), title, journal, and date) and/or contact information for the primary investigator. Additionally, please provide original supporting or supplemental information files relevant to the published information (*i.e.,* data tables, data sets, or data files, etc.).

Nominations were received via the EPA docket (Docket ID No. EPA-HQ-OW-2022-0946) on the Federal eRulemaking Portal (http://www.regulations.gov) and were also accepted by mail or hand delivery. The EPA compiled and reviewed the information to identify the contaminants nominated and any supporting data submitted that could supplement data gathered by the EPA to inform the selection of the draft CCL 6.

3.6.2 Summary of CCL 6 Public Nominations

EPA received public nominations for three individual chemicals. These nominations include lithium, manganese, and perchlorate, as shown in Table 9 below:

Table 9. Publicly Nominated Chemicals

Chemical Name	CASRN	DTXSID
Lithium	7439-93-2	DTXSID5036761
Manganese	7439–96–5	DTXSID2024169
Microplastics	Multiple	Multiple
Perchlorate	14797-73-0	DTXSID6024252
PFAS	Multiple	Multiple
Pharmaceutical Waste (specifically estrogenic compounds)	Multiple	Multiple

In addition to individually nominated chemicals, the EPA also received three nominations for chemical groups, including “pharmaceutical waste (specifically estrogenic compounds),” “microplastics,” and PFAS. Publicly nominated microbial contaminants are discussed in the Technical Support Document for the Draft Sixth Candidate List (CCL 6) – Microbial Contaminants (USEPA, 2026c).

All submission letters for public nominations can be viewed on the CCL 6 docket (Docket ID No. EPA-HQ-OW-2022-0946) at https://www.regulations.gov.

3.6.3 Outcome of Publicly Nominated Chemicals for the PCCL

The EPA reviewed the publicly nominated chemical contaminants and identified the chemicals that were not already included in the top scoring selection (see Section 3.5) and not subject to proposed or promulgated NPDWRs and therefore needed to be considered for further analysis.

Of the three nominated individual chemicals, all were already included in the CCL 6 Universe from available main data sources (see Section 2.2.1) and included among the top scoring chemicals prioritized through points assignment onto the base PCCL (see Section 3.5). One of the chemicals – perchlorate – was undergoing a pending regulatory action to establish a national drinking water standard at the time of draft CCL 6 development and was consequently removed from further evaluation. The other two chemicals, lithium and manganese, were retained on the PCCL for further evaluation under Step 3 (Classification).

For the three publicly nominated chemical groups, all were determined to represent Agency priorities based on feedback received through the nomination materials and through ongoing communication from stakeholders. Following established CCL protocol, any individual chemicals screened from the Universe to the PCCL that fell under the definition for the chemical groups were separated from further evaluation in Step 3 (Classification) and listed instead under their relevant chemical group. Two of the publicly nominated chemical groups – microplastics and PFAS – are listed explicitly on the draft CCL 6. The third nominated chemical group, "pharmaceutical waste (specifically estrogenic compounds)" is considered listed by proxy through the pharmaceutical chemical group. Descriptions and additional Agency considerations for listing these chemical groups on the draft CCL 6 are provided in more detail under Section 4.8.

3.7 Chemicals Excluded from the PCCL 6

3.7.1 Regulatory Determination 5 Contaminants

In January 2025, under the Regulatory Determination 5 (RD 5) process, the EPA made preliminary regulatory determinations for nine chemicals: 2-aminotoluene, cylindrospermopsin, ethoprop, microcystins, molybdenum, permethrin, profenofos, tebuconazole, and tribufos (90 FR 3830, (USEPA, 2025a)). One of these chemicals, 2-aminotoluene, was not included on the PCCL 6 due to a lack of sufficient points assigned through the screening process. The remaining eight RD 5 chemicals received sufficient points to be included on the PCCL 6 but were excluded due to the recent preliminary determination not to regulate these contaminants.

3.7.2 Canceled Pesticides

The list of top scoring chemicals pulled from the Universe for deriving the PCCL 6 contained a total of 31 canceled pesticides with no reported alternative uses. The EPA identified canceled pesticides by checking the status of all chemicals on the PCCL 6 in the EPA's Pesticide Product and Label System (PPLS). The EPA did not exclude any chemicals from the CCL process that have a canceled pesticide registration but have other uses that could contribute to their occurrence in the environment. For example, acrylonitrile has a registration status of canceled in PPLS, but its recent chemical production data shows that it is still widely used in manufacturing.

To exclude any canceled pesticides that are not persistent in the environment, the EPA evaluated the persistence and occurrence of these canceled pesticides (e.g., biodegradation half-life (rounded), end-of-use date, and monitoring data in finished and/or ambient water) using the following five-step protocol:

1. Canceled pesticides were assigned a persistence score based on the EPA's 2012 TSCA Work Plan Chemicals: Methods document (USEPA, 2012), according to the biodegradation half-life in air, water, soil and sediment for a pesticide.
2. End-of-use dates were used to determine when the canceled pesticides were last allowed to be used in the environment specifically as a pesticide.
3. Occurrence monitoring data collected after the end-of-use dates were used to determine if a canceled pesticide had any detections and/or data spikes that would pose a public health concern.
4. Canceled pesticides that were assigned a persistence score of 3 were included on the PCCL 6.
5. Canceled pesticides that were assigned a score of 1 or 2 but had detections in finished water were included on the PCCL 6, while those that had no or very few detections in ambient water were excluded.

Step 1. Canceled pesticides were assigned a persistence score based on their biodegradation half-life in the environment using the information in the EPA's CompTox Chemicals Dashboard. If the biodegradation half-life was greater than 6 months, a canceled pesticide was assigned a persistence score of 3. If the half-life was greater than or equal to 2 months, a canceled pesticide was assigned a persistence score of 2. If the half-life was less than 2 months, then a canceled pesticide was assigned a persistence score of 1 (see Table 10).

Table 10. Summary of Persistence Ranking Scores

Persistence	
Ranking Score	**Criterion**
3	Half-life > 6 months
2	Half-life ≥ 2 months
1	Half-life < 2 months

Step 2. Using the PPLS end-of-use dates were used to determine when the canceled pesticides were last allowed to be used in the environment. For CCL 6, the EPA did not use pesticide cancellation dates to assess the persistence in the environment because, when a pesticide registration is canceled, the EPA determines whether there is any significant potential risk associated with the use of the pesticide. If there is such concern, the EPA generally makes a case-by-case determination about allowing continued distribution, sale, or use of existing stocks of the canceled pesticide (USEPA, 1991).

Step 3. The EPA compared dates of occurrence monitoring data to end-of-use dates to determine if a canceled pesticide continued to have any detections and/or data spikes, or additional uses that would pose a public health concern. The data sources used for monitoring-include NAWQA, UCMR 1-5, UCM-State Rounds 1 and 2, NWIS, Pesticide Application Data, CDR, and TRI.

Step 4. The EPA included canceled pesticides in the CCL process that were assigned a persistence score of 3.

Step 5. The EPA evaluated canceled pesticides that received a persistence score of 1 or 2. If these canceled pesticides had detections in finished water, they were included on the PCCL 6. If they had no or few detections in ambient water, they were excluded from the PCCL 6.

The EPA assessed a total of 31 canceled pesticides for persistence. Five pesticides – aldrin, chlordecone (kepone), dieldrin, ethion, and mirex – were assigned a persistence score of 3 and showed detections in finished and/or ambient water and were included on the PCCL 6. Alpha-1,2,3,4,5,6-hexachlorocyclohexane was also included on the PCCL 6 because it showed finished water detections in the UCMR 4 occurrence data (collected 2018-2020). This chemical, in addition to the canceled insecticide lindane (gamma-1,2,3,4,5,6-hexachlorocyclohexane) that undergoes production that forms alpha-1,2,3,4,5,6-hexachlorocyclohexane as a byproduct, are organochlorides and isomers of hexachlorocyclohexane.

The remaining 25 pesticides were excluded from the PCCL 6 because they were assigned a score of 1 or 2 and showed either no detections in finished water, or few to no detections in ambient water. Finished or ambient water monitoring data were consistent with the end-of-use date and persistence hierarchy, indicating a low likelihood of public health concern. Table 11 shows the canceled pesticides the EPA assessed and ranked.

Table 11. Canceled Pesticides Assessed for Exclusion from PCCL 6

Chemical Name	CASRN	DTXSID	Half-Life (days)	TSCA Persistence Score	Last End Use Date
2,4,5-Trichlorophenoxyacetic acid	93-76-5	DTXSID5021388	4	1	1/22/1991
2,4,6-Trichlorophenol	88-06-2	DTXSID5021386	9	1	10/10/1989
Aldrin*	309-00-2	DTXSID8020040	331	3	5/15/1987
alpha-1,2,3,4,5,6-Hexachlorocyclohexane*	319-84-6	DTXSID2020684	19	1	10/1/2009
Azinphos-methyl	86-50-0	DTXSID3020122	95	2	12/13/2013
Benomyl	17804-35-2	DTXSID5023900	5	1	4/18/2005
Chlordecone (Kepone)*	143-50-0	DTXSID1020770	914	3	4/4/1977
Chlorthal-dimethyl	1861-32-1	DTXSID0024000	6	1	7/27/2005
Cyanazine	21725-46-2	DTXSID1023990	5	1	12/31/2002
Dichlorprop	120-36-5	DTXSID0020440	4	1	10/15/2004
Dicofol	115-32-2	DTXSID4020450	21	1	1/31/2013
Dieldrin*	60-57-1	DTXSID9020453	333	3	5/15/1987
Disulfoton	298-040-4	DTXSID0022018	143	2	12/31/2014
Endosulfan	115-29-7	DTXSID1020560	16	1	7/31/2016
Ethion*	563-12-2	DTXSID2024086	478	3	12/31/2004
Fenamiphos	22224-92-6	DTXSID3024102	5	1	10/6/2017
Fenthion	55-38-9	DTXSID8020620	5	1	12/31/2000
Isofenphos	25311-71-1	DTXSID8032417	3	1	1/26/2007
Methamidophos	10265-92-6	DTXSID6024177	5	1	12/31/2010
Methidathion	950-37-8	DTXSID5020819	141	2	12/30/2012
Methyl parathion	298-00-0	DTXSID1020855	5	1	12/31/2013
Mevinphos	7786-34-7	DTXSID2032683	4	1	7/1/1994
Mirex*	2385-85-5	DTXSID7020895	832	3	10/10/1989
Molinate	2212-67-1	DTXSID6024206	4	1	8/31/2009
Monuron	150-68-5	DTXSID0020311	5	1	9/30/1991
p,p'-DDD	72-54-8	DTXSID4020373	12	1	6/14/1972
p,p'-DDE	72-55-9	DTXSID9020374	17	1	10/10/1989 (DDT)
p,p’-DDT	50-29-3	DTXSID4020375	20	1	6/14/1972
Propachlor	1918-16-7	DTXSID4024274	4	1	10/4/2010
Propazine	139-40-2	DTXSID3021196	5	1	6/8/2021
Parathion	56-38-2	DTXSID7021100	5	1	10/31/2003

* included on the PCCL 6.

3.8 Summary of the PCCL 6

The resulting base PCCL 6 includes a total of 240 chemicals. As shown in Table 12, this includes the highest scoring chemicals, considers any public nominations, and excludes chemicals due to ongoing

Agency actions under the Regulatory Determination 5 (RD 5) process or pending regulatory actions, as well as any non-persistent canceled pesticides with no reported active alternative uses. The PCCL 6 also includes four chemical groups – disinfection byproducts (DBPs), microplastics, per- and polyfluoroalkyl substances (PFAS), and pharmaceuticals. See Appendix D for a complete listing of all expert-evaluated chemicals on the PCCL 6.

Table 12. PCCL 6 Accounting

Summation Steps	Chemical Counts	PCCL Chemical Totals
(+) Add the highest scoring chemicals from Universe	274	240 chemicals (base PCCL 6)
(+) Add any public nominated chemicals not already prioritized through screening points assignment	0	
(-) Exclude chemicals with preliminary regulatory determinations under the Reg Det 5 process or pending regulatory actions	9[a]	
(-) Exclude canceled non-persistent pesticides with no reported alternative uses	25[b]	
(-) Separate DBPs (listed under the chemical group instead)	8[c]	213 chemicals (expert-evaluated PCCL 6)
(-) Separate PFAS (listed under the chemical group instead)	4[d]	
(-) Separate pharmaceuticals with no reported alternative uses (listed under the chemical group instead)	15[e]	

[a] Cylindrospermopsin; Ethoprop; Microcystins; Molybdenum; Perchlorate; Permethrin; Profenofos; Tebuconazole; Tribufos.

[b] 2,4,5-Trichlorophenoxyacetic acid; 2,4,6-Trichlorophenol; Azinphos-methyl; Benomyl; Chlorthal-dimethyl; Cyanazine; Dichlorprop; Dicofol; Disulfoton; Endosulfan; Fenamiphos; Fenthion; Isofenphos; Methamidophos; Methidathion; Methyl parathion; Mevinphos; Molinate; Monuron; p,p'-DDD; p,p'-DDE; p,p'-DDT; Parathion; Propachlor; Propazine.

[c] Chloral hydrate; Chlorate; Chloropicrin; Dibromoacetonitrile; Dichloroacetonitrile; Formaldehyde; N-Nitrosodiethylamine; N-Nitrosodimethylamine.

[d] PFBA; PFBS; PFHxA; PFHxS.

[e] 17alpha-Ethinylestradiol; 17beta-Estradiol; 4-Androstene-3,17-dione; 5,5-Diphenylhydantoin; Acetaminophen; Acyclovir; Carbamazepine; Fluconazole; Hydrochlorothiazide; Lamivudine; Loratadine; Metformin; Metoprolol; Tamoxifen; Theophylline.

Chapter 4 Classification of PCCL Chemicals to Select a CCL

4.1 Overview

The purpose of Step 3 of the CCL 6 development process was to narrow down the PCCL 6 chemicals to a draft CCL 6 through a classification process conducted by EPA scientists, referred to as "chemical evaluators". The chemical evaluators assessed the available health and occurrence data for the PCCL 6 chemical contaminants and reached a consensus on whether to recommend listing them on the draft CCL 6. For the CCL 6 classification step, the EPA executed additional analyses to examine the evaluator recommendations and made refinements to the list based on the history of each contaminant under the SDWA regulatory determination process and whether new data are available since previous negative regulatory determinations were made.

As was the case with past CCLs, the CCL 6 classification process adhered to principles that reflect the critical goals of the CCL:

- Classification must consider chemicals for listing based on their potential for occurrence in finished drinking water and their potential for causing adverse health effects.
- Data supporting the decision to list or not list must be linked back to these criteria. The most relevant data used for the classification process are health data that indicate adverse effects associated with chronic oral exposure and occurrence data that indicate the prevalence and magnitude of potential occurrence in drinking water.
- The classification approach must be a transparent process that can be reviewed by external experts and the public. The attributes and data characterizing the contaminants should be easy to understand and the decision-making process to list or not list a particular chemical must be conveyed in a straightforward manner.

The EPA's first task in this step involved the collection of additional health effects and occurrence information for the top-scoring chemicals on the PCCL 6. The EPA used both main sources captured during Universe development, as well as supplemental sources that either were not identified during development of the Universe or were not available in a retrievable format. The EPA used this information to fill data gaps and calculate three types of data elements: health concentrations, final hazard quotients (fHQs), and attribute scores (all referred to as "calculated data elements"). The EPA then used these calculated data elements, along with relevant health effects and occurrence data metrics, to evaluate the contaminants on the PCCL 6 and summarize each in a standardized format called a Contaminant Information Sheet (CIS). More detail is available about the collection of supplemental data for the PCCL 6 chemicals in Section 4.2, calculated data elements in Section 4.3, and the CISs in Section 4.4.

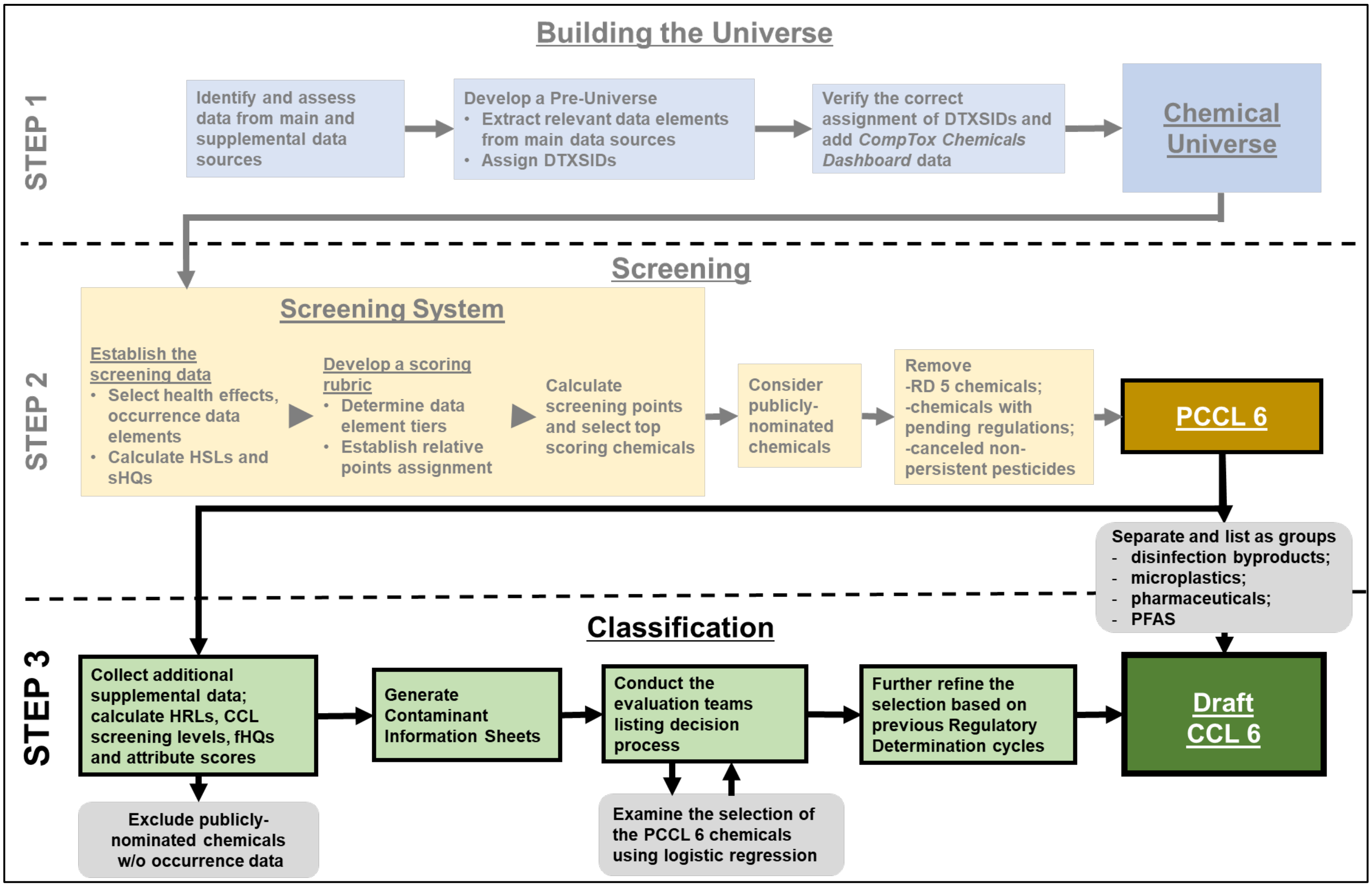
In the second task, the EPA formed two evaluation teams composed of chemical evaluators from multiple fields of specialization. These teams reviewed the occurrence and health effects information provided on the CISs and made recommendations on whether PCCL 6 chemicals should or should not be listed on the draft CCL 6. A more detailed explanation of the CCL 6 Evaluation Team process is provided in Section 4.5.

In the third task, to assess the accuracy and performance of the screening scores and other relevant variables as a predictor of listing outcomes and to validate the final number of chemicals to be reviewed by the CCL 6 Evaluation Team, the EPA developed logistic regression models. Further discussion on the logistic regression and its results is provided in Section 4.6.

Finally, the EPA reviewed the listing recommendations made by the evaluators. The EPA took a second look at contaminants that have previously been given negative regulatory determinations to evaluate whether there are any new data that may be considered in a future regulatory determination process. (further information is provided under Section 4.7).

Figure 10 illustrates the classification step for developing the draft CCL 6, shown in green:

Figure 10. Development Framework Step 3 – Classification



4.2 Classification Data Collection

4.2.1 Extracting Additional Occurrence Data Elements for Classification

Several additional data elements were extracted from ambient and finished water occurrence sources for classification. These data elements included the minimum, median, and 90th percentile concentrations of detections, as well as the total number of sites/samples and the number of sites/samples with detections. These additional data elements were extracted and saved using the simple file format described in Appendix M.3.

During screening, the detection rates for ambient and finished water sources received points and the maximum concentrations in finished water were used for calculating the sHQs. The maximum concentration was used during screening to provide the most health protective approach. During classification, the addition of the minimum, median, and 90th percentile concentration values allow the EPA to have a more nuanced understanding of the occurrence of contaminants in water.

Several of these data elements were used for the calculated elements described in Section 4.3 and all were presented on the CISs along with the data elements used for building the Universe (Chapter 2) and screening (Chapter 3).

Appendix M describes specifics about the pre-processing required to extract data elements from CCL 6 data sources used during classification, including how the EPA accessed the source data on the internet, when the EPA accessed the data, and any manipulation or calculations performed on the raw data.

4.2.2 Supplemental Data Collection

4.2.2.1 Occurrence Data

Systematic Occurrence Literature Review

The EPA's systematic literature reviews identified supplemental data to fill data gaps for PCCL 6 chemicals that required further evaluation. This included a search for additional peer-reviewed studies addressing the occurrence of chemicals in finished water or ambient water. For CCL 6, a systematic occurrence literature review was conducted for only one chemical, ethylene thiourea, that did not have any finished or ambient water data from the main sources. No additional studies were identified. The EPA continued to use the supplemental sources identified from the systematic occurrence literature reviews during CCL 5 (USEPA, 2022b).

The protocol used for conducting occurrence literature searches included performing an internet search, primarily through Google Scholar, using the contaminant name and keywords such as "drinking water," "occurrence," and "occurrence in water." The EPA also used Hazardous Substances Data Bank (HSDB) and an abstract sifter to identify occurrence information to fill data gaps.

Estimated Occurrence Concentrations

The EPA compiled estimated occurrence concentration data for pesticides on the PCCL 6 that lacked nationally representative finished and nationally representative ambient water data. These pesticides are registered through the EPA's Office of Pesticide Programs (OPP) and are the subject of risk assessments produced through the pesticide registration review process. These assessments often include modeled concentration estimates of acute and chronic drinking water risks that could result from oral exposure to contaminated surface water and groundwater. If no other occurrence data are available, these modeled concentrations, known as estimated drinking water concentrations (EDWCs), were provided as the occurrence concentration in place of finished or ambient water data. In some instances, OPP did not use models to estimate drinking water concentrations and instead used the limit of solubility in water as the estimated concentration. These modeled and estimated concentrations are considered conservative and often based on maximum use and application rates, which may overestimate actual environmental concentrations. For the pesticides without nationally representative water data, the EPA compared modeled data from OPP with the HRL. The modeled values allowed EPA to calculate the EDWC/HRL ratio for pesticides and/or their degradates.

If a pesticide had multiple estimated concentrations based on different lengths of exposure (e.g., acute, chronic, or lifetime exposure) or sources (e.g., surface water or groundwater), the EPA selected the higher value between the estimated surface and groundwater concentrations and selected the value that aligned with the critical effect and data element used to derive the health effect concentration for that chemical. For example, the health effect concentration for oxadiazon is a cancer-based value, with a critical effect of "increase of liver adenomas and/or carcinomas combined in males." Therefore, the EPA selected the surface water-chronic-cancer estimate as the occurrence concentration for oxadiazon rather than estimated peak, acute, or chronic non-cancer concentrations.

Specific information regarding OPP estimated occurrence concentrations can be found in the Occurrence page of the CISs for pesticides lacking other sources of occurrence data. The CISs contain descriptions of the type of estimations and models, the resulting estimated values, and notes about the selection of each value, among other relevant information. The estimated concentrations are also recorded on the Summary and Decision page of the CISs as the concentration in water used to derive the fHQ.

Additional Drinking Water Data Sources

The EPA occasionally conducts voluntary data requests that include information relevant to CCL. Data from the Third Six-Year Review (SYR 3) and Community Water System Survey (CWSS) (USEPA, 2009d); (USEPA, 2009e) were compiled during CCL 5 and continued to be used for CCL 6. For SYR 3, some primacy agencies submitted occurrence data for unregulated contaminants as well as the requested finished water compliance monitoring data collected from 2006 through 2011. For the 2006 CWSS, all systems serving more than 500,000 people were included and were asked about concentrations of unregulated contaminants in their raw and finished water. The SYR 3 ICR and CWSS data were included on the CISs. (Specific information on the SYR 3 ICR, state drinking water monitoring data, and CWSS data used in CCL 6 can be found in Appendix M.)

4.2.2.2 Health Effects Data

Health Assessment Search

As part of the classification step of CCL 6, the EPA searched for health assessments for each PCCL 6 chemical prioritized for evaluation (Appendix E). The EPA extracted toxicity values and other relevant health effects data within the data elements (e.g., cancer classifications) from each identified health assessment. These data informed subsequent steps of the CCL 6 process, including date limits for the rapid systematic review (RSR) (see Section 4.3 and Appendix E) and health concentration derivation (see Section 4.3 and Appendix F), and were provided to the CCL 6 Evaluation Teams as reference material (see Section 4.5.3).

To initiate this process, the EPA searched for qualifying health assessments for each PCCL 6 chemical. Qualifying health assessments include both EPA assessments and externally peer-reviewed, publicly available assessments published by other health agencies. These assessments generally follow methodologies consistent with the EPA's current health guidelines and guidance documents (see Appendix F). Qualifying health assessments were used to derive HRLs as health concentrations.

Some PCCL 6 chemicals do not have qualifying health assessments. In these cases, the EPA secondarily searched for non-qualifying health assessments relevant to these chemicals. Non-qualifying health assessments are also publicly available, are published by health agencies, and provide valuable health information, but they do not necessarily follow standard EPA methodologies and/or are not peer-reviewed by experts outside the publishing agency. Examples of non-qualifying health assessments include the Minnesota Department of Health Toxicological Summaries and the European Chemicals Agency (ECHA) dossiers. A complete list of non-qualifying health assessment sources is included in Appendix F. Non-qualifying health assessments could be used to derive CCL screening levels as health concentrations. Non-qualifying health assessments were also provided to the CCL 6 Evaluation Teams as reference material when qualifying assessments were not available.

Rapid Systematic Review

The EPA implemented a rapid systematic review (RSR) to identify supplemental health effects information for PCCL 6 chemicals prioritized during the CCL 6 screening process. This process was first developed under CCL 5 (USEPA, 2022b) and is further described in Appendix E. The RSR is aimed at identifying health effects data postdating the most recent qualifying health assessment. Specifically, for chemicals with qualifying health assessment(s), the date limit for RSR is set at one year prior to the publication date of the most recent qualifying health assessment. If a chemical does not have a qualifying health assessment, the literature search was not date limited. Rather than providing a comprehensive analysis, these "rapid" systematic reviews are designed to efficiently determine the quantity and types of health effects data available for each chemical. The RSR protocol includes identification of relevant health effects studies, evaluation of study confidence, and extraction of relevant health effects data (e.g., NOAELs and LOAELs) from animal toxicological studies. The RSR protocol also includes identification of relevant health effects studies from other evidence streams (e.g., epidemiological studies or physiologically based pharmacokinetic models), but these were not evaluated or extracted. Supplementary materials and literature search results for each chemical are accessible via the EPA docket (Docket ID No. EPA-HQ-OW-2022-0946).

The RSR protocol was developed under CCL 5 (USEPA, 2022b) and is designed to allow for screening and data synthesis for many chemicals in a relatively short time frame. The protocol comprises the following steps:

- Targeted literature search,
- Title-abstract screening to identify relevant literature, including use of machine learning techniques,
- Full text review to confirm relevant literature,
- Streamlined study confidence evaluation of relevant animal toxicological studies, and
- Targeted data extraction.

To increase efficiency and reduce redundancy of literature searches conducted by other offices and agencies, the EPA did not conduct a health effects RSR for the following groups of PCCL 6 chemicals:

- Chemical pesticides and pesticide metabolites or degradates registered under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), which regularly undergo literature searches through the registration review process of the Office of Pesticide Programs (OPP),

- Food and Drug Administration (FDA)-registered pharmaceuticals for which the EPA relied on lowest therapeutic doses extracted from FDA-approved labels,
- Essential nutrients for which the Institute of Medicine reports are regularly updated,
- Chemicals considered and listed as groups (i.e., DBPs, microplastics, PFAS, and pharmaceuticals), and
- Chemicals that underwent RSR as part of CCL 5 (see Appendix F of (USEPA, 2022b)).

Table E-1 in Appendix E lists the 21 PCCL 6 chemicals prioritized for the health effects RSR under CCL 6. The PCCL 6 also contained 40 chemicals that underwent RSR as part of CCL 5. Results of these RSR searches, including literature search dates, number of references identified, number of studies that passed title-abstract screening, and information related to the highest animal toxicological NOAEL and lowest LOAEL identified for each chemical (e.g., critical study, related health outcome) were populated on CISs and used as important supplemental data to inform the chemical evaluators (see section) of potential health effects that can result from chronic exposure to chemical contaminants. Individual files of chemicals undergoing RSR, which include additional information such as study confidence evaluation results for studies that pass full-text screening and more details of the health effects identified (e.g., age of the animals), were also provided to the chemical evaluators. In the case of PCCL 6 chemicals that already have data from RSR as part of CCL 5, the RSR results from CCL 5 were provided to the chemical evaluators. The full health effects RSR protocol is available in Appendix E.

4.3 Calculated Data Elements

4.3.1 Health Concentrations

Health reference levels (HRLs) and CCL screening levels, referred to collectively as health concentrations in this document, are non-regulatory health-based values that are expressed as concentrations of a chemical in drinking water at or below what a person could consume over a lifetime and be unlikely to experience adverse health effects. These health concentrations are derived for direct comparisons with occurrence concentrations to assess if contaminant levels in finished water suggest a potential risk to human health. Both HRLs and CCL screening levels are expressed in units of µg/L.

HRLs are typically derived from chronic oral toxicity values (e.g., RfDs, population-adjusted doses (PADs), CSFs) extracted from qualifying health assessments. The EPA considered toxicity values derived for shorter exposure durations (i.e., acute and subchronic, each defined consistent with the IRIS glossary (https://www.epa.gov/iris/iris-glossary)) for health concentration derivation if, for example, the health assessment considered these values to be protective of chronic exposure and may not have derived a chronic toxicity value regardless of the literature base on chronic health effects. As described previously, qualifying health assessments include both EPA assessments and externally peer-reviewed, publicly available assessments published by other health agencies that generally follow methodologies consistent with those of the EPA.

Generally, the EPA relied on the Agency's most recently published qualifying health assessment as the source of toxicity values to derive the HRL. The EPA relied on other sources if:

- No EPA health assessments were available for the chemical of interest or
- A qualifying health assessment from another source was published after the most recently published EPA health assessment and used new science (e.g., a critical study published after the

publication date of the EPA assessment, updated methods to derive the point of departure) to derive toxicity values.

When qualifying health assessments were not available for a contaminant, the EPA relied on RfD- and CSF-equivalents published in the most recently finalized non-qualifying health assessment to derive a CCL screening level. If non-qualifying health assessments were also not available for the contaminant of interest, toxicity values identified through the RSR could be used to derive a health concentration (see Section 4.2.2.2). An RfD can be calculated by extracting NOAELs and LOAELs from *medium* or *high* confidence peer-reviewed animal toxicology studies and dividing by the appropriate composite uncertainty factor (UF). Subsequently, this RfD can be used for CCL screening level derivation. Similar to the process for deriving HRLs, chronic toxicity values from non-qualifying assessments and studies identified through the RSR were typically used over toxicity values of shorter exposure duration to derive CCL screening levels.

Appendix F includes the procedures and formulae for calculating health concentrations. For carcinogens, the health concentration is the one-in-a-million (10^{-6}) cancer risk expressed as a drinking water concentration. The EPA applied age-dependent adjustment factors (ADAFs) to chemicals identified as having a mutagenic mode of action (e.g., 1,2,3-trichloropropane) to account for risks associated with early life exposure to mutagenic carcinogens (USEPA, 2005).

For most non-carcinogens, the health concentration is the toxicity value (RfD or equivalent) divided by an exposure factor (i.e., body weight-adjusted drinking water intake) (USEPA, 2019a) relevant to the target population and critical effect and then multiplied by a 20% relative source contribution (RSC) (USEPA, 2000b). An example of a critical effect corresponding to its target population is the case of increased risk of gestational diabetes in pregnant women. For some contaminants, the EPA has curated benchmarks or screening levels that are similar to the health concentrations derived for CCL 6. These benchmarks include the 2021 Human Health Benchmarks for Pesticides (https://www.epa.gov/sdwa/2021-human-health-benchmarks-pesticides) and the Human Health Benchmarks for Pharmaceuticals (HHB-Rx) in Drinking Water (https://www.epa.gov/sdwa/human-health-benchmarks). To ensure consistency across products, reduce redundancy, and address questions raised by public commenters during CCL 5 (USEPA, 2022c), the EPA used these previously curated values rather than re-derive health concentrations when possible. These decisions are documented in Appendix F.

If a chemical has toxicity values based on both cancer and non-cancer data presented within a single assessment, the EPA selected the endpoint that resulted in the most health-protective HRL or CCL screening level value as the final health concentration. The final health concentration is presented on the summary page of the CIS along with the critical effect and data source from which it was derived (see Section 4.4). Additionally, the EPA provided health concentrations derived from all available assessments in the health effects section of the CIS as an additional resource for the chemical evaluators.

4.3.2 Final Hazard Quotients

An important factor indicating potential for public health risk related to exposure from drinking water is the relationship between the chemical contaminant's relative toxicity and the concentrations at which it may be found in water. To assess this relationship, the EPA calculated a metric called the final hazard quotient (fHQ). An fHQ is the ratio of a the 90th percentile (of detections) concentration of a chemical in water to the health concentration (i.e., HRL or CCL screening level) at or below which no adverse effects are expected (as shown in Equation 2). When possible, this ratio was calculated for all PCCL 6 contaminants slated for review by the CCL 6 Evaluation Team.

Equation 2. Formula for Calculating Final Hazard Quotients

$$fHQ = \frac{90th\ percentile\ water\ concentration}{health\ concentration}$$

The fHQ is an important risk assessment tool that chemical evaluators can use to gauge the level of exposure concerns posed by each chemical in water. For the CCL 6, the EPA interpreted this ratio as follows:

- A value less than 0.1 indicates a water concentration less than 10% the health concentration value (lower concern).
- A value greater than 0.1 but less than 1.0 indicates a water concentration between 10% and 100% of the health concentration value (increased concern).
- A value greater than 1.0 indicates a water concentration exceeding the health concentration value (higher concern).

The EPA selected the 90th percentile (of detections) concentration in water as the point of comparison for the ratio, rather than the mean or median. The EPA can use the 90th percentile concentration level as a public health protective benchmark to identify a possible need for a health advisory for areas of the country that may have higher concentrations in finished water than others.

The EPA used a quality-based protocol (see Appendix G) to determine the data source for selecting the water concentration input across the different types of data available during the CCL 6 process. As in past iterations, the EPA prioritized the use of nationally representative finished water data, choosing from the UCMR, UCM, and NIRS datasets first, if available. Contaminants with finished drinking water data and/or large sample sizes as the basis for their fHQ calculations would present ratios that is a better indicator for the potential exposure at levels of public health concern in public water systems than contaminants with fHQ calculations based on ambient water monitoring data and/or small sample size of monitoring data.

For chemicals that lacked or had limited finished water data but had robust ambient water monitoring data such as NAWQA, the EPA used the ambient water concentration to develop the ratio. In rare cases where a chemical had data from a nationally representative ambient water dataset but few measured sites – for example, dithiopyr has one total NAWQA site - the EPA used the next best available data with more measured sites to provide as informative a fHQ calculation as possible. For pesticides with no measured nationally representative water data available, the EPA used modeled water data developed by its Office of Pesticide Programs (OPP), when available, to calculate the fHQ (Section 4.2.2.1). Note the use of modeled water data as the basis for estimating the concentration of occurrence is highly conservative and likely overestimates a contaminant's expected concentration in drinking water, thus fHQ's based on modeled data have a high degree of uncertainty. If a contaminant had no water data (either empirical or modeled), the fHQ would not be calculated and the entry would be left blank on the CIS.

The EPA preferentially selected HRLs as the input in the denominator of the fHQ ratio, as discussed in Section 4.3.1. If an HRL was not available, the EPA selected a CCL screening level derived from a non-qualifying assessment. If non-qualifying assessments were not available, the EPA selected a CCL screening level derived from data presented in studies identified during the rapid systematic literature review. For contaminants with no health concentrations from any source, the fHQ could not be calculated and the entry was left blank on the CIS document.

4.3.3 Attribute Scores

Attribute scores are numeric values the EPA assigned to characterize PCCL chemicals by their observed or predicted qualities or traits, which represent the health effects or anticipated occurrence of each contaminant. During the development of previous CCLs, the EPA recognized that a wide range of data elements would have to be used to characterize each attribute. The CCL process involves classifying relatively new and emerging contaminants, most of which will have incomplete databases and variation in the types of data available for unregulated chemical contaminants. To evaluate chemicals as potential CCL candidates, the EPA needs to establish consistent comparative frameworks for the different types of data representing measures of the attributes.

For CCL 6, the EPA used the attribute scoring protocol established during CCL 5 (USEPA, 2022b). This protocol adheres to the previous criteria for CCL 3 and CCL 4, based on external guidance from NRC and NDWAC:

- The scores for attributes that use numerical categorization should increase with concern (i.e., a 10 is of greater concern, 1 is of lesser concern).
- There should be enough scoring categories to capture the range of data and to discriminate among the data.
- The number of categories should not be so great that they create a false sense of precision.
- The possible range of the scores for a given attribute should be the same regardless of the data elements that are used to assign the score for that attribute.
- The data source and data element used for each attribute should consider more direct measures of occurrence or health effects before potential measures (e.g., peer-reviewed data before unpublished data, and measured data before modeled data).
- The calibration scale (i.e., the scale relating the range for a data element to the scoring categories) should be established using a representative "Universe" of data for each attribute to capture the potential range of values that might be encountered.
- The calibration scale must be set and remain constant throughout the operational process.
- The scoring approach should be as simple as possible, and data should be used with minimal transformations.

NRC recommended using the attributes potency and severity to describe health effects as well as prevalence and magnitude to describe occurrence during the development of CCL 3 (National Research Council (NRC), 2001). When occurrence data are not available, NRC also suggested that environmental fate properties (e.g., persistence and mobility) could be used as surrogates to estimate potential for occurrence. The EPA continues to agree that the recommended attributes are appropriate and consistent with data used in past decisions.

Each attribute as they relate to the CCL 6 data are described in the subsequent sections.

4.3.3.1 Potency

The potency attribute score quantifies the potential for a chemical to cause health effects based on the dose required to elicit the critical effect as identified in a health assessment. For CCL 6, the potency attribute score was quantified from the toxicity value (e.g., RfD, CSF) used to derive the health concentration (e.g., HRL or CCL screening level) for a specific chemical. Potency scores range from 1 to 10, with 10 corresponding to the greatest possible potency (i.e., the greatest potential to cause health effects at lower doses).

The CCL 6 protocol for assigning potency scores is the same as that used during CCL 5 (USEPA, 2022b). The potency scoring protocol requires calibration of a set of toxicity values to normalize a scale with a range from 1 to 10. Consistent with CCL 5, the potency score calibrations for CCL 6 incorporated all available toxicity values from the Universe - that is, a full range of potential potency from low to high toxicity - to establish a scale for potency scoring.

The EPA gathered CCL 6 Universe data and calibrated separate potency scoring scales for four types of toxicity values: CSFs (and equivalents), RfDs (and equivalents), NOAELs, and LOAELs. The EPA plotted the logarithmic distribution of these toxicity values (rounded to the nearest integer) to assess the normality of the distributions and to evaluate the possibility of developing a scale based on these measures. Distributions for each toxicity value type are shown in Figure 11.

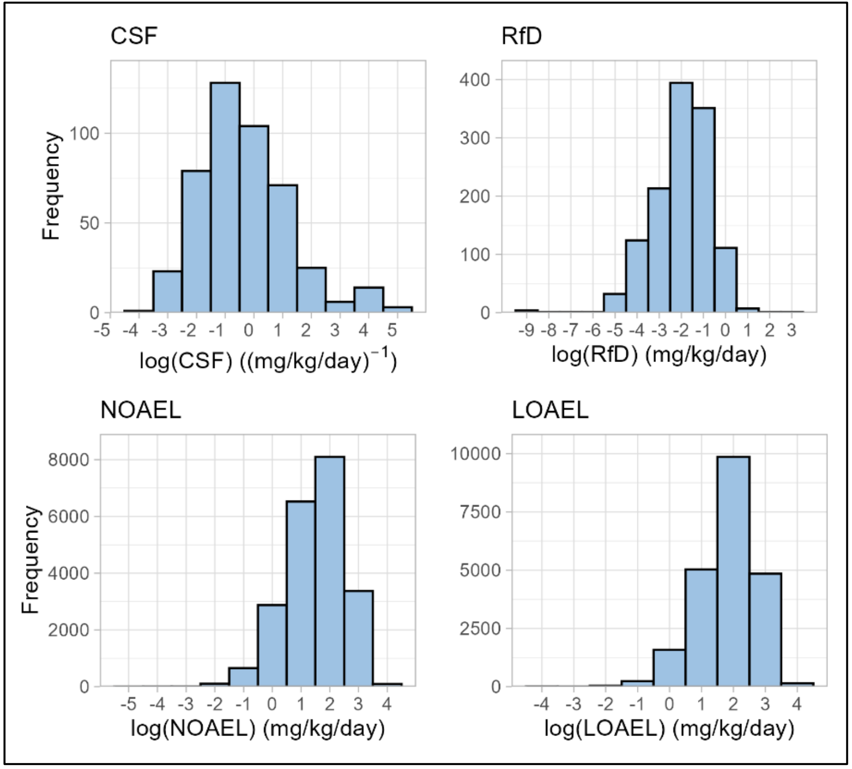


Figure 11. Rounded Logarithmic Distributions of CSFs, RfDs, NOAELs and LOAELs for the CCL 6 Universe

The logarithmic distributions for each type of toxicity value covered the entire range of potential values, with the most frequent value occurring in the middle of the distribution, making each curve approximately log normal. The values for each logarithmic distribution were normalized by binning the data into 10 levels with a center at level 5.

The logarithmic distribution, which is amenable to data with large ranges such as CSF, RfD, NOAEL, and LOAEL values, was used to establish a potency scoring scale equation for each measure of toxicity by identifying the median values in each logarithmic distribution. The results are shown in Table 13. The distribution for each type of toxicity value is different, which necessitated different calibrations for each measure.

Table 13. Median Logarithmic Distribution Values by Toxicity Measurement Type

Toxicity Value	$\log_{10}$(median value)
RfD	-1.82391
CSF	-0.5376
NOAEL	1.60206
LOAEL	1.907411

The median values can then be used in calculations of the potency score for individual chemicals based on the selected toxicity value. For RfDs, NOAELs, and LOAELs, the potency score equals the logarithm of the reported value for a chemical of interest, subtracted from the corresponding logarithmic median of all reported values in the Universe, plus 5, the centered point of the normalized distribution. For CSFs, the equation is similar; however, the properties of the value require the inverse of both the logarithmic median of all reported values and the logarithm of the reported value for the chemical of interest. The potency scoring equations corresponding to each type of toxicity value are listed in Table 14.

Table 14. Potency Scoring Equations by Toxicity Measurement Type

Toxicity Value	Potency Equation
RfD	$\text{Score} = -1.82391 - \log_{10}(\text{RfD}) + 5$
CSF	$\text{Score} = -(-0.5376) + \log_{10}(\text{CSF}) + 5$
NOAEL	$\text{Score} = 1.60206 - \log_{10}(\text{NOAEL}) + 5$
LOAEL	$\text{Score} = 1.907411 - \log_{10}(\text{LOAEL}) + 5$

As with the CCL 5 protocols, the resulting potency scores were rounded to the nearest whole number. Values above 10 were assigned a score of 10 and values below 1 were assigned a score of 1. It is important to note that due to differences in scale calibrations, potency scores derived for one type or category of toxicity value should be compared only to potency scores derived from that same type or category of toxicity value. Appendix H describes the steps required to derive the potency score for a chemical based on the available health information. The potency score associated with the toxicity value used to derive the health concentration is presented on the summary page of the CIS, along with the critical effect, the severity category, and the data source from which it was derived.

4.3.3.2 Severity

The data source used to describe a chemical's potency and health concentration is the same source that is used to describe its severity. Severity refers to the degree or seriousness of a critical physiological change caused by a chemical on the function or survival of a human or animal. Severity is a descriptive measure of the critical effect associated with the toxicity value (e.g., RfD, CSF) for a specific chemical. CCL 6 severity categories correspond with the type of outcome expected to occur at the LOAEL of a particular chemical.

Severity differs from the other attribute scores because it is a qualitative, not quantitative, categorical description of the effects, or lack thereof, of a chemical. As in CCL 5, the EPA elected to simplify categorization of severity for CCL 6 and retained categorical descriptions when referring to severity rather than attempting to quantitatively describe severity. The categories were chosen based on distinct differences in biological outcomes or distinct mechanisms leading to the outcomes. The eight qualitative severity categories used in CCL 6, which are the same as those used in CCL 5, are listed in Table 15. Interpretations of each severity category are presented in Appendix I.

Table 15. CCL 6 Severity Categories

Severity Categories
No adverse effects
Cosmetic effects
Non-cancer effects
Reproductive and developmental effects
Carcinogen with linear mode of action
Carcinogen with non-linear mode of action
Carcinogen with mutagenic mode of action
Reduced longevity

Like previous CCL iterations, the CCL 6 severity category application requires scientific judgment. Appendix I describes the steps required to identify the appropriate severity category for a chemical based on the availability and content of health information. The severity category associated with the health concentration is presented on the summary page of the CIS along with the critical effect and data source from which it was derived.

4.3.3.3 Prevalence and Magnitude

Prevalence and magnitude are the two attributes used to characterize actual or potential occurrence of chemicals in drinking water. Prevalence provides a measure of how widespread a chemical's occurrence is in the environment. Magnitude refers to the quantity of a chemical that is or may be in the environment. When measured or observed occurrence data are not available, persistence and mobility data can be used as surrogate indicators of potential occurrence of a chemical. Persistence and mobility are determined by chemical properties that indicate environmental fate characteristics of a chemical and affect their likelihood to occur in the water environment.

Like the health effects attributes, the occurrence attributes are interrelated. Prevalence and magnitude are linked to the same data source. Table 16 shows how each prevalence measure provides an indicator of how widely the contaminant may be present. The linked magnitude measure, on the other hand, indicates the median concentration of detections in water or the total pounds of the chemical released into the environment.

Table 16. Relationship Between Data Elements Used to Score Prevalence and Magnitude

Prevalence Data	Magnitude Data
Percent detections for a chemical in finished water (nationally)	Median concentration of detections for a chemical in finished water (nationally)
Percent detections for a chemical in ambient water (nationally)	Median concentration of detections for a chemical in ambient water (nationally)
Number of states reporting any releases of a chemical under the Toxics Release Inventory (TRI)	Amount of the total releases of a chemical by the states reporting under the TRI

Unlike the health effects attributes, the data elements used to characterize occurrence are not solely based on a disciplined progressive study of the contaminants. The availability of data from surveys of contaminants in ambient and finished water, detection limits of analytical methods, limitations in reporting requirements, and indirect measures of potential occurrence needed to be considered and evaluated. For the CCL 6, data sources that could provide occurrence data ranged from direct measures of concentrations in water to annual measures of environmental release or production.

The most relevant data elements for characterizing occurrence are measurements of nationally representative finished water taken at PWSs. The data sources for these elements are taken from monitoring studies. These sources include the following:

- Unregulated Contaminant Monitoring Rule (UCMR 1-5) datasets
- Unregulated Contaminant Monitoring-State Rounds 1 and 2 (UCM-State Rounds 1-2) datasets
- National Inorganic and Radionuclide Survey (NIRS)

In the absence of nationally representative finished water data, the next best data elements for characterizing the occurrence attributes are measurements of nationally representative ambient water. The data source for these elements provides a direct measure of chemical contaminants in potential source waters for PWSs and is indicative of possible occurrence in PWSs. The following is the data source used for this element:

- National Water-Quality Assessment (NAWQA) Project

Some chemicals evaluated through the CCL process did not have nationally representative finished or ambient water measurements. To fill this gap, the EPA relied on data elements for measures of pesticide application, chemical release and chemical production that could indicate potential drinking water exposure. The sources for these elements included the following:

- Estimated Annual Agricultural Pesticide Use dataset that provides state-level annual pesticide use estimates for the 48 contiguous states between 1992 and 2019
- Toxics Release Inventory (TRI) that reports annual volumes of chemicals released from industrial applications and the number of states in which those releases occur
- Chemical Data Reporting (CDR) results, which require manufacturers (including importers) to provide the Agency with information on the production and use of chemicals in commerce

4.3.3.4 Prevalence Scoring and Calibration

Prevalence scores are assigned to each PCCL chemical based on the highest ranked data element described in the previous section. The hierarchy of prevalence measures, shown in order from highest to lowest, are these:

1. Percent of PWSs with detections
2. Percent of ambient water sites with detections
3. Number of states reporting application of the chemical as a pesticide
4. Number of states reporting releases (total) of the chemical
5. Production volume in pounds per year

Each of these measures is described in the complete prevalence scoring protocol in Appendix J.

The CCL 6 prevalence scoring protocol is a carryover from the CCL 3 protocol (USEPA, 2009c). In CCL 3, developing the protocol required calibration of the measures for prevalence from the data sources shown in Section 4.3.3.3 to normalize a scale ranging from 1 (least prevalent) to 10 (most prevalent). The EPA compiled a learning dataset of 207 chemicals to develop and calibrate scales for scoring the magnitude and prevalence attributes. The EPA incorporated the full range of potential prevalence data (from low to high) and established an accurate scale to derive scores for the PCCL chemicals.

Scaling analyses focused on establishing chemical groups across the scoring scale. Creating 10 equal bins from the number of states with environmental releases resulted in a scale where a prevalence score of 10 meant releases had to have been reported from 45 or more states. The EPA revised the scale for release data so that if more than half the states (25) reported releases the chemical would receive a prevalence score of 10, which indicates the contaminant’s potential for occurrence was relatively high. The percentage of detections in finished and ambient water (i.e., percentage of systems/sites) was also adjusted to ensure that the most widely detected organic chemicals received more representative scores when compared to the naturally occurring inorganic compounds.

Among occurrence data elements, the link between the measures for prevalence and magnitude works well for the water measurements and environmental release measures. It does not work well when only annual production data are available. The production data provide a measure of pounds of a chemical product produced annually in the United States but do not provide a linked measure such as the number of states in which it is produced or used. This production rate represents the commercial importance of the chemical to some extent.

Since high production tonnage suggests a wide use of a commodity chemical, the EPA decided that production data would be used as a measure for likely prevalence across the country. For example, a chemical produced at a billion pounds per year is more likely to be used and released more widely than a chemical produced at only 10,000 pounds per year. In CCL 3, this hypothesis was supported by analyzing the correlation between a given chemical’s prevalence score based on measures of detections in water and the same chemical’s prevalence score based on the number of states receiving environmental releases based on production. Correlations were only fair to good but justified the use of production data as a measure of prevalence when other data on the spatial spread of a contaminant across the United States are not available.

4.3.3.5 Magnitude Scoring and Calibration

The magnitude scores are assigned to each PCCL chemical based on the highest ranked data element. The hierarchy for magnitude measures, shown in order from highest to lowest, are the following:

1. Median concentration of PWSs with detections
2. Median concentration of ambient water sites with detections
3. Application of the chemical as a pesticide in pounds
4. Total releases of the chemical in pounds
5. Persistence-mobility data

Each of these measures correspond to the complete magnitude scoring protocol in Appendix K.

As with prevalence scoring, the CCL 6 protocol to assign magnitude scores is a carryover from the CCL 3 protocol (USEPA, 2009c). Again, this method required calibration using the different occurrence values from the data sources shown in Section 4.3.3. During CCL 3, the EPA explored a variety of potential scales that could be applied to the finished water concentration data. The EPA converted the finished water data to a standard unit of measure (µg/L) and evaluated several ranges of concentrations to correspond to magnitude scores.

Scores were distributed across the range of values so organic contaminants as well as inorganic compounds could receive high scores. The EPA made comparisons and adjustments until the current protocols using a semi-logarithmic scale were selected. The methods explored and experiments used to calibrate and establish a scoring protocol for the magnitude attribute are further described in the classification document for CCL 3 (USEPA, 2009c).

4.3.3.6 Persistence-Mobility as a Surrogate Measure for Magnitude

If production data are the only measure of occurrence, scoring for prevalence and magnitude becomes difficult. In its report, “Classifying Drinking Water Contaminants for Regulatory Consideration,” NRC discusses persistence and mobility as a fifth attribute and suggests it could be used to predict possible occurrence if other direct measures were not available (National Research Council (NRC), 2001). NDWAC, in its review of the NRC recommendations, suggested that persistence and mobility could provide a surrogate measure of prevalence with production used as a measure of magnitude. The EPA examined the NDWAC proposal by conducting a series of exercises that examined magnitude scores derived from concentrations in drinking water and environmental releases to see if they correlated with production scores and persistence-mobility scores that were calculated using the scoring equation developed by NDWAC. In no case was correlation as good as one might desire, but it was apparent that the persistence-mobility approach showed a better correlation with the magnitude scores, based on the preferred data elements (concentration/release), rather than the production information recommended by NDWAC. Therefore, the EPA chose to use persistence-mobility as a surrogate measure for magnitude when production data were the only measure for scoring prevalence.

Persistence and mobility are environmental fate parameters and considered in combination as a measure of potential occurrence because both transport (i.e., mobility) and fate (i.e., persistence) are important when predicting whether a contaminant is likely to be found in water. Persistence is generally expressed as rate of degradation or half-life ($t_{1/2}$) indicating, in this case, the length of time required for the chemical to degrade to half its original concentration in the medium of interest (e.g., water). Mobility is a measure of a chemical's ability to be transported to and in water, affecting its potential to dissolve in source water and reach a PWS.

The physical/chemical parameters most relevant to a chemical's fate in drinking water are summarized in Table 17. The measure of persistence reflects the time the chemical will remain unchanged in the environment. The first three measures of mobility represent the equilibrium ratio for the partitioning of the contaminant from one medium to another: K_{oc} (sediment: water), K_{ow} (octanol: water) and K_H (air: water). K_{ow} is expressed as the log of the original measurement. For the fourth measure of mobility, solubility, a high solubility favors rapid dissolution of a chemical in the water body from a nearby source and potentially high concentrations if the water source is confined and the environmental release substantial.

The data elements for mobility listed in Table 17 are arranged in hierarchical order, with the most desirable at the top (i.e., the first data to be used if available).

Table 17. Data Elements Used to Score Persistence and Mobility

Persistence	Mobility
Biodegradation Half-Life[1]	Organic Carbon-Water Partition Coefficient (Koc) [1]
	Octanol-Water Partition Coefficient (Kow) [1]
	Henry's Law Coefficient (KH) [1]
	Water Solubility[2]

[1] The predicted biodegradation half-life, Koc, Kow, and KH parameters from the OPERA model (downloaded from EPA's CompTox Chemicals Dashboard)

[2] The predicted water solubility from the TEST or OPERA models (downloaded from EPA's CompTox Chemicals Dashboard)

4.3.3.7 Persistence-Mobility Data – Calibrating Scales and Scoring

Many measurements of environmental fate properties vary depending on the actual field or laboratory conditions. Some are reported in standard data sources only as ranges or categorical descriptions. Scoring was further complicated because two separate environmental fate parameters were used in the scoring of the one attribute. After experimenting with several approaches, the EPA selected the one proposed by NRC and supported by NDWAC by using the persistence and mobility information.
The persistence and mobility data were arrayed or partitioned into relatively simple low-medium-high categories, as suggested by NRC. Published definitions for the categories were used, such as the categories for the octanol/water partition coefficient (K_{ow}) from (Lyman, Reehl, & Rosenblatt, 1990). The categories are given values of 1, 2, or 3 based on the ranking of the measurement from low to high. The persistence value is averaged with the mobility value and a multiplier (10/3) is used to translate the score to a 10-point scale (see the persistence-mobility protocol in Appendix L for details).

The EPA recognized that the persistence-mobility protocol can result in relatively high scores (7 to 10) if more direct data elements for scoring are not available. However, given the uncertainty associated with some persistence-mobility data elements, the EPA decided the somewhat conservative scores were acceptable as surrogate measures for magnitude when only persistence and mobility data were available for scoring.

4.4 Contaminant Information Sheets (CISs)

The EPA developed a CIS for each chemical on the PCCL 6 that was evaluated by the chemical evaluators to make listing recommendations for the CCL 6. Each CIS presents a contaminant's health and occurrence data gathered from main and supplemental data sources along with health and occurrence statistical measures, and additional contextual information, such as previous CCL listing status. An annotated CIS Key and the CISs for the CCL 6 can be found in the CIS Technical Support Document.

Each CIS consists of four pages, including three pages of data and a fourth page for references. The first page provides the contaminant's identity information including name, DTXSID, and CASRN, as well as the contaminant's usage. This page also provides health and occurrence statistical measures such as the contaminant's HRL or CCL screening level (see Section 4.3.1), fHQ (see Section 4.3.2), and health and occurrence attribute scores (see Section 4.3.3). Additional information includes whether the contaminant was subject to past negative regulatory determinations, listed on past CCLs, and publicly nominated for the CCL 6. The first page also identifies whether the contaminant has been listed on the CCL 6; this information was added after the CCL 6 Evaluation Teams concluded their listing recommendations and the EPA conducted a further refinement. This page also indicates whether the contaminant is present on any health or occurrence-related lists (e.g., ATSDR CERCLA Substance Priority List).

The second page of the CIS provides the contaminant's health effects data, including RfDs, CSFs, cancer classifications extracted from health assessments, and other health data from main and supplemental data sources. The second page also summarizes results of the RSR of the health effects literature (Section 4.2.1). Data used to calculate statistical measures like attribute scores and HRLs or CCL screening levels are highlighted.

The third page[2] of the CIS provides the contaminant's occurrence data. This information includes nationally representative finished and ambient water data; application, release, and production data; biomonitoring data; predicted exposure data; and non-nationally representative finished and ambient water data from main and supplemental data sources. The third page also lists modeled environmental fate parameters for the contaminant. Data used to calculate statistical measures like attribute scores are highlighted.

4.5 Evaluation Teams Listing Recommendation Process

Fourteen EPA scientists, referred to as chemical evaluators, reviewed the PCCL contaminants in batches to determine which chemicals to recommend for listing on the draft CCL 6. Evaluation of each PCCL chemical involved the following:

---

[2] Some chemical contaminants have five-page CISs if the health effects information could not fit on one page.

- Review of all relevant health effects and occurrence data provided on the CISs and any available supplemental data and qualifying studies encountered during the additional data collection for PCCL chemicals
- Individual recommendations for chemical listing, with justification for the recommendation
- A facilitated discussion to reach a consensus on whether to recommend listing each chemical, if needed based on the individual recommendations

4.5.1 Evaluation Teams

The EPA divided the chemical evaluators into two seven-member teams to split the workload and expedite the listing recommendation process. The two teams had a similar composition of expertise and specialization. Participants included toxicologists, chemists, biologists, public health professionals, and researchers from the Office of Water, Office of Research and Development, Office of the Administrator (specifically the Office of Children's Health Protection), and Office of Chemical Safety and Pollution Prevention. The EPA also maintained a list of ten alternate chemical evaluators who could be called upon in the case of scheduling conflicts or absences among the primary group of evaluators.

Each team met virtually 12 times between mid-August and early November 2024. One team met for an additional meeting to finalize listing recommendations for chemicals that had been tabled in previous evaluation meetings. In these meetings, the chemical evaluators discussed their independent reviews of each PCCL chemical in the batch to arrive at a consensus on whether to recommend listing on the draft CCL. Batches ranged between seven to 21 chemicals, with a batch of 20 chemicals being the most common (i.e., 10 chemicals per evaluation team). Evaluation team meetings convened an average of one and a half hours per meeting.

4.5.2 Evaluator Training

Prior to chemical evaluators providing their review of PCCL contaminants, all evaluators participated in a training session to familiarize themselves with the background of CCL, the SDWA requirements, and the procedures to follow throughout the evaluations. The training introduced chemical evaluators to the process of taking chemicals from the Universe through classification (i.e., steps 1 to 3) and their role in recommending the draft CCL 6 chemicals. Some chemical evaluators had participated in the CCL 5 process and thus were already familiar with the general process.

At the training, chemical evaluators were also introduced to the internal website where CISs and supplemental health effects information were uploaded for each chemical, divided by their assigned team and batch number. For the CISs, an overview of the layout of the documents was provided with a focus on the calculated data elements such as the four attribute scores, HRLs, and fHQs. Chemical evaluators were also given an overview of the online survey tool they would use to provide written input for each chemical they reviewed independently. In addition to the training session, the evaluators were provided with additional tools - including an iterative FAQ document to ensure any question asked in one team could be available to the other team, a cancer classification reference guide, and a copy of the SDWA language describing the requirements of the CCL - to best support the CCL 6 evaluation process.

4.5.3 Independent Reviews

Before convening team meetings to discuss the chemicals in the weekly batch, the chemical evaluators conducted independent reviews of the chemicals. These reviews focused primarily on the health effects and occurrence information presented on the CISs and in the health effects supplemental information hosted on a SharePoint site created specifically for the evaluations. Upon completing their review of a batch of chemicals, evaluators completed a survey that asked for responses in two areas for each chemical in that batch:

- Provide a numeric ranking for the chemical based on your review of the supporting information [3]:
    - Not List – a score of 1
    - Not List? – a score of 2
    - List? – a score of 3
    - List – a score of 4

- Briefly describe the rationale behind your ranking in 1 to 3 sentences.

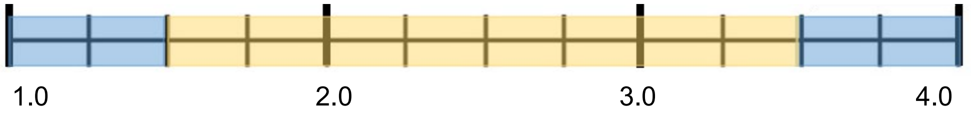
Based on the responses to the first question, the EPA calculated the simple average for the list recommendations across evaluators (between 1.00 and 4.00) for each chemical. Depending on the strength of the numerical listing average for a given chemical, the team would either forego discussion based on a strong consensus average or be required to discuss the chemical at the evaluation team meeting to finalize the list/not list recommendation. The thresholds for undertaking evaluation team discussion on a given chemical are shown in Table 18.

---

[3] A question mark (?) signified that the chemical evaluator was leaning either toward listing or toward not listing a chemical but with some uncertainty.

Table 18. Survey Ranking Outcomes

List Decision	Not List	Not List?	List?	List
Survey Average	(1.00 – 1.49)	(1.50 – 2.49)	(2.50 – 3.49)	(3.50 – 4.00)



Interpretation	Strong consensus	Weak consensus	Strong consensus
CCL 6 Outcome	Chemical not listed	Evaluation team discussion held to finalize the listing decision	Chemical listed

4.5.4 Evaluator Listing Recommendations

After receiving and tallying the chemical evaluators' survey responses for a given batch, the EPA prepared presentation slides with the evaluators' short rationales for each chemical to support any necessary discussion based on the numerical average ranking. The presentation slides helped chemical evaluators understand the range in rankings and justifications for the current batch of chemicals and were used by the meeting facilitator to guide discussions. The meeting facilitator was an EPA staff member with prior experience and certification in meeting facilitation.

At the meeting, the facilitator first summarized the average numerical list decision and range of individual list recommendations. The facilitator then asked each evaluator to explain the listing decision and justification for the chemical. Once all had shared their insights, the facilitator held a verbal roll call. If the team's listing average was within the range of a strong consensus to either list or not list (as shown in Table 17), the listing recommendation was considered final. If the consensus was weak, the outcome could be to go with the majority listing recommendation or table until a future team meeting pending further research. If evaluators required more data or information to make a recommendation - and the requested data or information were able to be obtained in a reasonable timeline - a chemical might be tabled for a later conversation.

Of the 274 top scoring Universe chemicals, the chemical evaluators reviewed 213 chemicals on PCCL 6; as shown previously in Section 3.8 under Table 12, the number 213 is derived by excluding 8 chemicals with preliminary regulatory determinations, one chemical with a pending regulatory action, 25 chemicals identified as canceled non-persistent pesticides with no alternative uses, 8 DBPs (listed as a group), 15 pharmaceuticals without alternative uses (listed as a group), and 4 PFAS (listed as a group).

Ultimately, 84 of the 213 evaluated PCCL 6 chemicals were recommended by evaluators for placement on the CCL 6. The Office of Water then reviewed these recommendations and further refined the list of 84 recommended chemicals. This review resulted in 75 chemicals proposed for the draft CCL 6. This refinement process is further detailed in Section 4.7 below.

4.6 Logistic Regression Analysis

4.6.1 Overview

Logistic regression is a generalized linear model used for binary classification (Kleinbaum & Klein, 2010). In logistic regression, the log-odds of a binary variable or outcome (0 or 1) is modeled by a linear combination of independent variables, or predictors and is used to calculate and predict probabilities between 0 and 1 and odds ratios (ORs) given a set of independent variables.

Simple logistic regression refers to one independent variable with one binary outcome of interest, whereas multiple logistic regression denotes one or more independent variables. An example of a binary classification problem is predicting whether a chemical is recommended for listing on the CCL 6. In CCL 6, the binary outcome of interest is the evaluation teams' list or not list decision. The independent variables or predictors that could influence listing recommendations are screening scores, attribute scores, fHQs, etc. Based on determinant analyses during CCL 5, screening score was utilized as the independent variable of choice for the logistic regression for CCL 6.

The PCCL 6 consisted of 213 chemicals screened from the CCL 6 Universe by a similar point-based screening process as CCL 5 (Chapter 3). To select chemicals for the CCL 6, two teams of chemical evaluators first reviewed the PCCL 6 chemicals (see Section 4.5). The EPA implemented a simple logistic regression model and a midway correlation analysis as a process control (Section 4.6.2). To ensure the efficacy of the screening process to generate a PCCL, the EPA conducted statistical analyses and ran the same logistic regression model developed from CCL 5 to validate selection of the top scoring chemicals for the PCCL 6 while the evaluation team reviews were ongoing. In addition, the EPA implemented a correlation analysis to determine any possible trends between the two evaluation teams and listing recommendations.

A general formulation of a simple logistic regression model with a single predictor expressed in terms of log-odds and probability is shown in Equation 3:

Equation 3. Simple Logistic Regression Model

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X$$

$$P(Y = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

Where X is the independent variable and Y is the dependent variable, or binary outcome of interest, when the outcome is positive (or 1). β_0 and β_1 are unknown model parameters, where $\beta_0\beta_1$ is a slope coefficient. These concepts are further explained in the following sections.

The EPA used data on the PCCL 6 chemicals reviewed by the evaluation teams in the logistic regression model and additional statistical analyses described in the next two sections. Listing recommendations were coded as a binary variable (0 = not list, 1 = list). For evaluated chemicals, screening scores ranged from 3,720 to 10,300.

The logistic regression classification models presented in this section were not used to categorize, prioritize, and/or classify PCCL 6 chemicals for inclusion on the CCL 6. The EPA developed the statistical models to assess the screening and classification processes of CCL 6. The next sections describe the statistical analyses the EPA conducted to investigate the evaluation process for the top scoring chemicals for the PCCL 6, to determine the efficacy of the point-based screening process, and to discover if there were any trends between process controls and listing recommendations in the classification step of CCL 6.

4.6.2 Logistic Regression Applied to Validate the Selection of the PCCL

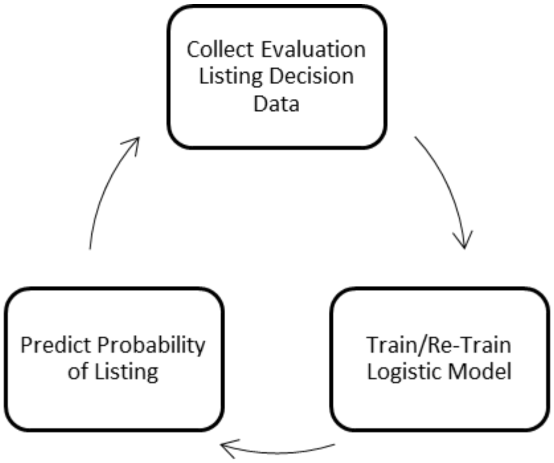
The screening scores prioritize the chemicals most relevant to drinking water exposure and with the potential for greatest public health concern for inclusion on the PCCL 6. The screening framework was designed to efficiently prioritize the entirety of the CCL 6 Universe of chemicals while limiting manual review and human bias. With over 25,000 chemicals in the CCL 6 Universe, the EPA used the screening scores to select and advance the top scoring chemicals for evaluation team review and potential inclusion on the CCL 6 (Section 3.5).

The EPA hypothesized that the screening scores had a positive association with listing outcomes and that the higher the screening score assigned to a chemical, the higher the probability it would be recommended for listing on the CCL 6 by the evaluation teams. To investigate this relationship, the EPA developed simple logistic regression models where screening scores were the sole predictor of listing decision outcomes. The goal of the simple logistic regression model is to use the model as a diagnostics tool during and after the evaluation teams' listing process to provide feedback on the selection of chemicals for the PCCL, as discussed in this section.

The EPA-developed simple logistic regression model provides iterative feedback during evaluation team reviews. The iterative modeling process, illustrated in Figure 12, consisted of three primary steps: Collect evaluation teams listing decision data, Train/re-train the logistic regression model, and Predict the probability of listing a chemical with the highest screening score (10,300) and lowest screening score (3,720). The score of 3,720 in the CCL 6 Universe was used as the cut-off for the top scoring chemicals.

To fit the logistic model according to screening scores and evaluation teams' listing recommendations, the model parameters (β_0 and β_1) need to be estimated. Fitting the model is referred to as the training phase of model development, and the dataset used during model fitting is referred to as the training dataset.

Figure 12. Flow Diagram of the Three -Step Iterative Process



Two teams evaluated PCCL 6 chemicals in 12 batches over several months. The iterative process began following completion of the fifth batch of chemical reviews and successively thereafter until all chemicals were evaluated. The 84 listing recommendations resulting from the first five batches provided a starting point to begin model training. The screening scores for the first five batches ranged from 3,920 to 10,300, which represents a reasonable initial training dataset to obtain probabilities of listing at the initial 3,920 screening score cut off. Upon completion of each subsequent batch of chemical reviews, the training dataset was updated with new listing recommendations, the logistic model was re-trained, and the logistic model was used to predict listing probabilities. The EPA monitored the listing probabilities and uncertainty in model parameter estimations during the training phase of model development. The remainder of this section details the modeling approach and results of the chemical evaluations.

The EPA used Bayesian methods for model parameter estimation of the simple logistic regression model. A Bayesian approach allows for characterization of uncertainty in the parameter estimates and predictions. Additional information on Bayesian statistical methods is provided in (Gelman, et al., 2013) and (Hoff, 2009). The training dataset is well suited for Bayesian logistic regression due to the EPA's need to quantify uncertainty in the predicted listing probabilities when analyzing screening scores. Screening scores are the primary driver deciding the composition of the PCCL 6 and, subsequently, which chemicals are candidates for the CCL 6.

An overview of the Bayesian simple logistic model developed for CCL 5 and re-implemented for CCL 6 is as follows: The binary response variable, the list or not list decision, is modeled as a Bernoulli distribution with a single continuous parameter p, the probability of a chemical being listed. The probability, p, is represented as the logistic model with parameters: β_0 (intercept) and β_1 (slope). The regression coefficients, β_0 and β_1, are related to the log-odds of the probability of a list decision. The EPA assigned uniform prior distributions on β_0 and β_1. The EPA used Markov Chain Monte Carlo (MCMC), which is a class of algorithms commonly used in Bayesian inference, to sample the posterior probability distribution of model parameters β_0 and β_1.[4] Table 19 shows the means, medians, and 95% credible intervals for the model parameters. β_1 is the slope parameter for screening score, and β_o is an intercept term.

Table 19. Summary Statistics for the MCMC Sample

Parameter	Mean	2.5%	Median	97.5%
β_o (intercept)	-3.07	-4.052	-3.043	-1.956
β_1 (slope)	4.859E-4	2.873E-4	4.818E-4	7.014E-4

If the estimated value for β_1 is positive, it indicates a positive association with the binary response variable. Examining the estimated mean value of the screening score slope parameter, β_1, chemicals with higher screening scores are more likely to be listed than those with lower screening scores. β_1 can also be expressed in terms of an odds ratio (OR). OR is a measure of association that represents the effect of a one-unit increase in the independent variable (screening score) on the dependent variable (listing decision outcome). The relationship between OR and a regression coefficient is $OR = e^{\beta_1}$. Therefore, the mean OR calculated from the MCMC sample is 1.000486 (Table 19). Further discussion on the statistical significance of screening scores as a predictor of listing outcomes is in Section 4.6.3.

After training the model, the EPA used the pair-wise samples of parameter values of the posterior distribution to calculate and predict probability of listing across the range of screening scores used in model training (3,720 to 10,300). The logistic model was used to calculate probability of listing at screening scores using the parameter values from the posterior distribution. The EPA focused on the probability of listing at the screening score of 10,300, the score associated with highest scored chemical in the Universe and on the PCCL 6, and the screening score of 3,720. Table 20 contains summary statistics for the probabilities of listing at the screening scores of 3,720 and 10,300 calculated from the MCMC sample.

Table 20. Summary Statistics of Probabilities of Listing at Screening Scores 3720 and 10300 Calculated from the MCMC Sample

Screening Score	Min	5%	Mean	95%	Max
3720 points	0.092	0.16	0.22	0.29	0.38
10300 points	0.53	0.75	0.86	0.95	0.99

Figure 13 illustrates the results of the Bayesian simple logistic model where all listing recommendations and the associated screening scores were used in model training. The x-axis is screening scores, and the y-axis is probability of listing a chemical based on screening score, where 1 is list and 0 is not list. The black line represents the mean probability of listing across the range of screening scores (3,720 to 10,300). The range of screening score values was discretized in evenly spaced steps of 10 to create a 1-

---

[4] To perform the MCMC sampling, the EPA used OpenBUGS (Bayesian inference Using Gibbs Sampling) version 3.2.3 rev. 1012 software (Lunn, Spiegelhalter, Thomas, & Best, 2009). Further analyses were conducted in R (R Core Team, 2025) in RStudio version 1.3.1056 using the CODA (Plummer, Best, Cowles, & Vines, 2006) and Tidyverse (Wickham, et al., 2019) packages. Three Markov chains were used to sample the posterior distribution; the chains were assigned dispersed initial parameter values, and each chain ran for 15,000 iterations. The EPA checked criteria for evidence of chain convergence, visually inspected convergence plots, and conducted posterior predictive checks. The 45,000 pair-wise samples of parameter values were retained.

dimensional grid of values. The result was a vector of equally spaced sequential screening score values that were used to make predictions. The light grey region around the mean curve represents the 90% highest density interval and illustrates how the probabilities vary as a function of screening score. The narrower the light grey band, the less uncertainty in the prediction, and vice versa. The training dataset, screening scores, and listing recommendations are indicated by the red or light blue dots located where the listing probability is 1 (list) or 0 (not list).

As indicated by Table 20 and Figure 13, screening scores have a positive association with listing outcomes, and the probability of listing increases as screening scores increase. The mean probability of listing at the top of the PCCL 6, or the screening score equal to 10,300, is 0.86. Conversely, the mean probability of listing at the screening score equal to 3,720 is 0.22, after the EPA evaluated an additional batch of 24 more chemicals past the score of 3,920. Midway through the evaluation process, the EPA added this additional batch of chemicals to be evaluated, lowering the screening point cut off from 3,920 to 3,720 and allowing the EPA to determine whether further lowering of the cut off would be beneficial for identifying further contaminants of concern in drinking water. The mean probability of listing a contaminant scoring 3,920 was determined to be 0.23. As shown in Section 4.6.3, additional batches would not be expected to generate major change in probability estimate to list lower scoring contaminants, as the estimates reached the lower tail of the logistic regression curve.

These results indicate the screening process achieved its intended goal to elevate chemicals for further review and inclusion on the CCL 6, based on data most relevant to drinking water exposure and potential for greatest health concern. With over 25,000 chemicals in the Universe, the EPA created a prioritization scheme that narrowed the focus of the evaluation teams' task of reviewing PCCL 6 chemicals for potential inclusion on the CCL 6.

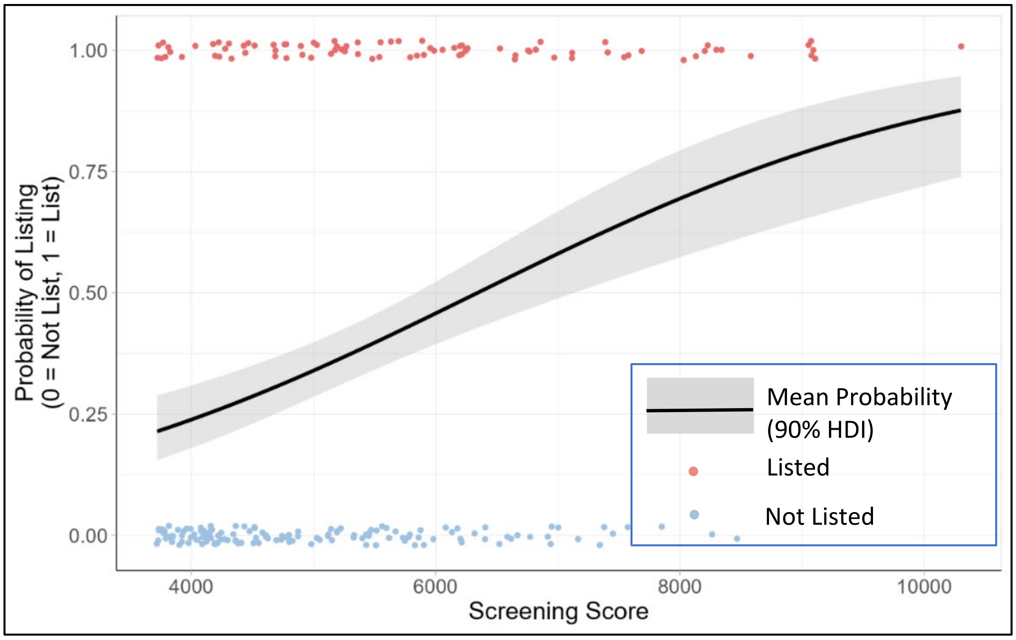


Figure 13. Results of the Bayesian Simple Logistic Model of Probability of Listing vs Screening Score

Although the screening scores incorporate health effects and occurrence data from main data sources, it is reasonable to assume not every contributing factor or determinant of a listing decision outcome is captured in the screening system. Other factors, such as attribute scores and other chemical properties that may impact listing recommendations, were unaccounted for in the simple logistic regression model, although previous CCL 5 results verified the use of screening score moving forward. During the evaluation team meetings, a chemical's screening score was not disclosed, and the data behind the screening scores represent a fraction of the information the chemical evaluators were provided when making listing recommendations. For this reason, there may be a disconnect between the screening scores and listing decision outcomes.

To evaluate potential biases, the EPA conducted further correlation analyses to look for trends between evaluation teams, process control for evaluation teams, and batch trends compared to the listing decision. If significant trends were noticed, it would raise concerns with biases in the evaluation process. Using the *PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis* (Peterson, et al., 2024) package in R, (R Core Team, 2025) batches 1-8 were evaluated for trends and significance level was tested (p-values 0, 0.001, 0.01, 0.05, 0.1, 1). There was no significant correlation between listing decision and evaluation team (0.066, $p>0.1$) or batch (-0.12, $p>0.1$). There was a significant, positive trend between screening score and listing decision (0.33, $p<0.001$) as expected.

4.6.3 Evaluating the Lower Tail of Probability to List

After 11 batches of PCCL 6 contaminants, the lowest scoring chemical on the PCCL had 3,920 screening points with a 0.23 mean probability to list chemicals at that score. The EPA evaluated an additional batch to determine how additional batches would impact the probability to list. The addition of the next 24 chemicals on the PCCL brought the lowest score to 3,720, and re-analysis with the logistic regression yielded a 0.22 mean probability to list. A probability matrix, shown in Figure 14, was developed using the training data through 3,720 to estimate probability for listing lower scores down to 1000, which helped the EPA determine when the evaluation efforts were complete. Figure 14 shows and extension of the lower tail of the estimated probability of listing down to a score of 1000 using the available listing recommendations. As discussed in Section 4.6.2, a positive association was established between the screening scores and listing recommendations. The higher a chemical's screening score, the higher its probability of being listed on the CCL 6. The EPA has reviewed the top scoring chemicals and has reached the lower tail of probability to list.

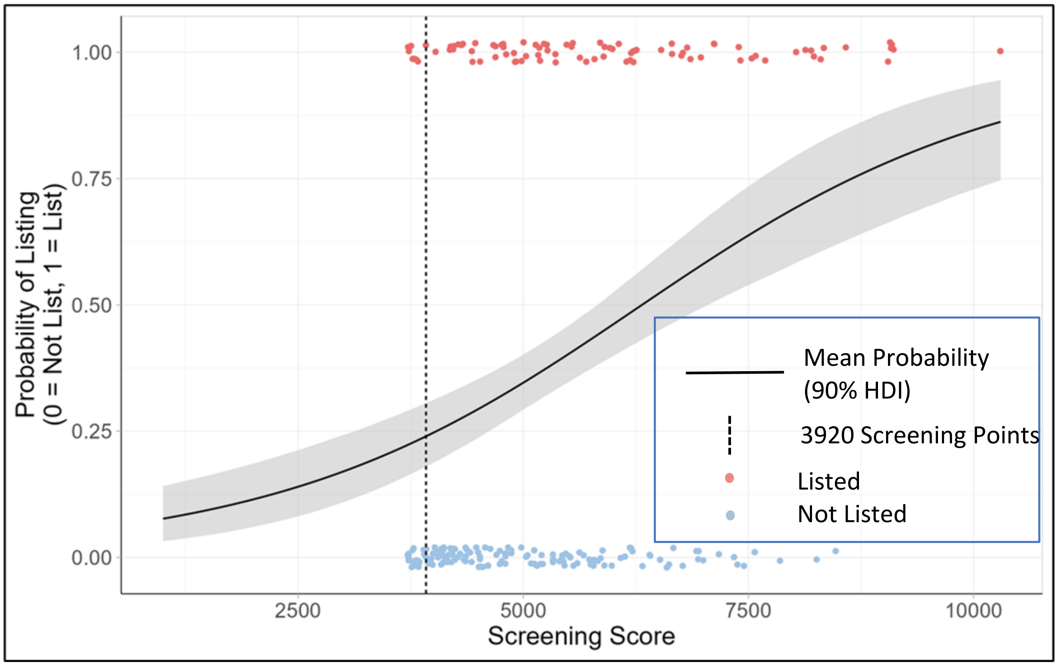


Figure 14. Results of the Bayesian Simple Logistic Model of Probability of Listing vs Screening Score (Extended)

4.7 Refinement for Chemicals with Previous Negative Regulatory Determinations

During the development of the draft CCL 6, the Agency introduced an additional consideration about the history of each contaminant under the SDWA regulatory determination process because this information is relevant to CCL.

Twelve of the 84 chemicals recommended for listing by the chemical evaluators had previously received negative regulatory determinations under the Agency’s Regulatory Determination process. For these twelve chemicals, the EPA assessed whether there was any new health and/or occurrence information to determine whether the new information could result in a different regulatory determination at this time, if listed on the CCL 6. Contaminants that did not have any changes in the health and occurrence information since the time of their previous negative regulatory determination were removed from consideration for listing on the draft CCL 6.

To evaluate whether there were changes in the health data between the prior negative regulatory determination and data collection for CCL 6, the EPA collected information about the HRLs, critical studies, and critical health effects used for the negative regulatory determinations. The EPA compared the information from the regulatory determinations to the data identified during CCL 6. The EPA assessed whether there was a substantive change in the health data, defined as a chemical-specific change (e.g., a change in the critical study or in the science underlying the HRL that is specific to a given chemical). Updates to an HRL because of change(s) that are not specific to a chemical were not considered substantive health changes (e.g., an exposure factor modification that is applied to multiple chemicals).

For all twelve chemicals, the EPA also evaluated whether the contaminants had new nationally representative finished water data since the negative regulatory determination. Manganese was the only contaminant with new UCMR data since the negative regulatory determination. For contaminants with substantive changes to the HRL, the occurrence data were also reanalyzed for any detections in drinking water above the new HRL.

Seven of the twelve chemicals (1,1-dichlorothane; 1,1,2,2-tetrachloroethane; 2,4-dinitrotoluene; aldrin; boron; dieldrin; RDX) lacked substantive changes related to their calculated HRL values and did not have new nationally representative finished water data since the original negative regulatory determinations and were consequently removed from consideration on the draft CCL 6. For example, the HRL for 1,1,2,2-tetrachloroethane decreased from the regulatory determination (RD 2) (0.4 µg/L) to CCL 6 (0.1 µg/L). However, upon further evaluation, the change in the HRL was due to changes in the methodology and calculation of the CSF and an updated exposure factor, both of which are not changes related to health effect information that is chemical-specific, have the potential to be applied for many chemicals, and do not impact the underlying science of health effects attributable to 1,1,2,2-tetrachloroethane. The chemical-specific critical study and critical effect selected was the same in CCL 6 as in the regulatory determination. Therefore, the EPA determined the change in HRL was not a substantive change in the underlying science. Ultimately, because the national finished water occurrence datasets show few exceedances above the CCL 6 HRL and were similar to those evaluated at the time of the regulatory determination, it is unlikely that the new information would result in a different regulatory determination decision.

Two of the chemicals (metribuzin; naphthalene) did have updated health effects information, resulting in substantively different HRL values for CCL 6 that the Agency compared against all occurrence data and noted no reported exceedances. Therefore, these chemicals were also removed from listing consideration on the draft CCL 6. The HRL calculated for naphthalene in CCL 6 (600 µg/L) was greater than the HRL calculated in the regulatory determination (RD 1) (140 µg/L), indicating a lower risk for public health concerns. With the increased HRL, there is no occurrence in the nationally representative drinking water data above the CCL 6 HRL. The lower risk for public health concerns and lack of new occurrence data suggests that the EPA would likely make the same regulatory determination decision with the new information. For metribuzin, the HRL calculated for CCL 6 (7.7 µg/L) was lower than HRL calculated for the regulatory determination (RD 1) (91 µg/L) due to additional considerations for protecting sensitive populations such as the application of the Food Quality Protection Act (FQPA) Safety Factor. With the decreased HRL, the occurrence in available finished drinking water datasets is below the CCL 6 HRL, suggesting the EPA would make the same regulatory determination decision.

Three of the chemicals showed substantive changes related to their calculated HRLs since the original negative determinations which resulted in exceedances or uncertainty in exceedances in the occurrence data (Table 21). Manganese had updated health effects information underlying the HRL, resulting in a lower HRL value, and has occurrence data showing detections above the HRL. This chemical was kept on the draft CCL 6 due to the potential for a different decision under a future regulatory determination. The other two chemicals (2,6-dinitrotoluene; terbufos) also had updated health data resulting in lower HRL values. The updated HRLs (0.02 µg/L and 0.01 µg/L for 2,6-dinitrotoluene and terbufos, respectively) are below UCMR minimum reporting levels (0.5 µg/L), meaning the Agency does not have occurrence data at the level of the HRL and cannot determine if there is occurrence in the range between the MRL and HRL. Due to the uncertainty in the occurrence data, these two chemicals remain on CCL 6.

Table 21 Data Review of PCCL Chemicals with Previous Negative Regulatory Determinations Recommended for Listing

Chemical Name	Regulatory Determination Cycle	New UCMR Data	Substantive Change to Health Data	Occurrence Above the HRL[1]	Listed on the Draft CCL 6?
Manganese	1	Yes	Yes, indicating increased potency	Yes	Yes
2,6-Dinitrotoluene	2	No	Yes, indicating increased potency	Uncertain (MRL[1]>HRL)	Yes
Terbufos	3	No	Yes, indicating increased potency	Uncertain (MRL[1]>HRL)	Yes
Metribuzin	1	No	Yes, indicating increased potency	No	No
Naphthalene	1	No	Yes, indicating decreased potency	No	No
1,1,2,2-Tetrachloroethane	2	No	No	N/A	No
1,1-Dichloroethane	4	No	No	N/A	No
2,4-Dinitrotoluene	2	No	No	N/A	No
Aldrin	1	No	No	N/A	No
Boron	2	No	No	N/A	No
Dieldrin	1	No	No	N/A	No
RDX	4	No	No	N/A	No

[1] Applies only to chemicals that met the criteria for new UCMR data and substantive changes to health data.
[2] MRL = minimum reporting level

4.8 Draft CCL 6 Chemical Selection

The CCL 6 is comprised of four chemical groups (Section 4.8.1) and 75 individual chemicals (Section 4.8.2).

4.8.1 Group Listings

DBPs, microplastics, PFAS, and pharmaceuticals have been identified as Agency priorities and contaminants of concern for drinking water under other Agency actions. Listing these four chemical groups on the CCL 6 does not necessarily mean that the EPA will make subsequent regulatory decisions for the entire group. Rather, the EPA will evaluate additional data on the listed groups, subgroups, and individual contaminants to inform any regulatory determinations. Contaminants from the CCL with sufficient health effects and occurrence-information are considered for regulatory determinations and rulemaking under SDWA. For example, the contaminants in the CCL 5 cyanotoxin group were assessed individually during the regulatory determination process and two contaminants from the group had sufficient data to make preliminary regulatory determinations (USEPA, 2025)).

4.8.1.1 Disinfection Byproducts (DBPs)

DBPs are a class of compounds formed in water when disinfectants, used for purposes of antimicrobial treatment, react with naturally occurring or man-made materials. Over 600 DBPs have been identified since their discovery ((Richardson & Postigo, The Next Generation of Drinking Water Disinfection By-Products: Occurrence, Formation, Toxicity, and New Links with Human Epidemiology, 2015)). DBPs are known to co-occur in distribution systems and have the potential for adverse health effects ((Villanueva, et al., 2004); (Richardson, Plewa, Wagner, Schoeny, & Demarini, 2007); (Bove, Rogerson, & Vena, 2010); (Rahman, Driscoll, Cowie, & Armstrong, 2010); (Kalita, Kamilaris, Havinga, & Reva, 2024)). In 2006, the EPA's Stage 2 Disinfectants and Disinfection Byproducts Rule resulted in the regulation of eleven DBPs from three subgroups that include four trihalomethanes, five haloacetic acids, and two inorganic compounds (bromate and chlorite) (Stage 2 Disinfection Byproducts Requirements, 40 CFR Part 141 Subpart V, 2006). Later, under the SYR 3, the EPA identified ten of these DBPs (with the exception of bromate) as candidates for revision (USEPA, 2017).

For CCL 6, the EPA is listing 27 unregulated DBPs as a group, shown in Table 22. Listing unregulated DBPs as a group on CCL 6 is consistent with the EPA's identification of several microbial and disinfection byproduct drinking water regulations as candidates for revision under the SYR 3 and acknowledges DBPs as a continued Agency priority.

For the draft CCL 6, the individual DBPs listed under this chemical group fall under one or more of the following categories (identified in Table 22):

- Are among the top scoring Universe chemicals (separated from the PCCL 6 for inclusion in the chemical group and excluded from the evaluation teams listing recommendation process described in Section 4.5).
- Are recommended for inclusion by Agency subject matter experts,
- Are brought forward from the CCL 5 DBP chemical group.

Table 22. Unregulated DBPs in the Draft CCL 6 DBP Group

Chemical Name	CASRN	DTXSID
Brominated Haloacetic Acids		
Bromochloroacetic acid (BCAA) ‡	5589-96-8	DTXSID4024642
Bromodichloroacetic acid (BDCAA) ‡	71133-14-7	DTXSID4024644
Dibromochloroacetic acid (DBCAA) ‡	631-64-1	DTXSID3031151
Tribromoacetic acid (TBAA) ‡	75-96-7	DTXSID6021668
Haloacetonitriles		
Bromochloroacetonitrile (BCAN) †	83463-62-1	DTXSID9021500
Dibromoacetonitrile (DBAN)* ‡	3252-43-5	DTXSID3024940
Dichloroacetonitrile (DCAN)* ‡	3018-12-0	DTXSID3021562
Trichloroacetonitrile (TCAN) †	545-06-2	DTXSID0021672
Halonitromethanes		
Bromodichloronitromethane (BDCNM) ‡	918-01-4	DTXSID4021509
Chloropicrin (trichloronitromethane, TCNM)* ‡	76-96-2	DTXSID0020315
Dibromochloronitromethane (DBCNM) ‡	1184-89-0	DTXSID00152114
Iodinated Trihalomethanes		
Bromochloroiodomethane (BCIM) ‡	34970-00-8	DTXSID4021503
Bromodiiodomethane (BDIM) ‡	557-95-9	DTXSID70204235
Chlorodiiodomethane (CDIM) ‡	638-73-3	DTXSID20213251
Dibromoiodomethane (DBIM) ‡	557-68-6	DTXSID60208040
Dichloroiodomethane (DCIM) ‡	594-04-7	DTXSID7021570
Iodoform (triiodomethane, TIM) ‡	75-47-8	DTXSID4020743
Nitrosamines		
Nitrosodibutylamine (NDBA) ‡	924-16-3	DTXSID2021026
N-Nitrosodiethylamine (NDEA)* ‡	55-18-5	DTXSID2021028
N-Nitrosodimethylamine (NDMA)* ‡	62-75-9	DTXSID7021029
N-Nitrosodi-n-propylamine (NDPA) ‡	621-64-7	DTXSID6021032
N-Nitrosodiphenylamine (NDPhA) ‡	86-30-6	DTXSID6021030
Nitrosopyrrolidine (NPYR) ‡	930-55-2	DTXSID8021062
Others		
Chloral hydrate*†	302-17-0	DTXSID7020261
Chlorate*‡	14866-68-3	DTXSID3073137
Chloronitramide anion†	58999-86-3	DTXSID701357782
Formaldehyde*‡	50-00-0	DTXSID7020637

*Top scoring Universe chemicals
†Recommended for inclusion by Agency subject matter experts
‡Brought forward from the CCL 5 DBP chemical group

4.8.1.2 Microplastics

The EPA acknowledges the interest in and concern for microplastics in sources of drinking water. The Agency received a public nomination for including microplastics on CCL 6 that was accompanied by three data sources ((Miller, et al., 2021); (Ragusa, et al., 2021); and (Zarus, Muianga, Hunter, & Pappas, 2021)), all indicating potential concern for exposure to microplastics. In the Science Advisory Board's recommendations for the draft CCL 5, the SAB encouraged the EPA to consider the assessment and inclusion of microplastics on future CCLs ((USEPA, 2022b)). The Agency is including these contaminants as a group on the draft CCL 6 as a first step to define and better understand the potential public health risk from exposure via drinking water. It is important to note that this is the first time that a group that contains primarily polymers, instead of chemicals, is listed on the draft CCL. Research is needed to determine the adverse health effects and the characteristics of the microplastics (i.e., size, polymer type, etc.) that are factors leading to adverse health effects from exposure via drinking water. Research is also needed to develop robust and validated analytical methods for measuring microplastics in drinking water that may be used to standardize data collection and analysis in the future. Similarly, although some studies reported adverse health effects from exposure to microplastic constituents in drinking water, the available data need to be evaluated for the purpose of health-based risk level derivation by the EPA and additional research is needed to inform any form of regulatory health-based risk level (Coffin, et al., 2022), which would be necessary to derive a certain level for public health concern.

At the time of the publication for the draft CCL 6, significant data gaps for health effects and occurrence of microplastics that require further research before the Agency can fully understand the health risks associated with microplastics in drinking water. The data gaps include (but are not limited to) the following:

1. Factors associated with health effects: Determination of the characteristics of the microplastics (i.e., colors, polymers, shapes, sizes, additional chemical components, etc.) that are consistently associated with adverse health effects in humans from exposure in drinking water.
2. Detection technology: A validated, reliable analytical method with the proper quality control data, accuracy, and precision is needed to allow the EPA to be able to detect and analyze the concentrations of microplastics occurring in drinking water.
3. Combination with other substances: the need to better understand how microplastics occurring with and within other substances may impact detection of specific microplastics and identifying their health risks.
4. Sources in drinking water: the need to better understand all potential sources of plastic pollution that contribute to the formation of microplastics in sources of drinking water.

4.8.1.3 Per- and Polyfluoroalkyl Substances (PFAS)

PFAS are a class of synthetic chemicals commonly used to make products resistant to water, heat, and stains and are consequently found in industrial and consumer products like clothing, food packaging, cookware, cosmetics, carpeting, and fire-fighting foam (Cohen, 2020); (USEPA, 2019a). More than 4,000 PFAS have been manufactured and used globally since the 1940s (USEPA, 2019b), which would make listing PFAS individually on the CCL 6 a challenge. The EPA is listing PFAS as a group inclusive of any PFAS that fit the CCL 6 structural definition, which was developed under the CCL 5 (USEPA, 2022b), except those that are already covered under an NPDWR (National Primary Drinking Water Regulations, Subpart Z – Control of Per- and Polyfluoroalkyl Substances (PFAS), 40 C.F.R. § 141.900 – 141.905). The proposal to list PFAS as a chemical group is responsive to public nominations and is consistent with the approach taken for CCL 5. For CCL 6, the same structural definition of per- and polyfluoroalkyl substances (PFAS) as developed for the final CCL 5, includes chemicals that contain at least one of these three structures:

1) R-(CF2)-CF(R′)R″, where both the CF2 and CF moieties are saturated carbons, and none of the R groups can be hydrogen
2) R-CF2OCF2-R′, where both the CF2 moieties are saturated carbons, and none of the R groups can be hydrogen
3) CF3C(CF3)RR′, where all the carbons are saturated, and none of the R groups can be hydrogen.

The EPA is committed to addressing PFAS in drinking water and the environment, guided by the principles of strengthening the science, fulfilling statutory obligations and enhancing communication, and building partnerships (https://www.epa.gov/newsreleases/administrator-zeldin-announces-major-epa-actions-combat-pfas-contamination).
Listing PFAS as a group on the draft CCL 6 is in keeping with the Agency's commitment to better understand and ultimately reduce the potential risks caused by this broad class of chemicals. Including the group of PFAS on the draft CCL 6 demonstrates the Agency's commitment to prioritizing and building a strong foundation of science on PFAS.

4.8.1.4 Pharmaceuticals

For over a decade, public concern about the presence of pharmaceutical substances in sources of drinking water has been a recurring topic of discussion for the Agency's prioritization of contaminants under SDWA. Since 2012, the EPA has led a federal workgroup on pharmaceuticals in water alongside USDA, FDA, and USGS to exchange information on pharmaceuticals in the environment and to support the coordination of joint studies. Additionally, the EPA has listed individual pharmaceuticals on previous CCLs and has required monitoring for some pharmaceuticals on previous UCMR cycles.
The EPA committed to understanding contaminants in drinking water and has identified pharmaceuticals as an Agency priority based on feedback received through the public nominations process. The EPA received a public nomination in CCL 6 for "pharmaceutical waste (specifically estrogenic compounds)." For CCL 6, the Agency incorporated new data sources ((Schaider, et al., 2014) and (Battaglin, et al., 2018)) along with previous main sources from CCL 5 (e.g. (Glassmeyer, et al., 2017)) that provided additional information about the occurrence of pharmaceutical products in water; for health data, the Agency added a new data source in CCL 6 to identify chemicals with modeled and quantified estrogen receptor bioactivity and continued to rely on sources describing potential toxicity of pharmaceutical compounds as identified in CCL 5 (e.g., maximum recommended daily therapeutic doses published by FDA) (USEPA, 2023b). The Agency also developed the Human Health Benchmarks for Pharmaceuticals (HHB-Rx) in Drinking Water (https://www.epa.gov/sdwa/human-health-benchmarks), non-enforceable drinking water levels that provide information about adverse health effects from drinking water exposure to contaminants that have no drinking water standards or health advisories. Benchmarks were developed for pharmaceuticals that met certain criteria; there are not benchmarks for all pharmaceuticals. The benchmarks, based on potential health effects from exposure via drinking water, informed the screening of pharmaceuticals and identification of the top scoring pharmaceuticals and provided the EPA with an updated perspective on current gaps in toxicological research for this broad class of chemicals.

The Agency is proposing the inclusion of a pharmaceutical group on the draft CCL 6 to provide support for the areas of research and information needed to identify which specific pharmaceuticals are occurring in drinking water and may be of greatest public health concern to inform further action under the SDWA. For the purposes of the draft CCL 6, the EPA considers pharmaceuticals to include any substances defined as a “drug” under the FFDCA (Federal Food, Drug, and Cosmetic Act. 21 U.S.C § 321, 1938).[5]

4.8.2 Individual Listings

In addition to the four chemical groups described in Section 4.8.1, the draft CCL 6 contains 75 individual chemicals recommended for listing by the evaluation teams and further refined by the EPA based on prior history of each contaminant under the SDWA regulatory determination process. (discussed previously in Sections 4.5 and Section 4.7, respectively). Table 23 contains a complete list of the individual draft CCL 6 chemicals.

Table 23. Individual Chemical Contaminants on the Draft CCL 6

Chemical Name	CASRN	DTXSID
1,2,3-Trichloropropane	96-18-4	DTXSID9021390
1,2,4-Triazole	288-88-0	DTXSID6027131
1,2,4-Trimethylbenzene	95-63-6	DTXSID6021402
1,2-Diphenylhydrazine	122-66-7	DTXSID7020710
1,4-Dioxane	123-91-1	DTXSID4020533
1-Methylnaphthalene	90-12-0	DTXSID9020877
2,4,6-Trinitrotoluene	118-96-7	DTXSID7024372
2,6-Dinitrotoluene	606-20-2	DTXSID5020528
4-tert-Octylphenol	140-66-9	DTXSID9022360
Acephate	30560-19-1	DTXSID8023846
Acrylonitrile	107-13-1	DTXSID5020029
alpha-1,2,3,4,5,6-Hexachlorocyclohexane	319-84-6	DTXSID2020684
Aluminum	7429-90-5	DTXSID3040273
Anthraquinone	84-65-1	DTXSID3020095
Bensulide	741-58-2	DTXSID9032329
Benzyl butyl phthalate	85-68-7	DTXSID3020205
Bisphenol A	80-05-7	DTXSID7020182
Bromoxynil	1689-84-5	DTXSID3022162
Carbaryl	63-25-2	DTXSID9020247
Carbendazim	10605-21-7	DTXSID4024729
Chlordecone (Kepone)	143-50-0	DTXSID1020770
Chloromethane	74-87-3	DTXSID0021541
Chlorothalonil	1897-45-6	DTXSID0020319
Chlorpyrifos	2921-88-2	DTXSID4020458
Clothianidin	210880-92-5	DTXSID2034465
Cobalt	7440-48-4	DTXSID1031040

---

[5] The term “drug” means (A) articles recognized in the official United States Pharmacopoeia, official Homoeopathic Pharmacopoeia of the United States, or official National Formulary, or any supplement to any of them; and (B) articles intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease in man or other animals; and (C) articles (other than food) intended to affect the structure or any function of the body of man or other animals; and (D) articles intended for use as a component of any article specified in clause (A), (B), or (C).

Chemical Name	CASRN	DTXSID
Diazinon	333-41-5	DTXSID9020407
Dicamba	1918-00-9	DTXSID4024018
Dichlorvos	62-73-7	DTXSID5020449
Dicrotophos	141-66-2	DTXSID9023914
Dithiopyr	97886-45-8	DTXSID9032379
Diuron	330-54-1	DTXSID0020446
Ethalfluralin	55283-68-6	DTXSID8032386
Ethylene thiourea	96-45-7	DTXSID5020601
Fenbuconazole	114369-43-6	DTXSID8032548
Fipronil	120068-37-3	DTXSID4034609
Flufenacet	142459-58-3	DTXSID2032552
Fluometuron	2164-17-2	DTXSID8020628
Fluoranthene	206-44-0	DTXSID3024104
Imazalil	35554-44-0	DTXSID8024151
Iodide	20461-54-5	DTXSID80912339
Iprodione	36734-19-7	DTXSID3024154
Isophorone	78-59-1	DTXSID8020759
Lithium	7439-93-2	DTXSID5036761
Malaoxon	1634-78-2	DTXSID9020790
Malathion	121-75-5	DTXSID4020791
Manganese	7439-96-5	DTXSID2024169
Methomyl	16752-77-5	DTXSID1022267
Methyl mercury	22967-92-6	DTXSID9024198
Methyl tert-butyl ether (MTBE)	1634-04-4	DTXSID3020833
Nicotine	54-11-5	DTXSID1020930
Nitroglycerin	55-63-0	DTXSID1021407
Nonylphenol	25154-52-3	DTXSID3021857
Oryzalin	19044-88-3	DTXSID8024238
Oxadiazon	19666-30-9	DTXSID3024239
Oxyfluorfen	42874-03-3	DTXSID7024241
p-Cresol	106-44-5	DTXSID7021869
Phorate	298-02-2	DTXSID4032459
Phosmet	732-11-6	DTXSID5024261
Propargite	2312-35-8	DTXSID4024276
Quinoline	91-22-5	DTXSID1021798
Silver	7440-22-4	DTXSID4024305
Strontium	7440-24-6	DTXSID3024312
Terbufos	13071-79-9	DTXSID2022254
tert-Butyl alcohol	75-65-0	DTXSID8020204
Thiamethoxam	153719-23-4	DTXSID2034962
Tolyltriazole	29385-43-1	DTXSID0026171
Tri-allate	2302-17-5	DTXSID5024344
Tributyl phosphate	126-73-8	DTXSID3021986
Trifluralin	1582-09-8	DTXSID4021395
Triphenyl phosphate	115-86-6	DTXSID1021952

Chemical Name	CASRN	DTXSID
Tris(2-chloroethyl) phosphate	115-96-8	DTXSID5021411
Tungsten	7440-33-7	DTXSID8052481
Vanadium	7440-62-2	DTXSID2040282
Zinc	7440-66-6	DTXSID7035012

Chapter 5 CCL 6 Data Availability Assessment

5.1 Overview

The CCL 6 development process included assessing the current availability of data for the chemical contaminants listed on the draft CCL 6 and PCCL 6. In later steps, upon finalizing the CCL 6, the EPA will assess the data needs and evaluate and identify future research priorities, including efforts such as evaluating a chemical contaminant for potential monitoring under the UCMR program or identifying contaminants in need of health assessment revisions or development.

5.2 Data Availability for Draft CCL 6 Chemicals

The EPA provides the initial assessment of the current data availability of chemical contaminants on the draft CCL 6 in Table 24. Chemicals are categorized depending on availability of their occurrence and health effects data. This list is a starting point for identifying the data needs of the Final CCL 6 contaminants.

Contaminants in Group A have nationally representative finished water data and qualifying health assessments that derive oral toxicity value(s). The Group A contaminants currently meet the data needs to proceed for further evaluation by the EPA during the Regulatory Determination process. Contaminants in Group B have finished water data that are not nationally representative and have qualifying health assessments. The Group B contaminants could be considered by the EPA for potential inclusion on future UCMR data collections to fill the remaining data gap. Contaminants in groups C and D have more substantial data needs because these contaminants lack either finished water data or a qualifying health assessment that derives oral toxicity value(s). Contaminants in Group C are anticipated to occur in public water systems based on the available ambient water data but have a need for finished water data before the EPA can fully examine the public health concern related to exposure from drinking water. Contaminants in Group D have various occurrence data gaps (for example, 4-tert-octylphenol is lacking finished water data and has the same needs as Group C). The contaminants in Group D do not currently have a qualifying health assessment (i.e., a peer-reviewed, publicly available health assessment developed by the EPA or other health agencies that follow similar guidance and methods), which are preferred for evaluation of contaminants during the Regulatory Determination process. The EPA did not assess data availability for the DBPs and PFAS groups because the availability of health effects and occurrence data varies with individual chemicals in each group. The EPA is addressing these groups broadly in drinking water based on a subset of chemicals in these groups that are known to occur in PWSs and cause adverse health effects.

Table 24. Data Availability for Draft CCL 6 Chemicals

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Qualifying Health Assessment	Drinking Water Analytical Method[1]
A. Contaminants with Nationally Representative Finished Water Occurrence Data and Qualifying Health Assessment(s) with Derivation of Oral Toxicity Value(s)					
1,2,3-Trichloropropane	96-18-4	DTXSID9021390	Finished National	Yes	EPA 551.1, EPA 502.2, EPA 504.1, EPA 524.3, EPA 524.4, ASTM D5790-18, SM 6200 B, SM 6200 C

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Qualifying Health Assessment	Drinking Water Analytical Method[1]
1,2,4-Trimethylbenzene	95-63-6	DTXSID6021402	Finished National	Yes	EPA 502.2, EPA 524.2, EPA 524.3, EPA 524.4, ASTM D5790-18, SM 6200 B, SM 6200 C
1,2-Diphenylhydrazine	122-66-7	DTXSID7020710	Finished National	Yes	EPA 526
1,4-Dioxane	123-91-1	DTXSID4020533	Finished National	Yes	EPA 522, EPA 541
2,4,6-Trinitrotoluene	118-96-7	DTXSID7024372	Finished National	Yes	EPA 529
2,6-Dinitrotoluene	606-20-2	DTXSID5020528	Finished National	Yes	EPA 525.2, EPA 525.3, EPA 529
alpha-1,2,3,4,5,6-Hexachlorocyclohexane	319-84-6	DTXSID2020684	Finished National	Yes	EPA 508, EPA 508.1, EPA 525.2, EPA 525.3
Aluminum	7429-90-5	DTXSID3040273	Finished National	Yes	EPA 200.5, EPA 200.7, EPA 200.8, EPA 200.9, SM 3111 D, SM 3113 B, SM 3120 B, ASTM D857-17
Carbaryl	63-25-2	DTXSID9020247	Finished National	Yes	EPA 531.1, EPA 531.2, ASTM D5315-04, SM 6610 B
Chlorpyrifos	2921-88-2	DTXSID4020458	Finished National	Yes	EPA 525.2, EPA 525.3, EPA 527
Cobalt	7440-48-4	DTXSID1031040	Finished National	Yes	EPA 200.7, EPA 200.8, EPA 200.9, ASTM D3558-15 A, ASTM D3558-15 B, SM 3111 B, SM 3111 C, SM 3113 B, SM 3120 B
Diazinon	333-41-5	DTXSID9020407	Finished National	Yes	EPA 526
Dicamba	1918-00-9	DTXSID4024018	Finished National	Yes	EPA 555, SM 6640 B
Diuron	330-54-1	DTXSID0020446	Finished National	Yes	EPA 532
Lithium	7439-93-2	DTXSID5036761	Finished National	Yes	EPA 200.7, ASTM D1976-20, SM 3111 B, SM 3120 B, SM 3500-Li B
Manganese	7439-96-5	DTXSID2024169	Finished National	Yes	EPA 200.5, EPA 200.7, EPA 200.8, EPA 200.9, SM 3111 B, SM 3111 C, SM 3113 B, SM 3120 B, SM 3500-Mn B
Methomyl	16752-77-5	DTXSID1022267	Finished National	Yes	EPA 531.1, EPA 531.2, EPA 540, ASTM D5315-04, SM 6610 B
Methyl tert-butyl ether (MTBE)	1634-04-4	DTXSID3020833	Finished National	Yes	EPA 524.2, EPA 524.3, EPA 524.4, ASTM D5790-18, SM 6200 B, SM 6200 C
Oxyfluorfen	42874-03-3	DTXSID7024241	Finished National	Yes	EPA 525.3

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Qualifying Health Assessment	Drinking Water Analytical Method[1]
Quinoline	91-22-5	DTXSID1021798	Finished National	Yes	EPA 530, EPA 538
Silver	7440-22-4	DTXSID4024305	Finished National	Yes	EPA 200.5, EPA 200.7, EPA 200.8, EPA 200.9, SM 3120 B, SM 3111 B, SM 3113 B, ASTM D3866-18 A, ASTM D3866-18 C
Strontium	7440-24-6	DTXSID3024312	Finished National	Yes	EPA 200.7
Terbufos	13071-79-9	DTXSID2022254	Finished National	Yes	EPA 526
Vanadium	7440-62-2	DTXSID2040282	Finished National	Yes	EPA 200.5, EPA 200.7, EPA 200.8, SM 3111 D, SM 3120 B, SM 3500-V B
Zinc	7440-66-6	DTXSID7035012	Finished National	Yes	EPA 200.5, EPA 200.7, EPA 200.8, SM 3111 B, SM 3120 B, ASTM D1691-17
B. Contaminants with Non-Nationally Representative Finished Water Occurrence Data and Qualifying Health Assessment(s) with Derivation of Oral Toxicity Value(s)					
1,2,4-Triazole	288-88-0	DTXSID6027131	Finished Non-National	Yes	No
1-Methylnaphthalene	90-12-0	DTXSID9020877	Finished Non-National	Yes	No
Acrylonitrile	107-13-1	DTXSID5020029	Finished Non-National	Yes	EPA 524.2
Anthraquinone	84-65-1	DTXSID3020095	Finished Non-National	Yes	No
Benzyl butyl phthalate	85-68-7	DTXSID3020205	Finished Non-National	Yes	EPA 506, EPA 525.2, EPA 525.3
Bisphenol A	80-05-7	DTXSID7020182	Finished Non-National	Yes	SM 6810 B
Bromoxynil	1689-84-5	DTXSID3022162	Finished Non-National	Yes	EPA Method in Development 562
Carbendazim (MBC)	10605-21-7	DTXSID4024729	Finished Non-National	Yes	EPA Method in Development 562
Chlorothalonil	1897-45-6	DTXSID0020319	Finished Non-National	Yes	EPA 508, EPA 508.1, EPA 525.2, EPA 525.3
Clothianidin	210880-92-5	DTXSID2034465	Finished Non-National	Yes	EPA Method in Development 562

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Qualifying Health Assessment	Drinking Water Analytical Method[1]
Dichlorvos	62-73-7	DTXSID5020449	Finished Non-National	Yes	EPA 507, EPA 525.2, EPA 525.3
Dicrotophos	141-66-2	DTXSID9023914	Finished Non-National	Yes	EPA 538
Ethalfluralin	55283-68-6	DTXSID8032386	Finished Non-National	Yes	No
Fenbuconazole	114369-43-6	DTXSID8032548	Finished Non-National	Yes	No
Fipronil	120068-37-3	DTXSID4034609	Finished Non-National	Yes	EPA Method in Development 562
Flufenacet	142459-58-3	DTXSID2032552	Finished Non-National	Yes	EPA Method in Development 562
Fluometuron	2164-17-2	DTXSID8020628	Finished Non-National	Yes	EPA 532
Fluoranthene	206-44-0	DTXSID3024104	Finished Non-National	Yes	EPA 550, EPA 550.1
Imazalil	35554-44-0	DTXSID8024151	Finished Non-National	Yes	No
Iodide	20461-54-5	DTXSID80912339	Finished Non-National	Yes	No
Iprodione	36734-19-7	DTXSID3024154	Finished Non-National	Yes	No
Isophorone	78-59-1	DTXSID8020759	Finished Non-National	Yes	EPA 525.2, EPA 525.3
Malaoxon	1634-78-2	DTXSID9020790	Finished Non-National	Yes	No
Malathion	121-75-5	DTXSID4020791	Finished Non-National	Yes	EPA 527
Methylmercury	22967-92-6	DTXSID9024198	Finished Non-National	Yes	No
Oryzalin	19044-88-3	DTXSID8024238	Finished Non-National	Yes	No

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Qualifying Health Assessment	Drinking Water Analytical Method[1]
Oxadiazon	19666-30-9	DTXSID3024239	Finished Non-National	Yes	No
p-Cresol	106-44-5	DTXSID7021869	Finished Non-National	Yes	No
Phorate	298-02-2	DTXSID4032459	Finished Non-National	Yes	EPA 525.3
Phosmet	732-11-6	DTXSID5024261	Finished Non-National	Yes	No
Propargite	2312-35-8	DTXSID4024276	Finished Non-National	Yes	No
tert-Butyl alcohol	75-65-0	DTXSID8020204	Finished Non-National	Yes	EPA 524.3, EPA 524.4
Thiamethoxam	153719-23-4	DTXSID2034962	Finished Non-National	Yes	EPA Method in Development 562
Tri-allate	2302-17-5	DTXSID5024344	Finished Non-National	Yes	No
Tributyl phosphate	126-73-8	DTXSID3021986	Finished Non-National	Yes	No
Trifluralin	1582-09-8	DTXSID4021395	Finished Non-National	Yes	EPA 551.1, EPA 508, EPA 508.1, EPA 525.2, EPA 525.3
Tris(2-chloroethyl) phosphate	115-96-8	DTXSID5021411	Finished Non-National	Yes	No
Tungsten	7440-33-7	DTXSID8052481	Finished Non-National	Yes	No
C. Contaminants with Qualifying Health Assessment(s) with Derivation of Oral Toxicity Value(s) but Lacking Finished Water Occurrence Data					
Acephate	30560-19-1	DTXSID8023846	Ambient National	Yes	EPA 538
Bensulide	741-58-2	DTXSID9032329	Ambient Non-National	Yes	EPA 540, EPA 543
Chlordecone (Kepone)	143-50-0	DTXSID1020770	Ambient Non-National	Yes	EPA 527, EPA Method in Development 562
Dithiopyr	97886-45-8	DTXSID9032379	Ambient National	Yes	No
Ethylene thiourea	96-45-7	DTXSID5020601	Environmental Release	Yes	No

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Qualifying Health Assessment	Drinking Water Analytical Method[1]
Nitroglycerin	55-63-0	DTXSID1021407	Ambient Non-National	Yes	No
D. Contaminants Lacking Qualifying Health Assessment(s) with Derivation of Oral Toxicity Value(s) (The parentheses in the qualifying health assessment column refer to the best available health effects data)					
4-tert-Octylphenol	140-66-9	DTXSID9022360	Ambient Non-National	No (CCL SL)	EPA 559
Chloromethane	74-87-3	DTXSID0021541	Finished National	No (T 3 data)	EPA 502.2, EPA 524.2, EPA 524.3, EPA 524.4
Nicotine	54-11-5	DTXSID1020930	Finished Non-National	No (CCL SL)	No
Nonylphenol	104-40-5	DTXSID3021857	Finished Non-National	No (CCL SL)	EPA 559
Tolytriazole	29385-43-1	DTXSID0026171	Finished Non-National	No (CCL SL)	No
Triphenyl phosphate	115-86-6	DTXSID1021952	Finished Non-National	No (CCL SL)	No

National = Occurrence data that are nationally representative are available
Non-National = Occurrence data that are not nationally representative are available
Note: Data availability was not assessed for cyanotoxins, DBPs and PFAS.
[1] Published methods are listed by EPA number or voluntary consensus standard bodies (VCSB) number. Methods in development by the EPA, for which research is still needed, are also identified.

The occurrence and health effects data used to categorize data availability can be found on the CISs (USEPA, 2025b). The following sections describe the types of data or information gaps listed in Table 24 and provide examples of contaminants that fall into each group.

5.2.1 Occurrence

Under the regulatory determination process, the occurrence data availability assessment is used to identify contaminants that have sufficient data and information to characterize whether the contaminant is known to occur or there is substantial likelihood for the contaminant to occur in PWSs. However, for the CCL development, the EPA is required to identify contaminants that were known or anticipated to occur in PWSs. The EPA used nationally representative finished water data as the best available occurrence information. However, in the absence of nationally representative finished water data, non-nationally representative finished drinking water occurrence data were also used. The EPA then evaluated additional sources of information such as ambient/source water occurrence, production/use, and environmental release data. To identify data availability, as shown in Table 24, the EPA categorized occurrence data needs as follows:

- Finished water occurrence data that are nationally representative. Data sources may include:
  - UCMRs (i.e., UCMR 1 - 5), the Unregulated Contaminant Monitoring – State (Round 1 and Round 2) and NIRS.

- Finished water occurrence data that are not nationally representative. These data may include:
  - Finished water assessments by federal agencies (e.g., the EPA, the USDA and the USGS). These may include assessments that are geographically distributed across the nation but are not intended to be statistically representative of the nation.
  - State-level finished water monitoring data.
  - Research performed by institutions and universities (e.g., scientific literature), including targeted or local monitoring studies.
- Finished water occurrence data are not available. The best available data sources may include:
  - Ambient/source water data.
  - Environmental release data (such as TRI data or pesticide application data).

5.2.2 Health Effects

Under the regulatory determination process, the EPA generally relies on final peer-reviewed health assessments to determine if and at what level a contaminant may have an adverse effect on the health of persons. Health effects data sources evaluated for the most recent regulatory determination (RD 5) included EPA health assessments, health assessments developed by authoritative sources such as the Agency for Toxic Substances and Disease Registry (ATSDR), World Health Organization (WHO), and Health Canada, as well as assessments developed by other state and global organizations such as the European Food Safety Authority and the California Environmental Protection Agency's (CalEPA) Office of Environmental Health Hazard Assessment (OEHHA). The health assessment must have been publicly available, final, peer-reviewed, and must have used comparable methods, standards, and guidelines to an EPA health assessment.

For the CCL 6, as shown in Table 24, the EPA categorized the health effects data availability in the following way:

- Health effects data are available. A peer-reviewed qualifying health assessment is available or is in the process of being revised.
- Qualifying health effects data currently not available. A peer-reviewed qualifying health assessment is not available, or existing qualifying health assessments do not include the derivation of oral toxicity values. The next best available health effects data is specified.

5.2.3 Analytical Methods

To conduct nationally representative drinking water occurrence studies that could support a regulatory determination, the EPA must have an analytical method suitable for the drinking water matrix and robust enough to be used by many laboratories to conduct national studies and/or compliance monitoring. For CCL 6, the EPA assessed the status of the development of analytical methods for drinking water.

Although many methods for monitoring the CCL 6 chemical contaminants are available from scientific papers and consensus organizations, not all may be appropriate for use in drinking water or for a national monitoring effort. The status of drinking water analytical methods for the CCL chemical contaminants is presented in Table 24. For CCL 6 method availability assessment, the EPA only considered EPA validated drinking water methods.

5.3 Data Availability for PCCL 6 Chemicals not on the Draft CCL 6

To evaluate the type of information researchers or agencies could collect or provide to fill data gaps relevant to CCL listing considerations, the EPA also assessed the data availability of PCCL 6 chemicals not included on the draft CCL 6. The data files for occurrence and health effects were assessed to identify the best available occurrence and health effects data of these chemicals. The occurrence data identified are listed here from the most relevant to drinking water exposure to least relevant:

- Nationally representative finished water monitoring data
- Non-nationally representative finished water monitoring data
- Nationally representative ambient water monitoring data
- Non-nationally representative ambient water monitoring data
- Pesticide application data
- Production and release data

In Table 25, the EPA noted PCCL 6 chemicals not listed on the draft CCL 6 that had a health concentration (HRL or CCL screening level (CCL SL)) derived during the CCL 6 classification process (see Chapter 4). For other chemicals, the EPA provided the screening tier (see Chapter 3) for the best available health effects data. The health effects tiers established during the CCL 6 screening are:

- Tier 1 (T 1) health effects data including RfDs, CSFs, and health-based concentrations such as a chronic benchmark
- Tier 2 (T 2) health effects data including chronic NOAELs and chronic LOAELs
- Tier 3 (T 3) health effects data including cancer classifications, subchronic RfDs, and subchronic health-based concentrations
- Tier 4 (T 4) health effects data including acute health-based concentrations, acute RfDs, subchronic LOAELs, subchronic NOAELs, MRDDs, or a chemical is present on a list of known neurotoxicants
- Tier 5 (T 5) health effects data including TD_{50}, LD_{50}, ToxCast assay percent active, androgen or estrogen bioactivity, and number of PubMed articles

Additionally, chemicals are categorized into three occurrence groups:

- Group A contaminants have nationally representative finished water data.
- Group B contaminants have non-nationally representative finished water data.
- Group C contaminants lack any finished water data.

The occurrence and health effects data listed in Table 25 are the best available data for that chemical contaminant at the time of the draft CCL 6 development.

Table 25. Data Availability for PCCL 6 Chemicals not on the Draft CCL 6

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Best Available Health Effects Data
Group A. Contaminants with Nationally Representative Finished Water Data				
1,1,2,2-Tetrachloroethane	79-34-5	DTXSID7021318	Finished National	HRL
1,1-Dichloroethane	75-34-3	DTXSID1020437	Finished National	HRL
1,3-Butadiene	106-99-0	DTXSID3020203	Finished National	T 1
1,3-Dichloropropene	542-75-6	DTXSID1022057	Finished National	HRL
2,4-Dichlorophenol	120-83-2	DTXSID1020439	Finished National	HRL
2,4-Dinitrotoluene	121-14-2	DTXSID0020529	Finished National	HRL
Acetochlor	34256-82-1	DTXSID8023848	Finished National	HRL
Acetochlor ESA	187022-11-3	DTXSID6037483	Finished National	CCL SL
Acetochlor OA	194992-44-4	DTXSID1037484	Finished National	CCL SL
Alachlor ESA	142363-53-9	DTXSID6037485	Finished National	CCL SL
Alachlor OA	171262-17-2	DTXSID1037486	Finished National	CCL SL
Aldrin	309-00-2	DTXSID8020040	Finished National	HRL
Boron	7440-42-8	DTXSID3023922	Finished National	HRL
Bromochloromethane	74-97-5	DTXSID4021503	Finished National	HRL
Calcium	7440-70-2	DTXSID9050484	Finished National	T 5
Chlorodifluoromethane	75-45-6	DTXSID6020301	Finished National	T 3
Dieldrin	60-57-1	DTXSID9020453	Finished National	HRL
Dimethoate	60-51-5	DTXSID7020479	Finished National	HRL
EPTC	759-94-4	DTXSID1024091	Finished National	HRL
Fonofos	944-22-9	DTXSID2024113	Finished National	HRL
Iron	7439-89-6	DTXSID5043710	Finished National	HRL

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Best Available Health Effects Data
Linuron	330-55-2	DTXSID2024163	Finished National	HRL
Magnesium	7439-95-4	DTXSID0049658	Finished National	HRL
Methyl bromide	74-83-9	DTXSID8020832	Finished National	HRL
Metolachlor	51218-45-2	DTXSID4022448	Finished National	HRL
Metolachlor ESA	171118-09-5	DTXSID1037567	Finished National	CCL SL
Metolachlor OA	152019-73-3	DTXSID6037568	Finished National	CCL SL
Metribuzin	21087-64-9	DTXSID6024204	Finished National	HRL
Naphthalene	91-20-3	DTXSID8020913	Finished National	HRL
Nitrobenzene	98-95-3	DTXSID3020964	Finished National	HRL
Phosphorus	7723-14-0	DTXSID1024382	Finished National	T 4
Potassium	7440-09-7	DTXSID9049748	Finished National	T 4
Prometon	1610-18-0	DTXSID6022341	Finished National	HRL
RDX (Hexahydro-1,3,5-trinitro-1,3,5-triazine)	121-82-4	DTXSID9024142	Finished National	HRL
Silicon	7440-21-3	DTXSID0051441	Finished National	T 4
Sodium	7440-23-5	DTXSID1049774	Finished National	HRL
Terbacil	5902-51-2	DTXSID8024317	Finished National	HRL
Group B. Contaminants with Non-Nationally Representative Finished Water Data				
2,4-Dichlorophenoxybutyric acid	94-82-6	DTXSID7024035	Finished Non-National	HRL
2-Chloro-4,6-diamino-1,3,5-triazine	3397-62-4	DTXSID1037806	Finished Non-National	HRL
2-Hydroxyatrazine	2163-68-0	DTXSID6037807	Finished Non-National	HRL
2-Methylnaphthalene	91-57-6	DTXSID4020878	Finished Non-National	HRL
5-Methyl-1H-benzotriazole	136-85-6	DTXSID1038743	Finished Non-National	CCL SL
6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	21145-77-7	DTXSID7041544	Finished Non-National	CCL SL
Acetamiprid	135410-20-7	DTXSID0034300	Finished Non-National	HRL

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Best Available Health Effects Data
Acetone	67-64-1	DTXSID8021482	Finished Non-National	HRL
Ametryn	834-12-8	DTXSID1023869	Finished Non-National	HRL
Ammonia	7664-41-7	DTXSID0023872	Finished Non-National	HRL
Argon	7440-37-1	DTXSID3052482	Finished Non-National	T 5
Azoxystrobin	131860-33-8	DTXSID0032520	Finished Non-National	HRL
Benfluralin	1861-40-1	DTXSID3023899	Finished Non-National	HRL
Bentazone	25057-89-0	DTXSID0023901	Finished Non-National	HRL
Benzophenone	119-61-9	DTXSID0021961	Finished Non-National	HRL
Bicarbonate	71-52-3	DTXSID2049921	Finished Non-National	T 5
Bifenthrin	82657-04-3	DTXSID9020160	Finished Non-National	HRL
Boscalid	188425-85-6	DTXSID6034392	Finished Non-National	HRL
Bromacil	314-40-9	DTXSID4022020	Finished Non-National	HRL
Bromide	24959-67-9	DTXSID6043967	Finished Non-National	HRL
Caffeine	58-08-2	DTXSID0020232	Finished Non-National	CCL SL
Carbon	7440-44-0	DTXSID9027651	Finished Non-National	T 5
Carbon dioxide	124-38-9	DTXSID4027028	Finished Non-National	T 4
Carbon disulfide	75-15-0	DTXSID6023947	Finished Non-National	HRL
Chlorimuron-ethyl	90982-32-4	DTXSID0023955	Finished Non-National	HRL
Chlorsulfuron	64902-72-3	DTXSID7023980	Finished Non-National	HRL
Clopyralid	1702-17-6	DTXSID9029221	Finished Non-National	HRL
Coumaphos	56-72-4	DTXSID2020347	Finished Non-National	HRL
Cyfluthrin	68359-37-5	DTXSID5035957	Finished Non-National	HRL
Cypermethrin	52315-07-8	DTXSID1023998	Finished Non-National	HRL
Cyproconazole	94361-06-5	DTXSID0032601	Finished Non-National	HRL

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Best Available Health Effects Data
DEET	134-62-3	DTXSID2021995	Finished Non-National	HRL
Desethylatrazine	6190-65-4	DTXSID5037494	Finished Non-National	HRL
Deisopropylatrazine	1007-28-9	DTXSID0037495	Finished Non-National	HRL
Deltamethrin	52918-63-5	DTXSID8020381	Finished Non-National	HRL
Dibutyl phthalate	84-74-2	DTXSID2021781	Finished Non-National	HRL
Dichlobenil	1194-65-6	DTXSID5032365	Finished Non-National	HRL
Diethyl phthalate	84-66-2	DTXSID7021780	Finished Non-National	HRL
Difenoconazole	119446-68-3	DTXSID4032372	Finished Non-National	HRL
Dimethenamid	87674-68-8	DTXSID4032376	Finished Non-National	HRL
Dinotefuran	165252-70-0	DTXSID7034549	Finished Non-National	HRL
Endosulfan sulfate	1031-07-8	DTXSID3037541	Finished Non-National	HRL
Epoxiconazole	133855-98-8	DTXSID1040372	Finished Non-National	HRL
Esfenvalerate	66230-04-4	DTXSID4032667	Finished Non-National	HRL
Ethion	563-12-2	DTXSID2024086	Finished Non-National	HRL
Ethofumesate	26225-79-6	DTXSID8034580	Finished Non-National	HRL
Fenitrothion	122-14-5	DTXSID4032613	Finished Non-National	HRL
Fenpropathrin	39515-41-8	DTXSID0024002	Finished Non-National	HRL
Fludioxonil	131341-86-1	DTXSID2032398	Finished Non-National	HRL
Flumetsulam	98967-40-9	DTXSID4032615	Finished Non-National	HRL
Galaxolide	1222-05-5	DTXSID8027373	Finished Non-National	CCL SL
Hexazinone	51235-04-2	DTXSID4024145	Finished Non-National	HRL
Imazaquin	81335-37-7	DTXSID3024152	Finished Non-National	HRL
Imazethapyr	81335-77-5	DTXSID3024287	Finished Non-National	HRL
Imidacloprid	138261-41-3	DTXSID5032442	Finished Non-National	HRL

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Best Available Health Effects Data
lambda-Cyhalothrin	91465-08-6	DTXSID7032559	Finished Non-National	HRL
MCPA	94-74-6	DTXSID4024195	Finished Non-National	HRL
Mecoprop	93-65-2	DTXSID9024194	Finished Non-National	HRL
Metalaxyl	57837-19-1	DTXSID6024175	Finished Non-National	HRL
Methane	74-82-8	DTXSID8025545	Finished Non-National	T 4
Metsulfuron-methyl	74223-64-6	DTXSID6023864	Finished Non-National	HRL
Mirex	2385-85-5	DTXSID7020895	Finished Non-National	HRL
Myclobutanil	88671-89-0	DTXSID8024315	Finished Non-National	HRL
Nicosulfuron	111991-09-4	DTXSID6034764	Finished Non-National	HRL
Norflurazon	27314-13-2	DTXSID8024234	Finished Non-National	HRL
Pendimethalin	40487-42-1	DTXSID7024245	Finished Non-National	HRL
Pentachloronitrobenzene	82-68-8	DTXSID2021105	Finished Non-National	HRL
Phenanthrene	85-01-8	DTXSID6024254	Finished Non-National	T 3
Phenol	108-95-2	DTXSID5021124	Finished Non-National	HRL
Phosphate/Orthophosphate	14265-44-2	DTXSID7039672	Finished Non-National	HRL
Piperonyl butoxide	51-03-6	DTXSID1021166	Finished Non-National	HRL
Prometryn	7287-19-6	DTXSID4024272	Finished Non-National	HRL
Propanil	709-98-8	DTXSID8022111	Finished Non-National	HRL
Propetamphos	31218-83-4	DTXSID7032470	Finished Non-National	HRL
Propiconazole	60207-90-1	DTXSID8024280	Finished Non-National	HRL
Propoxur	114-26-1	DTXSID7021948	Finished Non-National	HRL
Propyzamide	23950-58-5	DTXSID2020420	Finished Non-National	HRL
Prosulfuron	94125-34-5	DTXSID9034868	Finished Non-National	HRL
Pyrene	129-00-0	DTXSID3024289	Finished Non-National	HRL

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Best Available Health Effects Data
Silica	7631-86-9	DTXSID1029677	Finished Non-National	T 3
TDCPP	13674-87-8	DTXSID9026261	Finished Non-National	HRL
Tebupirimfos	96182-53-5	DTXSID1032482	Finished Non-National	HRL
Tebuthiuron	34014-18-1	DTXSID3024316	Finished Non-National	HRL
Terbuthylazine	5915-41-3	DTXSID4027608	Finished Non-National	HRL
Tetraconazole	112281-77-3	DTXSID8034956	Finished Non-National	HRL
Thiabendazole	148-79-8	DTXSID0021337	Finished Non-National	HRL
Thiobencarb	28249-77-6	DTXSID6024337	Finished Non-National	HRL
Triadimefon	43121-43-3	DTXSID3023897	Finished Non-National	HRL
Triadimenol	55219-65-3	DTXSID0032493	Finished Non-National	HRL
Triclopyr	55335-06-3	DTXSID0032497	Finished Non-National	HRL
Triclosan	3380-34-5	DTXSID5032498	Finished Non-National	HRL
Tris(2-butoxylethyl) phosphate	78-51-3	DTXSID5021758	Finished Non-National	HRL
Group C. Contaminants Lacking Finished Water Data				
Cyhalothrin	68085-85-8	DTXSID6023997	Ambient National	HRL
Fluridone	59756-60-4	DTXSID8024107	Ambient National	HRL
Hydramethylnon	67485-29-4	DTXSID6023868	Ambient Non-National	HRL
Indoxacarb	173584-44-6	DTXSID1032690	Ambient National	HRL
Pyraclostrobin	175013-18-0	DTXSID7032638	Ambient National	HRL
Sulfentrazone	122836-35-5	DTXSID6032645	Ambient National	HRL
Thiram	137-26-8	DTXSID5021332	Ambient Non-National	HRL
Trifloxystrobin	141517-21-7	DTXSID4032580	Ambient National	HRL
Triticonazole	131983-72-7	DTXSID0032655	Ambient National	HRL

Chapter 6 Data Management and Quality Assurance

6.1 Overview

All steps of the CCL 6 development process underwent quality assurance/quality control (QA/QC) activities to ensure the integrity of the data and calculations used to generate the draft CCL 6. The process consisted of two phases: QA/QC of the PCCL 6 development (Section 6.2) and QA/QC of the CIS development (Section 6.3). The QA/QC activities generally fell into review of five categories: input data, output data, code, DTXSID assignments, and CISs.

The CCL 6 Universe file, the screening code, and the CISs were developed primarily using the R or supplemental data sources, as well as the program and code developed to generate CISs, was subject to at least one review. In addition, the screening code was independently reviewed. Code developed and QA'd during CCL 5 was used when possible (with updates incorporated as needed). As part of building the CCL 6 Universe, as explained in Sections 2.3 and 2.4, the EPA conducted input checks, such as verifying that the original source data matched the data contained in the simple files used to build the Pre-Universe and Universe files. To check the accuracy of the screening code and ensure screening points were assigned correctly, the EPA also conducted output checks. This entailed review of screening point assignments for a selection of 8 chemicals, which together represent all data elements involved in screening, to confirm the expected screening scores. See Section 3.3 for details on the screening point assignments. The CISs underwent rounds of QA/QC in which data values on the CISs were spot-checked against the data in the CIS input files. Further details about QA/QC of the PCCL 6 development and CISs are described in the following sections.

6.2 Quality Assurance of PCCL 6 Development

6.2.1 Overview

For the PCCL development process, the EPA wrote code using R programming language (version 4.0.2) (R Core Team, 2025) and documented it using R Markdown files (version 4.4.2) (Allaire, et al., 2020). This allowed for transparent documentation and organization of the PCCL development process and QA/QC activities. The EPA developed R Markdown files, which documented the PCCL 6 process, including the following:

- A set of individual R Markdown files used to import and process data from the CCL 6 main data sources (referred to as "pre-processing code" hereafter).
  - The pre-processing code extracted and transformed relevant data from the main data sources to a simple data format. Details on the simple data format are described in Section 2.3.4 and Appendix M.
- Three R Markdown files were used to develop the PCCL 6 from the resulting simple data files provided by the pre-processing code.
  - The first file, "Simple File Combine," aggregated the simple data files produced from the pre-processing code into the Pre-Universe file described in Section 2.3.
  - The second file, "ID and Screen," standardized units, performed chemical groupings as necessary, and added supplemental data from the CompTox Chemicals Dashboard (Williams, et al., 2017). The output of the second R Markdown code was the Universe file described in Section 2.4.1.

- The third file, "Prioritize" (referred to as screening code hereafter), assigned screening points according to the screening point assignment hierarchy described in Chapter 3, and calculated the screening score for each compound in the Universe. The output of this code was the Scored Universe file.

The following sections describe the QA/QC activities conducted during the PCCL development process.

6.2.2 QA/QC for the Pre-Universe Input Data Files

The first QA/QC activity was reviewing the input data used to build the Pre-Universe. The EPA used R code to pull data elements from main data sources into the simple file format described in Appendix M.3. This QA/QC activity involved verifying the data extracted from the main data sources and ultimately used to build the Pre-Universe. For each main data source, the EPA accessed the original data while reviewing the simple file output. The EPA verified that all the analytes and correct data elements were extracted, and then randomly selected 10-15 analytes to check any CASRN identifiers and any calculations or data manipulations that were performed. When issues were identified, R code was updated and rerun, and the simple files were rechecked to ensure the issue had been addressed. For example, an issue that was identified was the unnecessary rounding of values. For each data source, there was a primary reviewer who completed the QA/QC checks and a secondary reviewer who reviewed the results of the QA/QC check and confirmed identified issues. More details about the screening pre-processing steps for each data source are described in Appendix M.2. The QA/QC of the simple files allowed the EPA to check for any edits that needed to be made to the R code that created the simple files and to check that the data used to build the universe is accurate and complete. These simple files were used as the input for building the Pre-Universe.

6.2.3 QA/QC for the Pre-Universe File

The second QA/QC activity involved cleaning up the Pre-Universe file prior to conducting the curation work for DTXSID assignments. The EPA reviewed the Pre-Universe to identify any entries that could not be identified or classified as unique chemicals. The Pre-Universe file entries excluded at this step generally fit under the following categories:

- Ambiguous names
- Exposure routes
- Manufacturing processes
- Mixtures of two of more chemicals
- Viruses and bacteria
- Water quality parameters

The cleanup work involved an iterative application of key search terms to filter, identify, and remove non-chemical entries. The resulting outcome of this work was the removal of approximately 3,000 "junk" entries across several of the main data sources compiled into the Pre-Universe file.

6.2.4 QA/QC for DTXSID Assignments

The third QA/QC activity of PCCL 6 development was a review of DTXSID assignments to data entries in the cleaned Pre-Universe file described in Section 6.2.3. The EPA used an iterative process to determine the optimal approach for assigning DTXSIDs to data entries and screen out contaminants not of interest to CCL (i.e., data entries not associated with chemical substances or cannot be confidently identified as a single chemical). The purpose of this QA activity was to ensure the following:

- Correct DTXSID had been assigned
- Data entry was not describing a mixture of substances
- Data entry was describing a chemical substance rather than a microbe or physical characteristic
- Data entry was clear as to the specific chemical substance measured

CCL development staff worked closely with the EPA's Office of Research and Development (ORD) staff managing the Computational Toxicology (CompTox) Chemicals Dashboard (https://comptox.epa.gov/dashboard) to help efficiently assign missing DTXSIDs and correct any misassigned DTXSIDs originating from the simple files used to generate the Pre-Universe file. The EPA changed the method of assigning DTXSIDs from using the mapping file used during the CCL 5 process, to using a more efficient batch search function in the CompTox Chemicals Dashboard. The CompTox Chemicals Dashboard is continuously being updated and improved. However, the mapping file used during CCL 5 was static (i.e., not updated over the course of the PCCL development process) and could not reflect subsequent updates to the CompTox Chemicals Dashboard. The EPA determined that a more efficient approach for CCL 6 would be to amend downloaded data containing chemicals names and CASRNs (when provided in the original sources) with DTXSIDs downloaded from the CompTox Chemicals Dashboard. Any entry that could not be assigned a DTXSID using this process was temporarily designated "NA" for its unique identifier key field in the Pre-Universe file. Data entries with "NA" were subsequently reviewed on an individual basis for follow-up curation. If no DTXSID could be assigned after this additional manual step, those entries were removed from the finalized Pre-Universe file (as discussed earlier in Section 2.4.2).

As a result of QA checks from both CCL staff and CompTox staff, the EPA determined that the method used to remove data entries from the Pre-Universe was uniformly applied.

Using a batch search function to assign DTXSIDs based on provided names and CASRN identifiers from the data sources, nearly all the entries in the Pre-Universe file were correctly mapped; the resulting Pre-Universe file had 130k rows of data, relevant to approximately 25k chemicals that ultimately constituted the CCL 6 Universe of chemicals. Roughly seven hundred fifty entries still needed manual curation performed by CompTox staff. The EPA identified 185 data entries with 89 unique labels that could not be matched to an existing DTXSID in the CompTox Chemicals Dashboard. In these cases, the CompTox curators noted the reason for these omissions, including ambiguous or unregistered names, classes of chemicals, radiation, and biological entries like spores and moths. The EPA ultimately excluded these 185 entries before finalizing the Pre-Universe file.

6.2.5 QA/QC for the Universe File

The fourth QA/QC activity included reviewing the final Universe file to verify that no errors had been introduced when aggregating the simple files to create the Universe file. The EPA randomly sampled 400 data entries from the Pre-Universe file and checked them against the simple files. The EPA ensured all main data sources were represented in the input data review. The goal of the input data review was to ensure that the error rate in the input data was less than 1%. The null hypothesis of this scenario was that the error rate is 1% or greater. The EPA assessed the error rate using the beta distribution in R by checking N entries and finding K of them defective (errors), which would lead to calculation of the estimated error rate as K/N. An upper 95% confidence interval for the error rate estimate was calculated using the beta distribution: qbeta (0.95, K +0.5, N-K+0.5), where 0.5 are shape parameters for the beta distribution. With 400 data entries (N = 400) reviewed, if zero or one errors were found, the error rate would be less than 1% and the null hypothesis could be rejected. If more than one error was found, the error rate would be above the threshold of 1% or less error.

Upon completing the review of the 400 random samples of input data from the Universe file, the EPA did not identify any errors when comparing data entries to the input simple files. This confirmed that the process used to develop the Universe file from the simple files likely did not introduce new errors and indicated the targeted error rate of < 1% was successfully achieved.

The QA/QC of the 400 entries from the Universe file did identify seven contaminants with DTXSID assignment errors. Subsequently, the DTXSID assignments for these chemicals were addressed and an additional 200 data entries from a cleaned Universe file were randomly sampled, including entries from each data source, and the same QA/QC of checking entries against the simple files was conducted. No errors were identified during the review of the 200 data entries.

6.2.6 QA/QC for the Screening Code

The fifth QA/QC activity was a rigorous review of all R code and R Markdown files written for the PCCL development process. The screening code review was conducted by EPA staff who were not the primary code developers. Generally, the review consisted of checking the following:

- If the code achieved its intended goal
- For coding errors
- For correct transformations of the original data
- If calculations followed best statistical practices
- Overall code structure and style

Each file described in Section 6.2.1 was inspected line by line for functionality and observations. The “ID and Screen” code was flagged for modifications and removals, as well as QA checks. Example QA checks include confirming the chemical groupings, DTXSID corrections, and conversion of data source concentration units. The “Prioritize” code, which generates the scored universe was flagged for modifications and QA checks. Example modifications include file paths, titles, and other general edits, as well as updating code to import new health source data such as androgen and estrogen receptor data and organic carbon-water partition coefficient (Koc) from the CompTox Dashboard. Code was modified to remove chemicals with existing NPDWRs and to update with the public nominations received for CCL 6. Example QA checks include confirming input files exist and are properly formatted, using RStudio commands like *levels()* and *mutate()* within *(TRUE)* loops to check for style violations, variable counts, unused variables, and similar issues. An input file for the screening code that contains the list of chemicals with NPDWRs to be removed from the CCL process was found to incorrectly contain chemicals with Secondary MCLs. These chemicals were manually added back onto the PCCL 6.
The reviewers suggested several improvements to the coding style and efficiency of the scripts to incorporate CCL 6 process changes while generating a comparable Pre-universe and Universe to CCL 5**.** Subsequently, the EPA incorporated improvements to the respective R Markdown files for CCL 6.

6.2.7 QA/QC for the Scored Universe File

The sixth QA activity was a review of the output of the screening code, the Scored Universe. The EPA reviewed data entries for eight chemicals on the PCCL 6 that were identified as being data rich based on the number of unique data sources and data elements in the CCL 6 Universe for the chemical. Data entries in the Scored Universe were compared against the updated points assignment rubric (Table 7 and Table 8) to ensure chemicals were being scored as expected. This QA/QC review identified errors like the EDSP androgen and estrogen receptor data not being scored. Corrections were made to the code, and the screening code was rerun. This QA/QC activity helped to ensure that the scores used to determine the chemicals on the PCCL were accurate and complete.

6.2.8 QA/QC for the Composition of the PCCL 6

The final QA/QC activity for the development of the PCCL was a review of the output of the screening code that contained the 250 highest scoring contaminants. The composition of the PCCL 6 was evaluated in comparison to the PCCL 5. This included analyzing the chemicals that were added and removed compared to the PCCL 5. Chemicals with large points changes compared to CCL 5 were also closely reviewed. Chemicals from CCL 5 that did not make the cut off for CCL 6 and chemicals that gained more than 1000 points or lost more than 500 points were assessed by determining where the distribution in health effects and occurrence-based points changed from CCL 5 to CCL 6. Using the CCL 6 Scored Universe and CCL 5 Scored Universe, the data elements for the chemical with changes in points were identified and reviewed to ensure that the changes were accurate based on the CCL 6 data. This QA/QC activity was the final check of the composition of the PCCL 6.

6.3 Quality Assurance of CIS Development

6.3.1 Overview

This section describes the data management and QA/QC activities used to produce the CISs. As described in Chapter 4, to generate CCL 6 the EPA identified and gathered data on the health effects and occurrence of each of the PCCL 6 chemicals evaluated, then summarized this information on the CISs. This section also describes the EPA's procedures to compile and structure occurrence and health effects data, which use a variety of data sources to generate the CISs. The following sections describe the data management and QA/QC activities for each of these efforts in greater detail.

6.3.2 Preparing Health Effects Data for CISs

The EPA prepared data collected from supplemental data sources (i.e., qualifying and non-qualifying health assessments and rapid systematic review (RSR) results) for the CISs in Excel workbooks which were formatted onto the CISs using R scripts. The file containing data extracted from qualifying and non-qualifying health assessments was redesigned for CCL 6, based on experience from CCL 5, to be more compatible with formatting via R. The update simplified the structure of the file and eliminated the need for several interim data cleaning steps and is presented in Appendix F. The data file reporting results from the RSR used the same structure as was developed for CCL 5.

The EPA compiled health assessment data on all PCCL 6 contaminants in an Excel workbook, including data on the RfDs, CSFs, critical effects, target population, potency score, severity category, and health concentration. When possible, if the data remain up to date from previous EPA extraction efforts, data were copied and pasted from files developed for CCL 5 and Regulatory Determination 5, which had already undergone extensive QA/QC processes (USEPA, 2022b); (USEPA, 2025). For newly identified assessments, the EPA assigned one individual to conduct the primary extraction of information from that assessment. A second individual confirmed that the information was extracted correctly and reviewed the calculated and assigned data elements (i.e., target population, drinking water intake rate, health concentration, potency, and severity) to ensure that the selections or calculations followed the policies and methods outlined in Appendix F. If there were any conflicts between the two individuals, the issue was discussed, and a consensus was reached by a broader group of 4-5 individuals. Once QA/QC activities were complete, the file was used to populate information on the Summary + Decision and Health Effects tabs of the CIS. Once populated onto the CIS, the EPA reviewed each cell and confirmed that the R code correctly formatted the data as intended.

The EPA initiated the development of the RSR results data file for CCL 6 by starting from the file developed under CCL 5 that had previously undergone QA/QC. For new data collected under CCL 6, data were compiled using data extraction source files produced during RSR data extraction and underwent QC as described in Appendix E. QA/QC of the file was performed by checking data points against the original data extraction file by two reviewers. The data collected under CCL 5 and CCL 6 were reviewed by two reviewers and updated for consistency (e.g., orientation of results layout, use of "NOAEL/LOAEL" to indicate RfDs instead of "TRUE/FALSE", data columns order, and figure legends).

The data used to populate the "Other Health Data" section of the Health Effects tab of the CISs came directly from the Universe file. QA/QC was not conducted on the data itself as this was completed when building the Universe. However, the EPA did conduct QA/QC to ensure that the data were correctly displayed on the CIS. This included converting the data element names used for scoring back to the data element names as they are commonly known and presented by the data source. For example, in the CCL 6 Universe, the EPA has data from ATSDR reported as RfDs. The conventional terminology used by ATSDR for an RfD is a "chronic oral minimal risk level," which was the data element name used on the CISs. Additionally, several data elements, such as LD_{50}s, TD_{50}s, and others had several values in the Universe file from an individual source. As was done for CCL 5, the EPA presented these types of health effect data as ranges (i.e., min value – max value) to reduce the space needed on the CIS. For some contaminants, this resulted in the reduction of dozens of rows on the Health Effects tab. Scripts were updated to automate the printing of the words minimum value and maximum value on the CISs; however, the selection of the minimum and maximum value for each data element from each data source was manual. This assignment was reviewed during QA/QC of the populated CISs.

6.3.3 Preparing Occurrence Data for CISs

The EPA prepared occurrence data for the CISs which were formatted onto the CISs using R scripts. To extract and compile water occurrence data in support of the classification step (Chapter 4), the EPA wrote code using R Studio, referred to as R scripts, which was documented in R Markdown. The EPA extracted occurrence-related data elements, such as detection and concentration statistics, for CIS development, occurrence attribute scoring, and fHQ calculations. This section describes the procedures the EPA undertook to extract and gather occurrence data from main data sources and supplemental occurrence data sources. This section also describes the QA/QC activities implemented during this process. See Appendix M for specific data processing information for data sources.

The EPA developed a series of R scripts to extract and transform classification relevant data elements from occurrence main data sources. These data elements included the minimum, median and 90th percentile concentration of detections, the total number of sites/samples and the number of sites/samples with detections. The R scripts were documented in R Markdown (Allaire, et al., 2020). The outputs from the R scripts are a series of files containing classification relevant data elements in the simple format. Details on the simple data format and classification pre-processing steps are described in Appendix M. Source and contaminant metadata such as water type and monitoring year ranges, were also included where possible. For sources that were used in CCL 5 and did not have updated data, the EPA relied on the QA/QC completed during CCL 5. The data files from CCL 5 were combined with the data newly processed for CCL 6.

Two EPA staff completed QA/QC of the classification simple file outputs. For each data source, three analytes were randomly selected and the minimum, median and 90th percentile concentrations, total site count and number of sites with detections were manually checked against the original data source. If any errors were identified, the R code was updated and rerun. The simple files were then rechecked to ensure the error had been addressed. For NAWQA and NWIS, the original data file was too large to easily access the data for individual contaminants, so the QA/QC involved checking the functionality of the code.

Once QA/QC of the classification simple files was completed, the data were combined with the maximum concentration and percent detection values from the Universe file and with the classification data entries from static main sources that had undergone QA/QC during CCL 5. Duplicate entries (entries that had the same source and data type) were flagged and inspected manually for inclusion in the CISs. The combined data created the occurrence input file used to populate data onto the CISs.

DTXSIDs were added to the classification data entries using the DTXSID Lookup Key. An inline code QA checked that every line of data had a DTXSID; any contaminant names in the data that were missing in the Contaminant DTXSID Lookup Key were flagged in this script and added to the Contaminant DTXSID Lookup Key. The Contaminant DTXSID Lookup Key is an Excel file containing all contaminant names (e.g., synonyms) in all sources matched to their DTXSIDs, CASRNs, and preferred names, thereby standardizing contaminant names and IDs. DTXSID and CASRN for contaminants were obtained using the batch search function in the EPA's CompTox Chemistry Dashboard.

After all occurrence data from main sources and supplemental sources were compiled in the simple file format, the EPA used this information to calculate the occurrence attribute scores (prevalence and magnitude) and fHQs (see Chapter 4). The EPA compiled occurrence attribute scoring information and fHQs in Excel workbooks used as input files for the CISs. Prevalence and magnitude attribute scores were manually calculated separately by two EPA staff. Any differences in attribute scores were reviewed to identify the correct value to present on the CIS. The EPA documented and corrected all errors identified during the QA/QC of occurrence attribute scores and fHQ values.

CIS development was performed using R scripts.[6] The EPA spot-checked the scripts that read in the data, checked that formatting and data processing were consistent, and checked that there were no duplicates or contradictory data.

Other relevant summary data (e.g., previous CCL listing decisions, previous Regulatory Determination decisions) were also converted to simple format using an R script where necessary. QA/QC was performed by checking data points from the original files against the data produced in the simple format. Special attention was paid to dates and any errors that could be introduced by changes in source data column name changes.

6.3.4 Data Management and QA/QC of CISs

Once the occurrence and health effects data had passed through the QA process and were prepared in the simple format, the next step was to generate the CISs. Several supporting lookup keys were updated to assist with formatting all CIS data onto the Excel workbooks. An R script pulled in all the re-formatted data source files and generated CISs. Using the "openxlsx" package (Schauberger & Walker, 2025) in R, CISs were created in Excel workbooks for each PCCL 6 chemical to be reviewed by the CCL 6 Evaluation teams.

Once all data were pasted into their respective locations in the CIS Excel workbooks, formatting was applied. Formatting styles were created in the R script using the "openxlsx" package then applied according to an index file, which assigns a row and column location and cell formatting type to all data to be added into the CIS. Column widths and Excel theme were copied from the blank CIS template (see CIS Technical Support Document (USEPA, 2026b) accessible via the EPA docket (Docket ID No. EPA-HQ-OW-2022-0946), and row heights and cell borders were added.

The CISs then underwent QA/QC. The QA/QC reviewers checked the data on the CIS against the data inputs in the R script. Formatting was also reviewed visually to check for any errors, including if an Excel cell size was too small for data, that all sections that were expected to have data from the input files had data, or whether the highlighting matched the scoring data correctly. If a chemical had previously been given a negative regulatory determination, the rationale for this determination was added to the CIS. The rationales were reviewed by multiple reviewers prior to being added to the CIS. The reviewers also conducted calculation cross-checks, looked for errors in rounding or significant figures, checked DTXSID hyperlinks, verified unit conversions, and scanned for missing data. These checks were critical in ensuring important measures such as the attribute scores and fHQ were accurate before undergoing review by the two evaluation teams in the classification step. For example, it was discovered that the fHQ for the chemical 1,2,4-triazole was not showing up on the CIS and was manually added. Once the QA/QC of CISs for each chemical had been performed, CISs were ready to be reviewed by the chemical evaluators during the classification step (Chapter 4).

---

[6] CIS development was conducted using R (mainly 4.4.2 - the version varies as software updates were released). The primary script was written in R Markdown to aid in documentation and organization of the scripts, with the exception of a small supporting script that was sourced at the beginning of each of the other scripts to set file directories and to load packages, or to produce intermediate data files. R script version control was maintained through a repository on GitLab, a web-based software development and IT operations lifecycle tool.

CISs were regenerated after evaluation meetings to include information not available at the time of the ongoing evaluation meetings. This information included list/not list recommendations subject to change based on post-evaluation refinement, updated information about rationales for negative regulatory determinations, as well as updated references with assigned numbers to be inclusive of all contaminants. QA/QC of these regenerated CISs after the evaluation meetings consisted of a focused check of the above items as well as all additional content presented in the CISs as well as fixing any errors, such as inappropriate number of significant figures. If any errors or CIS updates compared to the evaluator-evaluated CISs were discovered that have any potential to impact the initial listing recommendations, the EPA reconnected with the evaluators to evaluate if their listing recommendations would change. No changes in listing recommendations were made from this procedure.

Chapter 7 References

Allaire, J. J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., . . . Iannone, R. (2020). *rmarkdown: Dynamic Documents for R*. Retrieved from R package version 2.5: https://github.com/rstudio/rmarkdown

Aschner, M., Ceccatelli, S., Daneshian, M., Fritsche, E., Hasiwa, N., Hartung, T., . . . Lein, P. J. (2017). Reference compounds for alternative test methods to indicate developmental neurotoxicity (DNT) potential of chemicals: example lists and criteria for their selection and use. *ALTEX - Alternatives to animal experimentation, 34*(1), 49-74. doi:10.14573/altex.1604201

Batt, A. L., Kincaid, T. M., Kostich, M. S., Lazorchak, J. M., & Olsen, A. R. (2016). Evaluating the extent of pharmaceuticals in surface waters of the United States using a National-scale Rivers and Streams Assessment survey. *Environmental toxicology and chemistry, 35*(4), 874-881. doi:https://doi.org/10.1002/etc.3161

Battaglin, W. A., Bradley, P. M., Iwanowicz, L., Journey, C. A., Walsh, H. L., & Blazer, V. S. (2018). Pharmaceuticals, hormones, pesticides, and other bioactive contaminants in water, sediment, and tissue from Rocky Mountain National Park, 2012–2013. *Science of the Total Environment, 643*, 651-673.

Bove, G. E., Rogerson, P. A., & Vena, J. E. (2010). Case-Control Study of the Effects of Trihalomethanes on Urinary Bladder Cancer Risk. *Archives of Environmental & Occupational Health, 62*(1), 39-47. doi:10.3200/AEOH.62.1.39-47

Bradley, P. M., Journey, C. A., Romanok, K. M., Barber, L. B., Buxton, H. T., Foreman, W. T., . . . Villeneuve, D. L. (2017). Expanded Target-Chemical Analysis Reveals Extensive Mixed-Organic-Contaminant Exposure in U.S. Streams. *Environmental science & technology, 51*(9), 4792-4802. doi:10.1021/acs.est.7b00012

Bradley, P. M., Kolpin, D. W., Romanok, K. R., Smalling, K. L., Focazio, M. J., Brown, J. B., . . . Wilson, V. S. (2018). Reconnaissance of Mixed Organic and Inorganic Chemicals in Private and Public Supply Tapwaters at Selected Residential and Workplace Sites in the United States. *Environmental Science & Technology, 52*(23), 13972-13985. doi:10.1021/acs.est.8b04622

CalEPA. (n.d.). *Office of Environmental Health Hazard Assessment (OEHHA)*. Retrieved May 2019, from Chemicals: https://oehha.ca.gov/library/chemicals

CDC. (n.d.). *Agency for Toxic Substances and Disease Registry (ATSDR)*. Retrieved April 2018, from Minimal Risk Levels (MRLs) for Hazardous Substances: https://www.cdc.gov/TSP/MRLS/mrlsListing.aspx

CDPR. (n.d.). *Surface Water Database (SURF)*. Retrieved April 2019, from https://www.cdpr.ca.gov/docs/emon/surfwtr/surfdata.htm

Coffin, S., Bouwmeester, H., Brander, S., Damdimopoulou, P., Gouin, T., Hermabessiere, L., . . . Wright, S. (2022). Development and application of a health-based framework for informing regulatory action in relation to exposure of microplastic particles in California drinking water. *Microplastics & Nanoplastics, 2*(12).

Cohen, A. D. (2020). Summit briefs policy-makers on drinking water safety. *Science, 368*, 379-380. doi:10.1126/science.368.6489.379

Executive Office of the President. (2025). Executive Order 14303: Restoring Gold Standard Science. Federal Register, Vol. 90, No. 102, 23 May 2025, pp. 22601–606, https://www.govinfo.gov/content/pkg/FR-2025-05-29/pdf/2025-09802.pdf.

Federal Food, Drug, and Cosmetic Act. 21 U.S.C § 321. (1938).

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). New York: Chapman and Hall/CRC. doi:10.1201/b16018

Glassmeyer, S. T., Furlong, E. T., Kolpin, D. W., Batt, A. L., Benson, R., Boone, J. S., . . . Wilson, V. S. (2017). Nationwide reconnaissance of contaminants of emerging concern in source and treated drinking waters of the United States. *Science of The Total Environment, 581-582*, 909-922. doi:10.1016/j.scitotenv.2016.12.004

Grandjean, P., & Landrigan, P. J. (2006). Developmental neurotoxicity of industrial chemicals. *The Lancet, 368*(9553), 2167-2178. doi:10.1016/S0140-6736(06)69665-7

He, K., Hain, E., Timm, A., Tarnowski, M., & Blaney, L. (2019). Occurrence of antibiotics, estrogenic hormones, and UV-filters in water, sediment, and oyster tissue from the Chesapeake Bay. *Science of The Total Environment, 650*, 3101-3109. doi:10.1016/j.scitotenv.2018.10.021

HHS. (n.d.). *National Library of Medicine. Hazardous Substances Data Bank (HSDB)*. Retrieved April 2019, from National Institutes of Health (NIH): https://www.nlm.nih.gov/toxnet/index.html

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods.* Springer, New York: Springer Texts in Statistics. doi:10.1007/978-0-387-92407-6

Kalita, I., Kamilaris, A., Havinga, P., & Reva, I. (2024). Assessing the Health Impact of Disinfection Byproducts in Drinking Water. *ACS ES&T Water, 4*(4), 1564-1578. doi:https://doi.org/10.1021/acsestwater.3c00664

Kleinbaum, D. G., & Klein, M. (2010). *Logistic Regression: A Self-Learning Text* (3rd ed.). New York, NY: Springer-Verlag. doi:10.2307/41305046

Kleinstreuer, N. C., Ceger, P., Watt, E. D., Martin, M., Houck, K., Browne, P., . . . Judson, R. (2017). Development and Validation of a Computational Model for Androgen Receptor Activity. *Chemical Research in Toxicology, 30*(4), 946-964. doi:10.1021/acs.chemrestox.6b00347

Kostich, M. S., Batt, A. L., & Lazorchak, J. M. (2014). Concentrations of prioritized pharmaceuticals in effluents from 50 large wastewater treatment plants in the US and implications for risk estimation. *Environmental Pollution, 184*, 354-359. doi:10.1016/j.envpol.2013.09.013

Kostich, M. S., Batt, A. L., Glassmeyer, S. T., & Lazorchak, J. M. (2010). Predicting variability of aquatic concentrations of human pharmaceuticals. *Science of The Total Environment, 408*(20), 4504-4510. doi:10.1016/j.scitotenv.2010.06.015

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine, 28*, 3049-3067. doi:10.1002/sim.3680

Lyman, W. J., Reehl, W. F., & Rosenblatt, D. H. (1990). *Handbook of Chemical Property Estimation Methods.* Washington, DC: American Chemical Society.

Maruya, K. A., Dodder, N. G., Sengupta, A., Smith, D. J., Lyons, J. M., Heil, A. T., & Drewes, J. E. (2016). Multimedia screening of contaminants of emerging concern (CECS) in coastal urban watersheds in southern California (USA). *Environmental Toxicology and Chemistry, 35*, 1986-1994. doi:10.1002/etc.3348

Masoner, J. R., Kolpin, D. W., Cozzarelli, I. M., Barber, L. B., Burden, D. S., Foreman, W. T., . . . Bradley, P. M. (2019). Urban Stormwater: An Overlooked Pathway of Extensive Mixed Contaminants to Surface and Groundwaters in the United States. *Environmental Science and Technology, 53*(17), 10070-10081. doi:10.1021/acs.est.9b02867

Miller, E., Sedlak, M., Lin, D., Box, C., Holleman, C., Rochman, C. M., & Sutton, R. (2021). Recommended best practices for collecting, analyzing, and reporting microplastics in environmental media: Lessons learned from comprehensive monitoring of San Francisco Bay. *Journal of Hazardous Materials, 409*.

Mundy, W. R., Padilla, S., Breier, J. M., Crofton, K. M., Gilbert, M. E., Herr, D. W., . . . Cowden, J. (2015). Expanding the test set: Chemicals with potential to disrupt mammalian brain development. *Neurotoxicology and Teratology, 52A*, 25-35. doi:10.1016/j.ntt.2015.10.001

National Primary Drinking Water Regulations, Subpart Z – Control of Per- and Polyfluoroalkyl Substances (PFAS), 40 C.F.R. § 141.900 – 141.905. (n.d.). Retrieved from https://www.ecfr.gov/current/title-40/chapter-I/subchapter-D/part-141/subpart-Z

National Research Council (NRC). (2001). *Classifying Drinking Water Contaminants for Regulatory Consideration.* Washington DC: National Academy Press. Retrieved from https://nap.nationalacademies.org/catalog/10080/classifying-drinking-water-contaminants-for-regulatory-consideration

NDWAC. (2004). *National Drinking Water Advisory Council (NDWAC) Report on the CCL Classification Process to the U. S. Environmental Protection Agency.* Retrieved from https://www.epa.gov/sites/default/files/2015-11/documents/report_ccl_ndwac_07-06-04.pdf

Newton, S. R., McMahen, R. L., Sobus, J. R., Mansouri, K., Williams, A. J., McEachran, A. D., & Strynar, M. J. (2018). Suspect screening and non-targeted analysis of drinking water using point-of-use filters. *Environmental Pollution, 234*, 297-306. doi:10.1016/j.envpol.2017.11.033

Peterson, B. G., Carl, P., Boudt, K., Bennett, R., Ulrich, J., Zivot, E., . . . Jain, D. (2024). *PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis (Version 2.0.8) [R package]*. Retrieved from Comprehensive R Archive Network (CRAN): https://cran.r-project.org/web/packages/PerformanceAnalytics/index.html

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News, 6*(1), 7-11. Retrieved from https://journal.r-project.org/articles/RN-2006-002/

R Core Team. (2025). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ragusa, A., Svelato, A., Criselda, S., Catalano, P., Notarstefano, V., Carnevali, O., . . . Giorgini, E. (2021). Plasticenta: First evidence of microplastics in human placenta. *Environmental International, 146*.

Rahman, M. B., Driscoll, T., Cowie, C., & Armstrong, B. K. (2010). Disinfection by-products in drinking water and colorectal cancer: a meta-analysis. *International Journal of Epidemiology, 39*(3), 733-745. doi:10.1093/ije/dyp371

Richardson, S. D., & Postigo, C. (2015). The Next Generation of Drinking Water Disinfection By-Products: Occurrence, Formation, Toxicity, and New Links with Human Epidemiology. In K. C. Thompson, S. Gillespie, & E. Goslan (Eds.), *Disinfection By-products in Drinking Water* (eBook ed., pp. 1-13). Royal Society of Chemistry. doi:10.1039/9781782622710-00001

Richardson, S. D., Plewa, M. J., Wagner, E. D., Schoeny, R., & Demarini, D. M. (2007). Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: a review and roadmap for research. *Mutation Research, 636*(1-3), 178-242. doi:10.1016/j.mrrev.2007.09.001

Schaider, L. A., Rudel, R. A., Ackerman, J. M., Dunagan, S. C., Brody, J. G., , , & , . (2014). Pharmaceuticals, perfluorosurfactants, and other organic wastewater compounds in public drinking water wells in a shallow sand and gravel aquifer. *Science of the Total Environment, 468-469*, 384-393. doi:10.1016/j.scitotenv.2013.08.067

Schauberger, P., & Walker, A. (2025). *openxlsx: Read, Write and Edit xlsx Files.* R package version 4.2.8, https://github.com/ycphs/openxlsx. Retrieved from https://ycphs.github.io/openxlsx/index.html

Scott, T. M., Phillips, P. J., Kolpin, D. W., Colella, K. M., Furlong, E. T., Foreman, W. T., & Gray, J. L. (2018). Pharmaceutical manufacturing facility discharges can substantially increase the pharmaceutical load to U.S. wastewaters. *636*, 69-79. doi:10.1016/j.scitotenv.2018.04.160

Stage 2 Disinfection Byproducts Requirements, 40 CFR Part 141 Subpart V. (2006). Retrieved from https://www.ecfr.gov/current/title-40/chapter-I/subchapter-D/part-141/subpart-V. Retrieved from https://www.govinfo.gov/content/pkg/CFR-2025-title40-vol25/pdf/CFR-2025-title40-vol25-part141.pdf

Sun, M., Arevalo, E., Strynar, M., Lindstrom, A., Richardson, M., Kearns, B., . . . Knappe, D. R. (2016). Legacy and Emerging Perfluoroalkyl Substances Are Important Drinking Water Contaminants in the Cape Fear River Watershed of North Carolina. *Environmental Science & Technology Letters, 3*(12), 415-419. doi:10.1021/acs.estlett.6b00398

USEPA. (1991). Existing Stocks of Pesticides Products; Statement of Policy. *Federal Register, 56*(123), 29362.

USEPA. (1998). Announcement of the Drinking Water Contaminant Candidate List; Notice. *Federal Register, 63*(40), 10274.

USEPA. (2000a). *Database of Sources of Environmental Releases of Dioxin-Like Compounds in the United States.* Retrieved from https://cfpub.epa.gov/ncea/dioxin/recordisplay.cfm?deid=20797

USEPA. (2000b). *Methodology for Deriving Ambient Water Quality Criteria for the Protection of Human Health (2000).* Washington, D.C.: U.S. Environmental Protection Agency, Office of Water, Office of Science and Technology. Retrieved from https://www.epa.gov/sites/production/files/2018-10/documents/methodology-wqc-protection-hh-2000.pdf

USEPA. (2005). *Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens.* Washington, D.C.: U.S. Environmental Protection Agency, Office of Research and Development.

USEPA. (2005b). Notice - Drinking Water Contaminant Candidate List 2; Final Notice. *Federal Register, 70*(36), 9071.

USEPA. (2009a). *Final Contaminant Candidate List 3 Chemicals: Identifying the Universe.* Washington, D.C.: U.S. Environmental Protection Agency, Office of Water, Office of Groundwater and Drinking Water.

USEPA. (2009b). *Final Contaminant Candidate List 3 Chemicals: Screening to a PCCL.* Washington, D.C.: U.S. Environmental Protection Agency, Office of Water, Office of Groundwater and Drinking Water.

USEPA. (2009c). Final Contaminant Candidate List 3 Chemicals: Classification of the PCCL to CCL.

USEPA. (2009d). Community Water System Survey 2006. Volume 1: Overview. *EPA 815-R-09-001*.

USEPA. (2009e). Community Water System Survey 2006. Volume II: Detailed Tables and Survey Methodology. *EPA 815-R-09-002*.

USEPA. (2009f). Drinking Water Contaminant Candidate List 3 - Final. *Federal Register, 74*(194), 51850.

USEPA. (2012). *TSCA Work Plan Chemicals: Methods Document.* Washington, D.C.: USEPA. Retrieved from https://www.epa.gov/sites/production/files/2014-03/documents/work_plan_methods_document_web_final.pdf

USEPA. (2014a). *Child-Specific Exposure Scenarios Examples.* Washington D.C.: Office of Research and Development, National Center for Environmental Assessment.

USEPA. (2014b). *Guidance for Applying Quantitative Data to Develop Data-Derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation.* Washington D.C.: Office of the Science Advisor, Risk Assessment Forum.

doi:https://19january2017snapshot.epa.gov/sites/production/files/2015-01/documents/ddef-final.pdf

USEPA. (2014c). *Framework for Human Health Risk Assessment to Inform Decision Making.* Washington D.C.: Office of the Science Advisor, Risk Assessment Forum.

USEPA. (2016a). Drinking Water Contaminant Candidate List 4-Final. *Federal Register, 81*(222), 81099.

USEPA. (2016c). *User's Guide for T.E.S.T. (version 4.2) (Toxicity Estimation Software Tool): A Program to Estimate Toxicity from Molecular Structure.* Retrieved from https://19january2021snapshot.epa.gov/sites/static/files/2016-05/documents/600r16058.pdf

USEPA. (2017). National Primary Drinking Water Regulations; Announcement of the Results of EPA's Review of Existing Drinking Water Standards and Request for Public Comment and/or Information on Related Issues. *Federal Register, 82*(3518), 3518.

USEPA. (2018). Request for Nominations of Drinking Water Contaminants for the Fifth Contaminant Candidate List. *Federal Register, 83*(194), 50364.

USEPA. (2019a). *Update for Chapter 3 of the Exposure Factors Handbook. Ingestion of Water and Other Select Liquids.* Washington, D.C.: U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment. Retrieved from https://www.epa.gov/sites/production/files/2019-02/documents/efh_-_chapter_3_update.pdf

USEPA. (2019b). EPA's Per- and Polyfluoroalkyl Substances (PFAS) Action Plan. *EPA 823R18004*.

USEPA. (2020). Chemical Data Reporting (CDR) Results. Retrieved July 31, 2023, from https://www.epa.gov/chemical-data-reporting/access-cdr-data#2020

USEPA. (2021a). Prepublication Notice. Proposed Rule. TSCA Section 8(a)(7) Reporting and Recordkeeping Requirements for Perfluoroalkyl and Polyfluoroalkyl Substances. *EPA-HQ-OPPT-2020-0549*.

USEPA. (2021b). Announcement of Final Regulatory Determinations for Contaminants on the Fourth Drinking Water Contaminant Candidate List. *Federal Register, 86*(40), 12272.

USEPA. (2022a). Drinking Water Contaminant Candidate List 5-Final. *Federal Resister, 87*(218), 68060.

USEPA. (2022b). Technical Support Document for the Final Fifth Candidate List (CCL 5) – Chemical Contaminants. *EPA 815-R-22-002*.

USEPA. (2022c). Response to Comments Document on the Draft Fifth Contaminant Candidate List (CCL 5). *EPA 815-R-22-001*.

USEPA. (2023a). Drinking Water Contaminant Candidate List 6- Nominations. *Federal Register, 88*(33), 10316.

USEPA. (2023b, June). *Endocrine Disruptor Screening Program (EDSP) Estrogen Receptor Bioactivity.* Retrieved from https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-estrogen-receptor-bioactivity

USEPA. (2024). PFAS National Primary Drinking Water Regulation. *Federal Register, 89*(82), 32532.

USEPA. (2025). Announcement of Preliminary Regulatory Determinations for Contaminants on the Fifth Drinking Water Contaminant Candidate List. *Federal Register, 90*(9), 3830.

USEPA. (2026a). U.S. Environmental Protection Agency Implementation of Gold Standard Science. https://www.epa.gov/system/files/documents/2026-01/final-epa-gss-report.pdf.

USEPA. (2026b). Technical Support Document for the Draft Sixth Contaminant Candidate List (CCL 6) – Contaminant Information Sheets (CISs). *EPA 815-R-26-005*.

USEPA. (2026c). Technical Support Document for the Draft Sixth Contaminant Candidate List (CCL 6) - Microbial Contaminants. *EPA 815-R-26-006*.

USEPA-n.d.-a. (n.d.). Integrated Risk Information System (IRIS). IRIS Advanced Search. Retrieved May 17, 2019, from https://cfpub.epa.gov/ncea/iris/search/index.cfm?keyword

USEPA-n.d.-b. (n.d.). Toxics Release Inventory (TRI) Program. Retrieved from https://www.epa.gov/toxics-release-inventory-tri-program

Villanueva, C. M., Cantor, K. P., Cordier, S., Jaakkola, J. J., King, W. D., Lynch, C. F., . . . Kogevinas, M. (2004). Disinfection byproducts and bladder cancer: a pooled analysis. *Epidemiology, 15*(3), 357-367. doi:10.1097/01.ede.0000121380.02594.fc

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. doi:10.21105/joss.01686

Williams, A. J., Grulke, C. M., Edwards, J., McEachran, A. D., Mansouri, K., Baker, N. C., . . . Richard, A. M. (2017). The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *Journal of Cheminformatics, 9*(1), 61. doi:10.1186/s13321-017-0247-6

Zarus, G. M., Muianga, C., Hunter, C. M., & Pappas, R. S. (2021). A Review of Data for Quantifying Human Exposures to Micro and Nanoplastics and Potential Risks. *Science of the Total Environment, 756*.

Zhang, H., West, D., Shi, H., Ma, Y., Adams, C., & Eichholz, T. (2019). Simultaneous Determination of Selected Trace Contaminants in Drinking Water Using Solid-Phase Extraction-High Performance Liquid Chromatography-Tandem Mass Spectrometry. *Water, Air,& Soil Pollution, 230*(28). doi:10.1007/s11270-018-4066-9

List of Appendices

- Appendix A – Main Data Source Descriptions
- Appendix B – Supplemental Data Sources
- Appendix C – CCL 5 Chemicals Not Listed on the Draft CCL 6
- Appendix D – PCCL Chemical Contaminants
- Appendix E – Protocol for the Rapid Systematic Health Effects Literature Review
- Appendix F – Protocol to Derive Health Concentrations
- Appendix G – Protocol to Select Water Concentrations Used for Calculating Final Hazard Quotients for PCCL Chemicals
- Appendix H – Protocol to Determine Potency Attribute Scores
- Appendix I – Protocol to Determine Severity Attribute Scores
- Appendix J – Protocol to Determine Prevalence Attribute Scores
- Appendix K – Protocol to Determine Magnitude Attribute Scores
- Appendix L – Protocol to Determine Magnitude Attribute Scores from Persistence-Mobility
- Appendix M - Data Management for CCL 6

Appendix A – Main Data Source Descriptions

This appendix includes descriptions of the main sources of health effects and occurrence data that form the CCL 6 chemical Pre-Universe and Universe.

Main Sources of Health Effects Data

1. Agency for Toxic Substances and Disease Registry (ATSDR) Minimal Risk Levels (MRLs) – Centers for Disease Control and Prevention (CDC)

MRLs are substance-specific health guidance levels developed by the ATSDR. They are estimates of the level of daily exposure to a hazardous substance that is likely associated with no significant risk of adverse non-cancer health effects in humans. MRLs are derived for acute, intermediate, and chronic durations of exposure and can be derived for both inhalation and oral exposure pathways.

2. Cancer Potency Data Bank – National Library of Medicine, U.S. Department of Health and Human Services (HHS)

The Cancer Potency Data Bank provided results from 45 years of long-term animal cancer tests, including data on the carcinogenic potency (TD_{50}) of different chemicals. The Cancer Potency Data Bank has since been replaced with the Carcinogenic Potency Database and was removed from the web. The Carcinogenic Potency Database was last updated on July 10, 2025 and is currently in archived status.

3. *"Development and Validation of a Computational Model for Androgen Receptor Activity" – Kleinstreuer et al. 2017*

This peer-reviewed publication provided modeled androgen receptor (AR) bioactivity data for 1855 commercial and environmental chemicals. Modeled activity integrated *in vitro* high-throughput screening assays in multiple cell types that reported alterations in the androgen pathway at various levels, including androgen receptor binding, coregulator recruitment, gene transcription, and protein production.

4. Drinking Water Standards and Health Advisories (DWSHA) Tables – EPA

The DWSHA Tables provide the EPA's drinking water regulations, health advisories, RfDs, and cancer risk values for drinking water contaminants. The tables are revised periodically. The 2018 edition, which is the most recently published version, was used in the development of the CCL 6.

5. Endocrine Disruptor Screening Program (EDSP) – Estrogen Receptor Bioactivity – EPA

The EPA's EDSP has published results from modeling of quantitative *in vitro* Estrogen Receptor Bioactivity based on the ToxCast(TM) "Endocrine Receptor (ER) ER Model."

6. Guidelines for Canadian Drinking Water Quality – Health Canada

The Guidelines for Canadian Drinking Water Quality provide health-based guidelines developed based on reviews of contaminant health effects, exposure levels, and availability of treatment and analytical technologies. Guidance values are developed for contaminants that may have adverse health effects in humans and frequently occur or are expected to occur in drinking water supplies in Canada at a level of possible human health concern.

7. Guidelines for Drinking-Water Quality – WHO

The WHO Guidelines for Drinking-Water Quality (2022, 4th ed., incorporating the first and second addenda) provide guideline values for chemicals that, according to international risk assessments, show evidence of occurrence in drinking water and actual or potential health effects.

8. Hazardous Substances Data Bank (HSDB) – National Library of Medicine, HHS

The HSDB provides peer-reviewed toxicology data on potentially hazardous chemicals compiled from books, government documents, technical reports, and primary journal literature. The HSDB is a data rich source, and the only source of LD_{50}s for the CCL 6 process. The EPA retrieved this data source for CCL 5 and continues to use that downloaded data for CCL 6.

9. Health-Based Screening Levels (HBSLs) – U.S. Geological Survey (USGS)

At the time of CCL 6 Universe development, the USGS's HBSL dataset contained water quality data for 808 contaminants. HBSLs are non-enforceable water-quality benchmarks that were developed by the USGS National Water-Quality Assessment (NAWQA) Project for contaminants without EPA Maximum Contaminant Levels (MCLs) or Human Health Benchmarks for Pesticides (HHBPs).

10. Human Health-Based Water Guidance Table – Minnesota Department of Health

The Human Health-Based Water Guidance Table provides health-based rules and guidance developed by the Minnesota Department of Health to evaluate potential human health risks from exposures to chemicals in groundwater. At the time of CCL 6 Universe development, the dataset contained acute, short-term, subchronic, chronic, and cancer health risk limits, health-based values, or risk assessment advice for 203 contaminants.

11. Human Health Benchmarks for Pesticides – EPA

The EPA has developed human health benchmarks for 430 pesticides. These include benchmarks for acute and chronic exposures for the general population and sensitive populations (i.e., children and females 13-49 years of age) from exposure to pesticides that may be found in drinking water or sources of drinking water.

The dataset also includes benchmarks for pesticides in drinking water that have the potential for cancer risk and for pesticide active ingredients for which Health Advisories or enforceable National Primary Drinking Water Regulations (e.g., maximum contaminant levels) have not been developed.

12. Human Health Benchmarks for Pharmaceuticals (HHB-Rx) in Drinking Water – EPA

The EPA has developed human health drinking water benchmarks for 374 pharmaceuticals. The HHB-Rx are non-regulatory and non-enforceable and are intended to provide information to help states, tribes, and water systems better characterize potential health risks associated with the occurrence of pharmaceuticals in drinking water and prioritize pharmaceuticals that are known to or have the potential to occur in source waters and treated drinking water for additional monitoring and/or health effects research. Two HHB-Rx were derived for each pharmaceutical: one for the general population (all ages) and one for infants (birth to <1 year).

13. Integrated Risk Information System (IRIS) – EPA

The EPA's IRIS database contains toxicity data from assessments of hundreds of contaminants, including toxicity values (e.g., RfD, oral slope factor) for health effects resulting from chronic exposure to chemicals, as well as cancer classifications.

14. International Agency for Research on Cancer (IARC) Cancer Classifications – World Health Organization (WHO)

Since 1969, the IARC has led evaluation of the carcinogenic risk of chemicals to humans with the help of international working groups of experts in carcinogenesis and related fields. This dataset contains cancer classifications for over 1,000 contaminants.

15. Maximum Recommended Daily Dose (MRDD) Database – U.S. Food and Drug Administration (FDA)

The FDA Center for Drug Evaluation and Research's Maximum Recommended Daily Dose database contains values for 1,216 pharmaceuticals listed in Martindale: The Extra Pharmacopoeia (1973, 1983, and 1993) and The Physicians' Desk Reference (1995 and 1999).

16. National Recommended Water Quality Criteria - Human Health Criteria – EPA

The Human Health Ambient Water Quality Criteria contains recommended water quality criteria for human health for 128 pollutants. These are specific levels of chemicals or conditions in a water body that are not expected to cause adverse human health effects. Most of the criteria have been updated in 2015 to reflect the latest scientific information and EPA policies, including updated fish consumption rate, body weight, drinking water intake, health toxicity values, bioaccumulation factors, and relative source contributions (RSCs).

17. National Toxicology Program (NTP) Cancer Classifications – HHS

At the time of CCL 6 Universe development, the NTP cancer classification dataset was compiled from a list of almost 600 NTP peer-reviewed technical reports and includes cancer classifications for each contaminant based on short-term and long-term studies on rats and mice.

18. Provisional Peer-Reviewed Toxicity Values (PPRTVs) – EPA

PPRTVs, which may include provisional RfDs, CSFs, and cancer classifications, are developed for compounds that lack IRIS assessments or that lack a quantified toxicity value in their IRIS assessment. PPRTVs may be derived for acute, subchronic, and chronic exposure scenarios and for exposure via inhalation or oral routes.

19. Toxicity Criteria Database – California Environmental Protection Agency (CalEPA) Office of Environmental Health Hazard Assessment (OEHHA)

CalEPA's Toxicity Criteria Database provides health hazard information developed by the CalEPA OEHHA, including drinking water public health goals and cancer potency data such as cancer slope factors (CSFs).

20. Toxicity Reference Database (ToxRefDB) version 2.1 – EPA

The ToxRefDB contains decades of results from approximately 5,900 *in vivo* animal toxicity studies on over 1,100 chemicals, which follow relatively strict methodology guidelines set by the EPA and NTP. No new studies were added in version 2.1 compared to version 1.0 used in CCL 5 (USEPA, 2022), but this update recovered inadvertently omitted data and improved functionality of the database.

Main Sources of Occurrence Data

1. ATSDR Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) Substance Priority List – CDC

The Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) requires that ATSDR and EPA publish, every two years, a list of substances that are most commonly found at facilities on the National Priorities List (NPL) and that are deemed to present the greatest potential threat to human health, based on their frequency of occurrence, toxicity, and potential for human exposure at NPL sites. SDWA Section 1412(b)(1) requires EPA to consider the contaminants in this CERCLA priority list in the development of the CCL.

2. Chemical Data Reporting (CDR) Results – EPA

Under the CDR rule requirements described in section 8 of the Toxic Substances Control Act (TSCA), EPA collects commercial manufacturing, processing, and use information for chemicals throughout the United States, including production volume data.

3. "Concentrations of prioritized pharmaceuticals in effluents from 50 large wastewater treatment plants in the US and implications for risk estimation" – Kostich et al. 2014

This EPA study measured concentrations of 56 active pharmaceutical ingredients in effluent of 50 large wastewater treatment plants in the U.S. in 2011.

4. Disinfection By-products Information Collection Rule (DBP ICR) – EPA

The Disinfection By-products Information Collection Rule (DBP ICR) "Aux 1" Database contains monitoring data from large public water systems (PWSs) (serving a population greater than or equal to 100,000) from July 1997 to December 1998. A total of 296 water systems reported data, including monitoring results for microbial contaminants and disinfectant byproducts.

5. "Evaluating the extent of pharmaceuticals in surface waters of the United States using a National-scale Rivers and Streams Assessment survey" – Batt et al. 2016

This EPA study examined occurrence of active pharmaceutical ingredients and risks to aquatic life by sampling 182 sites in rivers within close proximity to urban streams in 2008-2009.

6. "Expanded Target-Chemical Analysis Reveals Extensive Mixed-Organic-Contaminant Exposure in U.S. Streams" – Bradley et al. 2017

This study provides surface water data on 719 compounds measured in 38 streams across the U.S., including a mixture of urban and agricultural watersheds.

7. Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) registered pesticides and pesticide ingredients - EPA

In the development of the CCL, EPA is required by SDWA Section 1412(b)(1) to consider substances registered as pesticides under the FIFRA. The FIFRA list contains 4,038 registered substances used in the production of pesticide products in the U.S. as part of federally mandated reporting under this act.

8. “Legacy and emerging perfluoroalkyl substances are important emerging water contaminants in the Cape Fear River Watershed of North Carolina” – Sun et al. 2016

This dataset provides concentrations of per- and polyfluoroalkyl substances (PFAS) and more recently discovered perfluoroalkyl ether carboxylic acids (PFECA) in source water of three drinking water treatment plants in the Cape Fear River watershed of North Carolina monitored for over six months in 2013.

9. “Multimedia screening of contaminants of emerging concern (CECS) in coastal urban watersheds in southern California (USA).” – Maruya et al. 2016

This source provided concentration data for 13 contaminants of emerging concern including pharmaceuticals, personal care products, commercial and household chemicals, and pesticides. Samples were taken from the Santa Clara River in October 2013. The Santa Clara River receives treated wastewater from municipal wastewater treatment plants.

10. National Aquatic Resource Surveys, 2017 National Lakes Assessment - EPA

The 2017 National Lakes Assessment reports on the condition of lakes and reservoirs in the United States for 2017. The data includes concentration data for cyanotoxins, atrazine, and *E. coli*.

11. National Aquatic Resource Surveys, Rivers and Streams 2018-2019 Algal Toxin - EPA

This national survey collects data on biological, chemical, and physical indicators of the ecological condition of streams and rivers. For CCL, this source provides concentration data from algal toxin sampling. Samples were collected between 2018 – 2019.

12. National Health and Nutrition Examination Survey (NHANES) Biospecimen Program – CDC

The CDC’s “National Report on Human Exposure to Environmental Chemicals, Updated” provides nationally representative, cumulative biomonitoring data for chemicals and metabolites measured in blood, serum, and urine samples from random subsamples collected continuously in NHANES from 1999–2000 through 2017-2018.

13. National Inorganics and Radionuclides Survey (NIRS) – EPA

The National Inorganics and Radionuclides Survey (NIRS) provides 1984-1986 occurrence data on radionuclides and inorganic contaminants being considered for national primary drinking water regulations from a group of randomly selected, nationally representative PWSs served by ground water in 49 States and Puerto Rico from 1984 through 1986 (USEPA, 2008). NIRS data are available in the docket for Regulatory Determination 4 at https://www.regulations.gov/document/EPA-HQ-OW-2019-0583-0290.

14. National Water Information System (NWIS) – Water Quality Portal (WQP) – USGS

The Water Quality Portal (WQP) is housed in EPA's National Contaminant Occurrence Database and is a cooperative service sponsored by the USGS, EPA, and National Water Quality Monitoring Council. The WQP houses the NWIS and includes nationally representative National Water-Quality Assessment (NAWQA) data as well as non-nationally representative data. This source provides summary detection information on contaminants in surface water and ground water, collected since 1991 by over 400 state, federal, tribal, and local agencies.

15. National Water-Quality Assessment (NAWQA) – WQP – USGS

Refer to description of National Water Information System (NWIS) Water Quality Portal – (WQP) above for more information on the Water Quality Portal and the data it provides.

16. "Nationwide reconnaissance of contaminants of emerging concern in source and treated drinking waters of the United States" – Glassmeyer et al. 2017

This joint USGS-EPA, two-part study conducted between 2007 and 2012 examined 25 drinking water treatment plants across the U.S. with probable wastewater inputs to their source waters to assess the prevalence of a wide range of analytes (e.g., pharmaceuticals, anthropogenic waste indicators, PFAS, inorganic chemicals, microbes) in source waters and identify those that persist after treatment.

17. "Nationwide reconnaissance of contaminants of emerging concern in source and treated drinking waters of the United States: Pharmaceuticals" – Furlong et al. 2017

This joint USGS-EPA, two-part study conducted between 2007 and 2012 examined 25 drinking water treatment plants across the U.S. with probable wastewater inputs to their source waters to assess the prevalence of a wide range of pharmaceuticals in source waters and identify those that persist after treatment.

18. NEWTS Coal Mine Drainage Dataset from Cravotta Brady (2015)

Data from this source include water quality results for inorganic compounds in untreated and treated streams that are affected by coal mine discharges or are near coal processing locations. Data are from 2011.

19. NEWTS Department of Energy (DOE) Geothermal Technologies Office Rare Earth Element Assessment Report Database

The Geothermal Database presents the concentration of aqueous rare earth elements in geothermal produced waters. There were 224 samples of produced waters from sources such as wells, libraries, and rock samples. The data were collected 2016 – 2018.

20. NEWTS EPA Ash Database

Data from the US EPA Ash Database containing ash pond effluent data from selected U.S. coal power plants. Data are from 1973 – 2014.

21. NEWTS EPA Flue Gas Desulfurization (FGD) Effluent Database

The US EPA FGD Database contains FGD (Flue-gas desulfurization) effluent data from selected U.S. coal power plants. Data are from 2007 – 2014.

22. NEWTS EPA Leachate Dataset

Concentration data in this source are from select US coal power plant ash landfill leachate. Data are from 2010.

23. NEWTS USGS Brackish Water Database

Data are from brackish water compositions sources across the US. Samples are from 1901-2013.

24. NEWTS USGS National Produced Waters Geochemical Database v2.3

The produced water database contains concentration data from waters produced alongside oil and gas. Data are from 1905 – 2014.

25. "Occurrence of antibiotics, estrogenic hormones, and ultraviolet (UV) filters in water, sediment, and oyster tissue from the Chesapeake Bay." He et al. 2019

The study of contaminants of emerging concern in the Chesapeake Bay included 43 antibiotics, 3 estrogenic hormones, and 5 UV-filters found in personal care products. Data are from 2016 and 2017.

26. Pesticide Data Program (PDP) – USDA

USGS monitors pesticide residues in food as well as in finished water, untreated water, and ground water. This database contains over 31.3 million pesticide residue findings, including both positive detections and non-detects, for the 255,061 samples tested by the Pesticide Data Program (PDP) from 1994 through 2013.

27. Pesticide Use Estimates – USGS

This dataset provides state-level annual pesticide use estimates for the 48 states comprising the contiguous U.S., collected between 1992 and 2019.

28. "Pharmaceuticals, hormones, pesticides, and other bioactive contaminants in water, sediment, and tissue from Rocky Mountain National Park" Battaglin et al. 2018

The study examined the occurrence of pharmaceuticals, hormones, pesticides, and other bioactive contaminants at 20 sites in Rocky Mountain National Park (RMNP), Colorado, in 2012–2013. Samples were analyzed for 369 parameters including 149 pharmaceuticals, 22 hormones, 137 pesticides, and 61 other chemicals to provide a representative assessment of bioactive contaminant occurrence within RMNP.

29. "Pharmaceutical manufacturing facility discharges can substantially increase the pharmaceutical load to U.S. wastewaters" Scott et al. 2018

This study provides data on concentrations of 120 pharmaceuticals and pharmaceutical degradates in treated wastewater effluent samples at various treatment plants, including some that received discharges from pharmaceutical manufacturing facilities and others that did not. In addition to pharmaceuticals, the survey also analyzed samples for 13 natural and synthetic hormones, 32 domestic use products, 7 plant and animal biochemicals, and 27 other organic chemicals including pesticides. Data were collected from 2004-2013, 2011-2012, and 2016--2017.

30. “Pharmaceuticals, perfluorosurfactants, and other organic wastewater compounds in public drinking water wells in a shallow sand and gravel aquifer.” Schaider et al. 2014

The study includes organic wastewater compounds tested in 20 public drinking water supply wells on Cape Cod, Massachusetts.

31. “Predicting variability of aquatic concentrations of human pharmaceuticals” – Kostich et al. 2010

This EPA study predicts pharmaceutical concentrations in surface water. To derive predicted environmental concentrations, the study compiled measured environmental concentrations from wastewater, surface water, ground water, and other sources reported in other peer-reviewed publications.

32. “Reconnaissance of mixed organic and inorganic chemicals in private and public supply tapwaters at selected residential and workplace sites in the United States” – Bradley et al. 2018

In this study, USGS, NIH, and EPA scientists measured 482 organic and 19 inorganic chemicals in finished tap water from 13 home (7 public supply, 6 private supply) and 12 workplace (public supply) sites in 11 states across the U.S., in May-September 2016.

33. “Simultaneous Determination of Selected Trace Contaminants in Drinking Water Using Solid-Phase Extraction-High Performance Liquid Chromatography-Tandem Mass Spectrometry” Zhang et al. 2019

Occurrence of selected contaminants in Missouri State drinking water system during hot and cold seasons.

34. Surface Water Database (SURF) – California Department of Pesticide Regulation (CDPR)

CDPR maintains the SURF database which contains data from 614 environmental monitoring studies testing for the presence of pesticides in statewide surface waters dating back to 1925.

35. “Suspect screening and non-targeted analysis of drinking water using point-of-use filters” – Newton et al. 2018

This is a pilot study on the use of point-of-use water filtration devices for screening and non-targeted analysis of drinking water. The filtration devices (Brita brand commercial filters) were employed to collect time-integrated drinking water samples for nine North Carolina homes. From these samples, a suspect screening analysis was performed by matching high resolution mass spectra of unknown features to molecular formulas from EPA's DSSTox database.

36. Sustaining the Earth’s Watersheds - Agricultural Research Database System (STEWARDS) - United States Department of Agriculture

ARS contains water quality monitoring data collected from water resource management groups across the country and details on collecting methods, type of contaminant found, and samples collected croplands, with grazing lands and wetlands. Data are from 2000 – 2015.

37. Toxics Release Inventory (TRI) – EPA

The TRI is a public database provided by EPA to track chemical releases and pollution prevention activities reported by industrial and federal facilities across the United States. The 2021 TRI dataset includes environmental release data on 531 on-site and off-site chemicals reported, disposed of or otherwise released in 2021.

38. Unregulated Contaminant Monitoring Rule (UCMR) Cycles 1-4 – EPA

Every five years, the EPA develops a list of contaminants that PWSs must monitor as part of the UCMR program. The EPA uses UCMR to collect nationally representative data to understand the frequency and level of occurrence of unregulated contaminants in the nation's PWSs. These data are collected from both large PWSs which serve more than 10,000 people as well as representative samples from small PWSs which serve fewer than or equal to 10,000 people. UCMR data are provided in the EPA's National Contaminant Occurrence Database. This monitoring program provides a basis for future regulatory actions to protect public health.

39. UCMR Cycle 5 – EPA

UCMR 5 requires monitoring for 30 chemical contaminants between 2023 and 2025 using analytical methods developed by the EPA and consensus organizations. Refer to description of UCMR 1-4 above for more information on data collection for the UCMR process.

40. Unregulated Contaminant Monitoring-State (UCM-State) Rounds 1 and 2 – EPA

The UCM-State Round 1 and 2 datasets contain PWS monitoring results collected by states and primacy entities in 1988-1992 and 1993-1997, respectively, of then-unregulated contaminants.

41. "Urban Stormwater: An Overlooked Pathway of Extensive Mixed Contaminants to Surface and Groundwaters in the United States" Masoner et al. 2019

The study has collected data from 50 stormwater runoff events across 17 states between August 2016-December 2017.

References

References for data sources are provided in Appendix M. Other references cited here are listed below.

https://www.accessdata.fda.gov/scripts/cder/daf/

https://dailymed.nlm.nih.gov/dailymed/

USEPA. 2008. The Analysis of Occurrence Data from the Unregulated Contaminant Monitoring (UCM) Program and National Inorganics and Radionuclides Survey (NIRS) in Support of Regulatory Determinations for the Second Drinking Water Contaminant Candidate List. EPA 815-R-08-012.

Appendix B – Supplemental Data Sources

This appendix lists all the supplemental data sources that were used for filling data gaps or were included in public nominations in the CCL 6 process. This list includes sources used for health assessment searches and supplemental data sources identified in CCL 5, as well as sources from CCL 6 public nominations. For full citations, refer to the CIS Technical Support Document (USEPA, 2026b).

Supplemental Health Sources

1. "A Review of Data for Quantifying Human Exposures to Micro and Nanoplastics and Potential Health Risks" – Zarus et al. 2021[1]
2. CompTox Chemicals Dashboard – EPA
3. Dietary Reference Intake documents – National Academy of Medicine
4. Guideline Technical Documents – Health Canada
5. Guidelines for Drinking-Water Quality documents – WHO
6. Health Advisory supporting documents – EPA Office of Water (OW)
7. Health Effects Support Documents (HESDs) – EPA OW
8. Human Health Benchmarks for Pharmaceuticals (HHB-Rx) in Drinking Water – EPA
9. Integrated Risk Information System (IRIS) Chemical Assessment Summaries and Toxicological Reviews – EPA
10. Peer-reviewed studies identified through the health effects rapid systematic literature review (see Section 4.2.2 and Appendix E for more details and the spreadsheet titled "CCL 6 Rapid Systematic Literature Review Results" in the docket for a full list of references)
11. "Plasticenta: First evidence of microplastics in human placenta." Ragusa et al. – 2021[1]
12. Provisional Peer-Reviewed Toxicity Value (PPRTV) support documents – EPA ORD
13. Public Health Goal support documents – California Environmental Protection Agency (CalEPA) Office of Environmental Health Hazard Assessment (OEHHA)
14. Registration Dossiers – European Chemicals Agency (ECHA)
15. Reregistration Eligibility Decision (RED) documents and Human Health Risk Assessments – EPA Office of Pesticides Program
16. Risk Evaluation support documents – EPA Office of Chemical Safety and Pollution Prevention
17. Scientific Output publications – European Food Safety Authority (EFSA)
18. Toxicological Profiles – Centers for Disease Control and Prevention (CDC), Agency for Toxic Substances and Disease Registry (ATSDR)
19. Toxicological Summaries – Minnesota Department of Health

Supplemental Occurrence Sources

1. 2006 Community Water System Survey (CWSS) – EPA
2. "A survey of occurrence and risk assessment of pharmaceutical substances in the Great Lakes Basin" – Uslu et al. 2013[2]
3. CompTox Chemicals Dashboard – EPA
4. Dieldrin and Drinking Water – Minnesota Department of Health 2016[2]
5. "Groundwater quality data from the National Water-Quality Assessment Project, May 2012 through December 2013 (ver. 1.1, November 2016): U.S. Geological Survey Data Series 997" – Arnold et al. 2016[2]
6. "Hormones and pharmaceuticals in groundwater used as a source of drinking water across the United States" – Bexfield et al. 2019[2]
7. "Human health risk assessment of pharmaceuticals in water: An uncertainty analysis for meprobamate, carbamazepine, and phenytoin" – Kumar & Xagoraraki 2010[2]
8. "Occurrence of anthropogenic organic compounds and nutrients in source and finished water in the Sioux Falls area, South Dakota, 2009–10": U.S. Geological Survey Scientific Investigations Report 2012–5098, 21 p. plus appendices. [2]
9. "Occurrence of neonicotinoid insecticides in finished drinking water and fate during drinking water treatment" – Klarich et al. 2017[2]
10. "Recommended best practices for collecting, analyzing, and reporting microplastics in environmental media: Lessons learned from comprehensive monitoring of San Francisco Bay." – Ezra Miller, et al. 2021[1]
11. "Reconnaissance of land-use sources of pesticides in drinking water, McKenzie River, Oregon" – Kelly, Anderson, & Morgenstern 2012[2]
12. Reregistration Eligibility Decision documents, Human Health Risk Assessments, and Dietary Exposure and Risk Assessments (used to identify Estimated Drinking Water Concentrations (EDWCs), as described in Section 4.2.2) – EPA Office of Pesticides Program
13. Six Year Review (SYR) 3 State Data on Unregulated Contaminants (includes data from Maine, Pennsylvania, DC, and Washington)
14. Southeast Stream Quality Assessment (SESQA) – USGS[2]
15. State drinking water monitoring data for unregulated contaminants/contaminants of emerging concern (CECs) that are accessible online
    a. Water Quality Analyses Database Files – California Water Boards
    b. Energy and Environmental Affairs Data Portal – Commonwealth of Massachusetts Executive Office of Energy and Environmental Affairs
    c. Drinking Water Data Base – Florida Department of Environmental Protection
    d. Public Drinking Water System Data – Wisconsin Department of Natural Resources Drinking Water

16. "Trace levels of dieldrin and bromacil in two Oahu Water Systems" – State of Hawaii Department of Health 2015[2]
17. Village Creek Dieldrin Screening: Final Report – EPA Region 4 2015[2]
18. "Year-long evaluation on the occurrence and fate of pharmaceuticals, personal care products, and endocrine disrupting chemicals in an urban drinking water treatment plant" – Padhye et al. 2014[2]

[1] Data sources cited in public nominations for CCL 6.

[2] Data sources identified through the CCL 5 occurrence literature review

Appendix C – CCL 5 Chemicals Not Listed on the Draft CCL 6

Chemical Name	CASRN	DTXSID
A. CCL 5 Chemicals Below the Screening Points Threshold for Inclusion on PCCL 6		
2-Aminotoluene	95-53-4	DTXSID1026164
2,4-Dinitrophenol	51-28-5	DTXSID0020523
Acrolein	107-02-8	DTXSID5020023
Desvenlafaxine	93413-62-8	DTXSID40869118
B. CCL 5 Chemicals Screened to PCCL 6, “Not List” Recommendations by the Chemical Evaluators		
2-Hydroxyatrazine	2163-68-0	DTXSID6037807
6-Chloro-1,3,5-triazine-2,4-diamine	3397-62-4	DTXSID1037806
Deethylatrazine	6190-65-4	DTXSID5037494
Desisopropyl atrazine	1007-28-9	DTXSID0037495
Dimethoate	60-51-5	DTXSID7020479
Norflurazon	27314-13-2	DTXSID8024234
Propanil	709-98-8	DTXSID8022111
Propoxur	114-26-1	DTXSID7021948
Tebupirimfos	96182-53-5	DTXSID1032482
C. CCL 5 Chemicals Screened to PCCL 6, Removed as Canceled Non-Persistent Pesticides		
Propachlor	1918-16-7	DTXSID4024274
Propazine	139-40-2	DTXSID3021196
D. CCL 5 Chemicals Screened to PCCL 6, Removed for Ongoing Regulatory Determination 5 Status		
Cylindrospermopsin	143545-90-8	DTXSID2031083
Ethoprop	13194-48-4	DTXSID4032611
Microcystin	77238-39-2	DTXSID001024118
Molybdenum	7439-98-7	DTXSID1024207
Permethrin	52645-53-1	DTXSID8022292
Profenofos	41198-08-7	DTXSID3032464
Tebuconazole	107534-96-3	DTXSID9032113
Tribufos	78-48-8	DTXSID1024174
E. CCL 5 Chemicals Screened to PCCL 6, “List” Decision by Chemical Evaluators, Removed during Refinement Step of CCL 6 Process		
Boron	7440-42-8	DTXSID3023922
Dieldrin	60-57-1	DTXSID9020453

*Note: 17-alpha ethynyl estradiol and fluconazole were listed on CCL 5 and are included on the draft CCL 6 via the pharmaceutical chemical group.

Appendix D – PCCL Chemical Contaminants

Chemical Name	CASRN	DTXSID
1,1,2,2-Tetrachloroethane	79-34-5	DTXSID7021318
1,1-Dichloroethane	74-34-3	DTXSID1020437
1,2,3-Trichloropropane	96-18-4	DTXSID9021390
1,2,4-Triazole	288-88-0	DTXSID6027131
1,2,4-Trimethylbenzene	95-63-6	DTXSID6021402
1,2-Diphenylhydrazine	122-66-7	DTXSID7020710
1,3-Butadiene	106-99-0	DTXSID3020203
1,3-Dichloropropene	542-75-6	DTXSID1022057
1,4-Dioxane	123-91-1	DTXSID4020533
1-Methylnaphthalene	90-12-0	DTXSID9020877
2,4,6-Trinitrotoluene	118-96-7	DTXSID7024372
2,4-Dichlorophenoxybutyric acid	94-82-6	DTXSID7024035
2,4-Dichlorophenol	120-83-2	DTXSID1020439
2,4-Dinitrotoluene	121-14-2	DTXSID0020529
2,6-Dinitrotoluene	606-20-2	DTXSID5020528
2-Chloro-4,6-diamino-1,3,5-triazine	3397-62-4	DTXSID1037806
2-Hydroxyatrazine	2163-68-0	DTXSID6037807
2-Methylnaphthalene	91-57-6	DTXSID4020878
4-tert-Octylphenol	140-66-9	DTXSID9022360
5-Methyl-1H-benzotriazole	136-85-6	DTXSID1038743
6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	21145-77-7	DTXSID7041544
Acephate	30560-19-1	DTXSID8023846
Acetamiprid	135410-20-7	DTXSID0034300
Acetochlor	34256-82-1	DTXSID8023848
Acetochlor ESA	187022-11-3	DTXSID6037483
Acetochlor OA	194992-44-4	DTXSID1037484
Acetone	67-64-1	DTXSID8021482
Acrylonitrile	107-13-1	DTXSID5020029
Alachlor ESA	142363-53-9	DTXSID6037485
Alachlor OA	171262-17-2	DTXSID1037486
Aldrin	309-00-2	DTXSID8020040
alpha-1,2,3,4,5,6-Hexachlorocyclohexane	319-84-6	DTXSID2020684
Aluminum	7429-90-5	DTXSID3040273
Ametryn	834-12-8	DTXSID1023869
Ammonia	7664-41-7	DTXSID0023872
Anthraquinone	84-65-1	DTXSID3020095
Argon	7440-37-1	DTXSID3052482

Chemical Name	CASRN	DTXSID
Azoxystrobin	131860-33-8	DTXSID0032520
Benfluralin	1861-40-1	DTXSID3023899
Bensulide	741-58-2	DTXSID9032329
Bentazone	25057-89-0	DTXSID0023901
Benzophenone	119-61-9	DTXSID0021961
Benzyl butyl phthalate	85-68-7	DTXSID3020205
Bicarbonate	71-52-3	DTXSID2049921
Bifenthrin	82657-04-3	DTXSID9020160
Bisphenol A	80-05-7	DTXSID7020182
Boron	7440-42-8	DTXSID3023922
Boscalid	188425-85-6	DTXSID6034392
Bromacil	314-40-9	DTXSID4022020
Bromide	7726-95-6	DTXSID6043967
Bromochloromethane	74-97-5	DTXSID4021503
Bromoxynil	1689-84-5	DTXSID3022162
Caffeine	58-08-2	DTXSID0020232
Calcium	7440-70-2	DTXSID9050484
Carbaryl	63-25-2	DTXSID9020247
Carbendazim	10605-21-7	DTXSID4024729
Carbon	7440-44-0	DTXSID9027651
Carbon dioxide	124-38-9	DTXSID4027028
Carbon disulfide	75-15-0	DTXSID6023947
Chlorimuron-ethyl	90982-32-4	DTXSID0023955
Chlordecone (Kepone)	143-50-0	DTXSID1020770
Chlorodifluoromethane	75-45-6	DTXSID6020301
Chloromethane	74-87-3	DTXSID0021541
Chlorothalonil	1897-45-6	DTXSID0020319
Chlorpyrifos	2921-88-2	DTXSID4020458
Chlorsulfuron	64902-72-3	DTXSID7023980
Clopyralid	1702-17-6	DTXSID9029221
Clothianidin	210880-92-5	DTXSID2034465
Cobalt	7440-48-4	DTXSID1031040
Coumaphos	56-72-4	DTXSID2020347
Cyfluthrin	68359-37-5	DTXSID5035957
Cyhalothrin	68085-85-8	DTXSID6023997
Cypermethrin	52315-07-8	DTXSID1023998
Cyproconazole	94361-06-5	DTXSID0032601
DEET	134-62-3	DTXSID2021995
Deethylatrazine	6190-65-4	DTXSID5037494
Deisopropylatrazine	1007-28-9	DTXSID0037495

Chemical Name	CASRN	DTXSID
Deltamethrin	52918-63-5	DTXSID8020381
Diazinon	333-41-5	DTXSID9020407
Dibutyl phthalate	84-74-2	DTXSID2021781
Dicamba	1918-00-9	DTXSID4024018
Dichlobenil	1194-65-6	DTXSID5032365
Dichlorvos	62-73-7	DTXSID5020449
Dicrotophos	141-66-2	DTXSID9023914
Dieldrin	60-57-1	DTXSID9020453
Diethyl phthalate	84-66-2	DTXSID7021780
Difenoconazole	119446-68-3	DTXSID4032372
Dimethenamid	87674-68-8	DTXSID4032376
Dimethoate	60-51-5	DTXSID7020479
Dinotefuran	165252-70-0	DTXSID7034549
Dithiopyr	97886-45-8	DTXSID9032379
Diuron	330-54-1	DTXSID0020446
Endosulfan sulfate	1031-07-8	DTXSID3037541
Epoxiconazole	133855-98-8	DTXSID1040372
EPTC	759-94-4	DTXSID1024091
Esfenvalerate	66230-04-4	DTXSID4032667
Ethalfluralin	55283-68-6	DTXSID8032386
Ethion	563-12-2	DTXSID2024086
Ethofumesate	26225-79-6	DTXSID8034580
Ethylene thiourea	96-45-7	DTXSID5020601
Fenbuconazole	114369-43-6	DTXSID8032548
Fenitrothion	122-14-5	DTXSID4032613
Fenpropathrin	39515-41-8	DTXSID0024002
Fipronil	120068-37-3	DTXSID4034609
Fludioxonil	131341-86-1	DTXSID2032398
Flufenacet	142459-58-3	DTXSID2032552
Flumetsulam	98967-40-9	DTXSID4032615
Fluometuron	2164-17-2	DTXSID8020628
Fluoranthene	206-44-0	DTXSID3024104
Fluridone	59756-60-4	DTXSID8024107
Fonofos	944-22-9	DTXSID2024113
Galaxolide	1222-05-5	DTXSID8027373
Hexazinone	51235-04-2	DTXSID4024145
Hydramethylnon	67485-29-4	DTXSID6023868
Imazalil	35554-44-0	DTXSID8024151
Imazaquin	81335-37-7	DTXSID3024152
Imazethapyr	81335-77-5	DTXSID3024287

Chemical Name	CASRN	DTXSID
Imidacloprid	138261-41-3	DTXSID5032442
Indoxacarb	173584-44-6	DTXSID1032690
Iodide	7553-56-2	DTXSID80912339
Iprodione	36734-19-7	DTXSID3024154
Iron	7439-89-6	DTXSID5043710
Isophorone	78-59-1	DTXSID8020759
lambda-Cyhalothrin	91465-08-6	DTXSID7032559
Linuron	330-55-2	DTXSID2024163
Lithium	7439-93-2	DTXSID5036761
Magnesium	7439-95-4	DTXSID0049658
Malaoxon	1634-78-2	DTXSID9020790
Malathion	121-75-5	DTXSID4020791
Manganese	7439-96-5	DTXSID2024169
MCPA	94-74-6	DTXSID4024195
Mecoprop	93-65-2	DTXSID9024194
Metalaxyl	57837-19-1	DTXSID6024175
Methane	74-82-8	DTXSID8025545
Methomyl	16752-77-5	DTXSID1022267
Methyl bromide	74-83-9	DTXSID8020832
Methyl mercury	22967-92-6	DTXSID9024198
Methyl tert-butyl ether	1634-04-4	DTXSID3020833
Metolachlor	51218-45-2	DTXSID4022448
Metolachlor ESA	171118-09-5	DTXSID1037567
Metolachlor OA	152019-73-3	DTXSID6037568
Metribuzin	21087-64-9	DTXSID6024204
Metsulfuron-methyl	74223-64-6	DTXSID6023864
Mirex	2385-85-5	DTXSID7020895
Myclobutanil	88671-89-0	DTXSID8024315
Naphthalene	91-20-3	DTXSID8020913
Nicosulfuron	11991-09-4	DTXSID6034764
Nicotine	54-11-5	DTXSID1020930
Nitrobenzene	98-95-3	DTXSID3020964
Nitroglycerin	55-63-0	DTXSID1021407
Nonylphenol	25154-52-3	DTXSID3021857
Norflurazon	27314-13-2	DTXSID8024234
Oryzalin	19044-88-3	DTXSID8024238
Oxadiazon	19666-30-9	DTXSID3024239
Oxyfluorfen	42874-03-3	DTXSID7024241
p-Cresol	106-44-5	DTXSID7021869
Pendimethalin	40487-42-1	DTXSID7024245

Chemical Name	CASRN	DTXSID
Pentachloronitrobenzene	82-68-8	DTXSID2021105
Phenanthrene	85-01-8	DTXSID6024254
Phenol	108-95-2	DTXSID5021124
Phorate	298-02-2	DTXSID4032459
Phosmet	732-11-6	DTXSID5024261
Phosphate/Orthophosphate	14265-44-2	DTXSID7039672
Phosphorus	7723-14-0	DTXSID1024382
Piperonyl butoxide	51-03-6	DTXSID1021166
Potassium	7440-09-7	DTXSID9049748
Prometon	1610-18-0	DTXSID6022341
Prometryn	7287-19-6	DTXSID4024272
Propanil	709-98-8	DTXSID8022111
Propargite	2312-35-8	DTXSID4024276
Propetamphos	31218-83-4	DTXSID7032470
Propiconazole	60207-90-1	DTXSID8024280
Propoxur	114-26-1	DTXSID7021948
Propyzamide	23950-58-5	DTXSID2020420
Prosulfuron	94125-34-5	DTXSID9034868
Pyraclostrobin	175013-18-0	DTXSID7032638
Pyrene	129-00-0	DTXSID3024289
Quinoline	91-22-5	DTXSID1021798
RDX	121-82-4	DTXSID9024142
Silica	7631-86-9	DTXSID1029677
Silicon	7440-21-3	DTXSID0051441
Silver	7440-22-4	DTXSID4024305
Sodium	7440-23-5	DTXSID1049774
Strontium	7440-24-6	DTXSID3024312
Sulfentrazone	122836-35-5	DTXSID6032645
TDCPP	13674-87-8	DTXSID9026261
Tebupirimfos	96182-53-5	DTXSID1032482
Tebuthiuron	34014-18-1	DTXSID3024316
Terbacil	5902-51-2	DTXSID8024317
Terbufos	13071-79-9	DTXSID2022254
Terbuthylazine	5915-41-3	DTXSID4027608
tert-Butyl alcohol	75-65-0	DTXSID8020204
Tetraconazole	112281-77-3	DTXSID8034956
Thiabendazole	148-79-8	DTXSID0021337
Thiamethoxam	153719-23-4	DTXSID2034962
Thiobencarb	28249-77-6	DTXSID6024337
Thiram	137-26-8	DTXSID5021332

Chemical Name	CASRN	DTXSID
Tolyltriazole	29385-43-1	DTXSID0026171
Triadimefon	43121-43-3	DTXSID3023897
Triadimenol	55219-65-3	DTXSID0032493
Tri-allate	2303-17-5	DTXSID5024344
Tributyl phosphate	126-73-8	DTXSID3021986
Triclopyr	55335-06-3	DTXSID0032497
Triclosan	3380-34-5	DTXSID5032498
Trifloxystrobin	141517-21-7	DTXSID4032580
Trifluralin	1582-09-8	DTXSID4021395
Triphenyl phosphate	115-86-6	DTXSID1021952
Tris(2-butoxylethyl) phosphate	78-51-3	DTXSID5021758
Tris(2-chloroethyl) phosphate	115-96-8	DTXSID5021411
Triticonazole	131983-72-7	DTXSID0032655
Tungsten	7440-33-7	DTXSID8052481
Vanadium	7440-62-2	DTXSID2040282
Zinc	7440-66-6	DTXSID7035012

Note: Asterisk (*) indicates publicly nominated chemical contaminants. See Sections 3.3.2 and 3.4 for a description of the point assignment process and calculation of screening scores for each chemical.

Appendix E – Protocol for the Rapid Systematic Health Effects Literature Review

The CCL 6 health effects rapid systematic review (RSR) was based on the RSR first developed as part of the CCL 5 process (USEPA, 2022a) and peer-reviewed by the EPA Science Advisory Board (SAB) (USEPA, 2022b). The focus of this RSR was on identifying animal toxicity studies with dose-response data relevant to oral exposure to chemical contaminants. The RSR for supplemental health effects information for CCL 6 was divided into five steps:

Step 1: Literature identification
Step 2: Title-abstract screening
Step 3: Full text review
Step 4: Study quality evaluation
Step 5: Data extraction

Depending on the available literature for a chemical (i.e., if no studies met the inclusion criteria described in steps 2 and 3 below), the RSR process could be concluded after steps 2, 3, 4, or 5. A conclusion of the RSR process at any stage of the RSR protocol does not equate to the removal of the chemical from consideration for listing on the CCL 6; any RSR data and all available data from other portions of the CCL process are still presented to chemical evaluators for their use when considering chemicals for potential listing on the CCL.

The CCL 6 RSR was conducted for non-pesticide, non-pharmaceutical, and non-nutrient contaminants that were not listed on the PCCL 5 (i.e., did not have RSR results collected under the CCL 5 process). The following protocol outlines the systematic identification of supplemental health effects information as part of the classification process of CCL 6. Refer to Section (4.2.2) of the text for additional information.

Step 1: Literature identification

Health Assessment Identification

The RSR process leveraged publicly available peer-reviewed toxicity assessments to set the literature search date for PCCL 6 chemicals.

The EPA identified the most recently published qualifying health assessments that provided information on health effects resulting from oral exposures. Qualifying health assessments, as defined in Section 4.2.2.2, for PCCL 6 chemicals included:

- Agency for Toxic Substances and Disease Registry (ATSDR) – Toxicological Profiles
- California Environmental Protection Agency (CalEPA) Office of Environmental Health Hazard Assessment (OEHHA) – Public Health Goals
- European Food Safety Authority (EFSA) – Scientific Output Publications
- EPA Office of Pollution Prevention and Toxics (OPPT) – Risk Evaluations
- EPA Office of Research and Development (ORD), Integrated Risk Information System (IRIS) – Chemical Assessment Summaries or Toxicological Reviews

- EPA Office of Water (OW) – Drinking Water Health Advisories or Health Effects Support Documents
- EPA ORD – Provisional Peer-Reviewed Toxicity Values (PPRTVs)
- Health Canada (HC) – Guidelines for Canadian Drinking Water Quality
- World Health Organization (WHO) – Guidelines for Drinking Water-Quality

If a chemical had at least one assessment from one of the agencies listed above, the date limit for the peer-reviewed literature search for that chemical was set to one year prior to the publication date of the most recent assessment. Literature searches for chemicals without an existing assessment from at least one of the agencies listed above were not date limited.

The relevant risk assessment documents and search date limits for chemicals that underwent the RSR process under CCL 6 are provided in Table E-1.

Table E-1. Date limits for CCL 6 RSR literature searches

Chemical	Search Date	Search Date Limit	Most Recent Qualifying Assessment
1,1-Dichloroethane	7/10/2024	8/1/2014	ATSDR, August 2015
1,2-Diphenylhydrazine	8/8/2024	10/1/2019	ATSDR, October 2020
1-Methylnaphthalene	8/12/2024	3/1/2023	PPRTV, March 2024
2,4,6-Trinitrotoluene	8/12/2024	6/1/1994	ATSDR, June 1995
5-Methyl-1H-Benzotriazole	8/9/2024	Unlimited	None
6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	8/27/2024	Unlimited	None
Acrylonitrile	1/16/2025	11/1/1990	IRIS, November 1991
Aluminum	9/3/2024	3/1/2020	HC, March 2021
Argon	9/5/2024	Unlimited	None
Bicarbonate	9/30/2024	5/1/2013	EFSA, May 2014
Bromide	8/9/2024	1/1/2008	WHO, January 2009
Carbon	8/9/2024	Unlimited	None
Carbon Dioxide	8/9/2024	Unlimited	None
Methane	8/26/2024	Unlimited	None
Nitrobenzene	1/16/2025	1/1/2023	ATSDR, January 2024
Phosphate/Orthophosphate	1/16/2025	6/1/2022	PPRTV, June 2023
RDX (Hexahydro-1,3,5-trinitro-1,3,5-triazine)	7/23/2024	11/1/2019	HC, November 2020
Silver	9/4/2024	7/1/2020	EFSA, July 2021; WHO, July 2021
Strontium	7/17/2024	5/1/2018	HC, May 2019
Tert-butyl alcohol	8/22/2024	8/1/2020	IRIS, August 2021
Triphenyl phosphate	12/6/2024	8/1/2019	OPPT scoping document, August 2020

Identification of Primary Literature

The next portion of the literature identification step included searches for primary literature of human and animal studies that is peer-reviewed and related to oral exposure to the PCCL 6 chemicals of interest. To ensure that all relevant literature for each chemical was captured, the EPA first curated a list of search synonyms for each chemical using two databases: the CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard) and PubChem (https://pubchem.ncbi.nlm.nih.gov/). The EPA used search terms previously curated for the CCL 5 process, when possible. The Chemicals Dashboard was searched using DSSToxIDs previously assigned for each chemical (Section 2.3.3). The active CASRN retrieved from this DSSToxID search was then used to search the PubChem database. All available synonyms from both databases were collected and considered for inclusion in the search string. Only synonyms classified as "valid" or "good" according to criteria defined by (Williams, et al., 2017), which involve manual curation or algorithmic generation by systematic naming software in the case of "valid" and consensus across public databases in the case of "good," were included in the search string. Duplicate and ambiguous synonyms were reviewed by a subject matter expert and removed prior to conducting the literature search.

A comprehensive search of peer-reviewed literature was conducted in PubMed and Web of Science using the search terms curated for each chemical. The "tox" filter in PubMed was used to target studies with health effects data in humans and animals. Corresponding search strings were developed for Web of Science searches and were limited to relevant research areas to reduce off-topic hits. These research areas included:

• Allergy • Anatomy & morphology • Audiology & speech language pathology • Behavioral sciences • Cardiovascular system & cardiology • Critical care medicine • Dentistry, oral surgery & medicine • Dermatology • Developmental biology • Emergency medicine • Endocrinology & metabolism • Gastroenterology & hepatology • General & internal medicine • Genetics & heredity • Geriatrics & gerontology • Hematology • Immunology • Infectious diseases • Neurosciences & neurology • Nutrition & dietetics • Obstetrics & gynecology • Oncology • Ophthalmology • Orthopedics • Otorhinolaryngology • Pathology • Physiology • Psychiatry • Public, environmental & occupational health • Reproductive biology • Respiratory system • Rheumatology • Toxicology • Urology & nephrology

Filters for English references were used for searches conducted in both databases. An example search string for 6-acetyl-1,1,2,4,4,7-hexamethyl is provided in Table E-2. Duplicate references across the two databases were removed.

Table E-2. Example RSR peer-review search criteria for 6-Acetyl-1,1,2,4,4,7-hexamethyl

Search date	Description	PubMed	Web of Science	Total Unique
8/27/24	6-Acetyl-1,1,2,4,4,7-hexamethyl AND Tox Filter	326	66	333

Set	Search strategy for PubMed	Results
6-Acetyl-1,1,2,4,4,7-hexamethyl	("21145-77-7"[rn] OR "AHMT" OR "Fixolide" OR "Tonalid" OR "Tonalide" OR "acetyl methyl tetramethyl tetralin"[supplementary concept] OR "AHTN musk"[tiab:~5])	384
	Limits: No date limit; English only	
Set	**Search strategy for Web of Science**	**Results**
6-Acetyl-1,1,2,4,4,7-hexamethyl	("1-(3,5,5,6,8,8-Hexamethyl-5,6,7,8-tetrahydronaphthalen-2-yl)ethan-1-one" OR "1-(3,5,5,6,8,8-Hexamethyl-5,6,7,8-tetrahydronaphthalen-2-yl)ethanone" OR "1-(5,6,7,8-tetrahidro-3,5,5,6,8,8-hexametil-2-naftil)etan-1-ona" OR "1-(5,6,7,8-Tetrahydro-3,5,5,6,8,8-hexamethyl-2-naphthyl)ethan-1-on" OR "1-(5,6,7,8-tetrahydro-3,5,5,6,8,8-hexamethyl-2-naphthyl)ethan-1-one" OR "1-(5,6,7,8-Tetrahydro-3,5,5,6,8,8-hexamethyl-2-naphtyl)ethan-1-one" OR "1-(5,6,7,8-tetrahydro-3,5,5,6,8,8-hexamethyl-2-naphtyl)ethane-1-one" OR "21145-77-7" OR "6-Acetyl-1,1,2,4,4,7-hexamethyltetralin" OR "6-Acetyl-1,1,2,4,4,7-hexamethyltetraline" OR "7-Acetyl-1,1,3,4,4,6-hexamethyl-1,2,3,4-tetra-hydronaphthalene" OR "7-Acetyl-1,1,3,4,4,6-hexamethyltetralin" OR "7-Acetyl-1,1,3,4,4,6-hexamethyltetraline" OR "Acetylhexamethyltetrahydronaphthalene" OR "AHMT" OR "Etanona, 1-(5,6,7,8-tetrahidro-3,5,5,6,8,8-hexametil-2-naftalenil)-" OR "Ethanone, 1-(5,6,7,8-tetrahydro-3,5,5,6,8,8-hexamethyl-2-naphthalenyl)-" OR "ETHANONE, 1-(5,6,7,8-TETRAHYDRO-3,5,5,6,8,8- HEXAMETHYL-2-NAPHTHALENYL)-" OR "Extralide" OR "Fixolide" OR "NSC 19550" OR "TETRAHYDRONAPHTHALENE, 7-ACETYL-1,1,3,4,4,6-HEXAMETHYL-" OR "Tonalid" OR "Tonalide" OR "DTXSID7041544" OR "Acetyl hexamethyl tetralin" OR "AHTN musk")	469
	Limits: No date limit; English only, select Research Areas	

Consistent with methods outlined in the *ORD Staff Handbook for Developing IRIS Assessments* (USEPA, 2022c), the EPA used SWIFT-Review, a software developed by Sciome (Howard et al., 2016; https://sciome.com/swift-review/), to refine the body of literature to only the most relevant studies based on evidence stream. This refinement included statistical text mining and machine learning methods (e.g., meta-data filters searching for terms based on MeSH Terms and MeSH Supplementary Concept Records) applied to the identified literature to categorize studies by human and animal evidence streams (i.e., studies tagged “human”, “animal (all)”, “animal (human health models)”, and “no tag”). Studies prioritized by SWIFT-Review were subject to title-abstract screening, described in Step 2.

The review was terminated for contaminants with greater than 50,000 literature search results prior to deduplication. The review was also terminated for contaminants with greater than 10,000 search results after deduplication and SWIFT Review filtering. For example, several PCCL 6 contaminants (i.e., carbon dioxide, carbon, methane, and silver) each returned more than 10,000 search results from one or more literature databases. For these contaminants, SWIFT Review filters were applied to all unique references and the post-SWIFT Review filtering literature search results remained over 10,000. Each review was terminated as a result. The EPA determined that the large number of search results, many of which were likely to not be relevant to the goals of this systematic review, were too high to be amenable to the screening, review, and extraction steps of the RSR protocol and exceeded the scope and purpose of this effort.

Step 2: Title-abstract screening

The EPA defined Population, Exposure, Control, and Outcome (PECO) criteria (Table E-3) to determine relevance for the title-abstract screening (Step 2) and full text reviews (Step 3). Epidemiologic studies with human health effects data were also identified in title-abstract screening and catalogued for future review but did not move forward to full text review. The number of epidemiologic studies identified during title-abstract screening was documented in the CIS as described in Section in humans associated with4.5.3. Studies solely describing health effects in humans associated with chemical exposure are not amenable to the RSR process due to the complexity of epidemiologic data and the level of effort required to perform study quality evaluation, extract relevant results, and derive toxicity values. Therefore, aside from identifying studies as epidemiologic, further review and evaluation of studies in humans were not conducted.

Table E-3. Population, Exposure, Comparator, Outcome (PECO) criteria for inclusion at the title-abstract and full text levels of the CCL 6 rapid systematic review (RSR)

PECO element	Evidence
Populations	Nonhuman mammalian animal species (whole organism) of any life stage (including preconception, in utero, lactation, peripubertal, and adult stages). Limited to the following mammalian species only: mice, rats, rabbits, guinea pigs, dogs, and monkeys. Studies evaluating hazard in animals with a gene knock-out should be included. Epidemiologic studies on the target chemical will be excluded and tracked separately.
Exposures	**Relevant Chemical Forms [a]:** **Animal:** Controlled exposure to the chemical of interest via oral routes. Any exposure length or route is acceptable for reproductive or developmental exposures. All other study designs require an exposure duration of 28 days or more (if not stated, include at title-abstract screening). Studies must include at least 2 chemical exposure levels not including the control group. Studies involving exposures to mixtures will be included only if animals are exposed exclusively to the relevant chemical with at least 2 exposure levels.
Comparators	**Animal:** A concurrent control group exposed to vehicle-only treatment or an untreated control.
Outcomes	All health outcomes (both cancer and non-cancer; see list under Step 5, including clinical chemistry endpoints). Meta-analysis presenting hazard findings from a compilation of existing literature should be included. Studies evaluating changes in organ morphology, even if the study design is targeted at evaluating protective effects, should be included.

[a] Relevant chemical forms = identifiers or synonyms for a specific chemical.

During the title-abstract screening, reviewers tagged references based on relevance to the animal hazard PECO statement. Two independent reviewers screened and tagged each reference. A senior tertiary reviewer resolved tagging conflicts between reviewers as needed. Studies with PECO-relevant animal hazard information were tagged as "include" and proceeded to full text review. The "supplemental" tag was applied when at least one reviewer tagged accordingly and was confirmed by the senior tertiary reviewer if there were discrepancies between the primary reviewers' supplemental tags. Tag categories and their descriptions are provided in Table E-4.

Table E-4. Tag descriptions for title-abstract screening

Category	Description
Included	**Animal Hazard:** Reference meets PECO criteria in Table E-3.
Excluded	Reference does not meet PECO. **Human Hazard:** Epidemiologic studies on the target chemical will be excluded and tracked separately (i.e., "Human Hazard").
Supplemental	Add this tag if the study contains any of the following types of information (supplemental tags may be applied to included or excluded studies, but some supplemental tags may lead to exclusion—e.g., not relevant literature): **Mechanistic, *in vitro*, and *in silico***: data on mode of action (e.g., oxidative stress, genotoxicity, DNA/RNA/protein inductions, bioinformatics), transgenic animal model or *in vitro* (e.g., cells, tissues, biochemical reactions) studies, and *in silico*/modeling toxicity studies. **Toxicokinetics (TK)**: includes TK or physiologically based pharmacokinetic (PBPK) models; data on mammalian absorption, distribution, metabolism, or excretion (ADME). **Non-PECO exposure**: non-PECO exposure route (e.g., inhalation, dermal, injection for non-developmental/reproductive studies) and duration < 28 days outside the context of reproductive or developmental exposures. Studies with the target chemical administered at a single exposure level, as a mixture, or in a non-PECO vertebrate. Do not use this tag if the reference is a non-oral exposure in a developmental/reproductive study. **Exposure only:** contains only data on human exposure (e.g., biological matrices, predicted or occupational exposures) or measures target chemical in relevant human exposure matrices (e.g., food, drinking water, air). **Not relevant literature**: case reports or case series that reports data for a small number of individuals without a comparison group (human only). Secondary data source (e.g., reviews, commentaries, editorials) with hazard data for humans or other mammals. Non-English language references and references that are abstracts only.

Reviewers prioritized studies during title-abstract screening using the SWIFT-Active machine learning tool (https://www.sciome.com/swift-activescreener/). This tool develops an algorithm to prioritize references based on title-abstract screening tag results (i.e., included studies) to push references with a higher likelihood of inclusion toward the top of the screening queue, which results in capturing at least 95% of the predicted relevant literature, similar to manual screening efforts (Howard et al., 2020). References most likely to be relevant to PECO criteria are prioritized and provided to screeners for review first. For this RSR, only references that met the animal hazard PECO criteria were used to train machine learning models. Screening was considered complete when one of the following conditions was met:

- SWIFT-Active predicted that 95% of relevant references were identified (for contaminants with greater than 1,000 search results),
- SWIFT-Active predicted that >80% of relevant references were identified and at least 300 references were screened as not relevant since the last include (for contaminants with greater than 1,000 search results),
- 1,000 references were screened without an identified PECO-relevant study, or
- All references were screened.

In special cases, screening may be stopped prior to meeting one of these conditions. For example, in CCL 5, screening of quinoline and 1-butanol were stopped when the machine learning algorithm reached a predictive plateau. Plateaus are characterized by rapidly diminishing returns with respect to the level of effort required to identify additional relevant references (i.e., there was a large number of unscreened references left, but the model predicted that very few relevant references remained). In another case, the review of bisphenol A, screening was abandoned at the title-abstract screening step because the inclusion rates were too high (e.g., approximately 230 PECO-relevant studies) to be amenable to the screening, review, and extraction steps of the RSR protocol. Similarly, in CCL 6, the review for phosphate/orthophosphate was temporarily halted because the original reference list was found to contain a high number of studies focused on biological forms of phosphate (e.g., adenosine triphosphate, inositol phosphate, etc.) or other commonly studied phosphate-containing substances (e.g., triphenyl phosphate, tris(chloropropyl) phosphate, etc.), indicating the literature search had inadvertently captured off-topic chemical terms. In the case of phosphate/orthophosphate, librarians employed the Keyword Analysis Tool (KAT), a tool developed by ICF, that is used when off-topic terms skew literature search results. The KAT allowed for the removal of approximately 800 references that were identified to only contain terms related to biological forms of phosphate or other phosphate-containing substances and screeners were able to begin title-abstract screening for relevant phosphate/orthophosphate references.

Step 3: Full text review

Full text reviews were conducted in EPA's Health Assessment Workspace Collaborative (HAWC) software, a modular web-based interface that facilitates development of human health assessments of chemicals (https://www.epa.gov/risk/health-assessment-workspace-collaborative-hawc). References identified as "include", or relevant, during title-abstract screening were subject to a full text review comprised of two primary reviewers, and conflicts were reviewed by a senior quality control reviewer. The animal hazard PECO criteria (Table E-3) were again used to confirm reference relevancy. EPA did not evaluate excluded studies identified as supplemental past the full text review phase. Included studies were eligible for supplemental tags when applicable.

Step 4: Study quality evaluation

EPA conducted study quality evaluations in HAWC for each reference determined to meet the animal hazard PECO criteria at the full text review step. Reviewers employed four metrics to evaluate study quality to ensure each reference used or had i) an accurate and relevant chemical exposure, ii) a non-biased and fully-reported outcome assessment, iii) minimal confounding factors, and iv) any additional concerns not covered by any other metric. These metrics were developed for CCL 5, prior to the publication of the final *ORD Staff Handbook for Developing IRIS Assessments* (USEPA, 2022c), but are abbreviated versions of the domains currently used by the Agency in systematic reviews. Reviewers scored each metric as either Good, Adequate, Deficient, or Critically Deficient and provided a justification highlighting major strengths and concerns for each study. A complete description of study quality metrics and scoring is provided in Table E-5.

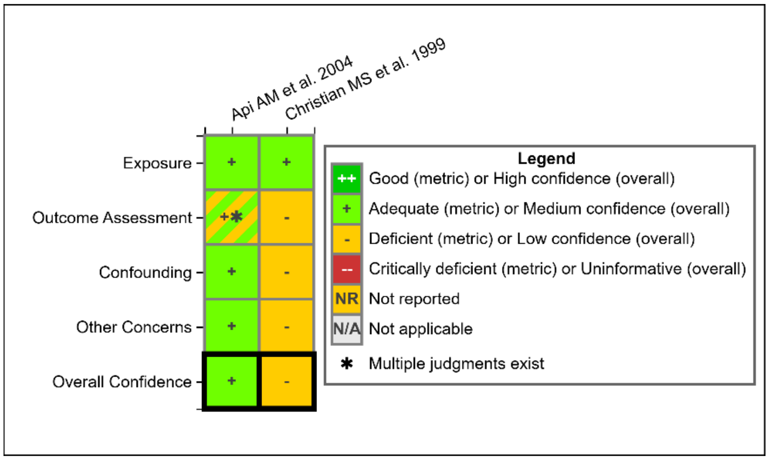
Table E-5. Study Quality Evaluation Metrics

Metric 1 – Exposure
References should be evaluated for the following components of exposure characterization, administration, and timing:
Exposure Characterization • Source and/or CAS registry number of the administered chemical is reported. • Purity of the chemical is reported.
Chemical Administration • Homogeneity and stability in the vehicle were reported or are not a concern. • Methods indicated the chemical was administered correctly and consistently within groups.
Timing • An appropriate window of exposure was used for the outcome of interest. • When animals were dosed through drinking water or diet, a rate of consumption was monitored or estimated.
Metric 2 – Outcome Assessment
References should be evaluated for the following components of outcome evaluation, reporting, and statistical analysis:
Outcome Evaluation • Methods of outcome assessment are well reported, sensitive, and appropriately applied. • Outcomes were assessed consistently across exposure groups. • Assessors were blinded to exposure status for subjective outcomes.
Results Presentation/Statistical Analysis • Number of animals used is presented for each exposure group and is sufficient to assess outcomes (typically 10 animals/group). High attrition in an exposure group is a concern when it results in an insufficient sample size for assessing an effect or implies that the outcome assessment may be impacted by severe toxicity (e.g., neurological evaluation on moribund animals). • Outcome data are presented with means and a measure of variance for continuous endpoints. For dichotomous endpoints, incidence is reported for each exposure group. • Results were presented separately for sex and age (if relevant).
Note: Examples of PECO-relevant health outcomes are listed in Table E-7, which include clinical chemistry and histopathology endpoints.
Metric 3 – Confounding
References should be evaluated for the following potential sources of confounding:
Animal Allocation and Attrition • A randomized, computerized, or weighted allocation method was used to assign animals to groups in an unbiased manner. • No concerns related to high attrition that indicate a health concern across the population (e.g., high attrition in controls indicating a virus in colony) or discrepancies in dose administration (e.g., gavage error deaths limited to a single dose group).

Animal Husbandry • Test animal characteristics were reported (e.g., species, age, weights) and consistent between controls and exposed animals. • Animal housing details were provided and indicated uniform conditions. • No concerns related to animal handling (e.g., lack of vehicle controls in a gavage study).
Metric 4 – Other Concerns
If there are other concerns regarding the study not covered by the above criteria, please state them here with a detailed description of the concern and the potential impact on confidence in the results of the study. If there are no concerns, select a score of **Adequate** and include a comment stating, "No other concerns."
Overall Score
Considering the identified strengths and limitations, provide an overall confidence rating for the study. The overall score should reflect overall confidence in the study as defined below, which is not a simple sum of individual metric scores. • A rating of **High** should be used when the study fully reports all information considered in Metrics 1-3 and presents no concerns or uncertainties. • A rating of **Medium** should be used when there are minor limitations or uncertainties, which could be reflected in a **deficient** score in one metric. Most studies are anticipated to have an overall rating of **Medium**. • A rating of **Low** should be used in cases where both Exposure and Outcome Assessment metrics were scored **deficient** or there were serious concerns about a single metric that call into question the reliability of the study. A **Low** overall score indicates that caution should be used when considering data from that study. • A score of **Uninformative** indicates that a study has serious flaws that make it not usable for the assessment. If any metric is rated **critically deficient** the overall score should be **Uninformative**.

Figure 1 is an example of the output of the study quality assessment conducted in HAWC for 6-acetyl-1,1,2,4,4,7-hexamethyl. In the case of 6-acetyl-1,1,2,4,4,7-hexamethyl, two studies passed the full text review and were evaluated for study quality. In the static versions, these heatmaps indicate the scores (i.e., Good (++), Adequate (+), Deficient (-), Not Reported (NR), and Critically Deficient (--)) for each study quality metric (Exposure, Outcome Assessment, Confounding, Other Concerns) as well as the Overall Confidence (i.e., High (++), Medium (+), Low (-), and Uninformative (--)) for each included study using different colors to visually represent the quality of the chemical's evidence base.

Figure 1. Example Study Quality Evaluation Heatmap from PCCL 6 chemical 6-Acetyl-1,1,2,4,4,7-hexamethyltetralin



Step 5: Data Extraction

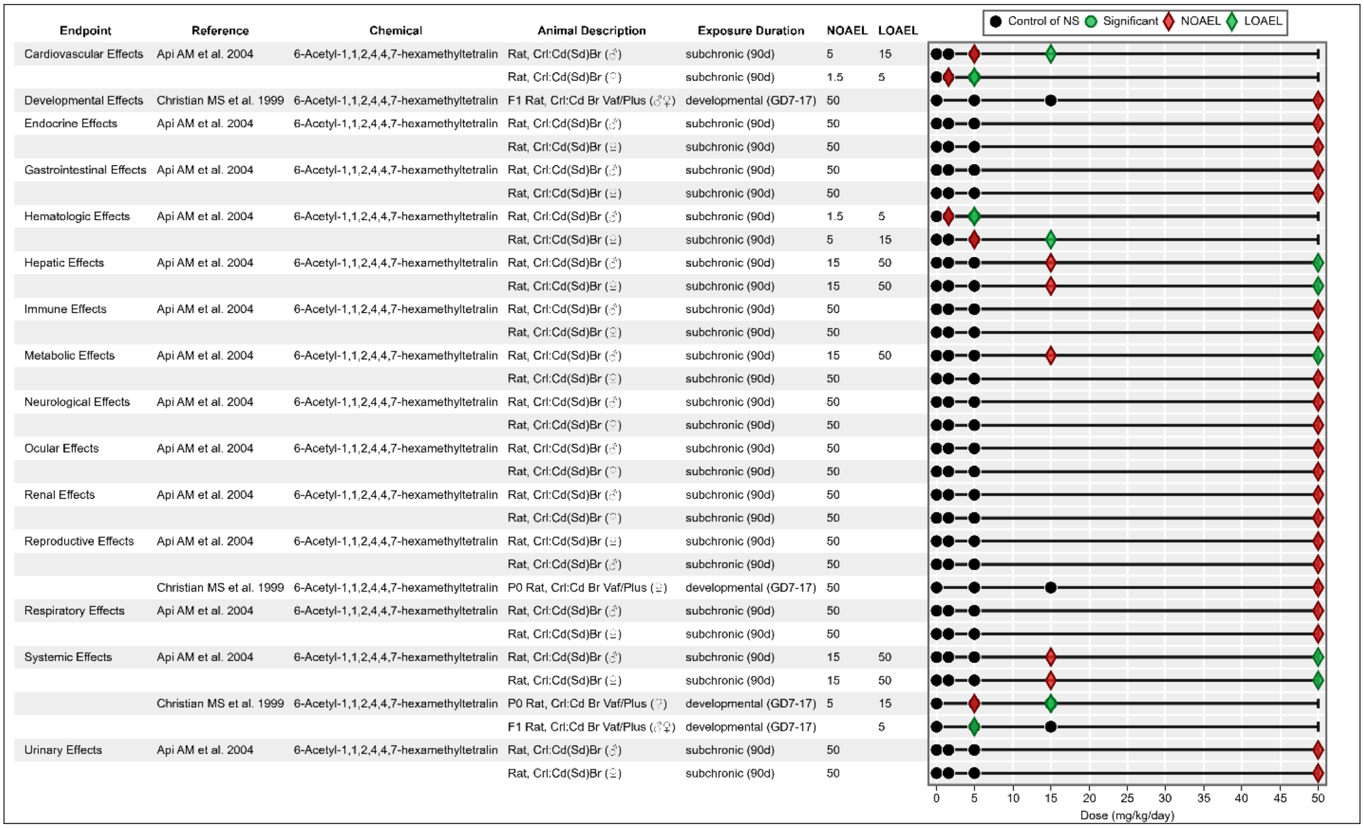
Studies with an overall score of High, Medium, or Low confidence proceeded to data extraction while studies with an overall score of Uninformative were removed from further review. A senior toxicologist reviewed and either confirmed or modified these scores prior to progression to the data extraction step.

EPA used HAWC to conduct a simple extraction of animal hazard data and capture the LOAEL and NOAEL at the "health outcome category" level. A senior toxicologist performed quality control and reviewed each extraction for accuracy and completion. Extractions were conducted at the health outcome category level so that all endpoints within a given health outcome were extracted collectively by animal group. The EPA considered mechanistic data as supplemental data and therefore did not complete data extraction for these endpoints. The possible health outcome categories are listed in Table E-7.

Reviewers also extracted details related to study design (i.e., species, strain, sex, generation, sample sizes, and lifestage of each treatment group), chemical exposure information (i.e., chemical source, purity, vehicle, route of exposure, controls, dose groups, and duration of exposure), target system/organs, and all associated endpoints. Within a health outcome, the lowest LOAEL and NOAEL across endpoints were quantified. For dietary or drinking water studies that ONLY provided the dose concentration in the oral exposure medium (e.g., mg/L, ppm), the dose concentrations as reported in the study were then converted into mg/kg-day dose concentrations using a conversion spreadsheet.

Data pivots were created in HAWC to summarize the findings across references for each chemical. See Table E-6 for an example pivot; HAWC data pivots typically provide high-level information on the test species and strain, exposure duration, endpoint(s), and doses administered in each included study. The graphic uses various symbols and colors to indicate doses and the significance of responses (e.g., a green diamond indicates a LOAEL, a black circle indicates a statistically non-significant response). Examples of extracted health outcomes are provided in Table E-7.

Table E-6. Example Health Effect Category Pivot Table from PCCL 6 Chemical 6-Acetyl-1,1,2,4,4,7-hexamethyltetralin



Endpoint	Reference	Chemical	Animal Description	Exposure Duration	NOAEL	LOAEL
Cardiovascular Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	5	15
			Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	1.5	5
Developmental Effects	Christian MS et al. 1999	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	F1 Rat, Crl:Cd Br Vaf/Plus (♂♀)	developmental (GD7-17)	50	
Endocrine Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	50	
			Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	50	
Gastrointestinal Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	50	
			Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	50	
Hematologic Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	1.5	5
			Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	5	15
Hepatic Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	15	50
			Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	15	50
Immune Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	50	
			Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	50	
Metabolic Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	15	50
			Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	50	
Neurological Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	50	
			Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	50	
Ocular Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	50	
			Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	50	
Renal Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	50	
			Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	50	
Reproductive Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	50	
			Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	50	
	Christian MS et al. 1999	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	P0 Rat, Crl:Cd Br Vaf/Plus (♀)	developmental (GD7-17)	50	
Respiratory Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	50	
			Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	50	
Systemic Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	15	50
			Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	15	50
	Christian MS et al. 1999	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	P0 Rat, Crl:Cd Br Vaf/Plus (♀)	developmental (GD7-17)	5	15
			F1 Rat, Crl:Cd Br Vaf/Plus (♂♀)	developmental (GD7-17)		5
Urinary Effects	Api AM et al. 2004	6-Acetyl-1,1,2,4,4,7-hexamethyltetralin	Rat, Crl:Cd(Sd)Br (♂)	subchronic (90d)	50	
			Rat, Crl:Cd(Sd)Br (♀)	subchronic (90d)	50	

Table E-7. Health Effect Categories and Examples of Endpoints

Health Outcome	Example Endpoints
Neoplasms	Histopathology
Cardiovascular	Serum lipids Histopathology Organ weight Liver function-blood clotting Blood pressure and heart rate
Dermal	Skin/dermal irritation Histopathology
Developmental	Fetal growth restriction Gestational duration Fetal loss Birth defects Maternal body weight Offspring body weight Offspring mortality

Health Outcome	Example Endpoints
	Placenta effects Offspring organ weight Developmental timing Structural abnormalities Mammary gland development Lactation index Organ maturation
Endocrine	Steroid and adrenal hormones Thyroid and thyroid-related hormones and thyroid disease Thyroid hormone antibodies Organ weights Histopathology
Gastrointestinal	Inflammatory bowel disease Organ weight Histopathology
Hematological	Anemia and whole blood hemoglobin Serum electrolytes Serum iron
Hepatic	Serum protein Serum biomarkers of hepatic injury Liver disease or injury Histopathology Organ weight
Immune	Autoimmune disease Immune hypersensitivity Organ weight Immune cellularity Immune response Histopathology
Metabolic Systemic	Insulin Adiponectin and leptin Metabolic syndrome Glucose homeostasis Body mass composition
Musculoskeletal	Size measures Bone health Muscular function Bone parameters
Nervous	Organ weight Behavior

Health Outcome	Example Endpoints
	Histopathology Neurotransmitters Electrophysiology
Ocular	Histopathology Eye disease Ocular Function
Renal	Urinary biomarkers Uric acid Glomerular pore size Glomerular filtration rate Chronic kidney disease Renal inflammation Kidney weight and size Histopathology Serum biomarkers of renal injury Urinalysis
Reproductive	Female reproductive hormones Anogenital distance Female fertility indicators Organ weight Histopathology Estrous cyclicity Ovarian function Female pubertal development Male pubertal development Male reproductive hormones Semen parameters Organ weight Histopathology Male mating and fertility
Respiratory	Pulmonary inflammation Histopathology Organ weight Lung function measures Obstructive disease Lung permeability
Systemic	Body weight Survival and mortality Clinical observations Food and water consumption

Data extracted from relevant studies are summarized on the health effects page of the CIS for each chemical (see Section 4.4 for further descriptions of CISs). For further details related to RSR results for individual chemicals, refer to the file titled "CCL 6 Rapid Systematic Literature Review Results" provided in the CCL 6 docket (EPA-HQ-OW-2022-0946).

References

Howard, B.E., J. Phillips, K. Miller, et al. 2016. SWIFT-Review: a text-mining workbench for systematic review. Syst Rev. 5:87. doi:10.1186/s13643-016-0263-z

Howard B.E., Phillips J., Tandon A., et al. 2020. SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation. Environ Int.;138:105623. doi:10.1016/j.envint.2020.105623.Sciome: SWIFT-Review [Computer software]. Available at: https://www.sciome.com/swift-review/

USEPA. 2022a. Technical Support Document for the Final Fifth Contaminant Candidate List (CCL 5) - Chemical Contaminants. U.S. Environmental Protection Agency, Office of Water, Washington, D.C. https://www.epa.gov/system/files/documents/2022-10/Technical%20Support%20Document%20for%20the%20Final%20CCL%205%20-%20Chemical%20Contaminants.pdf

USEPA. 2022b. EPA-SAB CCL5 Final Report -22-007 (2). Available at: https://www.regulations.gov/document/EPA-HQ-OW-2018-0594-0111

USEPA, 2022c. ORD Staff Handbook for Developing IRIS Assessments. U.S. Environmental Protection Agency, Office of Research and Development. https://iris.epa.gov/Document/&deid=356370#downloads

Williams, A.J., C.M. Grulke., J. Edwards, A.D. McEachran, K. Mansouri, N.C. Baker, I. Shah, J.F. Wambaugh, R.S. Judson, and A.S. Richard. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. J Cheminform 9(1): 61.

Appendix F – Protocol to Derive Health Concentrations

A health concentration is the concentration of a given contaminant in water in which exposure at or below is not expected to result in adverse effects on human health. Health concentrations derived for the CCL process include Health Reference Levels (HRLs) and CCL Screening Levels, which are described further below. HRLs are the optimal type of health concentration developed under CCL 6 because they are derived from toxicity values extracted from qualifying health assessments, which are further described under Step 1 below. CCL Screening Levels are developed only when qualifying health assessments are not available to serve as the basis of an HRL. CCL Screening Levels are derived from toxicity values extracted from either non-qualifying health assessments or information collected during the rapid systematic review (RSR) process, as described in Step 2 below.

The protocol to derive the appropriate health concentration for chemicals of interest is based upon the following three steps:

Step 1: The EPA identified relevant qualifying health assessments and selected the appropriate toxicity value for derivation of the HRL. For CCL 6, qualifying health assessments are those that apply standard methodologies consistent with current EPA guidelines and guidance documents to derive toxicity values for chemical contaminants. Current acceptable human health guidelines and methodologies are found at the EPA's Risk Assessment Guidance webpage (https://www.epa.gov/risk/risk-assessment-guidance#tab-1) and in the specific resources listed below:

- Guidelines for Developmental Toxicity Risk Assessment (USEPA, 1991)
- Guidelines for Reproductive Toxicity Risk Assessment (USEPA, 1996)
- Guidelines for Neurotoxicity Risk Assessment (USEPA, 1998)
- A Review of the Reference Dose and Reference Concentration Processes (USEPA, 2002)
- Guidelines for Carcinogen Risk Assessment (USEPA, 2005a)
- Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens (USEPA, 2005b)
- A Framework for Assessing Health Risks of Environmental Exposures to Children (USEPA, 2006)
- EPA's Exposure Factors Handbook (USEPA 2011a; including individual chapters subsequently updated),
- Recommended Use of Body Weight$^{3/4}$ as the Default Method in Derivation of the Oral Reference Dose (USEPA, 2011b)
- Benchmark Dose Technical Guidance Document (USEPA, 2012)
- Child-Specific Exposure Scenarios Examples (USEPA, 2014a)
- Guidance for Applying Quantitative Data to Develop Data-Derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation (USEPA, 2014b)
- Framework for Human Health Risk Assessment to Inform Decision Making (USEPA, 2014c)
- ORD Staff Handbook for Developing IRIS Assessments (USEPA, 2022a)

For PCCL 6 contaminants, the EPA considered the following internal or external agency health assessments as qualifying assessments:

- EPA Office of Research and Development (ORD) Integrated Risk Information System (IRIS) Chemical Assessment Summaries and Toxicological Reviews
- EPA Office of Water health assessments - Health Advisory (HA) documents and Health Effects Support Documents (HESDs)
- EPA ORD Provisional Peer-Reviewed Toxicity Value (PPRTV) support documents
- EPA Office of Pollution Prevention and Toxics (OPPT) Risk Evaluations
- EPA Office of Pesticide Programs (OPP) Reregistration Eligibility Decision (RED) documents, Interim Reregistration Eligibility Decisions (IREDs), Tolerance Reassessment Progress and Risk Management Decisions (TREDs), Health Effects Division Human Health Risk Assessments (HED HHRAs)
- California EPA (CalEPA) Public Health Goal support documents
- Health Canada (HC) Drinking Water Guidelines support documents
- World Health Organization (WHO) Drinking Water Quality Guidelines documents
- Agency for Toxic Substances and Disease Registry (ATSDR) Toxicological Profiles
- European Food Safety Authority (EFSA) Scientific Output Publications

Websites for these types of health assessments are listed in the references section of this appendix.

For most contaminants, the EPA searched for assessments from all of the sources listed above. For some types of contaminants, the EPA has curated benchmarks that are similar to the health concentrations derived for CCL 6. To ensure consistency across products, reduce redundancy, and address questions raised by public commenters during CCL 5 (USEPA, 2022b), the EPA used the previously curated values rather than re-derive health concentrations.

This was the case for certain pesticides regulated under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA). For all registered pesticides, the EPA relied on the most recent publicly available EPA OPP health assessment and used the population-adjusted dose (PAD, which is a (RfD) equivalent) or CSF provided in that assessment to derive the HRL. However, for pesticides with existing Human Health Benchmarks for Pesticides (HHBPs; https://www.epa.gov/sdwa/2021-human-health-benchmarks-pesticides), the EPA relied on the existing carcinogenic or non-cancer HHBPs as the health concentration rather than calculating a new value. The exception was if a more recent health assessment had been published by the EPA-OPP and had updated toxicity values, in which case the EPA derived an HRL from the newer information. For all pesticides, chronic PADs, if available, were preferentially used over steady-state PADs. An acute PAD could be used if a chronic or steady-state PAD were not available and the health assessment stated that the acute PAD was expected to be protective of chronic toxicity. Justification from the assessment for the use of steady-state PADs or acute PADs was noted on the CISs for the evaluators (see Section 4.4).

If the chemical was a registered pesticide with all uses in the United States canceled but was determined to be persistent per the procedure for TSCA persistence review (see Section 3.7.2), the EPA searched and identified all qualifying health assessment sources for data extraction. Pesticide metabolites or degradates were treated as pesticides only if an OPP assessment assigned or derived toxicity values for those chemicals. If not, the EPA identified toxicity values from other qualifying health assessments, if available, for the derivation of health concentrations.

If a chemical had a single qualifying assessment that provided a toxicity value relevant to chronic oral exposure (e.g., RfDs, CSFs, or equivalents), that assessment was selected as the source of toxicity values for HRL derivation. If a chemical was the subject of multiple qualifying health assessments, the EPA derived HRLs from all assessment sources. The HRL from the most recent EPA assessment was selected as the overall HRL used to derive the final hazard quotient (fHQ) unless a qualifying assessment from another source was more recent and incorporated more current science (e.g., critical studies or newer methods published after the EPA's most recent assessment) in their toxicity value derivations. If multiple assessments were available from the EPA, the most recent assessment was selected. If there were no assessments published by the EPA or if there were multiple assessments presenting more current science than the most recent EPA assessment, the EPA selected the most recent published assessment to derive the HRL.

If the EPA identified a qualifying health assessment containing quantitative toxicity data for a contaminant, the Agency proceeded to Step 3. If no qualifying health assessments were available for a chemical, the EPA proceeded to Step 2.

Step 2: If qualifying health assessments were not available, the EPA identified non-qualifying health assessments and selected the toxicity value most appropriate for derivation of a CCL Screening Level. Alternatively, if neither type of health assessment is available, the EPA identified relevant peer-reviewed studies to use as a source of toxicity values for derivation of a CCL Screening Level.

To differentiate between health concentrations derived from non-qualifying assessment toxicity values (or peer-reviewed studies) and qualifying health assessment toxicity values, the EPA refers to concentrations calculated from non-qualifying health assessments as CCL Screening Levels rather than HRLs. A non-qualifying health assessment is a publicly available assessment published by a health agency that provides relevant health information but does not necessarily follow standard EPA methodologies and/or is not externally peer-reviewed by subject matter experts. The EPA generally does not consider these assessments for regulatory purposes but recognizes that they may provide valuable toxicity information for CCL purposes for chemicals that have no relevant qualifying health assessments available. Both CCL Screening Levels and HRLs can be used to derive the final Hazard Quotient (fHQ) (see Section 4.3.2).

To derive CCL Screening Levels, the EPA searched for any of the following toxicity values in corresponding publicly available non-qualifying health assessments for the chemical of interest:

- RfDs from Minnesota Department of Health Toxicological Summaries and
- Derived No Effect Levels (DNELs) from European Chemicals Agency (ECHA) Registration Dossiers

Websites for these types of health assessments are listed in the references section of this appendix.

Similar to the HHBPs described above, for pharmaceutical contaminants, the EPA has curated benchmarks that are similar to the health concentrations derived for CCL 6, known as the Human Health Benchmarks for Pharmaceuticals (HHB-Rx) in Drinking Water. To ensure consistency across products and reduce redundancy, the EPA used these newly curated values rather than re-derive health concentrations. For pharmaceuticals with existing HHB-Rx (https://www.epa.gov/sdwa/human-health-benchmarks) and no qualifying health assessment available, the EPA relied on the existing non-cancer HHB-Rx for infants as the health concentration rather than calculating a new value.

For non-pharmaceuticals, if a chemical had a single non-qualifying assessment that provided a toxicity value relevant to chronic oral exposure, that assessment was selected as the source of toxicity values for CCL Screening Level derivation. If a chemical had multiple non-qualifying health assessments available, the EPA selected chronic non-cancer or cancer toxicity values (i.e., RfDs, CSFs, or equivalents) from the most recent published assessment for derivation of the CCL Screening Level. If chronic toxicity values were not available, a subchronic or acute non-cancer toxicity value could be used if the assessment provides rationale stating that this value was expected to be protective of chronic effects.

If no non-qualifying health assessments were available, the EPA referenced toxicity values extracted during the rapid systematic literature review (see Section 4.2.2 and Appendix E). During this literature review, the EPA identified No Observed Adverse Effect Levels (NOAELs) extracted from available PECO-relevant studies and extracted NOAELs and their associated critical effects. If no NOAELs were identified, the EPA extracted the overall lowest LOAEL and its associated critical effect. Similar to previous CCL protocols (USEPA, 2009; USEPA, 2022c), a default composite uncertainty factor (UF) of 1,000 was applied to NOAELs and a default composite UF of 3,000 was applied to LOAELs – these values were then used as surrogate RfDs for derivation of CCL Screening Levels.

Step 3: Derive the health concentration.

The process used to derive health concentrations was generally the same as the process used for CCL 5 (USEPA, 2022c). For carcinogens, the health concentration was the one-in-a-million (10^{-6}) cancer risk expressed as a drinking water concentration. For non-carcinogens, health concentrations were obtained by dividing the RfD (or equivalent) by an exposure factor, also known as the bodyweight-adjusted drinking water intake (DWI), relevant to the target population and critical effect (USEPA, 2019a), and multiplying by a relative source contribution (RSC) to account for non-water sources of exposure (USEPA, 2000).

-Table F-1 exhibits the formulae used to derive health concentrations from various data elements. All health concentrations were converted to units of µg/L to compare with CCL 6 occurrence concentrations. If a chemical had no available qualifying or non-qualifying health assessments or studies identified through the rapid systematic review (RSR) process, or if the available health assessments elect not to derive toxicity values, the EPA did not derive a health concentration.

Table F--1. Health Concentration Formulae

Non-Cancer Equations	
From RfD or Equivalent	$\boldsymbol{HRL} = \left(\frac{RfD}{DWI}\right) * RSC$
From NOAEL	$\boldsymbol{HRL} = \left(\frac{NOAEL/1000}{DWI}\right) * RSC$
From LOAEL	$\boldsymbol{HRL} = \left(\frac{LOAEL/3000}{DWI}\right) * RSC$
Cancer Equations	
Linear Carcinogen	$\boldsymbol{HRL} = \frac{1x10^{-6}}{CSF * DWI}$
Non-Linear Carcinogen	HRL derived from non-cancer RfD (or equivalent) is protective of carcinogenicity
Mutagenic Carcinogen	$\boldsymbol{HRL} = \frac{1x10^{-6}}{\sum_i (CSF_i * ADAF_i * DWI_i * F_i)}$

HRL = Health Reference Level; RfD = reference dose; DWI = body weight-adjusted drinking water intake; RSC = relative source contribution; NOAEL = no observed adverse effect level; LOAEL = lowest observed adverse effect level; CSF = cancer slope factor; ADAF = age-dependent adjustment factor; F = fraction of life spent in each age group (i).
Note: concentrations derived from NOAELs and LOAELs extracted from peer-reviewed studies are considered CCL Screening Levels, but are identified here as HRLs for clarity and consistency; final health concentrations are converted to units of µg/L.

Relevant target populations or life stages and their corresponding DWI are presented in Table F-2.

Table F--2. Exposure Factors Used for Derivation of Health Concentrations

Target Population/ Life Stage	DWI (mL/kg-day)	Description of Exposure Metric	Citation
General population	33.8	90th percentile direct and indirect consumption of community water, consumer-only 2-day average, all ages	2019 Exposure Factors Handbook Chapter 3, Table 3-21, NHANES 2005-2010
All infants	143	90th percentile combined direct and indirect drinking water consumption of community water, consumers-only 2-day average, birth to <1 year	2019 Exposure Factors Handbook Chapter 3, Table 3-21, NHANES 2005-2010
Pregnant women	33.3	90th percentile combined direct and indirect drinking water intake of community water, consumers-only 2-day average	2019 Exposure Factors Handbook Chapter 3, Table 3-63, NHANES 2005-2010
Lactating women	46.9	90th percentile combined direct and indirect drinking water intake of community water, consumers-only 2-day average	2019 Exposure Factors Handbook Chapter 3, Table 3-63, NHANES 2005-2010
Females of reproductive age	35.4	90th percentile combined direct and indirect drinking water intake of community water, consumers-only 2-day average, 13 to <50 years	2019 Exposure Factors Handbook Chapter 3, Table 3-63, NHANES 2005-2010

DWI = bodyweight-adjusted drinking water intake; NHANES = national health and nutrition examination survey.

For some types of contaminants, the EPA-OW has curated benchmarks or screening levels that are similar to the health concentrations derived for CCL 6. To ensure consistency across products, reduce redundancy, and address questions raised by public commenters during CCL 5 (USEPA, 2022a), the EPA used these previously curated values to compare to occurrence concentrations rather than re-derive health concentrations.

For pesticides with existing HHBPs (https://www.epa.gov/sdwa/2021-human-health-benchmarks-pesticides), the EPA relied on the existing carcinogenic or non-cancer HHBPs as the health concentration for CCL 6. The exception was if a more recent health assessment had been published by the EPA-OPP and had updated toxicity values, in which case the EPA derived an HRL from the newer information. For pharmaceuticals with existing HHB-Rx (https://www.epa.gov/sdwa/human-health-benchmarks) and no available qualifying health assessment, the EPA relied on the existing non-cancer HHB-Rx for infants as the health concentration for CCL 6. Overall, the methodologies used to derive health concentrations in the CCL process and to derive HHBPs or HHB-Rx are the same, although slight differences exist (e.g., use of updated drinking water consumption rates in CCL 6, different RSC value applied in HHB-Rx) (USEPA, 2021; USEPA, 2026).

In some cases, a PCCL 6 contaminant was considered an essential nutrient by the Institute of Medicine (IOM). These chemicals often do not have qualifying or non-qualifying health assessments and did not fall into the scope of the CCL 6 rapid systematic review. For these contaminants, the EPA identified the relevant IOM Dietary Reference Intake document and extracted the tolerable upper intake level (or equivalent) from that document to present to evaluators. The EPA did not use health concentrations calculated from these values to derive final Hazard Quotients but did provide these values and documents to evaluation teams for their reference.

Toxicity values identified through this process were used to inform and derive several other health effects metrics, including the potency attribute score, and severity category (Section 4.3.3). For each PCCL 6 chemical, all health-related information was compiled and presented on the corresponding CIS. Adaptations of this compilation of health effects data, edited to fit this technical support document, are presented in Table F-3 (non-cancer effects) and Table F-4 (cancer effects). Table F-3 depicts an example health assessment data compilation and derivation of an RfD-based (non-cancer) HRL for tris(1,3-dichloro-2-propyl) phosphate (TDCPP) while Table F-4 depicts an example health assessment data compilation and derivation of a CSF-based (cancer) HRL for triallate. The health concentrations are presented in columns 10 and 11 of Table F-3 and Table F-4, respectively.

Table F--3. Example Health Assessment Data Compilation for Non-Cancer Effects of Tris(1,3-dichloro-2-propyl) phosphate (TDCPP)

1 Name	2 DTXSID	3 Assessment Source	4 Assessment Title (date)	5 RfD (or equivalent) value	6 RfD critical study	7 Critical effect	8 Target population	9 Exposure factor (mL/kg-day)	10 HRL (µg/L, rounded)	11 Notes on non-cancer HRL
Tris(1,3-dichloro-2-propyl) phosphate	DTXSID9026261	ATSDR	Toxicological Profile for Phosphate Ester Flame Retardants. (2012).	0.02 mg/kg-day	Stauffer Chemical Co. 1981	renal tubule epithelial hyperplasia	general population	33.8	100	Chronic oral MRL based on $BMDL_{10}$ of 1.94 mg/kg-day with an uncertainty factor of 10 for animal to human extrapolation and an uncertainty factor of 10 for human variability.

ATSDR = Agency for Toxic Substances and Disease Registry; RfD = Reference dose; HRL = Health reference level; MRL = Minimal risk level; $BMDL_{10}$ = Benchmark dose level associated with a 10% increase in the risk of an effect

Table F--4. Example Health Assessment Data Compilation for Cancer Effects of Triallate

1 Name	2 DTXSID	3 Assessment Source	4 Assessment Title (date)	5 CSF	6 CSF critical study	7 Tumor types or locations	8 Cancer classification	9 Target population	10 Exposure factor (mL/kg-day)	11 HRL (µg/L, rounded)	12 Notes on cancer HRL
Triallate	DTXSID5024344	OPP	Triallate. Human Health Draft Risk Assessment in Support of Registration Review. (2019).	0.0717 $(mg/kg\text{-}day)^{-1}$	Stout et al. 1983	hepatocellular carcinomas in male mice, renal tubular cell adenomas in rats	C	general population	33.8	0.413	Note: The HRL is based on the EPA's HHBP (see https://www.epa.gov/sdwa/2021-human-health-benchmarks-pesticides for more information).

OPP = Office of Pesticide Programs; CSF = cancer slope factor; Cancer classification "C" = possible human carcinogen; HRL = Health Reference Level

In some cases, the health assessment selected as the appropriate source for health concentration derivation provided both cancer and non-cancer toxicity values (e.g., an RfD and CSF). When this situation occurred, the EPA derived health concentrations based on both data elements and selected the most health protective (i.e., lowest value) to serve as the final health concentration to be presented on the summary page of the CIS. Health concentrations on the summary page were subsequently used for derivation of the final hazard quotient (fHQ) as a means of comparing health data to corresponding occurrence data. This process is described in Section 4.3.2 of the main document.

References

Within this section, references that have no associated date (n.d.) contain links to websites that provide relevant documents and health assessments.

ATSDR. 2012. Toxicological Profile for Phosphate Ester Flame Retardants. United States Department of Health & Human Services, Public Health Service, Agency for Toxic Substances and Disease Registry (ATSDR). https://www.atsdr.cdc.gov/toxprofiles/tp202.pdf.

ATSDR. n.d. Toxicological Profiles. United States Department of Health & Human Services, Center for Disease Control, Agency for Toxic Substances and Disease Registry (ATSDR). https://www.atsdr.cdc.gov/toxicological-profiles/about/?CDC_AAref_Val=https://www.atsdr.cdc.gov/toxprofiledocs/

California Environmental Protection Agency (CalEPA). n.d. Public Health Goals. Office of Environmental Health Hazard Assessment. https://oehha.ca.gov/water/public-health-goals-phgs

European Union, European Chemicals Agency (ECHA). n.d. Registration Substances Dossiers. https://chem.echa.europa.eu/

European Union, European Food Safety Authority (EFSA). n.d. Scientific Output Publications. https://www.efsa.europa.eu/en/publications?sort=computed_sort_date&order=desc

Health Canada (HC). n.d. Guidelines for Canadian Drinking Water Quality. Water and Air Quality Bureau, Healthy Environments and Consumer Safety Branch. https://www.canada.ca/en/health-canada/services/environmental-workplace-health/reports-publications/water-quality.html#tech_doc

Institute of Medicine (IOM). n.d. Dietary Reference Intakes. National Academy of Sciences, Engineering, and Medicine, Food and Nutrition Board. https://ods.od.nih.gov/Health_Information/Dietary_Reference_Intakes.aspx.

Minnesota Department of Health (MDH). n.d. Toxicological Summaries. Environmental Health Division, Health Risk Assessment Unit, Human Health-Based Guidance for Water. https://www.health.state.mn.us/communities/environment/risk/guidance/gw/table.html.

USEPA. 1991. Guidelines for Developmental Toxicity Risk Assessment. Risk Assessment Forum. https://www.epa.gov/sites/production/files/2014-11/documents/dev_tox.pdf.

USEPA. 1996. Guidelines for Reproductive Toxicity Risk Assessment. Risk Assessment Forum. https://www.epa.gov/sites/production/files/2014-11/documents/guidelines_repro_toxicity.pdf.

USEPA. 1998. Guidelines for Neurotoxicity Risk Assessment. Risk Assessment Forum. https://www.epa.gov/sites/production/files/2014-11/documents/neuro_tox.pdf.

USEPA. 2000. Methodology for Deriving Ambient Water Quality Criteria for the Protection of Human Health (2000). Office of Water. https://www.epa.gov/sites/default/files/2018-10/documents/methodology-wqc-protection-hh-2000.pdf

USEPA. 2002. A Review of the Reference Dose and Reference Concentration Processes. Risk Assessment Forum. https://www.epa.gov/sites/production/files/2014-12/documents/rfd-final.pdf.

USEPA. 2005a. Guidelines for Carcinogen Risk Assessment. Risk Assessment Forum. https://www.epa.gov/sites/default/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf

USEPA. 2005b. Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens. Risk Assessment Forum. https://www.epa.gov/sites/default/files/2013-09/documents/childrens_supplement_final.pdf.

USEPA. 2006. A Framework for Assessing Health Risks of Environmental Exposures to Children. Office of Research and Development. https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=158363.

USEPA. 2009. Final Contaminant Candidate List 3 Chemicals: Classification of the PCCL to CCL. Office of Water. https://www.epa.gov/sites/default/files/2014-05/documents/ccl3_pccltoccl_08-31-09_508.pdf

USEPA. 2011a. Recommended Use of Body Weight3/4 as the Default Method in Derivation of the Oral Reference Dose. Office of the Science Advisory, Risk Assessment Forum. https://www.epa.gov/sites/production/files/2013-09/documents/recommended-use-of-bw34.pdf.

USEPA. 2011b. Exposure Factors Handbook. Office of Research and Development. https://www.epa.gov/expobox/about-exposure-factors-handbook.

USEPA. 2012. Benchmark Dose Technical Guidance. Risk Assessment Forum. https://www.epa.gov/sites/production/files/2015-01/documents/benchmark_dose_guidance.pdf.

USEPA. 2014a. Child-Specific Exposure Scenarios Examples. Office of Research and Development. https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=262211.

USEPA. 2014b. Guidance for Applying Quantitative Data to Develop Data-Derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation. Office of the Science Advisor, Risk Assessment Forum. https://www.epa.gov/sites/default/files/2015-01/documents/ddef-final.pdf

USEPA. 2014c. Framework for Human Health Risk Assessment to Inform Decision making. Office of the Science Advisor, Risk Assessment Forum. https://www.epa.gov/sites/default/files/2014-12/documents/hhra-framework-final-2014.pdf.

USEPA. 2018. 2018 Edition of the Drinking Water Standards and Health Advisories Table. Office of Water. https://www.epa.gov/system/files/documents/2022-01/dwtable2018.pdf

USEPA. 2019a. Update for Chapter 3 of the Exposure Factors Handbook. Ingestion of Water and Other Select Liquids. Office of Research and Development. https://www.epa.gov/sites/production/files/2019-02/documents/efh_-_chapter_3_update.pdf.

USEPA. 2019b. Triallate. Human Health Draft Risk Assessment in Support of Registration Review. Office of Pesticide Programs. EPA-HQ-OPP-2014-0573-0015.

USEPA, 2021. Human Health Benchmarks for Pesticides: Updated 2021 Technical Document. U.S. Environmental Protection Agency. https://www.epa.gov/system/files/documents/2021-07/hh-benchmarks-technical-document-2021.pdf

USEPA, 2022a. ORD Staff Handbook for Developing IRIS Assessments. U.S. Environmental Protection Agency, Office of Research and Development. https://iris.epa.gov/Document/&deid=356370#downloads

USEPA. 2022b. Response to Comments Document on the Draft Fifth Contaminant Candidate List (CCL 5). U.S. Environmental Protection Agency, Office of Water, Washington, D.C. EPA-815-R-22-001. https://www.epa.gov/system/files/documents/2022-10/Response%20to%20Comments%20Document%20on%20the%20Draft%20Fifth%20Contaminant%20Candidate%20List%20%28CCL%205%29.pdf

USEPA. 2022c. Technical Support Document for the Final Fifth Contaminant Candidate List (CCL 5) - Chemical Contaminants. U.S. Environmental Protection Agency, Office of Water, Washington, D.C. https://www.epa.gov/system/files/documents/2022-10/Technical%20Support%20Document%20for%20the%20Final%20CCL%205%20-%20Chemical%20Contaminants.pdf

USEPA. 2026. Human Health Benchmarks for Pharmaceuticals (HHB-Rx) in Drinking Water. https://www.epa.gov/sdwa/human-health-benchmarks.

USEPA. n.d.-b. Integrated Risk Information System (IRIS). Office of Research and Development. https://cfpub.epa.gov/ncea/iris_drafts/atoz.cfm?list_type=alpha.

USEPA. n.d.-c. Provisional Peer-Reviewed Toxicity Values (PPRTVs). Office of Research and Development. https://www.epa.gov/pprtv

USEPA. n.d.-d. Chemicals Undergoing Risk Evaluation under the Toxic Substances Control Act (TSCA). Office of Chemical Safety and Pollution Prevention. https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/ongoing-and-completed-chemical-risk-evaluations-under.

World Health Organization (WHO). 2022. Guidelines for Drinking-Water Quality (4th Edition incorporating the first and second addenda).
https://www.who.int/publications/i/item/9789240045064

Appendix G – Protocol to Select Water Concentrations Used for Calculating Final Hazard Quotients for PCCL Chemicals

A. Does the chemical have UCMR 1-5 data with greater than 0 detects?

1. Yes → select the concentration in the following order, depending upon availability (the highest concentration is selected if multiple UCMR monitoring results are available):
    - i. 90th percentile
    - ii. maximum
2. No → move on to item B.

B. Does the chemical have NIRS data with greater than 0 detects?

1. Yes → select concentration in the following order, depending on availability:
    - i. 90th percentile
    - ii. maximum
2. No → move on to item C

C. Does the chemical have UCM-State Round 1 or Round 2 data with greater than 0 detects (the highest concentration is selected if multiple Rounds results are available)?

1. Yes → select concentration in the following order, depending upon availability:
    - i. 90th percentile
    - ii. maximum
2. No → move on to item D.

D. Does the chemical have NAWQA (ambient water), NLA or NRSA concentration data with greater than 0 detects?

1. Yes → select highest concentration, depending upon availability:[7]
    - i. 90th percentile
    - ii. maximum
2. No → move on to item E.

E. Is the chemical a pesticide with modeled concentration data from an OPP evaluation?

1. Yes → identify the Estimated Drinking Water Concentration (EDWCs) values that coincide with the critical health effect (acute, chronic, or cancer) and select the value from the use scenario with the highest modeled concentration (surface water or groundwater)

---

[6] In the few cases where NAWQA had very small site counts, NWIS data was used instead if it had a larger site count.

2. No → move on to item F.

F. Does the chemical have non-national finished water concentration data with greater than 0 detects?

1. Yes → select concentration in the following order, depending upon availability:
    a. State/SYR data

        i. If there is data available from multiple states, select the 90th percentile concentration from the state with the most recent data.

        ii. If 90th percentile is not available, choose the maximum.
        iii. If there are multiple states with overlapping monitoring periods (data available from the same period), select the highest 90th percentile concentration value.
        iv. If 90th percentile is not available, choose the maximum.
    b. Disinfection By-products Information Collection Rule (DBP ICR)
        i. 90th percentile
        ii. maximum
    c. NAWQA/NWIS (finished water)
        i. 90th percentile
        ii. maximum
    d. USDA PDP
        i. 90th percentile
        ii. maximum
    e. Individual studies that are main data sources (Bradley et al. 2018, Furlong et al. 2017, Glassmeyer et al. 2017, Zhang et al. 2019) and results from the literature search
        i. If multiple studies provide concentration data for a compound, choose the highest 90th percentile concentration available.
        ii. If 90th concentrations are not available, choose the maximum.
    f. Community Water Systems Survey (CWSS)
        i. 90th percentile
        ii. maximum
2. No → move on to item G

G. Does the chemical have non-national ambient concentration data (surface, ground, source, and untreated water types) with greater than 0 detects?

1. Yes → select concentration in the following order, depending upon availability:
    a. NWIS (ambient water)
        i. 90th percentile
        ii. maximum
    b. CA SURF

        - i. 90th percentile
        - ii. maximum
    - c. USDA PDP
        - i. 90th percentile
        - ii. maximum
    - d. USGS Brackish
        - i. 90th percentile
        - ii. maximum
    - e. Individual studies that are main data sources and results from the literature search
        - i. If multiple studies provide concentration data for a compound, choose the highest 90th percentile concentration available.
        - ii. If the 90th percentile concentrations are not available, choose the maximum value.
2. No → move on to item H.

H. Does the chemical have effluent concentration data from main data sources?

1. Yes → select concentration in the following order, depending upon availability:
    - i. 90th percentile; select the highest 90th percentile concentration if multiple studies are available.
    - ii. maximum; select the highest maximum concentration if multiple studies are available.
2. No → move on to item I.

I. If none of the concentrations listed above are available, a final hazard quotient is not calculated and the fHQ cell on the Summary + Decision tab page of the CIS is left blank.

Appendix H – Protocol to Determine Potency Attribute Scores

This appendix outlines the three-step protocol used to derive and select the final potency score for a chemical of interest. Table H-1 is referenced throughout this protocol and provides an example of the information gathered during the identification of health effects data relevant to the potency attribute score for epoxiconazole.

Table H--1. Health Assessment Data Relevant to the Potency Score Extracted for Epoxiconazole

Chemical	DTXSID	Assessment Source	Toxicity Value (type)	Critical Study	Health Concentration	Potency Equation	Potency Score
Epoxiconazole	DTXSID 1040372	OPP (2005)	0.02 mg/kg-day (RfD)	Mellert et al. (1992)	100 µg/L	Score = -1.82391 – $\log_{10}$(RfD) + 5	5
Epoxiconazole	DTXSID 1040372	OPP (2005)	0.0304 (mg/kg-day)$^{-1}$ (CSF)	Mellert et al. (1992)	0.973 µg/L	Score = -(-0.5376) + $\log_{10}$(CSF) + 5	4

OPP = Office of Pesticide Programs; RfD = Reference dose; CSF = Cancer slope factor

Step 1: Identify toxicity values from available sources of health effects information.

The identification of toxicity values for each contaminant is covered in detail in Appendix F. If available, the EPA first identified toxicity values RfDs, CSFs, etc.) extracted from all published health assessments. When no health assessments existed, the EPA extracted toxicity values (NOAELs or LOAELs) from studies identified through the rapid systematic literature review (see Section 4.2.2). The EPA compiled toxicity values from these sources in a table along with other relevant health information (e.g., critical studies, critical effects). An adapted version of this table depicting the available health information with columns related to the potency of epoxiconazole is provided above (Table H-1).

Step 2: Select the relevant potency equation and derive potency scores for each extracted toxicity value.

The EPA derived a potency score for each toxicity value using the equations listed in Table 14 in Section 4.3.3 of this document. Each type of toxicity value has a different potency scoring equation based on the distribution of available values for that toxicity value in the Universe and subsequent calibration of the equations based on the median of each distribution. The EPA selected the appropriate equation for derivation based on the type of toxicity value presented (i.e., the EPA derived a potency score for an RfD using the potency score equation calibrated for RfDs, which would be different from the potency score equations calibrated for CSFs, NOAELs, or LOAELs; see Section 4.3.3). The EPA did not derive a potency score if toxicity values were not identified for the chemical of interest. Additionally, the EPA did not derive a potency score for the few cases where a health concentration was available without a RfD, CSF, or equivalents (e.g., maximum allowable or threshold concentrations from assessments from Health Canada in which there was no associated RfD or equivalent). Table H-1 lists the available toxicity values for epoxiconazole, each associated potency equation, and the subsequently derived scores.

Step 3: Determine the final potency score.

The EPA selected the potency score that corresponded with the toxicity value used to derive the selected health concentration (i.e., HRL or CCL screening level) to present on the summary page of the CIS. The final potency score might not necessarily be the highest score available for a chemical because the final potency score is associated with the health concentration selected to derive the final hazard quotient (fHQ), the process for which is described in Appendix F. For a given chemical, the final (i.e., selected) health concentration may not always be associated with the highest potency score for a range of reasons. First, the assessment selected as the basis for the final health concentration may not present the most potent toxicity value available for that chemical. The protocol for deriving and selecting health concentrations is available in Appendix F. Additionally, the existence of different potency score equations for each toxicity value type might also result in the selection of an apparent "lower" potency score for one toxicity value type that corresponds to the final health concentration that is more health protective than a health concentration derived from a different toxicity value that has a "higher" potency score.

The epoxiconazole entry presented in Table H-1 shows how a final potency score was selected while illustrating why potency scores from different toxicity value types cannot be compared with each other. In this example, a more health-protective health concentration, 0.973 µg/L derived from the CSF from the 2005 Office of Pesticide Programs (OPP) assessment, was presented over the 100 µg/L derived from the RfD from the same assessment. The CSF is associated with what appears to be a lower potency score (4) than for the RfD (5); however, these values are not comparable because they were derived based on different underlying data and scale calibrations. In this case, the potency score of 4 was selected as the final potency score because it corresponds with the toxicity value used to derive the final health concentration.

References

USEPA (OPP). 2005. Epoxiconazole: PP# 7E04885 and 0E6128. Tolerance on Bananas and Coffee Exported to the US. HED Risk Assessment. PC Code: 123909, Submission #: S581060, Case #: 292957, Decision #: 303871, DP #: 268360. U.S. Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention, Washington, D.C.

Appendix I – Protocol to Determine Severity Attribute Scores

This appendix outlines the steps required to identify the appropriate severity category for a chemical of interest. Table I-1 is referenced throughout the protocol and provides an example of the information gathered during the identification of health effects data relevant to the severity category for epoxiconazole.

Table I--1. Health assessment data extracted for epoxiconazole

Chemical	DTXSID	Assessment Source (date)	Toxicity Value (type)	Critical Study	Critical Effect	Severity Categories	Final Severity
Epoxiconazole	DTXSID 1040372	OPP (2005)	0.02 mg/kg-day (RfD)	Mellert et al. (1992)	Increased incidences of adrenal histopathological findings and increased incidences of ovarian cysts	Reproductive and developmental effects; Non-cancer effects	Reproductive and developmental effects
Epoxiconazole	DTXSID 1040372	OPP (2005)	0.0304 $(mg/kg\text{-}day)^{-1}$ (CSF)	Mellert et al. (1992)	Liver tumors in male and female mice, specifically male mouse liver adenoma and/or carcinoma combined tumor rates were used to determine the CSF	Carcinogen with linear MOA	Carcinogen with linear MOA

OPP = Office of Pesticide Programs; RfD = Reference dose; CSF = Cancer slope factor; MOA = Mode of action

Step 1: Identify the critical effect and corresponding toxicity value for each entry for the chemical contaminant of interest.

The EPA first identified the critical effect underlying the toxicity value of interest (e.g., RfD, CSF) as stated in each available health assessment or study. These critical effects correspond to the same toxicity value used to derive a health concentration (i.e., HRL or CCL screening level, see Section 4.3.1) and potency score (see Section 4.3.3) for that chemical. These critical effects were compiled into a health effects data table along with other relevant health and source document information. An adapted version of this table depicting the available health information of an example chemical, epoxiconazole, is provided above (Table I-1).

Step 2: Determine the appropriate severity category for each critical effect.

Based on the critical effect associated with the toxicity value, the EPA selected the appropriate severity category. Table I-2 lists the eight possible severity categories along with interpretations of the severity categories. The severity categories selected for each chemical and critical effect were reviewed for accuracy and consistency by EPA scientists from the Office of Water (OW). If the assessment or study lists multiple critical effects associated with the toxicity value, the EPA listed each applicable severity category. If there was no available toxicity value or no corresponding critical effect for a chemical, a severity category was not applied, and the entry was left blank. Because the classification into a severity category requires scientific judgement, any discrepancies during review were resolved by discussions among EPA scientists. Table I-1 lists the available toxicity values for epoxiconazole and each associated severity category.

Table I--2. CCL 6 Severity Categories

Severity Categories	Interpretations
No adverse effects	---
Cosmetic effects	Effects that alter appearance without affecting structure or function
Non-cancer effects	Includes transient or adaptive effects, risk factors or precursor effects, disorders in which the removal of exposure will restore health, and non-lethal persistent disorders that do not influence reproduction, development, or gestation
Reproductive and developmental effects	Effects on the developing fetus or infant or on a pregnant individual or effects that impact the ability of a population to reproduce
Carcinogen with linear mode of action	Effects resulting in any type of tumor, except those with a known mutagenic or non-linear mode of action
Carcinogen with non-linear mode of action	Effects resulting in any type of tumor with a known non-linear mode of action; Tumors are unlikely to occur below doses that result in non-carcinogenic effects
Carcinogen with mutagenic mode of action	Effects resulting in any type of tumor confirmed to result from chemical exposure-induced mutagenicity
Reduced longevity	Premature mortality

Generally, if a chemical is associated with effects unrelated to carcinogenicity and has co-critical effects that correspond with several severity categories, the EPA selected one category for that assessment based on the hierarchy of effects listed below:

reduced longevity > reproductive and developmental effects > non-cancer effects > cosmetic effects.

An example of this for epoxiconazole is depicted in Table I-1. In this example, multiple non-cancer co-critical effects are presented for epoxiconazole. One co-critical effect includes “increased incidences of ovarian cysts,” which corresponds with a severity category of “reproductive and developmental effects.” In this case, while there are also critical effects identified by this assessment that would fall into the severity category of “non-cancer effects,” the EPA selected “reproductive and developmental effects” as the severity category related to this assessment.

Step 3: Select the final severity category for the chemical.

In some cases, chemicals are associated with both cancer and non-cancer critical effects or chemicals have multiple assessments presenting different severity categories. For these instances, the severity category that corresponds to the critical effect and associated toxicity value selected to derive the CCL 6 health concentration (see Section 4.3.1) was the final severity category listed on the summary page of the CIS.

Generally, the final severity category corresponds to the most protective health concentration. In the example of epoxiconazole presented in Table I-1, the selected health concentration and potency score were based on the CSF from the 2005 Office of Pesticide Programs (OPP) health assessment (highlighted in yellow). This is attributable to the fact that the CSF resulted in a more health-protective, or lower, health concentration compared to the health concentration derived from the RfD. Therefore, the severity category listed on the summary page of the CIS was “carcinogen with a linear mode of action.”

References

USEPA (OPP). 2005. Epoxiconazole: PP# 7E04885 and 0E6128. Tolerance on Bananas and Coffee Exported to the US. HED Risk Assessment. PC Code: 123909, Submission #: S581060, Case #: 292957, Decision #: 303871, DP #: 268360. U.S. Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention, Washington, D.C.

Appendix J – Protocol to Determine Prevalence Attribute Scores

This section describes how to assign a numerical score for the prevalence attribute.

Step 1: Identify highest-ranked data value

When more than one data value is available for a particular contaminant candidate, use the hierarchy in Table J-1. Use the same type of data to score prevalence as for magnitude.

Table J--1. Hierarchy of Prevalence Data Elements

Rank	Prevalence Data Element	Type of Data
1	Percent of PWSs with detections	National scale / representative data (UCMR 1-5 has highest priority, then NIRS, then UCM-State Rounds 1-2) from the EPA.
2	Percent of ambient water sites or samples with detections	National scale / representative NAWQA data from USGS
3	Number of states reporting application of the chemical as a pesticide	Estimated Annual Agricultural Pesticide Use data from USGS
4	Number of states reporting releases (total) of the chemical	Toxics Release Inventory (TRI) Program data from the EPA
5	Production volume in pounds per year	Chemical Data Reporting (CDR) data from the EPA

Step 2: Use scoring table to find attribute score for value identified in Step 1.

For each element there is a corresponding column in the prevalence scoring table (see Table J-2), which contains a range of data values assigned to a numeric prevalence score between 1 and 10. Once a data value has been found for a particular element, look up the value in Table J-2 to determine the prevalence score. For CDR data, use the most recent year reported. For pesticides, if the compound is a degradate and does not have its own data, use the parent compound to score.

Table J--2. Prevalence Scoring Scales

Prevalence Score	1 % Finished Water with Detections (PWSs)	2 % Ambient Water with Detections (Sites/Samples)	3 # States Reporting Pesticide in Use	4 # States Reporting TRI Total Releases	5 Number of Pounds Produced
1	<= 0.10	<= 0.10	1	1	< 500,000
2	0.11 - 0.16	0.11 - 0.16	2	2	–-
3	0.17 - 0.25	0.17 - 0.25	3	3	>500,000 - 1,000,000
4	0.26 - 0.44	0.26 - 0.44	4	4	–-
5	0.45 - 0.61	0.45 - 0.61	5	5	>1,000,000 - 10,000,000
6	0.62 - 1.00	0.62 - 1.00	6	6	>10,000,000 - 50,000,000
7	1.01 - 1.30	1.01 - 1.30	7 - 10	7 - 10	>50,000,000 - 100,000,000
8	1.31 - 2.50	1.31 - 2.50	11 - 15	11 - 15	>100,000,000 - 500,000,000
9	2.51 - 10.00	2.51 - 10.00	16 - 25	16 - 25	>500,000,000 - 1,000,000,000
10	> 10.00	> 10.00	> 25	> 25	>1,000,000,000

Appendix K – Protocol to Determine Magnitude Attribute Scores

This section describes how to assign a numerical score for the magnitude attribute.

Step 1: Identify the highest-ranked data element

When more than one data element is available for a particular contaminant, use the hierarchy below to select the preferred element. Table K-1 presents the hierarchy of data elements to be used in the magnitude scoring process. Note that the magnitude element should be correlated with the value used to score the prevalence attribute, except when production data are used for prevalence and persistence-mobility is used for magnitude (see Appendix L).

Table K--1. Hierarchy of Magnitude Data Elements

Rank	Magnitude Data Element	Type of Data
1	Median concentration of PWSs with detection	National scale / representative data (UCMR 1-5 has highest priority, then NIRS, then UCM-State Rounds 1-2) from the EPA.
2	Median concentration of ambient water sites or samples with detections	National scale / representative NAWQA data from USGS
3	Application of the chemical as a pesticide in pounds	Estimated Annual Agricultural Pesticide Use data from USGS
4	Total releases of the chemical in pounds	Toxics Release Inventory (TRI) Program data from the EPA
5	Persistence-mobility	Empirical and modeled environmental fate data from the EPA

Step 2: Use scoring table to find attribute score for value identified in Step 1.

For each data element, there is a corresponding column in the magnitude scoring table (Table K-2), which contains a range of data values assigned to a numerical magnitude score. Locate the column in the table associated with the highest-ranking data element identified in step one. Use the information in the column to determine the numerical score associated with the data value for the chemical being scored. The number corresponding to each "score" is the maximum in that category, e.g., 0.1 µg/L for finished water scores 4, not 5. In cases where there are no data for scoring magnitude in Table K-2 (e.g., prevalence is scored using production volume data), use the Persistence-Mobility scoring approach to develop a magnitude score (see Appendix L).

Table K-2. Magnitude Scoring Scales

Magnitude Score	1	2	3	4	5
	Finished Water Median Concentration of Detections(µg/L)	Ambient Water Median Concentration of Detections(µg/L)	Pesticide Use (lbs/year)	TRI Total Releases (lbs/year)	Persistence-Mobility
1	<0.003	<0.003	<10,000	<300	Used when production data are used for prevalence score
2	0.003 - 0.01	0.003 - 0.01	–-	300 - 1,000	
3	>0.01 - 0.03	>0.01 - 0.03	10,000 - 30,000	>1,000 - 3,000	
4	>0.03 - 0.1	>0.03 - 0.1	>30,000 - 100,000	>3,000 - 10,000	
5	>0.1 - 0.3	>0.1 - 0.3	>100,000 - 300,000	>10,000 - 30,000	
6	>0.3 - 1	>0.3 - 1	>300,000 - 1,000,000	>30,000 - 100,000	
7	>1 - 3	>1 - 3	> 1,000,000 - 3,000,000	>100,000 - 300,000	
8	>3 - 10	>3 - 10	>3,000,000 - 10,000,000	>300,000 - 1,000,000	
9	>10 - 30	>10 - 30	>10,000,000 - 30,000,000	>1,000,000 - 3,000,000	
10	>30	>30	>30,000,000	>3,000,000	

Appendix L – Protocol to Determine Magnitude Attribute Scores from Persistence-Mobility

The approach for scoring persistence-mobility includes assigning two values, one for persistence and one for mobility, on a numeric scale of 1 through 3, representing low, medium, and high for each property as it favors the presence of the contaminant in water. Using a hierarchy of physical property data elements, each contaminant is scored for both persistence and mobility. The average of these two values is multiplied by 10/3 to normalize the score on a 1-10 scale for magnitude.

Step 1: Identify and select the highest-ranked data values to score Persistence and Mobility

Select the highest priority data element available for scoring (there is only one option in the case of persistence). When several values for a particular physical property are available, the highest scoring value should be used for scoring.

Step 2: Multiply the average of the persistence and mobility values by 10/3 to calculate a magnitude score.

Table L--1. Persistence-Mobility Scoring Scales

			Persistence Value		
		Units	**1** (Low)	**2** (Medium)	**3** (High)
1	Biodegradation Half-Life (OPERA QSAR)	time	days, days-weeks	weeks, weeks- months	months, recalcitrant

			Mobility Value		
		Units	**1** (Low)	**2** (Medium)	**3** (High)
1	Organic Carbon-Water Partitioning Coefficient (K_{OC})	L/kg	>1000	100 - 1000	<100
2	Log Octanol-Water Partitioning Coefficient (log K_{OW})	dimensionless	>4	1-4	<1
3	Henry's Law Coefficient (K_H)	atm*m^3/mol	>0.042	0.042 – 4.2x 10^{-6}	<4.2x10^{-6}
4	Solubility in water	µg/L	<1000	1000-1,000,000	>1,000,000

Appendix M - Data Management for CCL 6

Section M.1 Overview

The EPA documented all processes related to data management and decision-making in developing the CCL 6. This appendix describes the data management, processing, and extraction steps performed for the main and select supplemental data sources used in developing the CCL 6.

Section M.2 provides a brief description of each data source, references, download information, website addresses (if applicable), any data manipulation steps, and the extracted data elements. This section also describes different data processing steps that may have been required to extract data elements for the screening step versus the classification step.

Section M.3 provides details about the simple data format the EPA used to compile and structure data extracted for CCL 6.

Section M.4 provides the data elements and their descriptions extracted from the EPA's CompTox Chemicals Dashboard.

Section M.5 provides a list of data elements of the CCL 6 Universe file that were not assigned points in the screening step but were used as a resource by the evaluation teams during the classification step. Refer to Section 3.2.1, of the main document, for a list of data elements that were assigned screening points and details on the EPA's exclusion criteria for data element point assignment.

Section M.6 provides references for sections M.3, M.4, and M.5.

Section M.2 Data Source Descriptions and Pre-Processing Specifics for Main and Select Supplemental CCL 6 Data Sources

M.2.1 Health Sources

Agency for Toxic Substances and Disease Registry (ATSDR) Minimal Risk Levels (MRLs) and Toxicological Profiles –Centers for Disease Control and Prevention (CDC)

Data description: According to the Agency for Toxic Substances and Disease Registry (ATSDR), "[t]oxicological Profiles (Tox Profiles) are a unique compilation of toxicological information on a given hazardous substance. Each peer-reviewed Tox Profile reflects a comprehensive and extensive evaluation, summary, and interpretation of available toxicological and epidemiological information on a substance" (https://www.atsdr.cdc.gov/toxicological-profiles/about/index.html). Tox Profiles generally contain information pertaining to the derivation of Minimal Risk Levels (MRL). An MRL is "an estimate of the daily human exposure to a hazardous substance that is likely to be without appreciable risk of adverse non-cancer health effects over a specified duration of exposure. MRLs are derived when ATSDR determines that reliable and sufficient data exist to identify the target organ(s) of effect or the most sensitive health effect(s) for a specific duration for a given route of exposure to the substance" (https://www.atsdr.cdc.gov/minimal-risk-levels/about/index.html).
ATSDR develops MRLs for the oral and the inhalation route of exposure and for acute, intermediate, and chronic exposure durations. For screening and classification under CCL 6, ATSDR's chronic duration oral MRLs are considered comparable to EPA's RfDs, and the chronic duration inhalation MRLs are considered comparable to EPA's RfCs. Intermediate oral MRLs are considered comparable to subchronic RfDs, and acute duration oral MRLs are considered comparable to acute RfDs. Finally, intermediate inhalation MRLs are comparable to subchronic RfCs, and acute duration inhalation MRLs are considered comparable to acute RfCs. The MRL website was used as a main data source for CCL 6 and Toxicological Profiles were used as a supplemental data source for CCL 6.

Reference: Centers for Disease Control and Prevention (CDC). n.d. Agency for Toxic Substances and Disease Registry (ATSDR) Minimal Risk Levels (MRLs) for Hazardous Substances. https://wwwn.cdc.gov/TSP/MRLS/mrlsListing.aspx. Accessed July 2023.

Data processing for screening:

- **Data download:** The original data from the MRL website were copied and pasted into an Excel workbook.
- **Data manipulation:** Data manipulation was minimal and limited to altering the format of chemical identifiers (e.g., adding DTXSIDs).
- **Extracted data elements**: EPA wrote R code to extract all MRLs into simple files and code them in the Universe to the correct data element (oral MRLs equivalent to acute RfDs, subchronic RfDs, chronic RfDs; inhalation MRLs equivalent to acute reference concentrations (RfCs), subchronic RfCs, and chronic RfCs).

- **Use in screening:** EPA assigned screening points to oral MRLs of all durations (chronic MRL = Tier 1, subchronic MRL = Tier 3, acute MRL = Tier 4). Inhalation MRLs were not used for screening.

Data processing for classification:

- **Data download:** Toxicological Profiles for PCCL 6 contaminants were downloaded from the Tox Profile website.
- **Data manipulation:** Chronic and intermediate duration oral MRLs were converted to units of mg/kg/day, when necessary. In cases where ATSDR provided no chronic oral MRL but did provide an intermediate-duration MRL, EPA considered whether it was appropriate to add a subchronic-to-chronic uncertainty factor (UF) of 10 to the intermediate-duration MRL, in accordance with EPA guidance (USEPA, 2002). No additional data manipulation was needed for other intermediate and acute oral MRLs or any inhalation MRLs aside from what is described above for screening.
- **Extracted data elements:** In addition to the data extracted to support screening, EPA manually extracted data from Toxicological Profiles including the publication years, chronic oral MRLs (and intermediate-duration MRLs when needed), critical studies, and critical effects.
- **Use in classification:** EPA used chronic oral MRLs (and intermediate-duration oral MRLs, when applicable) to derive Health Reference Levels (HRLs) for CCL 6. EPA presented the remaining data elements extracted during the screening process in the Other Health Data section of the CIS, using the original names as presented by the data source.

Cancer Potency Data Bank (CPDB) – National Library of Medicine, U.S. Department of Health and Human Services (HHS)

Data description: The Cancer Potency Data Bank (CPDB) synthesized the results of 50 years of chronic, long-term carcinogenesis bioassays. Data were compiled into a common format from 6,540 experiments on 1,547 chemicals from the general literature and the Technical Reports of the National Cancer Institute/National Toxicology Program (NCI/NTP). Information recorded included the strain, sex, route of compound administration, target organ, histopathology, author's opinion about carcinogenicity, quantitative data on tumor incidence, dose-response, the tumorigenic dose-rate for 50% of experimental animals (TD_{50}), statistical significance of the dose-response, length of experiment, duration of dosing, and average daily dose-rate. This database was last updated in August 2007. This data source was used as a main data source for CCL 6. Because the database was not updated since CCL 5, EPA relied on the data extracted and processed previously to support screening and classification for CCL 6. This information is described below.

Reference: U.S. Department of Health and Human Services (HHS). n.d. National Institutes of Health (NIH). National Library of Medicine. TOXNET. Carcinogenic Potency Database (CPDB). https://www.nlm.nih.gov/toxnet/index.html. Accessed October 2018.

Data processing for screening:

- **Data download:** The original NIH-ToxNet website EPA accessed to download the CPDB is retired. The CPDB data was accessed through this link: https://www.nlm.nih.gov/databases/download/cpdb.html. The Cancer Potency Data Bank has since been replaced with the Carcinogenic Potency Database (through this link: and was removed from the web. The Cancer Potency Data Bank provided results from 45 years of long-term animal cancer tests, including data on the carcinogenic potency (TD*50*) of different chemicals. The Cancer Potency Data Bank has since been replaced with the Carcinogenic Potency Database and was removed from the web. This sourceThe Cancer Potency Data Bank has since been replaced with the Carcinogenic Potency Database (through this link: https://datadiscovery.nlm.nih.gov/Literature/Carcinogenic-Potency-Database-CPDB-Archived-Data/s5hf-hv94/about_data) and was removed from the web. The Carcinogenic Potency Database was last updated on July 10, 2025 and is currently in archived status.
- **Data manipulation:** The data manipulation for the CPDB data was minimal and was limited to altering the format of chemical identifiers (e.g., adding DTXSIDs). Additionally, chemicals reported as having no dose-related effects were assigned a value of 1.0E+31 in the pre-universe and universe files for coding purposes. These values were not reported on the CISs.
- **Extracted data elements**: The TD_{50} values were extracted for each entry.
- **Use in screening:** EPA assigned TD_{50} values screening points (Tier 5).

Data processing for classification:

- **Data download:** Data were downloaded as described for screening.
- **Data manipulation:** Data were manipulated as described for screening.

- **Extracted data elements**: Data were extracted as described for screening.
- **Use in classification:** EPA presented minimum and maximum TD_{50} values extracted during the screening process in the Other Health Data section of the CIS. “No dose-related effects” values were not reported on the CISs.

"Development and Validation of a Computational Model for Androgen Receptor Activity" – Kleinstreuer et al. 2017

Data description: This peer-reviewed publication provided modeled androgen receptor (AR) bioactivity for 1,855 commercial and environmental chemicals. Modeled activity integrated androgen receptor binding, coregulator recruitment, gene transcription, and protein production assays in multiple cell types. This data source was used as a main data source for CCL 6.

Reference: Kleinstreuer, N.C., P. Ceger, E.D. Watt, M. Martin, K. Houck, P. Browne, R.S. Thomas, W.M. Casey, D.J. Dix, D. Allen, S. Sakamuru, M. Xia, R. Huang, and R. Judson. 2017. Development and Validation of a Computational Model for Androgen Receptor Activity. Chem. Res. Toxicol. 30: 946-964. (https://pubs.acs.org/doi/10.1021/acs.chemrestox.6b00347)

Data processing for screening:

- **Data download:** EPA downloaded the publication and supporting information for 1,855 chemicals ("AUC values and associated confidence intervals for agonism, antagonism, and interference for all 1855 chemicals"; tx6b00347_si_004.xlsx).
- **Data manipulation:** The largest positive integer from the "AUC.Agonist" and "AUC.Antagonist" columns was selected to inform the overall value for each individual chemical. EPA wrote R code to extract the androgen receptor bioactivity area under the curve (AUC) values and tiered them into three categories (Category 1 – <0.001, Category 2 – 0.001 to 0.1; Category 3 – ≥ 0.1), based on information presented by Kleinstreuer et al. (2017).
- **Extracted data elements**: EPA extracted the categorized androgen receptor bioactivity values into the Universe.
- **Use in screening:** According to Kleinstreuer et al. (2017), chemicals with AUC values ≥ 0.1 likely have *in vivo* androgen receptor activity and chemicals with AUC values between 0.1-0.001 have potentially ambiguous androgen receptor activity; these categories received Tier 5 screening points in CCL 6. Chemicals with AUC values <0.001 are likely to be androgen receptor inactive *in vivo* and were not assigned screening points.

Data processing for classification:

- **Data download:** Data were downloaded as described for screening.
- **Data manipulation:** Data were manipulated as described for screening.
- **Extracted data elements**: Data were extracted as described for screening.
- **Use in classification**: EPA presented the category and range of AUC values described by the category, as well as biological interpretations of those values in the "Other Health Data" section of the CISs.

Drinking Water Standards and Health Advisories (DWSHA) Tables, Health Advisory Documents, and Health Effects Support Documents – Environmental Protection Agency

Data description: EPA's Drinking Water Standard and Health Advisories (DWSHA) table is a summary of Health Advisory values and EPA's National Primary Drinking Water Regulations (NPDWRs). This document is periodically updated to reflect changes in health advisory values or regulatory values. This data source was used as a main data source for CCL 6. Because the database was not updated since CCL 5, EPA relied on the data extracted and processed previously to support screening and classification for CCL 6. This information is described below. The EPA also publishes Health Advisory documents and Health Effects Support Documents (HESDs) that provide data and information supporting toxicity value and/or Health Advisory value development for individual contaminants. These documents were used as a supplemental source for CCL 6.

Reference: USEPA. 2018. Edition of the Drinking Water Standards and Health Advisories Tables. https://www.epa.gov/system/files/documents/2022-01/dwtable2018.pdf. Accessed November 2019.

USEPA. 2018. EPA Non-Regulatory Health-Based Drinking Water Levels. https://www.epa.gov/sdwa/epa-non-regulatory-health-based-drinking-water-levels. Accessed August 2024.

Data processing for screening:

- **Data download:** The data accessed via the DWSHA Tables link above was copied and pasted into an Excel file.
- **Data manipulation:** EPA converted cancer classifications from different sources to a comparable numeric scheme according to the same methodology used for CCL 3 (USEPA, 2009). This conversion is further explained in Section 2.4.4, of the main document. The DWSHA table includes cancer risk concentrations at the 10^{-4} cancer risk level. To allow comparison between cancer risk concentrations reported at different cancer risk levels, cancer risk concentrations are converted to the 10^{-6} cancer risk level. DTXSIDs were also assigned.
- **Extracted data elements**: Several relevant metrics were extracted from the DWSHA table. The 10-day Health Advisory values were extracted and categorized as acute benchmarks. Also extracted were the RfDs and CSFs, Lifetime Health Advisory values (considered chronic benchmarks), and cancer classifications.
- **Use in screening**: Lifetime Health Advisory values, RfDs, and CSFs were assigned Tier 1 screening points. Cancer classifications were assigned Tier 3 screening points. 10-day Health Advisory values were assigned Tier 4 screening points.

Data processing for classification:

- **Data download:** Health Advisory documents and HESDs were downloaded from the Non-Regulatory Health-Based Drinking Water Levels website in addition to the data elements downloaded as described for screening.
- **Data manipulation:** Data manipulation was not required for classification.

- **Extracted data elements**: The EPA manually extracted data from Health Advisory documents and HESDs including the publication years, RfDs, CSFs, cancer classifications, critical studies, and critical effects.
- **Use in classification**: EPA presented 10-day and Lifetime Health Advisory values extracted during the screening process in the Other Health Data section of the CIS. If the information presented by the HA documents and HESDs was not redundant with information presented by other sources (e.g., IRIS, OPP), RfDs, CSFs, cancer classifications, critical studies, and critical effects were presented under the Qualifying Health Assessment section of the CISs and could be used to derive CCL 6 HRLs.

Endocrine Disruptor Screening Program (EDSP) – Estrogen Receptor Bioactivity

Data description: This dataset contains the results from the Endocrine Disruptor Screening Program (EDSP) Estrogen Receptor Bioactivity Based on ToxCast(TM) "Endocrine Receptor (ER) ER Model" as of June 1, 2015. This data source was used as a main data source for CCL 6.

References:

USEPA. February 2023. Endocrine Disruptor Screening Program (EDSP) Estrogen Receptor Bioactivity. https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-estrogen-receptor-bioactivity. Accessed June 2023.

Judson, R.S., F.A. Magpantay, V. Chickarmane, C. Haskell, N. Tania, J. Taylor, M. Xia, R. Huang, D.M. Rotroff, D.L. Filer, K.A. Houck, M.T. Martin, N. Sipes, A.M. Richard, K. Mansouri, R.W. Setzer, T.B. Knudsen, K.M. Crofton, and R.S. Thomas. 2015. Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 In Vitro High-Throughput Screening Assays for the Estrogen Receptor. 148(1): 137-154. https://doi.org/10.1093/toxsci/kfv168. (https://academic.oup.com/toxsci/article/148/1/137/1659792)

Data processing for screening:

- **Data download:** EPA copied and pasted the data table (all columns and rows) into a CSV file.
- **Data manipulation:** EPA wrote R code to categorize the ER.Bioactivity area under the curve (AUC) column for each contaminant into three categories: Category 1 – <0.01, Category 2 – 0.01 to <0.1; Category 3 – ≥ 0.1 based on information presented in Judson et al. (2015).
- **Extracted data elements**: EPA extracted the categorized estrogen receptor bioactivity values into the Universe.
- **Use in screening:** According to Judson et al. (2015), chemicals with AUC values ≥ 0.1 likely have *in vivo* estrogen receptor activity and chemicals with AUC values between 0.1-0.01 have potentially ambiguous estrogen receptor activity; these categories received Tier 5 screening points in CCL 6. Chemicals with AUC values <0.01 are likely to be estrogen receptor inactive *in vivo* and were not assigned screening points.

Data processing for classification:

- **Data download:** Data were downloaded as described for screening.
- **Data manipulation:** Data were manipulated as described for screening.
- **Extracted data elements**: Data were extracted as described for screening.
- **Use in classification:** EPA presented the category and range of AUC values described by that category, as well as biological interpretations of those values on the CISs in the Other Health Data section.

Guidelines for Canadian Drinking Water Quality – Health Canada

Data description: Health Canada, in collaboration with the Federal-Provincial-Territorial Committee on Drinking Water of the Federal-Provincial-Territorial Committee on Health and the Environment, calculates maximum acceptable concentrations (MACs) for chemical and physical parameters in drinking water. MACs are drinking water guideline values that define maximum levels of contaminants in drinking water based on a comprehensive review of the known health effects associated with each contaminant. This data source was used as a main data source for CCL 6. Health Canada publishes technical documents that provide data and information supporting MAC development for individual contaminants. These documents were used as a supplemental source for CCL 6.

References:

Health Canada (2022) Guidelines for Canadian Drinking Water Quality-Summary Tables. Water and Air Quality Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario. https://www.canada.ca/en/health-canada/services/environmental-workplace-health/reports-publications/water-quality/guidelines-canadian-drinking-water-quality-summary-table.html#t2. Accessed July 2023.

Health Canada (n.d.) Water Quality – Reports and Publications (Guidelines and Technical Documents). Health Canada, Ottawa, Ontario. https://www.canada.ca/en/health-canada/services/environmental-workplace-health/reports-publications/water-quality.html#guide. Accessed August 2024.

Data processing for screening:

- **Data download:** EPA copied and pasted the data table (all columns and rows) from the summary table link above into an Excel workbook.
- **Data manipulation:** Raw data underwent light cleaning after copy and paste for readability in R. Specifically, EPA separated out year(s) from chemical name, set MAC values to NA if the value was "None required" or blank in the original data, removed trailing zeroes to facilitate csv formatting, removed unneeded columns, and assigned DTXSIDs.
- **Extracted data elements**: EPA wrote R code to extract MAC values.
- **Use in screening**: MAC values were considered chronic benchmarks and were assigned Tier 1 screening points.

Data processing for classification:

- **Data download**: MAC values were downloaded as described for screening. Guideline documents for PCCL 6 contaminants were also downloaded from the Reports and Publications website above.
- **Data manipulation:** TDIs were converted to units of mg/kg/day, when necessary. Data manipulations for MACs was as described for screening.
- **Extracted data elements:** In addition to the data extracted to support screening, EPA manually extracted data from guideline documents including the publication years, Tolerable Daily Intake (TDI) levels (considered equivalent to RfDs), critical studies, and critical effects.

- **Use in classification:** EPA used TDIs to derive Health Reference Levels (HRLs) for CCL 6. EPA presented MACs in the Other Health Data section of the CIS, using the original name as presented by the data source.

Guidelines for Drinking-Water Quality – World Health Organization (WHO)

Data description: The World Health Organization (WHO) publishes health-based guidance values for drinking water. The fourth edition of the Guidelines for Drinking-Water Quality (GDWQ), with the first and second addenda, was published in 2022. This data source was used as a main data source for CCL 6. WHO publishes background documents that provide data and information supporting guideline development for individual contaminants. These documents were used as a supplemental source for CCL 6.

Reference: World Health Organization (WHO). 2022. Guidelines for drinking-water quality. Fourth edition, incorporating the first and second addenda. https://www.who.int/publications/i/item/9789240045064. Accessed June 2023.

WHO. N.d. Chemical Hazards in Drinking-Water. https://www.who.int/teams/environment-climate-change-and-health/water-sanitation-and-health/chemical-hazards-in-drinking-water. Accessed August 2024.

Data processing for screening:

- **Data download:** EPA downloaded the PDF document and copy/pasted the mg/L values in Table A3.3 (Annex 3) into an Excel workbook.
- **Data manipulation:** Data manipulation was minimal and limited to altering the format of chemical identifiers (e.g., adding DTXSIDs).
- **Extracted data elements**: Guideline values for all chemical contaminants were extracted. Values noted to be derived from “short-term exposure” were considered subchronic benchmarks. All other guideline values were considered chronic benchmarks.
- **Use in screening**: Guideline values considered to be subchronic benchmarks were assigned Tier 3 screening points. All other values were considered chronic benchmarks and were assigned Tier 1 screening points.

Data processing for classification:

- **Data download:** Guideline documents for PCCL 6 contaminants were downloaded from the WHO website. Guideline values were downloaded as described for screening.
- **Data manipulation:** TDIs were converted to units of mg/kg/day, when necessary. Data manipulation for guideline values was as described for screening.
- **Extracted data elements:** In addition to the data extracted to support screening, EPA manually extracted data from guideline documents including the publication years, Tolerable Daily Intake (TDI) levels (considered equivalent to RfDs), critical studies, and critical effects.
- **Use in classification:** EPA used TDIs to derive Health Reference Levels (HRLs) for CCL 6. EPA presented guideline values in the Other Health Data section of the CIS, using the original name as presented by the data source.

Hazardous Substances Data Bank (HSDB) – National Library of Medicine, U.S. Department of Health and Human Services

Data description: The Hazardous Substances Data Bank (HSDB) is a toxicology database that includes information on human exposure, industrial hygiene, emergency handling procedures, environmental fate, regulatory requirements, toxicity values, and other information. The information in HSDB has been assessed by a Scientific Review Panel. This source was used as a main data source for CCL 6 as it is data-rich and the only source of LD_{50} for the CCL 6 process. Because the database was not updated since CCL 5, EPA relied on the data extracted and processed previously to support screening and classification for CCL 6. This information is described below.

Reference: HHS. n.d. National Institutes of Health (NIH). National Library of Medicine. Hazardous Substances Databank (HSDB). https://www.nlm.nih.gov/databases/download/hsdb.html. (Now part of PubChem.) Accessed April 2019.

Data processing for screening:

- **Data download:** EPA downloaded the HSDB data as an XML file.
- **Data manipulation:** Fields containing oral toxicity values based on animal studies are extracted from the large HSDB XML file. Regular expressions (regex) are used to extract LD_{50}s, NOAELs, LOAELs, and the corresponding units of measure from the text fields describing the toxicity studies. DTXSIDs were also added. Data manipulation steps were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract LD_{50}s, NOAELs, and LOAELs. EPA presented only the minimum and maximum LD_{50} values on CISs for chemicals with multiple entries.
- **Use in screening:** LD_{50}s were assigned Tier 5 screening points.

Data processing for classification:

- **Data download:** Data were downloaded as described for screening.
- **Data manipulation:** Data were manipulated as described for screening.
- **Extracted data elements**: EPA presented only the minimum and maximum LD_{50} values on CISs for chemicals with multiple entries.
- **Use in classification:** EPA presented min and max LD_{50} values extracted during the screening process in the Other Health Data section of the CIS.

Health-Based Screening Levels (HBSLs) – United States Geological Survey (USGS)

Data description: Health-based screening levels (HBSLs) are calculated by the USGS to help prioritize monitoring efforts and determine if concentrations of contaminants found in surface water or groundwater sources of drinking water may indicate a potential human health concern. HBSLs are calculated for non-cancer and cancer effects. This data source was used as a main data source for CCL 6. At the time of CCL 6 Universe data download (July 2023) the HBSL database had not been updated since CCL 5. Therefore, EPA relied on the data extracted and processed previously to support screening and classification for CCL 6. This information is described below. HBSLs have since been updated (January 2024) and may be different on the USGS website as compared to the data collected for CCL 6.

Reference: U.S. Geological Survey (USGS). n.d. Health-Based Screening Levels for Evaluating Water-Quality Data. https://water.usgs.gov/water-resources/hbsl/. Accessed July 2018.

Data processing for screening:

- **Data download:** EPA exported the HBSLs as a CSV file.
- **Data manipulation:** USGS provides HBSLs for cancer effects as a range of concentrations from the 10^{-6} to the 10^{-4} risk levels. To compare these values to other benchmarks, only HBSLs calculated using a 10^{-6} cancer risk level were extracted for screening. Other data manipulation for the HBSLs data was minimal and was limited to altering the format of chemical identifiers (e.g., adding DTXSIDs). Data manipulation steps were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract HBSLs. HBSLs were treated as chronic benchmarks.
- **Use in screening**: HBSLs were assigned Tier 1 screening points.

Data processing for classification:

- **Data download:** Data were downloaded as described for screening.
- **Data manipulation:** Data were manipulated as described for screening.
- **Extracted data elements**: Data were extracted as described for screening.
- **Use in Classification:** EPA presented HBSLs extracted during the screening process in the Other Health Data section of the CIS.

Human Health-Based Water Guidance Values – Minnesota Department of Health

Data description: The Minnesota Department of Health (MDH) develops health-based guidance values that can be used to help evaluate potential human health risks from exposures to chemicals in groundwater. The MDH calculates guidance values for cancer and non-cancer endpoints of various exposure durations including acute, short-term, subchronic, and chronic durations. This data source was used as a main data source for CCL 6. MDH publishes Toxicological Summary documents that provide data and information supporting guidance value development for individual contaminants. These documents were used as a supplemental source for CCL 6.

Reference: Minnesota Department of Health (MDH). n.d. Human Health-Based Water Guidance Table. https://www.health.state.mn.us/communities/environment/risk/guidance/gw/table.html. Accessed July 2023 (screening) and August 2024 (classification).

Data processing for screening:

- **Data download:** EPA downloaded the Excel MDH Water Guidance and Additivity Calculator (guidance.xlsx) from the webpage and used the “AllGuidance” tab to download the guidance values.
- **Data manipulation:** The benchmarks published by the MDH are at the 10^{-5} cancer-risk level. For cancer risk concentrations in the universe comparable, they were converted to the 10^{-6} cancer risk concentration. EPA wrote R code to categorize data types by 'Duration' as presented in the original data to acute, cancer, chronic, subchronic, and short term outputs. EPA updated the chemical identifiers to add DTXSIDs.
- **Extracted data elements**: EPA wrote R code to extract the acute, subchronic, cancer and chronic guidance values. Acute and short-term guidance values were considered acute benchmarks. Subchronic values were considered subchronic benchmarks. Chronic and cancer values were considered chronic benchmarks.
- **Use in screening:** Acute and short-term guidance values were assigned Tier 4 screening points. Subchronic guidance values were assigned Tier 3 screening points. Cancer and chronic (noncancer) guidance values were assigned Tier 1 screening points.

Data processing for classification:

- **Data download:** Toxicological Summary documents for PCCL 6 contaminants were downloaded from the MDH website. Guidance values were downloaded as described for screening.
- **Data manipulation:** RfDs were converted to units of mg/kg/day, when necessary. Data manipulation for guidance values was as described for screening.
- **Extracted data elements:** In addition to the data extracted to support screening, EPA manually extracted data from Toxicological Summary documents including the publication years, RfDs, critical studies, and critical effects.
- **Use in classification:** EPA used RfDs to derive CCL Screening Levels for CCL 6. EPA presented guidance values in the Other Health Data section of the CIS, using the original name as presented by the data source.

Human Health Benchmarks for Pesticides – Environmental Protection Agency

Data description: The Human Health Benchmarks for Pesticides are published by EPA and were updated in 2021. The purpose of the benchmarks is to determine whether the detection of a pesticide in drinking water or source waters for drinking water may indicate a potential health risk and help with EPA prioritization of monitoring efforts. There are benchmarks for acute and chronic exposure scenarios, cancer and non-cancer endpoints, and potentially sensitive populations. The 2021, HHBPs were derived for 430 pesticides. HHBPs are available for pesticide active ingredients for which Health Advisories or enforceable National Primary Drinking Water Regulations (e.g., maximum contaminant levels) have not been developed. This data source was used as a main data source for CCL 6. On the HHBP website, EPA provides links to Office of Pesticides Program (OPP) human health risk assessments that support HHBP derivation for individual contaminants. These documents were used as a supplemental source for CCL 6.

Reference: EPA. n.d. Human Health Benchmarks for Pesticides. https://www.epa.gov/system/files/documents/2021-07/hh-benchmarks-technical-document-2021.pdf. Accessed July 2023 and August 2024.

Data processing for screening:

- **Data download:** EPA copied and pasted HHBP data into a CSV file. Data were cleaned for superscripts and footnotes that would interfere with the R scripts and future formatting/manipulation. EPA consolidated compounds with multiple CASRNs associated into one entry that listed all CASRNs together.
- **Data manipulation:** Using R, the original column Carcinogenic HHBP (E-6 to E-4) (ppb) was split into two columns, one which contained Carcinogenic HHBP E-6 and one that contained Carcinogenic HHBP E-4 values. All blank cells, hyphens, or double hyphen values were converted to NA values. EPA updated the chemical identifiers to add DTXSIDs.
- **Extracted data elements**: EPA wrote R code to extract acute and chronic (cancer and non-cancer) benchmarks, acute and chronic PADs (treated as acute and chronic RfDs, respectively), and CSFs.
- **Use in screening**: Chronic benchmarks (cancer and non-cancer), chronic PADs, and CSFs were assigned Tier 1 screening points. Acute benchmarks and acute PADs were assigned Tier 4 screening points.

Data processing for classification:

- **Data download:** OPP risk assessment documents for PCCL 6 contaminants were downloaded from the HHBP website. Acute and chronic benchmarks, CSFs, and PADs were downloaded as described for screening.
- **Data manipulation:** Data manipulation was as described for screening.
- **Extracted data elements:** In addition to the data extracted to support screening, EPA manually extracted data from OPP risk assessments including the publication years, PADs, oral slope factors, cancer classifications, critical studies, and critical effects.
- **Use in classification:** When data were not available to support an updated HHBP value, EPA used the chronic HHBP (cancer and non-cancer) or acute HHBP (if a chronic HHBP

was not available) as the HRL that was presented on the CIS. EPA presented the cancer classification available in the accompanying OPP risk assessment on the CIS. EPA presented acute HHBPs and acute PADs in the Other Health Data section of the CIS.

Human Health Benchmarks for Pharmaceuticals (HHB-Rx) in Drinking Water – Environmental Protection Agency

Data description: The Human Health Benchmarks for Pharmaceuticals (HHB-Rx) in Drinking Water were published by EPA in 2026. The HHB-Rx are non-regulatory and non-enforceable and are intended to provide information to help states, tribes, and water systems better characterize potential health risks associated with the occurrence of pharmaceuticals in drinking water and prioritize pharmaceuticals for additional monitoring and/or health effects research. The 2026 HHB-Rx were derived for 374 pharmaceuticals currently used in U.S. Food and Drug Administration (FDA)-approved drug products that are known to, or have the potential to, occur in source waters (surface and groundwater) and treated drinking water. Two HHB-Rx were derived for each pharmaceutical: one for the general population (all ages) and one for infants (birth to <1 year). This data source was used as a main and supplemental data source for CCL 6.

Reference: EPA. n.d. Human Health Benchmarks for Pharmaceuticals in Drinking Water. https://www.epa.gov/sdwa/human-health-benchmarks. Accessed July 2023 and August 2024.

Data processing for screening:

- **Data download:** EPA copied and pasted relevant data from the HHB-Rx Excel files into Excel workbooks for CCL 6.
- **Data manipulation:** Raw data underwent manual light cleaning to remove extraneous characters. Minor formatting updates and the additional of chemical identifiers (e.g., DTXSIDs) were conducted using R.
- **Extracted data elements**: HHB-Rx were extracted and are considered chronic benchmarks for the screening step.
- **Use in screening**: HHB-Rx were assigned Tier 1 screening points.

Data processing for classification:

- **Data download:** Data were downloaded as described for screening.
- **Data manipulation:** Data were manipulated as described for screening.
- **Extracted data elements**: Data were extracted as described for screening.
- **Use in classification:** HHB-Rx were used as CCL Screening Levels for FDA-approved pharmaceutical contaminants with LTDs that did not have qualifying health assessments that provided an RfD and/or CSF.

Integrated Risk Information System (IRIS) – Environmental Protection Agency

Data description: EPA's Office of Research and Development (ORD) houses the IRIS program that supports the EPA by characterizing the toxicity of compounds. The oral toxicity values and cancer classifications derived by the IRIS program are highly relevant to the CCL 6 process. This data source was used as a main data source for CCL 6. The IRIS program develops Toxicological Assessments that provide data and information supporting assessment development for individual contaminants. These documents were used as a supplemental source for CCL 6.

Reference: USEPA. n.d. Integrated Risk Information System (IRIS). IRIS Advanced Search. https://cfpub.epa.gov/ncea/iris/search/index.cfm?keyword. Accessed November 2023.

USEPA. n.d. IRIS Assessments. https://iris.epa.gov/AtoZ/?list_type=alpha. Accessed August 2024.

Data processing for screening:

- **Data download:** EPA exported the 'Chemical Details,' 'RfD Toxicity Values', 'RfC Toxicity Values', and 'WOE_Details' tables to Excel.
- **Data manipulation:** EPA wrote code to combine the data in the four datasets. EPA separated units from values into individual columns and reformatted the toxicity values to scientific notation. EPA converted cancer classifications from different sources to a comparable numeric scheme according to the same methodology used for CCL 3. This conversion is further explained in Section 2.4.4, of the main document. EPA updated the chemical identifiers to add DTXSIDs.
- **Extracted data elements**: EPA extracted RfDs, oral slope factors (equivalent to CSFs), cancer classifications, inhalation unit risk values, RfCs, subchronic RfCs, and subchronic RfDs.
- **Use in screening:** Oral slope factors and chronic RfDs were assigned Tier 1 screening points. Cancer classifications and subchronic RfDs were assigned Tier 3 screening points. Inhalation data were not used for screening.

Data processing for classification:

- **Data download:** IRIS Toxicological Assessments were downloaded from the assessment website in addition to the data elements downloaded as described for screening.
- **Data manipulation:** Data manipulation was as described for screening.
- **Extracted data elements:** In addition to the data extracted to support screening, EPA manually extracted data from IRIS assessments including the publication years, RfDs, CSFs, cancer classifications, critical studies, and critical effects.
- **Use in classification**: EPA used RfDs and CSFs to derive HRLs for CCL 6. EPA presented cancer classifications along with these data. EPA presented subchronic RfDs as well as all inhalation values extracted for screening in the Other Health Data section of the CIS.

International Agency for Research on Cancer (IARC) Classifications – World Health Organization (WHO)

Data description: IARC classifies compounds into groups based on the available toxicity data. The dataset contains cancer classifications for over 1,000 contaminants. The IARC uses Group 1, carcinogenic to humans; Group 2A, probably carcinogenic to humans; Group 2B, possibly carcinogenic to humans; and Group 3, not classifiable as to its carcinogenicity to humans. This data source was used as a main data source for CCL 6.

Reference: World Health Organization (WHO). n.d. International Agency for Research on Cancer (IARC). IARC Monographs on the Identification of Carcinogenic Hazards to Humans. List of Classifications. https://monographs.iarc.who.int/list-of-classifications/. Accessed July 2023.

Data processing for screening:

- **Data download:** EPA downloaded the list of classifications (volumes 1-134) as a CSV file.
- **Data manipulation:** EPA selected and renamed data columns to fit the simple file format, and saved as a csv. EPA updated the chemical identifiers to add DTXSIDs.
- **Extracted data elements**: EPA wrote R code to extract the monograph conclusions (group 1, 2A, 2B, or 3), considered cancer classifications for screening purposes. EPA converted cancer classifications from different sources to a comparable numeric scheme according to the same methodology used for CCL 3. This conversion is further explained in Section 2.4.4, of the main document.
- **Use in screening**: Cancer classifications were assigned Tier 3 screening points.

Data processing for classification:

- **Data download:** Data were downloaded as described for screening.
- **Data manipulation:** Data were manipulated as described for screening.
- **Extracted data elements**: Data were extracted as described for screening.
- **Use in classification:** Monograph conclusions were presented in the Other Health Data section of the CIS.

Maximum Recommended Daily Dose (MRDD) Database – Food and Drug Administration (FDA)

Data description: The Food and Drug Administration (FDA) created the Maximum Recommended Daily Dose (MRDD) database, housed within the National Library of Medicine (DSSTox (FDAMDD) FDA Maximum (Recommended) Daily Dose Database), which includes MRDDs for over 1,200 pharmaceuticals included in Martindate: The Extra Pharmacopoeia (1973, 1983, 1993) and The Physicians' Desk Reference (1995 and 1999). This database was intended to serve as training data for QSAR modeling programs; therefore, some compounds were removed from the database because they are not suitable for most QSAR modeling programs. Some examples are inorganic compounds, high weight polymers, fibers, salts, or mixtures of compounds. MRDDs are not comparable to RfDs or LOAELs; however, this information is relevant for the screening step of CCL 6 due to the breadth of compounds included in the database and the inclusion of pharmaceutical chemicals with no or limited other sources of retrievable toxicity data. This data source was used as a main data source for CCL 6. Because the database was not updated since CCL 5, EPA relied on the data extracted and processed previously to support screening and classification for CCL 6. This information is described below.

Reference: Matthews, E.J., N.L. Kruhlak, R.D. Benz, and J.F. Contrera. 2004. Assessment of the health effects of chemicals in humans: I. QSAR estimation of the maximum recommended therapeutic dose (MRTD) and no effect level (NOEL) of organic chemicals based on clinical trial data. Current Drug Discovery Technologies, 1(1): 61-76.

Data processing for screening:

- **Data download:** EPA downloaded the data table containing MRDDs as a CSV file from the PubChem DSSTox (FDAMDD) FDA Maximum (Recommended) Daily Dose Database (housed by the National Institutes of Health (NIH), National Library of Medicine, National Center for Biotechnology Information at https://pubchem.ncbi.nlm.nih.gov/bioassay/1195).
- **Data manipulation:** EPA altered the format of chemical identifiers for each entry (e.g., added DTXSIDs). As described above, some compounds were removed from the database because they are not suitable for most QSAR modeling programs.
- **Extracted data elements**: EPA wrote R code to extract MRDD values from the data table.
- **Use in screening:** MRDDs are assigned Tier 4 screening points.

Data processing for classification:

- **Data download:** Data were downloaded as described for screening.
- **Data manipulation:** Data were manipulated as described for screening.
- **Extracted data elements**: Data were extracted as described for screening.
- **Use in classification:** MRDDs were presented in the Other Health Data section of the CIS.

National Recommended Water Quality Criteria - Human Health Criteria – Environmental Protection Agency

Data description: Human Health Criteria (HHC) are calculated by the EPA in accordance with the Clean Water Act. Criteria represent specific levels of chemicals or conditions in a water body that are not expected to cause adverse effects to human health. EPA calculates criteria for an exposure scenario, assuming the target population could be drinking contaminated water and consuming contaminated fish or could be consuming only contaminated fish. EPA provides recommendations for "water+organism" and "organism only" criteria for these two scenarios, respectively. HHC for carcinogens are calculated at the 10^{-6} cancer risk level. This data source was used as a main data source for CCL 6. Because the database was not updated since CCL 5, EPA relied on the data extracted and processed previously to support screening and classification for CCL 6. This information is described below.

Reference: USEPA. n.d. Office of Water National Recommended Water Quality Criteria - Human Health Criteria Table. https://www.epa.gov/wqc/national-recommended-water-quality-criteria-human-health-criteria-table. Accessed April 2018.

Data processing for screening:

- **Data download:** EPA copied and pasted the HHC data table into a CSV file.
- **Data manipulation:** Data manipulation was limited to the alteration of the format of chemical identifiers for each entry (e.g., added DTXSIDs).
- **Extracted data elements**: EPA wrote R code to extract HHC for the protection of water and organisms, considered chronic benchmarks for screening purposes.
- **Use in screening**: HHCs (cancer and non-cancer) were assigned Tier 1 screening points.

Data processing for classification:

- **Data download:** Data were downloaded as described for screening.
- **Data manipulation:** Data were manipulated as described for screening.
- **Extracted data elements**: Data were extracted as described for screening.
- **Use in classification:** HHCs were presented in the Other Health Data section of the CIS.

National Toxicology Program (NTP) Cancer Classifications – Department of Health and Human Services

Data description: The National Toxicology Program (NTP) publishes summaries of technical reports examining the carcinogenicity of compounds in mice and rats. The results of studies are classified as clear evidence (CE or P), some evidence (SE), equivocal evidence (EE or E), or no evidence (NE or N) of carcinogenicity. Other classifications include inadequate experiment (IS) and not tested (NT). This data source was used as a main data source for CCL 6. Because the database was not updated since CCL 5, EPA relied on the data extracted and processed previously to support screening and classification for CCL 6. This information is described below. The NTP cancer classification database has since been updated and may include additional contaminants compared to the data collected for CCL 5 and 6.

Reference: HHS. n.d. National Institutes of Health. National Institutes of Environmental Health Sciences. National Toxicology Program (NTP). NTP Technical Reports Index. https://ntp.niehs.nih.gov/data/tr/index.html. Accessed April 2018.

Data processing for screening:

- **Data download:** EPA copied and pasted the technical report results table into a CSV file.
- **Data manipulation:** EPA altered the format of chemical identifiers for each entry (e.g., added DTXSIDs). The species name and the study summary results code were joined into a single field (for example, a result of SE in Male Mice is written as Male.Mice SE). EPA also converted cancer classifications to a comparable numeric scheme according to the same methodology used for CCL 3. This conversion is further explained in Section 2.4.4 of the main document. Data manipulation steps were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract the combined species and study result information. This information is comparable to a cancer classification.
- **Use in screening**: Cancer classifications were assigned Tier 3 screening points.

Data processing for classification:

- **Data download:** Data were downloaded as described for screening.
- **Data manipulation:** Data were manipulated as described for screening.
- **Extracted data elements**: Data were extracted as described for screening.
- **Use in classification:** Cancer classifications were presented in the Other Health Data section of the CIS.

Provisional Peer-Reviewed Toxicity Values (PPRTVs) – Environmental Protection Agency

Data description: The Provisional Peer-Reviewed Toxicity Value (PPRTV) program supports EPA's Superfund program by generating health assessments for compounds not already assessed under EPA's IRIS program. The health assessments generate provisional toxicity values like p-RfDs and p-CSFs and include toxicity values for both subchronic and chronic exposure scenarios and cancer classifications. For the purpose of screening compounds from the universe to the PCCL, these provisional toxicity values are considered analogous to other EPA toxicity values. This data source was used as a main data source for CCL 6. The PPRTV program publishes assessments that provide data and information supporting toxicity value development for individual contaminants. These documents were used as a supplemental source for CCL 6.

Reference: Oak Ridge National Laboratory. n.d. Provisional Peer-Reviewed Toxicity Values (PPRTVs). https://hhpprtv.ornl.gov/quickview/pprtv_compare.php. Accessed July 2023.

USEPA. n.d. Provisional Peer-Reviewed Toxicity Values. https://www.epa.gov/pprtv/provisional-peer-reviewed-toxicity-values-pprtvs-assessments. Accessed August 2024.

Data processing for screening:

- **Data download:** EPA selected all the individual chemicals and categories for comparison of PPRTV data from the PPRTV Library housed by Oak Ridge National Laboratory and copied each table into an individual Excel file.
- **Data manipulation:** EPA altered the format of chemical identifiers for each entry (e.g., added DTXSIDs) and converted cancer classifications to a comparable numeric scheme according to the same methodology used for CCL 3. This conversion is further explained in Section 2.4.4 of the main document.
- **Extracted data elements**: Oral toxicity values including provisional RfDs, subchronic provisional RfDs, and provisional CSFs were extracted in addition to cancer classifications. Inhalation data including provisional RfCs, subchronic provisional RfCs, and provisional inhalation unit risks (IURs) were also extracted.
- **Use in screening**: Provisional RfDs and CSFs were assigned Tier 1 screening points. Cancer classifications and subchronic provisional RfDs were assigned Tier 3 screening points. Inhalation values were not assigned screening points in CCL 6.

Data processing for classification:

- **Data download:** PPRTV assessments were downloaded from the assessment website in addition to the data elements downloaded as described for screening.
- **Data manipulation:** Data manipulation was as described for screening.
- **Extracted data elements:** In addition to the data extracted to support screening, EPA manually extracted data from PPRTV assessments including the publication years, RfDs, CSFs, cancer classifications, critical studies, and critical effects.
- **Use in classification:** EPA used provisional RfDs and CSFs to derive HRLs for CCL 6. EPA presented cancer classifications along with these data. EPA presented subchronic

provisional RfDs as well as all inhalation values extracted for screening in the Other Health Data section of the CIS.

Toxicity Criteria Database – California Environmental Protection Agency (CalEPA) Office of Environmental Health Hazard Assessment (OEHHA)

Data description: CalEPA's Office of Environmental Health and Hazard Assessment's (OEHHA) Chemical Database contains all of California's toxicity criteria information developed for chemicals evaluated by OEHHA. This information includes reference exposure levels, California Public Health Goals (PHGs), notification levels, child-specific RfDs, Proposition 65 safe harbor numbers, soil-screening levels, and fish advisories. This data source was used as a main data source for CCL 6. CalEPA also publishes documents supporting PHG derivation that were used as supplemental data sources for CCL 6.

References: California Environmental Protection Agency (CalEPA). n.d. Office of Environmental Health Hazard Assessment (OEHHA) Chemicals. https://oehha.ca.gov/chemicals. Accessed July 2023.

CalEPA. n.d. Office of Environmental Health Hazard Assessment. Public Health Goals (PHGs). https://oehha.ca.gov/public-health-goals-phgs. Accessed August 2024.

Data processing for screening:

- **Data download:** The option to export database as a comma separated values (CSV) file was selected from the OEHHA Chemicals website.
- **Data manipulation:** Results reported in scientific notation were reformatted for the results to be recognized as numerical values in R. Other steps were taken to make the extracted data consistent with data from other sources. Additionally, DTXSIDs were added. Data manipulation steps were conducted using R.
- **Extracted data elements**: PHGs and notification levels were extracted and treated as chronic duration benchmarks and oral slope factors were extracted and coded as cancer slope factors (CSFs). MADLs for chemicals causing reproductive toxicity, inhalation unit risks (IURs), and RfCs were also extracted.
- **Use in screening**: PHGs, notification levels, and CSFs were assigned Tier 1 screening points. MADLs, IURs, and RfCs were not assigned screening points for CCL 6.

Data processing for classification:

- **Data download:** PHG support documents were downloaded from the CalEPA PHG website in addition to the data elements downloaded as described for screening.
- **Data manipulation:** Data manipulation was as described for screening.
- **Extracted data elements:** In addition to the data extracted to support screening, EPA manually extracted data from PHG support documents including the publication years, acceptable daily dose (ADD) levels (considered equivalent to RfDs for CCL 6), oral slope factors (considered equivalent to CSFs for CCL 6), critical studies, and critical effects.
- **Use in classification:** EPA used ADDs and oral slope factors to derive HRLs for CCL 6. EPA presented PHGs, notification levels, MADLs, and all inhalation values extracted for screening in the Other Health Data section of the CIS.

Toxicity Reference Database (ToxRefDB) – Environmental Protection Agency

Data description: The Toxicity Reference Database (ToxRefDB) contains *in vivo* study data from over 5,900 guideline or guideline-like studies for over 1,100 chemicals. This is largely comprised of curated animal study data from repeat dose studies conducted according to Health Effects Series 870 guidelines. By employing a controlled vocabulary for enhanced data quality, ToxRefDB serves as a resource for quantitative dose-response information with given guideline specifications. It includes information regarding study design, chemical treatment, dosing, treatment group parameters, and treatment-related and critical effects. This data source was used as a main data source for CCL 6.

References: U.S. EPA. 2022. Toxicity Reference tables from toxrefdb_2_1. Retrieved from https://clowder.edap-cluster.com/datasets/61147fefe4b0856fdc65639b#folderId=62c5cfebe4b01d27e3b2d851

on June 19, 2023.

Feshuk, M. Kolaczkowski, L. Watford, S. and Paul Friedman K. 2023. ToxRefDB v2.1: update to curated *in vivo* study data in the Toxicity Reference Database. *Frontiers in Toxicology* 5. doi.org/10.3389/ftox.2023.1260305.

Data processing for screening:

- **Data download:** EPA downloaded the ".sql" and ".dmp" file from the EPA Clowder workspace to obtain the MySQL database. Utilizing the R package mentioned in the User Guide (present at same link), EPA converted all tables in the database to CSV files for easier access.
- **Data manipulation:** EPA combined the downloaded chemical, study, and POD tables using R script. The POD table contains the NOAEL, LOAEL, and route of administration (e.g., oral, inhalation) information for extraction. The chemical table contains DTXSIDs and chemical names. The study table contains the "processed" column, where a value of 1 indicates a guideline or guideline-like study. Data was filtered to only include studies that had a value equal to 1, indicating it was conducted using guideline or guideline-like (i.e., best available) methods. The combined file was then organized based on POD type (LOAEL or NOAEL), study duration or type (chronic, subchronic, acute, reproductive, multigenerational, or developmental), and chemical administration route (oral or inhalation).
- **Extracted data elements**: EPA extracted all chronic, subchronic, and acute NOAELs and LOAELs from both oral and inhalation studies. Studies with reproductive, developmental, or multigenerational designs were considered chronic due to potential long-term consequences that may result from short-term exposure during sensitive windows of reproduction and development, in accordance with EPA guidance (USEPA, 1996).
- **Use in screening:** Chronic oral NOAELs and LOAELs (including oral reproductive, developmental, and multigenerational PODs) were assigned Tier 2 screening points. Subchronic oral NOAELs and LOAELs were assigned Tier 4 screening points. Acute oral NOAELs and LOAELs and all inhalation values were not assigned screening points in CCL 6.

Data processing for classification:

- **Data download:** Data were downloaded as described for screening.
- **Data manipulation:** Data were manipulated as described for screening.
- **Extracted data elements**: Data were extracted as described for screening.
- **Use in classification:** EPA presented only the minimum and maximum values for all extracted data elements in the Other Health Data section of the CIS.

M.2.2 Occurrence Data Sources

ATSDR Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) Substance Priority List – Centers for Disease Control and Prevention (CDC)

Data description: The Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) requires the Agency for Toxic Substances and Disease Registry (ATSDR) and EPA to prepare the Substance Priority List, in order of priority, of substances most commonly found at facilities on the National Priorities List (NPL) and that are determined to pose the most significant potential threat to human health due to their known or suspected toxicity and potential for human exposure at these NPL sites. SDWA requires that CERCLA priority substances be considered as part of the CCL development process. This data source was used as a main data source for CCL 6. (Description adapted from ATSDR's Substance Priority List website.)

Reference: Centers for Disease Control and Prevention (CDC). 2022. 2022 ATSDR Substance Priority List. https://www.atsdr.cdc.gov/programs/substance-priority-list.html. Accessed July 2023.

Data processing for screening:

- **Data download:** Data were copied and pasted from the online table into an Excel file.
- **Data manipulation:** EPA wrote R code to extract chemicals on the CERCLA list, and assigned each chemical a value of 1. CASRNs were extracted from the original data.
- **Extracted data elements**: EPA extracted chemicals identified as being on the Substance Priority List.

Data processing for classification:

- CERCLA data were not reformatted into an individual simple file for classification and was not specifically reprocessed beyond its inclusion in the Universe data for later use on the CIS sheets.

Chemical Data Reporting (CDR) Results – Environmental Protection Agency

Data description: The 2020 CDR data provides information collected by EPA under the Toxic Substances Control Act (TSCA) about chemical substances manufactured or imported in 2019. This data source was used as a main data source for CCL 6.

Reference: USEPA. 2020. 2020 Chemical Data Reporting (CDR) Results. https://www.epa.gov/chemical-data-reporting/access-cdr-data#2020. Accessed July 2023.

Data processing for screening:

- **Data download:** EPA downloaded the 2020 CDR Nationally Aggregated Production Volumes dataset for use in CCL 6.
- **Data manipulation:** EPA wrote R code to extract the most recent year of data for the nationally aggregated production volume for chemicals (2019). EPA added CASRNs to the simple file output. EPA inferred the unit of pounds based off the other reported PV columns.
- **Extracted data elements**: EPA extracted the nationally aggregated production volume for chemicals.

Data processing for classification:

- CDR data were not reformatted into an individual simple file for classification but were incorporated into a separate input file for inclusion in the CIS sheets.

Community Water System Survey (CWSS) – Environmental Protection Agency

Data description: The 2006 CWSS (USEPA, 2009) gathered data on the financial and operating characteristics of a random sample of community water systems (CWSs) nationwide. All systems serving more than 500,000 people (94 systems in 2006) were included in the survey, and systems in that size category were asked questions about concentrations of unregulated contaminants in their raw and finished water. Not all systems responded to the survey and, of the systems that responded, not all answered every question. EPA supplemented the dataset by gathering additional information about contaminant occurrence at the systems in this size category from publicly available sources (e.g., consumer confidence reports). Note that, because reported results are incomplete, they are only illustrative, not statistically representative, and used only as supplemental information. The EPA relied on the data processed and extracted during CCL 5 from this data source to support screening and classification for CCL 6. This information is described below. This data source was used as a supplemental data source for CCL 6.

References: USEPA. 2009. Community Water System Survey 2006. Volume 1: Overview. EPA 815-R-09- 001. February 2009. USEPA. 2009. Community Water System Survey 2006. Volume II: Detailed Tables and Survey Methodology. EPA 815-R-09-002. May 2009.

Data download: EPA extracted data from the publication and saved on two Excel spreadsheets.

Data manipulation: For concentrations reported in units other than parts per billion (ppb) (as noted in the raw data footnote), a column was added to denote what units the data were in. The raw and finished water data were in two separate sheets so they were combined, and a column was added to designate data as either finished or ambient water.

Extracted data elements: EPA wrote R code to extract the median and 90th percentiles of detections in addition to total number of systems, total number of samples, number of samples with detections, and percentage of samples with detections for each contaminant. Raw water data were classified as ambient water data. This data source was treated as a non-nationally representative occurrence water study providing ambient or finished water data, where appropriate.

CompTox Chemicals Dashboard – Environmental Protection Agency

Data description: The CompTox Chemicals Dashboard is a database developed by EPA that compiles information from many sites, databases, and sources into one web application. The database includes experimental, modeled, and use information for over 1,200,000 chemicals. This data source was used as a supplemental data source for CCL 6. The data pulled from this data source were added to the pre-universe to create the CCL 6 universe file.

Reference: Williams, A.J., C.M. Grulke, J. Edwards, A.D. McEachran, K. Mansouri, N.C. Baker, G. Patlewicz, I. Shah, J.F. Wambaugh, R.S. Judson, and A.M. Richard. 2017. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. Journal of Cheminformatics. 9:61. doi:10.1186/s13321-017-0247-6.

Data download: EPA downloaded CompTox Chemicals Dashboard data in May 2024 from the following website address: https://comptox.epa.gov/dashboard/. The EPA Office of Water with the assistance of Office of Research and Development (ORD) pulled downloaded data using the batch search tool for all unique DTXSIDs identified during pre-universe development.

Data manipulation: No data manipulation was required.

Extracted data elements: See Section N.4 for data elements and their descriptions extracted from the CompTox Chemicals Dashboard for use in CCL 6.

“Concentrations of prioritized pharmaceuticals in effluents from 50 large wastewater treatment plants in the US and implications for risk estimation” – Kostich et al. 2014

Data description: This is an EPA Office of Research and Development (ORD) publication that measures 56 active pharmaceutical ingredients in the effluents of 50 large wastewater treatment plants in the U.S. in 2011. The 50 plants sampled in this study discharge 6 billion gallons of effluent per day of water, which accounts for about 17% of all the wastewater produced by wastewater treatment plants in the country. The EPA relied on the data processed and extracted during CCL 5 from this data source to support screening and classification for CCL 6. This information is described below. This data source was used as a main data source for CCL 6. This data source was considered a non-nationally representative ambient occurrence study and the water data were further categorized as wastewater effluent for the classification step of CCL 6.

Reference: Kostich, M.S., A.L. Batt, and J.M. Lazorchak. 2014. Concentrations of prioritized pharmaceuticals in effluents from 50 large wastewater treatment plants in the US and implications for risk estimation. Environmental Pollution*.* 184: 354-359. https://doi.org/10.1016/j.envpol.2013.09.013.

Data processing for screening:

- **Data download:** EPA downloaded the publication and supplemental data file for use in CCL 6. Table 1 of the main text of the publication was copied into an Excel spreadsheet.
- **Data manipulation:** Percentage of detections was calculated and DTXSIDs were added. Data manipulations were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract maximum measured concentration and percentage of detections.

Data processing for classification:

- **Data download:** The supplemental data file downloaded in the pre-processing steps for screening above was used to extract data for the classification step.
- **Data manipulation:** Data denoted as "Censored" were removed and non-detections were reclassified as concentrations below the method reporting level. DTXSIDs were added. Data manipulation steps were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract the minimum, median, 90th percentile, and maximum concentration of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

Disinfection By-product Information Collection Rule (DBP ICR) – Environmental Protection Agency

Data description: The DBP ICR "Aux 1" database contains monitoring data from large public water systems (PWSs serving a population greater than or equal to 100,000) for the 18-month period of July 1997 to December 1998. A total of 296 water systems reported data; included in the database are monitoring results for microbials and disinfection byproducts (DBPs), plant treatment, source water characteristics and disinfectant type information. Summary reports on microbial and DBP data at national, state, and water system levels can be retrieved via the database. This data source was used as a main data source for CCL 6. The data was considered a non-nationally representative finished water occurrence source.

Reference: USEPA. 2000. ICR Auxiliary 1 Database. EPA 815-C-00-002.

Data processing for screening:

- **Data download:** EPA downloaded the DBP ICR Aux 1 Microsoft Access database on June 20, 2023 from the following website address: https://www.epa.gov/dwsixyearreview/supplemental-data-six-year-review-3.
- **Data manipulation:** Analyte ID and analyte results data were extracted from the Microsoft Access database, saved as comma separated values (CSV) files, then combined into one CSV file. Concentrations reported as -999 (non-detections) were converted to ½ the MRL. Maximum concentration of detections for each contaminant was calculated. Data manipulation steps were conducted using R. EPA extracted analyte results from multiple tables in the Access database.
- **Extracted data elements:** EPA wrote R code to calculate the count of sites, count of site with detections, percent detections, and maximum concentration of detections.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** The data flag for detections developed in prescreening to filter on detected data for summary statistics calculations was used in classification. The ICRPWSID column was used as a site ID for calculating site statistics. Analytes were presented with their full name except for some instances where no full name was provided from the analyte_id.csv crosswalk. NAs were used to replace data where there is no minimum or calculable value.
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

“Evaluating the extent of pharmaceuticals in the surface waters of the United States using a National-scale Rivers and Streams Assessment survey” – Batt et al. 2016

Data description: This is an EPA Office of Research and Development (ORD) publication focusing on active pharmaceutical ingredients and potential risks to aquatic life. The authors sampled 182 sites in rivers proximal to urban streams and measured the concentrations of 46 analytes representing many classes of active pharmaceutical ingredients. The EPA relied on the data processed and extracted during CCL 5 from this data source to support screening and classification for CCL 6. This information is described below. This data source was used as a main data source for CCL 6. The data source was treated as a non-nationally representative ambient water study.

Reference: Batt, A.L., T.M. Kincaid, M.S. Kostich, J.M. Lazorchak and A.R. Olsen. 2016. Evaluating the extent of pharmaceuticals in surface waters of the United States using a national-scale rivers and streams assessment survey. Environmental Toxicology and Chemistry. 35(4):874-81. https://doi.org/10.1002/etc.3161

Data processing for screening:

- **Data download:** EPA downloaded the publication and supplemental data file.
- **Data manipulation:** Data manipulation was minimal and limited to adding DTXSIDs.
- **Extracted data elements**: EPA wrote R code to extract maximum concentrations from Table S5 of the supplemental data file and percentage of sites with detections from Table 2 of the main text of the publication.

Data processing for classification:

- **Data download:** The supplemental data file downloaded for screening was used for extracting data used in classification.
- **Data manipulation:** Summary statistics were calculated from the data in Table S1. Full dataset in the supplemental data file. DTXSIDs were added. Data manipulations were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract minimum, median, 90th percentile, and maximum concentrations of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

“Expanded target-chemical analysis reveals extensive mixed-organic-contaminant exposure in U.S. streams” – Bradley et al. 2017

Data description: This publication, published by the United States Geological Survey (USGS) and the EPA’s Office of Research and Development (ORD), provides water data for 719 compounds sampled in 38 streams across the U.S. using 14 different methods. Study locations include a mixture of urban and agricultural watersheds. The EPA relied on the data processed and extracted during CCL 5 from this data source to support screening and classification for CCL 6. This information is described below. This data source was used as a main data source for CCL 6. The data was categorized as a non-nationally representative ambient occurrence source.

Reference: Bradley, P.M., C.A. Journey, K.M. Romanok, L.B. Barber, H.T. Buxton, W.T. Foreman, E.T. Furlong, S.T. Glassmeyer, M.L. Hladik, L.R. Iwanowicz, D.K. Jones, D.W. Kolpin, K.M. Kuivila, K.A. Loftin, M.A. Mills, M.T. Meyer, J.L. Orlando, T.J. Reilly, K.L. Smalling, and D.L. Villeneuve. 2017. Expanded Target-Chemical Analysis Reveals Extensive Mixed-Organic-Contaminant Exposure in U.S. Streams. Environmental Science & Technology. 51(9): 4792–4802. https://doi.org/10.1021/acs.est.7b00012.

Data processing for screening:

- **Data download:** EPA downloaded the publication and supplemental data files.
- **Data manipulation:** Data manipulation was minimal and restricted to adding DTXSIDs
- **Extracted data elements**: EPA wrote R code to extract maximum concentration data and percentage of detections from Table 3 of the supplemental data file.

Data processing for classification:

- **Data download:** The supplemental data file downloaded for screening was used for extracting data used in classification.
- **Data manipulation:** Summary statistics were calculated from data in Table S3 of the supplemental data file. DTXSIDs were added. Data manipulations were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract minimum, median, 90th percentile, and maximum concentration of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) registered pesticides and pesticide ingredients – Environmental Protection Agency

Data description: This list represents the active pesticide and pesticide ingredients currently registered by EPA in the US. SDWA requires that registered pesticides be considered in CCL development. This data source was used as a main data source for CCL 6.

Reference: USEPA. 2023. Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA). Office of Pesticide Programs. https://www.epa.gov/laws-regulations/summary-federal-insecticide-fungicide-and-rodenticide-act.

Data processing for screening:

- **Data download:** The EPA downloaded a list of entries that have been classified in the U.S. as pesticidal "active ingredients" (conventional, antimicrobial, or biopesticidal agents) The list of compounds included in the Pesticide Chemical Search Database was accessed on July 19, 2023 via the EPA's CompTox Chemicals Dashboard from the following website: https://comptox.epa.gov/dashboard/chemical-lists/EPAPCS. This list was last updated in 2023.
- **Data manipulation:** The EPA wrote R code to extract chemicals on the FIFRA list, assigned each chemical a value of 1, and extracted the DTXSIDs and CASRNs.
- **Extracted data elements**: EPA extracted chemicals identified as being a pesticide or pesticide ingredient registered on the FIFRA list.

Data processing for classification:

- FIFRA data was not reformatted into an individual simple file for classification and was not specifically reprocessed beyond its inclusion in the Universe data for later use on the CIS sheets.

"Legacy and emerging perfluoroalkyl substances are important drinking water contaminants in the Cape Fear River Watershed of North Carolina" – Sun et al. 2016

Data description: This is an EPA Office of Research and Development (ORD) and North Carolina State University publication focusing on short and long-chain per- and poly-fluoroalkyl substances in ambient water downstream and upstream of a fluorochemical manufacturing plant in the Cape Fear River watershed in North Carolina. Sampling occurred at three water treatment plants over a six-month period in 2013. Though this study sampled in one geographic region, the results are relevant to CCL development because they include ambient water monitoring concentrations of substances in an emerging class of compounds thought to be highly persistent in the environment and potentially harmful at low doses. The EPA relied on the data processed and extracted during CCL 5 from this data source to support screening and classification for CCL 6. This information is described below. This data source was used as a main data source for CCL 6. This data source was considered a non-nationally representative ambient water study.

Reference: Sun, M., E. Arevalo, M. Strynar, A. Lindstrom, M. Richardson, B. Kearns, A, Pickett, C. Smith, and D.R.U. Knappe. 2016. Legacy and emerging perfluoroalkyl substances are important emerging water contaminants in the Cape Fear River Watershed of North Carolina. Environmental Science & Technology Letters. 3(12): 415–419. https://doi.org/10.1021/acs.estlett.6b00398.

Data processing for screening:

- **Data download:** EPA downloaded the publication and supplemental data file.
- **Data manipulation:** Table S6 of the supplemental data file was copied and pasted into an Excel spreadsheet. Data manipulation was minimal and limited to adding DTXSIDs.
- **Extracted data elements**: EPA wrote R code to extract maximum concentrations.

Data processing for classification:

- **Data download:** The same publication and supplemental data file was used for extracting data elements for the classification step.
- **Data manipulation:** Data manipulation was minimal and limited to adding DTXSIDs.
- **Extracted data elements**: EPA wrote R code to extract minimum and maximum concentrations of detections, in addition to total number of sites, number of sites with detections, and percentage of sites with detections.

“Multimedia screening of contaminants of emerging concern (CECS) in coastal urban watersheds in southern California (USA).” – Maruya et al. 2016

Data description: Maruya et al. 2016 examined aqueous concentrations of contaminants of emerging concern (including pharmaceuticals, personal care products, commercial and household chemicals, and pesticides) sampled from the Santa Clara River watershed taken October 2013. The Santa Clara River receives treated wastewater from municipal wastewater treatment plants. This data source was used as a main data source for CCL 6. The data was categorized as a non-nationally representative ambient occurrence source.

Reference: Maruya, K.A., Dodder, N.G., Sengupta, A., Smith, D.J., Lyons, J.M., Heil, A.T. and Drewes, J.E., 2016. Multimedia screening of contaminants of emerging concern (CECS) in coastal urban watersheds in southern California (USA). Environmental toxicology and chemistry, 35(8), pp.1986-1994. https://doi.org/10.1002/etc.3348

Data processing for screening:

- **Data download:** EPA copied and pasted Table S6 from the Supplemental Information into a blank CSV file. EPA extracted Table S5 using power query in MS Excel and then saved it as a CSV.
- **Data manipulation:** Site ID columns were extracted from columns starting with SCR or SCE from tables S5 and S6. Entries with “BRL” were treated as non-detections. For chemicals where there were only non-detections, half the detection limit was substituted to calculate the maximum concentration.
- **Extracted data elements:** EPA wrote R code to extract the maximum ambient concentration, and the percent ambient detection by site.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** Code was written to filter on detected data (where value entries were not NA) for summary statistics calculations. The Site column was used as a site ID for calculating site statistics.
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

National Aquatic Resource Surveys, 2017 National Lakes Assessment - Environmental Protection Agency

Data description: The 2017 National Lakes Assessment reports on the condition of lakes and reservoirs in the conterminous United States for 2017. The datasets provide concentration data for cyanotoxins, atrazine, and E. coli. This data source was used as a main data source for CCL 6. The data was categorized as a nationally representative ambient occurrence source.

Reference National Lakes Assessment (NLA) Results. 2017. https://www.epa.gov/national-aquatic-resource-surveys/national-lakes-assessment-2017-results. Accessed October 2023.

Data processing for screening:

- **Data download:** EPA downloaded "National Lakes Assessment 2017 Data Files (zip)" from the reference link.
- **Data manipulation:** EPA wrote R code to conduct light data cleaning. Non-detections were replaced with half the RL or MDL.
- **Extracted data elements**: EPA wrote R code to extract the maximum concentrations and percent detections by site.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** Code was written to filter on detected data for summary statistics calculations using data flags for detected data developed in pre-processing.
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

National Aquatic Resource Surveys, Rivers and Streams 2018-2019 Algal Toxin - Environmental Protection Agency

Data description: The National Rivers and Streams Assessment is a national survey that collects data on biological, chemical, and physical indicators of the ecological condition of streams and rivers. The survey uses a probability-based sample design to assess the full range of flowing waters in the United States. The Algal toxin sampling data from the National Rivers and Streams Assessment 2018-2019 was used as a data source for CCL 6. This data source was used as a main data source for CCL 6. The data was categorized as a nationally representative ambient occurrence source.

Reference: U.S. EPA. 2022. National Rivers and Streams Assessment 2018-19: Environmental Protection Agency. https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys. Accessed October 2023.

Data processing for screening:

- **Data download:** EPA downloaded nrsa_1819_algal_toxin_-_data.csv and the algal toxin metadata from the reference link.
- **Data manipulation:** EPA wrote R code to change the Analyte abbreviations to full analyte names using the information in the metadata. Data was filtered to only include Cylindrospermopsin (CYLSPER), Microcystin (MICX), and Microcystin in legacy bottle type (MICZ). For the maximum concentration, non-detections were replaced with half the RL or MDL.
- **Extracted data elements**: EPA wrote R code to extract the maximum concentrations, and percent of sites with detections.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** EPA wrote code to filter on detected data for summary statistics calculations using data flags for detected data developed in pre-processing.
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant. Data was categorized as a national ambient occurrence source.

National Health and Nutrition Examination Survey (NHANES) – Centers for Disease Control and Prevention (CDC)

Data description: The National Report on Human Exposure to Environmental Chemicals offers a comprehensive set of biomonitoring data, collected continuously from the years 1999-2000 through 2017-2018, that is nationally representative. This report includes information summarizing the biomonitoring results of the National Health and Nutrition Examination Survey (NHANES). The purpose of the NHANES biospecimen program is to store and analyze biospecimens collected during the NHANES survey to help address future medical, environmental, and public health research questions. The stored specimen program includes samples of urine, plasma, serum and DNA that can be used by researchers. The CDC's National Report on Human Exposure to Environmental Chemicals summarizes the NHANES biomonitoring results for compounds that may be environmental contaminants. This data source was used as a main data source for CCL 6.

Reference: Centers for Disease Control and Prevention, U.S. Department of Health and Human Services. National Report on Human Exposure to Environmental Chemicals. Updated March 2022. Accessed September 2023. https://www.cdc.gov/exposurereport/

Data processing for screening:

- **Data download:** EPA downloaded individual tables from NHANES, selecting "Analysis of Whole Blood, Serum, and Urine Samples, NHANES 1999-2018" under step 1: sample types and selecting one terminal chemical sub option. EPA converted each PDF to Excel format. Light data cleaning was done.
- **Data manipulation:** EPA wrote R code to read Excel data into R and combine it in a preparatory step and cleaned the total population dataset. Data were labeled as urinary, blood, or serum based on the table name.
- **Extracted data elements**: 90th percentile total population from all relevant NHANES data rows.

Data processing for classification:

- NHANES data was not reformatted into an individual simple file for classification and was not specifically reprocessed beyond its inclusion in the Universe data for later use on the CISs.

National Inorganics and Radionuclides Survey (NIRS) – Environmental Protection Agency

Data description: In the mid-1980s, EPA implemented NIRS to provide a statistically representative sample of the national occurrence of select inorganic and radionuclide contaminants in community water systems (CWSs) served by groundwater. The survey is stratified based on system size (population served by the system). Most of the NIRS data are from smaller systems (92% from systems serving 3,300 persons or fewer). The NIRS database includes findings for 42 radionuclides and inorganic compounds (IOCs). NIRS provides contaminant occurrence data from 989 groundwater CWSs in 49 states (all except Hawaii) as well as Puerto Rico. Surface water systems were not included in the study, in part because IOCs tend to occur more frequently and at higher concentrations in groundwater than in surface water. Each of the 989 randomly selected CWSs was sampled once between 1984 and 1986. The NIRS data were collected in a randomly designed sample survey; therefore, the summary statistics are representative of national occurrence in groundwater CWSs. Information about NIRS monitoring and data analysis is available in Longtin (1988) and USEPA (2008). One limitation of the NIRS is a lack of occurrence data for surface water systems. The EPA relied on the data processed and extracted during CCL 5 from this data source to support screening and classification for CCL 6. This information is described below. This data source was used as a main data source for CCL 6. The data was categorized as a nationally representative finished water occurrence source.

References:

Longtin, J.P. 1988. Occurrence of Radon, Radium and Uranium in Groundwater. Journal of the American Water Works Association*.* 80(7): 84-93.

USEPA. 2008. The Analysis of Occurrence Data from the Unregulated Contaminant Monitoring (UCM) Program and National Inorganics and Radionuclides Survey (NIRS) in Support of Regulatory Determinations for the Second Drinking Water Contaminant Candidate List (CCL 2). EPA 815-R-08-014. June 2008.

Data processing for screening:

- **Data download:** NIRS data were originally stored in a Lotus 1-2-3 spreadsheet. Data were converted to Excel in the early 2000s. Data are in a horizontal format with one row per CWS sampled. The chemical concentration data are organized in columns.
- **Data manipulation:** DTXSIDs were added, and summary statistics were calculated in Excel.
- **Extracting relevant data elements**: EPA wrote R code to extract maximum concentration and percentage of detections.

Data processing for classification:

- **Data download:** The same data file used in the screening step was used for extracting data for classification.
- **Data manipulation:** No additional data manipulations were needed.
- **Extracted data elements**: EPA wrote R code to extract the minimum, median, 90th percentile, and maximum concentration of detections in addition to the minimum

sampling reporting level, total number of systems, number of systems with detections, and percentage of systems with detections for each chemical.

National Water Information System (NWIS) and National Water-Quality Assessment (NAWQA) Project – United States Geological Survey (USGS)

Data description: The Water Quality Portal (WQP) is the premiere source of discrete water-quality data in the United States and beyond. This cooperative service integrates publicly available water-quality data from the United States Geological Survey (USGS), the Environmental Protection Agency (EPA), and over 400 state, federal, tribal, and local agencies. The WQP provides National Water Information System (NWIS) data, the USGS repository of water resources data across the US. The WQP also provides National Water Quality Assessment (NAWQA) data, a repository of freshwater water quality data from USGS' National Water Quality Program. NWIS and NAWQA data include finished water, groundwater, and surface water data. NAWQA and NWIS were used as main data sources for CCL 6. NAWQA was categorized as a nationally representative ambient occurrence source and a non-nationally representative finished occurrence source. NWIS was categorized as a non-nationally representative ambient and finished occurrence source.

Reference: United States Geological Survey (USGS). n.d. National Water-Quality Assessment (NAWQA) Project. Accessed via the Water Quality Portal (WQP). URL: https://www.waterqualitydata.us/portal/. Accessed December 2023.

United States Geological Survey (USGS). n.d. National Water Information System (NWIS). USGS Water Data for the Nation. Accessed via the Water Quality Portal (WQP). URL: https://www.waterqualitydata.us/portal/. Accessed December 2023.

Data processing for screening:

- **Data download:** In the Water Quality Portal, EPA manually downloaded monitoring and station location data from January 1991 to December 2022.
- **Data manipulation:** EPA used R code to filter and clean the data (i.e., removing all negatives, zeros, and character data from results; selecting data with specific units; performing unit conversions; defining detections and non-detections). The data were split into NWIS and NAWQA subsets for ambient (defined as groundwater or surface water) and finished water data. There is no substitution of the detection limit for NWIS or NAWQA data. This means that some maximum concentrations are blank because there is no maximum of the data to report for that parameter.
- **Extracted data elements**: EPA wrote R code to extract to extract maximum concentration and percent detections by site for NWIS and NAWQA data, separated by ambient and finished water.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** The process for filtering and cleaning data developed for prescreening was used for classification. Data was separated into ambient (groundwater and surface water) and finished water summaries. The data flag for detections developed in prescreening was used to filter on detected data for summary statistics calculations.

- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

National Primary Drinking Water Regulations (NPDWRs) – Environmental Protection Agency

Data description: National Primary Drinking Water Regulations (NPDWRs) are legally enforceable primary standards and treatment techniques applicable to public water systems (PWSs). EPA publishes maximum contaminant levels (MCLs) and maximum contaminant level goals (MCLGs) as a means to protect public health by limiting the levels of contaminants in drinking water. While contaminants with MCLs/MCLGs are regulated and therefore not considered further in the CCL process, EPA collected these data to be used as reference for CCL 6.

Reference: USEPA. Office of Water. n.d. National Primary Drinking Water Regulations. https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations. Accessed July 2023.

Data processing for screening:

- **Data download:** NPDWRs were copied and pasted into a CSV file.
- **Data manipulation:** EPA added CASRNs to the dataset and matched CASRNs and names to the DTXSIDs.
- **Extracted data elements:** EPA wrote R code to extract Maximum Contaminant Levels (MCLs) and Maximum Contaminant Level Goals (MCLGs).

Data processing for classification:

- NPDWR data was not reformatted into an individual simple file for classification.

"Nationwide reconnaissance of contaminants of emerging concern in source and treated drinking waters of the United States" – Glassmeyer et al. 2017

Data description: This is an EPA Office of Research and Development (ORD) and USGS publication describing source water and drinking water concentrations of emerging contaminants. This was a two-phase study and sampling occurred between 2007 and 2012. Phase II of the study included more analytes and sometimes used more sensitive methods than Phase I. In Phase I, 87 compounds were monitored at nine treatment plants. In Phase II, 247 analytes were included at 25 drinking water treatment plants. The EPA relied on the data processed and extracted during CCL 5 from this data source to support screening and classification for CCL 6. This information is described below. This data source was used as a main data source for CCL 6. Source water data were categorized as non-nationally representative ambient occurrence data and treated water data were categorized as non-nationally representative finished occurrence data.

Reference: Glassmeyer, S.T., E.T. Furlong, D.W. Kolpin, A.L. Batt, R. Benson, J.S. Boone, O. Conerly, M.J. Donohue, D.N. King, M.S. Kostich, H.E. Mash, S.L. Pfaller, K.M. Schenck, J.E. Simmons, E.A. Varughese, S.J. Vesper, E.N. Villegas, and V.W. Wilson. 2017. Nationwide reconnaissance of contaminants of emerging concern in source and treated drinking waters of the United States. Science of The Total Environment*.* 581-582: 909-922. https://doi.org/10.1016/j.scitotenv.2016.12.004.

Data processing for screening:

- **Data download:** EPA downloaded the publication and supplemental data file. Table S2 of the supplemental data file was used to extract maximum concentration and detection information.
- **Data manipulation:** If a contaminant was measured in Phase I and Phase II of the study, the Phase II results were used. If a maximum concentration was reported as a non-detect, or "nd," the maximum concentration was replaced with 0. If a contaminant concentration was reported as "QL," or all measurements were qualitative, maximum concentrations were replaced with half of the reporting limit (RL) or half of the lowest concentration minimum reporting level (LCMRL). DTXSIDs were added. Data manipulation steps were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract maximum concentrations and qualitative detection rates for source and treated waters. Qualitative detection rates were used in the screening step as these metrics are a more conservative estimate of detection than are quantitative detection rates.

Data processing for classification:

- **Data download:** The publication and supplemental data files downloaded for screening were used to extract data used in classification.
- **Data manipulation:** Data manipulation was minimal and restricted to adding DTXSIDs. Concentration data as reported in the publication were used in the classification step.
- **Extracted data elements**: EPA wrote R code to extract median
- and maximum concentration of detections, total number of sites, and qualitative and quantitative site detection rates in source and treated waters. Quantitative detection rate

data are relevant to the classification step and included on the contaminant information sheets (CISs). Sampling year ranges for each study phase and reporting limits were also extracted.

"Nationwide reconnaissance of contaminants of emerging concern in source and treated drinking waters of the United States: Pharmaceuticals" – Furlong et al. 2017

Data description: This is an EPA Office of Research and Development (ORD) and USGS publication focusing on active pharmaceutical ingredients and their concentrations in water samples collected from 25 drinking water treatment plants (DWTPs) between 2007 and 2012. This was a two-phase study and includes sampling results in source water and finished drinking water. Phase II of the study included more analytes and sometimes used more sensitive methods than Phase I. There were 24 pharmaceuticals in Phase I and 118 in Phase II. This study is part of a series of papers published using the dataset of source and treated water samples from 25 DWTPs. The EPA relied on the data processed and extracted during CCL 5 from this data source to support screening and classification for CCL 6. This information is described below. This data source was used as a main data source for CCL 6. Source-water samples were categorized as non-nationally representative ambient occurrence data and treated-water samples were categorized as non-nationally representative finished occurrence data.

Reference: Furlong, E.T., A.L. Batt, S.T. Glassmeyer, N.C. Noriega, D.W. Kolpin, H. Mash, and K.M. Schenk. 2017. Nationwide reconnaissance of contaminants of emerging concern in source and treated drinking waters of the United States: Pharmaceuticals. Science of The Total Environment. 579: 1629-1642. https://doi.org/10.1016/j.scitotenv.2016.03.128.

Data processing for screening:

- **Data download:** EPA downloaded the publication and supplemental data file.
- **Data manipulation:** Tables 1 and 2 from the main text of the Furlong et al. 2017 publication were copied and pasted into an Excel spreadsheet. Some results reported in this publication are also published in Glassmeyer et al. 2017. Any results reported in both publications were considered as part of the Glassmeyer et al. 2017 data source to avoid duplication. If a contaminant was measured in Phase I and Phase II of the study, Phase II results were used. DTXSIDs were added. Data manipulation steps were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract maximum concentrations and qualitative percentage of detection data in finished and source waters. Qualitative detection frequencies were used in the screening step as these metrics are a more conservative estimate of detection than quantitative detection rates.

Data processing for classification:

- **Data download**: The publication and supplemental data files downloaded for screening were used to extract data used in classification.
- **Data manipulation:** The data manipulation steps described in the pre-processing steps for screening above were used to extract data for classification.
- **Extracted data elements**: EPA wrote R code to extract median and maximum concentration of detections and qualitative and quantitative site percentage of detection rates in finished and source waters. Quantitative detection rate data are relevant to the

classification step and included on the contaminant information sheets (CISs). Sampling year ranges for each study phase and reporting limits were also extracted.

National Energy Water Treatment and Speciation Database (NEWTS) Coal Mine Drainage Dataset from Cravotta Brady (2015) – Department of Energy (DOE)

Data description: This dataset contains information on water quality composition including inorganic compounds from untreated and treated streams of coal-mine discharge from coal mining and coal processing locations in Pennsylvania. This data source was used as a main data source for CCL 6. The data was classified as a non-nationally representative ambient occurrence source and the occurrence data were further categorized as wastewater effluent for the classification step of CCL 6.

Reference: Nicholas Siefert, Zineb Belarbi, Alison Fritz, Madison Wenzlick, NEWTS Coal Mine Drainage Dataset from Cravotta Brady (2015), 3/30/2023, https://edx.netl.doe.gov/dataset/newts-coal-mine-drainage-dataset-from-cravotta-brady-2015, DOI: 10.18141/1964003

Data processing for screening:

- **Data download:** EPA downloaded "cravottabrady2015_pa-amd_data_all-tabs.xlsx" from the reference link and used the "Original Stream Data_All" tab from the file for analysis.
- **Data manipulation:** A crosswalk was created to assign full names and units to abbreviated chemicals. EPA used "Total" data columns for chemicals when available. For chemicals with only non-detect data, maximum concentrations were reported as ½ the highest level of detection. Only effluent (outflow) data was used in analysis.
- **Extracted data elements**: EPA wrote R code to extract maximum concentrations, and percentage of sites with detections for each analyte.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** EPA wrote code to filter on detected data for summary statistics calculations using data flags for detected data developed in pre-processing. Similar data cleaning procedures from prescreening were applied to clean contaminant names for output.
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant for effluent data.

NEWTS Geothermal Technologies Office Rare Earth Element Assessment Report Database – Department of Energy (DOE)

Data description: This dataset represents produced water data, including oil & gas and coal bed methane produced waters, and geothermal waters. It includes information on rare earth elements and critical minerals. This data source was used as a main data source for CCL 6. The data was classified as a non-nationally representative ambient occurrence source and further categorized as wastewater effluent for the classification step of CCL 6.

Reference: Nicholas Siefert, Alison Fritz, Zineb Belarbi, Madison Wenzlick, Justin Mackey, Kathryn Smith, Burt Thomas, Quillinan, et al 2018 DOE Geothermal Technology Office REE Report for NEWTS Database and Case Studies, 2/13/2023, https://edx.netl.doe.gov/dataset/quillinan-et-al-2018-doe-geothermal-technology-office-ree-report-for-newts-database-and-case-studies, DOI: 10.18141/1924081

Data processing for screening:

- **Data download:** EPA downloaded quillinan_gto_ree_all_tabs.xlsx file from the reference link and used the "Original" tab for data calculations.
- **Data manipulation:** EPA created a crosswalk to clean parameter names, assign units, and label non-detections. For chemicals with only non-detections, maximum concentrations were reported as ½ the highest level of detection, if available.
- **Extracted data elements**: EPA wrote R code to extract maximum concentrations, and percentage detection across samples for each analyte.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** EPA wrote code to filter on detected data for summary statistics calculations using data flags for detected data developed in pre-processing.
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of samples, number of samples with detections, and percentage of samples with detections for each contaminant.

NEWTS Ash Database – Environmental Protection Agency

Data description: The U.S. EPA Ash Database contains ash pond effluent data from selected U.S. coal power plants. This data source was used as a main data source for CCL 6. The data was classified as a non-nationally representative ambient occurrence source and further categorized as wastewater effluent for the classification step of CCL 6.

Reference: Nicholas Siefert, Zineb Belarbi, Madison Wenzlick, Alison Fritz, NEWTS EPA Ash Dataset, 1/13/2023, https://edx.netl.doe.gov/dataset/newts-epa-ash-dataset, DOI: 10.18141/1909006

Data processing for screening:

- **Data download:** EPA downloaded the “epa-ash-data-all-tabs.xlsx” from the reference link.
- **Data manipulation:** EPA used the worksheet "Original EPA ERG Data" from “epa-ash-data-all-tabs.xlsx”. Data underwent reformatting in R to create simplified tables that contained the analyte names, concentration values, CASRNs, and if there was a detection.
- **Extracted data elements**: EPA wrote R code to extract the maximum concentration and percent site detection in wastewater and wastewater dissolved.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** EPA followed data cleaning procedures developed in preprocessing to extract data to calculate summary statistics.
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

NEWTS Flue Gas Desulfurization (FGD) Effluent Database – Environmental Protection Agency

Data description: The US EPA FGD Database contains FGD (Flue-gas desulfurization) effluent data from selected U.S. coal power plants. This data source was used as a main data source for CCL 6. The data was classified as a non-nationally representative ambient occurrence source and further categorized as wastewater effluent for the classification step of CCL 6.

Reference: Siefert, N., Belarbi, Z., Wenzlick, W. 2021. NEWTS EPA FGD Effluent Database: Environmental Protection Agency. https://edx.netl.doe.gov/dataset/newts-epa-fgd-effluent-database. Accessed October 2023.

Data processing for screening:

- **Data download:** EPA downloaded the epa-fgd-effluent_all-tabs.xlsx file from the reference link.
- **Data manipulation:** EPA used the worksheet "Original EPA ERG Data" for maximum concentration and percent detection calculations. EPA wrote R code to identify detections and standardized units. Data underwent reformatting in R to create simplified tables that contained the analyte names, concentration values, and CASRNs.
- **Extracted data elements**: EPA wrote R code to extract the maximum concentration and percent detections.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** EPA wrote code to filter on detected data for summary statistics calculations using data flags for detected data developed in pre-processing. The Plant.ID column was used as a site ID for calculating site statistics.
- **Extracted data elements**: EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

NEWTS Leachate Dataset – Environmental Protection Agency

Data description: The US EPA Leachate Database contains ash landfill leachate data from selected U.S. coal power plants. This data source was used as a main data source for CCL 6. The data was classified as a non-nationally representative ambient occurrence source and further categorized as wastewater effluent for the classification step of CCL 6.

Reference: Nicholas Siefert, Zineb Belarbi, Madison Wenzlick, Alison Fritz, NEWTS EPA Leachate Dataset, 1/13/2023, https://edx.netl.doe.gov/dataset/newts-epa-leachate-dataset, DOI: 10.18141/1909007

Data processing for screening:

- **Data download:** EPA downloaded the “epa-leachate-all-tabs.xlsx” from the reference link.
- **Data manipulation:** EPA used the worksheet "Original EPA ERG Data" for maximum concentration and percent detection calculations. Data underwent reformatting in R to create simplified tables that contained the analyte names, concentration values, CASRNs, and if there was a detection.
- **Extracted data elements**: EPA wrote R code to extract maximum concentrations across all sites, and percentage detection for each analyte across all sites.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** EPA wrote code to filter on detected data for summary statistics calculations using data flags for detected data developed in pre-processing. The Plant ID column was used as a site ID for calculating site statistics.
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

NEWTS Brackish Water Database – United States Geological Survey (USGS)

Data description: This database contains chemical data and aquifer information for brackish groundwater sources in the USA. Brackish groundwater is defined as having a dissolved-solids concentration between 1,000 and 10,000 mg/L. This data source was used as a main data source for CCL 6. The data was categorized as a non-nationally representative ambient occurrence source.

Reference: Nicholas Siefert, Zineb Belarbi, Madison Wenzlick, NEWTS USGS Brackish Water Database, 9/30/2022, https://edx.netl.doe.gov/dataset/newts-usgs-brackish-water-database, DOI: 10.18141/1890174

Data processing for screening:

- **Data download:** EPA downloaded both the usgs-brackish-water_all-tabs.xlsx and usgs_brackish-water_original-format_condensed.csv from the reference link. The excel file was used to generate a crosswalk to link the abbreviated contaminant names used in the CSV.
- **Data manipulation:** EPA wrote R code to develop a crosswalk for assigning full names and units to abbreviated chemicals, and to label detections and non-detections.
- **Extracted data elements**: EPA wrote R code to extract the maximum concentration and percent detections.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** Code was written to filter on detected data for summary statistics calculations using data flags for detected data developed in pre-processing.
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of samples, number of samples with detections, and percentage of samples with detections for each contaminant.

NEWTS National Produced Waters Geochemical Database v2.3 - United States Geological Survey (USGS)

Data description: The U.S. Geological Survey National Produced Waters Geochemical Database v2.3 is an updated compilation of geochemical and related information for water from oil and gas wells in the United States. It includes identification and location information, well descriptions, dates, rock properties, physical properties of the water, inorganic chemistry, organic chemistry, and isotopes. This data source was used as a main data source for CCL 6. The data was categorized as a non-nationally representative ambient occurrence source.

Reference: Blondes, M.S., Gans, K.D., Engle, M.A., Kharaka, Y.K., Reidy, M.E., Saraswathula, V., Thordsen, J.J., Rowan, E.L., and Morrissey, E.A., 2018, U.S. Geological Survey National Produced Waters Geochemical Database (ver. 2.3, January 2018): U.S. Geological Survey data release, https://doi.org/10.5066/F7J964W8. Accessed October 2023.

Data processing for screening:

- **Data download:** EPA downloaded the “USGSPWDBv2.3 Data Dictionary.csv” and the “USGSPWDBv2.3c.Rmd” from https://www.sciencebase.gov/catalog/item/59d25d63e4b05fe04cc235f9
- **Data manipulation:** EPA wrote R code to combine column names from the data dictionary (USGSPWDBv2.3 Data Dictionary.csv) to column names from the dataset download (USGSPWDBv2.3c.Rmd) and to identify detections and non-detections.
- **Extracted data elements**: EPA wrote R code to extract the maximum concentration and percent detections. Percent detection was found by sample data since there are no clean ‘site’ variables to group by.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** Code was written to filter on detected data for summary statistics calculations using data flags for detected data developed in pre-processing. EPA followed data cleaning procedures developed in preprocessing to extract data to calculate summary statistics.
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of samples, number of samples with detections, and percentage of samples with detections for each contaminant.

"Occurrence of antibiotics, estrogenic hormones, and ultraviolet (UV) filters in water, sediment, and oyster tissue from the Chesapeake Bay." – He et al. 2019

Data description: The study of contaminants of emerging concern in the Chesapeake Bay included 43 antibiotics, 3 estrogenic hormones, and 5 UV-filters found in personal care products. This data source was used as a main data source for CCL 6. The data was categorized as a non-nationally representative ambient occurrence source.

Reference: He, K., Hain, E., Timm, A., Tarnowski, M. and Blaney, L., 2019. Occurrence of antibiotics, estrogenic hormones, and UV-filters in water, sediment, and oyster tissue from the Chesapeake Bay. Science of The Total Environment, 650, pp.3101-3109.

Data processing for screening:

- **Data download:** EPA manually reconstructed Table 2 of the paper by copying and pasting individual cells into a CSV.
- **Data manipulation:** EPA used only the concentration value for calculation, removing the standard deviation and less than signs present in the source table. Contaminants that were below the level of detection at all sampling sites were not incorporated into the data used for CCL 6.
- **Extracted data elements:** EPA wrote R code to extract percent detection and maximum concentrations.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** EPA wrote code to filter on detected data (where concentration values were not NA) for summary statistics calculations. The sites column was used as a site ID for calculating site statistics.
- **Extracted data elements:** EPA wrote R code to extract the minimum detection in addition to total number of sites and number of sites with detections, and percentage of sites with detections for each contaminant.

Pesticide Data Program (PDP) – United States Department of Agriculture (USDA)

Data description: The USDA Pesticide Data Program (PDP) maintains a national pesticide residue database. PDP was initiated in 1991 to collect data on pesticide residues in food with sampling conducted on a statistically defensible representation of pesticide residuals in the U.S. food supply (USDA, 2018). Sampling and testing are conducted on fruits and vegetables, select grains, milk, and (as of 2001) finished water, untreated water, and ground water. The database contains over 31.3 million results.

The PDP drinking water program was initiated at CWSs in New York and California in 2001. Since then, the drinking water sampling program has expanded, though a somewhat changing mix of states is sampled each year. At one time or another, CWSs in 29 states and Washington, D.C., have contributed raw and/or finished water data to the program (USDA, 2018). The CWSs selected for sampling tend to be small- and medium-sized systems (primarily CWSs serving under 50,000), systems served by surface water, and systems located in regions of heavy agriculture. Sampling of untreated water in addition to treated water began in 2004; sampling continued until 2013 (USDA, 2018). Note that temporal trends cannot be evaluated based on these data since, with the exception of 2002 and 2003, samples were not collected from the same sites and states in consecutive years. This data source was used as a main data source for CCL 6. The untreated and groundwater data were categorized as a non-nationally representative ambient occurrence source and the finished water data was categorized as a non-nationally representative finished occurrence source.

Reference: United States Department of Agriculture (USDA). 2018. PDP Drinking Water Project (2001–2013). Available at: https://www.ams.usda.gov/datasets/pdp/pdp-drinking-water-project.

Data processing for screening:

- **Data download:** EPA downloaded the most recent 10 years (2004-2013) of occurrence data on untreated water, finished water, and groundwater on July 21, 2023, from the website address: https://apps.ams.usda.gov/pdp. The summary of findings option was selected for the output report. EPA downloaded the data manually by year and combined the data using a data prep script in R to prepare it for pre-processing.
- **Data manipulation:** Data manipulation was conducted using R. Percent detection rates were calculated using fields for the number of samples analyzed and number of samples with detections. Limit of detection (LOD) was reported as a range. If no detections were reported, ½ the highest LOD was used as the maximum concentration.
- **Extracted data elements**: EPA wrote R code to extract maximum concentrations and percentage of detection data for each pesticide/water type pair.

Data processing for classification:

- **Data download:** Zip folders containing PDP data from 1994 through 2013 were downloaded at this link (https://www.ams.usda.gov/datasets/pdp/pdpdata).
- **Data manipulation:** Data was filtered to include only results from a COMMOD of “WG” (groundwater), “WU” (untreated), or “WR” (finished water). Units were

standardized to ppb. CIS summary statistics and count statistics were grouped by analyte and water type.

- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to the total number of systems and number of systems with detections for each analyte/water type combination.

Pesticide Use Estimates – United States Geological Survey (USGS)

Data description: The USGS publishes estimates of pesticide application rates using projected county crop acres from the Census of Agriculture. The USGS generates high and low estimate application rates. For the low estimates, if there were missing data for a given county, the assumed pesticide use was 0 kg. For the high estimates, missing county data were estimated based on the surrounding county information. This data source was used as a main data source for CCL 6.

Reference: Wieben, C.M., 2021, Estimated annual agricultural pesticide use by major crop or crop group for states of the conterminous United States, 1992-2019 (including preliminary estimates for 2018-19): U.S. Geological Survey data release, https://doi.org/10.5066/P900FZ6Y. Accessed September 2023.

Data processing for screening:

- **Data download:** EPA downloaded the "High Estimate Agricultural Pesticide Use by Crop Group 1992-2019" dataset. The dataset was converted to a CSV file.
- **Data manipulation:** EPA calculated the total application rates for each pesticide for each year that data were available using R.
- **Extracted data elements**: EPA wrote R code to extract the total application rate for the most recent year for each pesticide.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** Data was summarized for the last reported year of contaminant data.
- **Extracted data elements:** EPA wrote R code to output a count of the number of states that each pesticide was used in for the more recent reported year of contaminant data and the associated year for each pesticide.

"Pharmaceuticals, hormones, pesticides, and other bioactive contaminants in water, sediment, and tissue from Rocky Mountain National Park" – Battaglin et al. 2018

Data description: The study examined the occurrence of pharmaceuticals, hormones, pesticides, and other bioactive contaminants at 20 sites in Rocky Mountain National Park (RMNP), Colorado, in 2012–2013. Samples were analyzed for 369 parameters including 149 pharmaceuticals, 22 hormones, 137 pesticides, and 61 other chemicals to provide a representative assessment of bioactive contaminant occurrence within RMNP. This data source was used as a main source in CCL 6. The data was categorized as a non-nationally representative ambient occurrence source.

Reference: Battaglin, W.A., Bradley, P.M., Iwanowicz, L., Journey, C.A., Walsh, H.L. and Blazer, V.S., 2018. Pharmaceuticals, hormones, pesticides, and other bioactive contaminants in water, sediment, and tissue from rocky mountain national park, 2012–2013. Science of the Total Environment, 643, pp.651-673.

Data processing for screening:

- **Data download:** EPA downloaded the publication.
- **Data manipulation:** EPA extracted data from Table 2 of the paper.
- **Extracted data elements**: EPA wrote R code to extract maximum concentrations and percent detections.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** There was insufficient information to calculate the minimum, median, and 90th percentile values for detected contaminant data.
- **Extracted data elements:** EPA wrote R code to extract total number of samples, number of samples with detections, and percentage of samples with detections for each contaminant.

"Pharmaceutical manufacturing facility discharges can substantially increase the pharmaceutical load to U.S. wastewaters" – Scott et al. 2018

Data description: This is a USGS publication measuring effluent from 20 wastewater treatment plants (WWTPs) around the U.S. that do and do not receive wastewater from pharmaceutical manufacturing facilities. In these samples, concentrations of 120 pharmaceutical and pharmaceutical degradate products were measured. The EPA relied on the data processed and extracted during CCL 5 from this data source to support screening and classification for CCL 6. This information is described below. This data source was used as a main data source for CCL 6. The data was classified as a non-nationally representative ambient occurrence source and further categorized as wastewater effluent for the classification step of CCL 6.

Reference: Scott, T.M., P.J. Phillips, D.W. Kolpin, K.M. Colella, E.T. Furlong, W.T. Foreman, and J.L. Gray. 2018. Pharmaceutical manufacturing facility discharges can substantially increase the pharmaceutical load to US wastewaters. Science of the Total Environment. 636:69-79. https://doi.org/10.1016/j.scitotenv.2018.04.160.

Data processing for screening:

- **Data download:** EPA downloaded the publication and supplemental data file. Tables S3 and S4 in the supplemental data file were exported to CSV files and used to easily access percent detection rate information.
- **Data manipulation:** Data manipulation was minimal and restricted to adding DTXSIDs.
- **Extracted data elements**: EPA wrote R code to extract percent detection information.

Data processing for classification:

- **Data download:** The same publication and supplemental data file was used for extracting data elements for the classification step.
- **Data manipulation:** Tables S5, S6, S7, and S8 were used to calculate summary concentration statistics and detection rate information. DTXSIDs were added. Data manipulation was conducted using R.
- **Extracted data elements**: EPA wrote R code to extract minimum, median, 90th percentile, and maximum concentration of detections, total sites with samples, number of sites with detections, and percentage of sites with detections.

"Pharmaceuticals, perfluorosurfactants, and other organic wastewater compounds in public drinking water wells in a shallow sand and gravel aquifer." – Schaider et al. 2014

Data description: The study includes organic wastewater compounds tested in 20 public drinking water supply wells on Cape Cod, Massachusetts. This data source was a main data source for CCL 6. The data was categorized as a non-nationally representative ambient occurrence source.

Reference: Schaider, L.A., Rudel, R.A., Ackerman, J.M., Dunagan, S.C. and Brody, J.G., 2014. Pharmaceuticals, perfluorosurfactants, and other organic wastewater compounds in public drinking water wells in a shallow sand and gravel aquifer. Science of the Total Environment, 468, pp.384-393. https://doi.org/10.1016/j.scitotenv.2013.08.067

Data processing for screening:

- **Data download:** EPA exported each table from the paper and the supplementary files to an excel file, and then transferred tables 1 and S2 to manually to a CSV. EPA also manually corrected errors from the PDF export process, including deleting extra headers and chemicals duplicated between the tables.
- **Data manipulation:** For chemicals with no maximum concentration EPA assigned 1/2 of the Method reporting limit.
- **Extracted data elements:** EPA wrote R code to extract the maximum concentration from groundwater, and the ambient detection rate.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** The total site count was back calculated from the percent detections provided by the source. It was assumed that contaminants without percent detections reported were sampled but had no detections for the 20 wells in the study.
- **Extracted data elements:** There was insufficient information to calculate minimum, median, and 90th percentile of detections. EPA wrote R code to extract the number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

"Predicting variability of aquatic concentrations of human pharmaceuticals" – Kostich et al. 2010

Data description: This is an EPA Office of Research and Development (ORD) study that derives predicted environmental concentrations of active pharmaceutical ingredients (APIs) and compares those predicted concentrations to measured environmental concentrations (MECs) published in the peer-reviewed literature. Peer-reviewed publications that report MECs for any API were identified via literature search. The search included studies that were conducted in the U.S., published between January 2001 and January 2009, and that reported mass spectrometry data. The EPA relied on the data processed and extracted during CCL 5 from this data source to support screening and classification for CCL 6. This information is described below. This data source was used as a main data source for CCL 6.

Reference: Kostich, M.S., A.L. Batt, S.T. Glassmeyer, and J.M. Lazorchak. 2010. Predicting variability of aquatic concentrations of human pharmaceuticals. Science of The Total Environment. 408(20):4504–4510. https://doi.org/10.1016/j.scitotenv.2010.06.015.

Data processing for screening:

- **Data download:** EPA downloaded the publication and supplemental data file. Appendix 2 in the supplemental data file contains maximum measured environmental concentrations (MECs) used in the screening step.
- **Data manipulation:** Data from studies measuring effluents from hospitals and drinking water treatment plants were excluded. DTXSIDs were added. Data manipulation steps were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract MECs. MECs were classified as maximum ambient concentrations in the screening step.

Data processing for classification:

- **Data download:** The supplemental data file downloaded in the pre-processing steps for screening above was used to extract data for the classification step.
- **Data manipulation:** This data source is a literature review and contains some data from other main data sources and data sources identified during the occurrence literature review process of the classification step. Duplicate data were removed. DTXSIDs were added. Data manipulation steps were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract MECs that were classified as maximum concentrations in either ambient or wastewater effluent, where appropriate. The original study references and MECs as reported in Kostich et al. 2010 were extracted and included on the contaminant information sheets (CISs).

"Reconnaissance of mixed organic and inorganic chemicals in private and public supply tapwaters at selected residential and workplace sites in the United States" – Bradley et al. 2018

Data description: This article was published by the United States Geological Survey (USGS), the National Institute of Health (NIH), and the EPA's Office of Research and Development (ORD). The authors sampled tap water from 13 homes and 12 workplaces across 11 states. The samples were analyzed for 482 organic compounds and 19 inorganic compounds. The EPA relied on the data processed and extracted during CCL 5 from this data source to support screening and classification for CCL 6. This information is described below. This data source was used as a main data source for CCL 6. The data was categorized as a non-nationally representative finished water occurrence source.

Reference: Bradley, P.M., D.W. Kolpin, K.M. Romanok, K.L. Smalling, M.J. Focazio, J.B. Brown, M.C. Cardon, K.D. Carpenter, S.R. Corsi, L.A. DeCicco, J.E. Dietze, N. Evans, E.T. Furlong, C.E. Givens, J.L. Gray, D.W. Griffin, C.P. Higgins, M.L. Hladik, L.R. Iwanowicz, C.A. Journey, K.M. Kuivila, J.R. Masoner, C.A. McDonough, M.T. Meyer, J.L. Orlando, M.J. Strynar, C.P. Weis, and V.W. Wilson. 2018. Reconnaissance of mixed organic and inorganic chemicals in private and public supply tapwaters at selected residential and workplace sites in the United States. Environmental Science & Technology. 52, 23:13972–13985. https://doi.org/10.1021/acs.est.8b04622.

Data processing for screening:

- **Data download:** EPA downloaded the publication and supplemental data files.
- **Data manipulation:** Maximum concentration data and percentage of detections were extracted from Tables S2 and S3 in the supplemental data files. This data source did not require additional calculations. The tables were reformatted from wide format into a long format and DTXSIDs were added.
- **Extracted data elements**: EPA wrote R code to extract maximum concentration of detections and percentage detections from Tables S2 and S3 in the supplemental data files.

Data processing for classification:

- **Data download:** The supplemental data files downloaded for screening were used for extracting data used in classification.
- **Data manipulation:** Summary statistics of concentration data and detection information were calculated from Tables S2 and S3 of the supplemental data files. DTXSIDs were added. Data manipulations were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract minimum, median, 90th percentile, and maximum concentration of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

"Simultaneous Determination of Selected Trace Contaminants in Drinking Water Using Solid-Phase Extraction-High Performance Liquid Chromatography-Tandem Mass Spectrometry" – Zhang et al. 2019

Data description: Occurrence of selected contaminants in Missouri State drinking water system during hot and cold seasons. This data source was used as a main data source for CCL 6. The data was categorized as a non-nationally representative ambient or finished occurrence source.

Reference: Zhang, H., West, D., Shi, H. et al. Simultaneous Determination of Selected Trace Contaminants in Drinking Water Using Solid-Phase Extraction-High Performance Liquid Chromatography-Tandem Mass Spectrometry. Water Air Soil Pollut 230, 28 (2019). https://doi.org/10.1007/s11270-018-4066-9

Data processing for screening:

- **Data download:** EPA retrieved the supplementary material from the Zhang et al. (2019) paper. EPA then copied and pasted the information from Tables S1 and S2 to a CSV, for reading into R. MDL values were retrieved from Table 4 in the Zhang et al. (2019) paper to replace non-detections with 1/2 the MDL. Table 4 used MDLs and table 1 for the contaminant names to make a contaminant crosswalk file for MDL to contaminant name.
- **Data manipulation:** EPA wrote R code to calculate the max concentration and percent detection in both ambient and finished water by contaminant. Non-detections were replaced by 1/2 the MDL for that contaminant. A simple summary was output that combined the hot and cold seasons. Results were separately calculated for water types of Source (ambient), and Drinking (finished).
- **Extracted data elements:** EPA wrote R code to extract the maximum concentration and percent detection rate in both ambient and finished water by contaminant.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** Code was written to filter on detected data for summary statistics calculations using data flags for detected data developed in pre-processing. Similar data cleaning procedures from prescreening were applied to process the CIS summary statistics. NAs were used to replace data where there is no minimum or calculable value.
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant. Data was summarized and categorized as ambient or finished

State Drinking Water Monitoring Datasets and EPA's Third Six-Year Review (SYR 3) – Environmental Protection Agency

Data description: There is no available national database that receives and stores all relevant data regarding the occurrence of regulated contaminants in public drinking water systems (PWS). Therefore, EPA conducts voluntary data requests from the states, territories, and tribes in support of national occurrence assessments as part of the Six-Year Review. For EPA's SYR 3 of drinking water regulations, some states submitted PWS occurrence data for unregulated contaminants along with the requested data on regulated contaminants. The EPA relied on the data processed and extracted during CCL 5 from these data sources to support screening and classification for CCL 6. This information is described below.

For SYR 3, the dataset of unregulated contaminant monitoring data included results from 14 states/entities. These unregulated data provide varying degrees of completeness in their coverage of the states/entities and are not necessarily representative of occurrence in those states/entities. For more details on the SYR 3 ICR dataset, refer to the EPA's SYR 3 occurrence analysis (USEPA, 2016a). For SYR 3, EPA requested (through an ICR) that primacy agencies voluntarily submit drinking water compliance occurrence data to EPA that were collected during 2006-2011. Six states (Massachusetts, Maine, Michigan, Pennsylvania, Tennessee, and Washington) plus Washington, D.C., American Samoa, Region 1 and 9 tribes, and Navajo Nation also submitted PWS occurrence data for unregulated contaminants in addition to the data for regulated contaminants. EPA was able to supplement these data on unregulated contaminants by downloading additional publicly available monitoring data from state websites (California, Florida, Massachusetts, and Wisconsin). The result was a collection of unregulated contaminant monitoring data from 14 states/entities; in this description of SYR 3 ICR and state drinking water monitoring datasets used in CCL 5, the term state is used for SDWA primacy entities. The 14 datasets vary in the range of monitoring dates (in some cases extending beyond the 2006-2011 period of interest for Six-Year Review), the number of contaminants monitored, the number of systems reporting monitoring, and the number of samples taken. The datasets vary widely in the number of PWSs sampled in each state relative to the total number of PWSs in that state. Hence, these data are used only to augment and complement any national drinking water data and to assess any unique occurrence that may suggest a need for further review. Both the SYR 3 and additional state data processed for CCL 5 were used in CCL 6.

For CCL 5, EPA extracted source and finished water data on PCCL 5 chemicals from the SYR 3 ICR Access database and occurrence monitoring data obtained through state websites (California, Florida, Massachusetts, and Wisconsin). Of the 14 datasets, eight datasets provided source or finished water data on PCCL 5 chemicals. The list of eight datasets used for CCL 5 include California, Washington, D.C., Florida, Massachusetts, Maine, Pennsylvania, Washington, and Wisconsin. These datasets were used as supplemental data sources for CCL 6 and included on the Contaminant Information Sheets (CISs). Detailed information on data downloads, data manipulation, and data element extraction for the California, Florida, Massachusetts, and Wisconsin datasets are described below. Data manipulation and data management for the SYR 3 ICR data can be found in USEPA (2016b).

References:

USEPA. 2016a. Analysis of Occurrence Data from the Third Six-Year Review of Existing National Primary Drinking Water Regulations: Chemical Phase Rules and Radionuclides Rules. EPA-810-R-16-014. December 2016.

USEPA. 2016b. The Data Management and Quality Assurance/Quality Control Process for the Third Six-Year Review Information Collection Rule Dataset. EPA-810-R-16-015. December 2016.

California Water Boards. n.d. Water Quality Analyses Database Files. California Division of Drinking Water. URL: https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/EDTlibrary.html. Accessed January 2020.

Commonwealth of Massachusetts Executive Office of Energy and Environmental Affairs. n.d. Energy and Environmental Affairs Data Portal. Massachusetts Office of Energy and Environmental Affairs (EEA). URL: https://eeaonline.eea.state.ma.us/portal#!/search/drinkingwater. Accessed January 2020.

Florida Department of Environmental Protection. n.d. Drinking Water Data Base. Florida Division of Water Resource Management. Source and Drinking Water Program. URL: https://floridadep.gov/water/source-drinking-water/content/information-drinking-water-database. Accessed January 2020.

Wisconsin Department of Natural Resources. n.d. Public Drinking Water System Data. Wisconsin Department of Natural Resources Drinking Water. URL: https://dnr.wisconsin.gov/topic/DrinkingWater/QualityData.html. Accessed January 2020.

California Drinking Water Monitoring Dataset:

- **Data download**: EPA downloaded unregulated contaminant monitoring data from the California State Water Resources Control Board, Division of Drinking Water, Water Quality Analyses database website. Drinking water analyses are reported directly into the database from laboratories. Data were downloaded manually as .dbf files then imported into Microsoft Access. Data were downloaded for 2006 through 2019. Supporting database files, including information on drinking water sources, systems, laboratories, and chemicals, were also downloaded.
- **Data manipulation and extracted data elements**: EPA extracted the relevant data elements for data analyses. EPA standardized the monitoring data to enable combining the monitoring data with data from other states. For example, in the source water type field, all instances of surface water or S were changed to SW. EPA determined how to identify analytical detections and non-detections. Contaminant monitoring data were restructured into a uniform structure to enable combining with monitoring data from other states. California inventory data (analyte name, PWSID, state, source type) and sample analytical result data (date, concentration, unit of measure, detect, detection limit value, detection limit unit) were mapped separately then combined into one file for analyses. EPA added DTXSIDs to each unique analyte. EPA performed a cursory review

for outliers or erroneous data. Records (approximately 2% of all records) were excluded from the analysis for the following reasons:

  - FINDING < 0
  - QMOD was equal to “Q” Or “I” Or “F” Or “0” Or “-“ (XMOD is the field to determine if a record is a detection or non-detection)
  - IF the water system status was equal to "MW" Or "AG" Or "DS" Or "AB" Or "WW" (i.e., did not represent a drinking water source)
  - EPA extracted minimum, median, 90th percentile and maximum concentration of detections as well as total number of systems, number of systems with detections, and percent of systems with detections for each PCCL 5 chemical, which were pulled in for CCL 6.

Florida Drinking Water Monitoring Dataset:

- **Data download:** EPA downloaded historical contaminant monitoring data from Florida’s Source and Drinking Water Program Chemical Data website by year for 2006 through 2018 (note monitoring data for PCCL 5 chemicals were available only for 2006-2011). Data were downloaded manually as Microsoft Excel files (.xlsx).
- **Data manipulation and extracted data elements:** EPA combined annual monitoring data into one file. EPA extracted the relevant data elements for data analyses. Minimal data manipulation was needed as the Florida data were organized in a simple, flat file. EPA standardized the monitoring data to enable combining the monitoring data with data from other states. For example, in the water type field, all instances of community water system or C were changed to CWS. EPA designated all data with RESULTS = 0 as non-detections and all data with RESULTS greater than 0 as detections. Contaminant monitoring data were restructured into a uniform structure to enable combining with monitoring data from other states. EPA added DTXSIDs to each unique analyte. EPA performed a cursory review for outliers or erroneous data. No analytical records were identified to exclude from the summary statistical analyses. EPA extracted minimum, median, 90th percentile and maximum concentration of detections as well as total number of systems, number of systems with detections, and percentage of systems with detections for each PCCL 5 chemical.

Massachusetts Drinking Water Monitoring Dataset:

- **Data download:** EPA downloaded unregulated contaminant monitoring data from the Massachusetts Office of Energy & Environmental Affairs Data Portal. Data were downloaded manually as a single Excel (xlsx) file for 2006 through 2020.
- **Data manipulation and extracted data elements:** Minimal data manipulation was needed as the monitoring data were organized in a simple, flat file. EPA extracted the relevant data elements for data analyses. EPA standardized the monitoring data to enable combining the monitoring data with data from other states. For example, in the source water type field, all instances of surface water or S were changed to SW. EPA determined how to identify analytical detections and non-detections. Contaminant monitoring data were restructured into a uniform structure to enable combining with monitoring data from other states. EPA added DTXSIDs to each unique analyte record. EPA performed a cursory review for outliers or erroneous data. No analytical records were identified to

exclude from the summary statistical analyses. EPA extracted minimum, median, 90th percentile and maximum concentration of detections as well as total number of systems, number of systems with detections, and percentage of systems with detections for each PCCL 5 chemical.

Wisconsin Drinking Water Monitoring Dataset:

- **Data download:** EPA downloaded unregulated contaminant monitoring data from the Public Drinking Water System database from the Wisconsin Department of Natural Resources. Contaminant monitoring data were searched, using the Find Contaminants in Public Water Supplies search function, and downloaded in batches by analyte for January 2006 through January 2020. Data were downloaded manually as a CSV file.
- **Data manipulation and extracted data elements:** Annual data files were combined into a single file. Minimal data manipulation was needed as the monitoring data were organized in a simple, flat file. EPA extracted the relevant data elements for data analyses. EPA standardized the monitoring data to enable combining the monitoring data with data from other states. For example, in the source water type field, all instances of surface water or S were changed to SW. EPA determined how to identify analytical detections and non-detections. Contaminant monitoring data were restructured into a uniform structure to enable combining with monitoring data from other states. EPA added DTXSIDs to each unique analyte. EPA performed a cursory review for outliers or erroneous data. Records (fewer than 1% of all records) were excluded from the analysis if Qualifier Code = “Unexplained” or Units were listed as something other than mg/L or ug/L. EPA extracted minimum, median, 90th percentile and maximum concentration of detections as well as total number of systems, number of systems with detections, and percent of systems with detections for each PCCL 5 chemical.

Surface Water Database (SURF) – California Department of Pesticide Regulation (CDPR)

Data description: CDPR Surface Water (SURF) Database was developed in 1997 to make information concerning the presence of pesticides in California surface waters available to the public. The database includes pesticide monitoring results from rivers, creeks, agricultural drains, urban streams, and estuaries in California. The database houses monitoring results collected by federal, state, and local agencies, private industry, and environmental groups. This data source contains monitoring information for 334 pesticides and pesticide metabolites. (Description adapted from DPR SURF website.) This data source was used as a main data source for CCL 6. The data was categorized as a non-nationally representative ambient occurrence source.

Reference: California Department of Pesticide Regulation (CDPR). n.d. Surface Water Database (SURF). https://www.cdpr.ca.gov/docs/emon/surfwtr/surfdata.htm. Accessed July 2023.

Data processing for screening:

- **Data download:** EPA downloaded the complete SURF database as a CSV file from the web page.
- **Data manipulation:** There are many samples in the SURF database collected by the United States Geological Survey (USGS). To alleviate concern for double-counting data from USGS's National Water Information System (NWIS) database and the SURF database, data in the SURF database that had been taken from USGS were removed.
- **Extracted data elements**: EPA wrote R code to extract the maximum concentration of detections and percentage of detections by sites.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** The data flag for detections developed in prescreening was used to filter on detected data for summary statistics calculations for classification. The Site_code column was used as a site ID for calculating site statistics.
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

"Suspect screening and non-targeted analysis of drinking water using point-of-use filters" – Newton et al. 2018

Data description: This EPA Office of Research and Development (ORD) publication discusses the results of a pilot study conducting non-targeted analysis of extracts from nine point-of-use drinking water filters in North Carolina. High resolution mass spectra of the filter extracts were matched to a library of chemical formulas, and 15 of the potential matches were confirmed with analytical standards. For unconfirmed compound matches, there is significant uncertainty in if the compound is truly present in the sample. This non-targeted approach is not designed to quantify concentrations of compounds but only to indicate if they are present in the sample. EPA considered Newton et al. (2018) as a case study of how a non-targeted analysis could be useful in drinking water contaminant prioritization. The EPA relied on the data processed and extracted from this data source during CCL 5 for CCL 6. This information is described below. This data source was considered as a main data source for CCL 6 as it met the four assessment factors and contaminants could have been added to the Pre-Universe as a result. However, detection frequencies were not included in the screening or classification steps because this study was not targeted, and the sample size was limited.

Reference: Newton, S.R., R.L. McMahen, J.R. Sobus, K. Mansouri, A.J. Williams, A.D. McEachran and M.J. Strynar. 2018. Suspect screening and non-targeted analysis of drinking water using point-of-use filters. Environmental Pollution. 234: 297-306. https://doi.org/10.1016/j.envpol.2017.11.033.

Data processing for screening:

- **Data download:** EPA downloaded the publication and supplemental data file.
- **Data manipulation:** No data manipulation was necessary.
- **Extracted data elements**: EPA wrote R code to extract "total detection frequency" data from the tab "candidate compounds" sheet in the supplemental data file.

Data processing for classification:

- Data was not separately processed as a standalone- simple file for classification.

Sustaining The Earth's Watersheds - Agricultural Research Database System (STEWARDS) – United States Department of Agriculture (USDA) – Agricultural Research Service

Data description: This data represents water quality monitoring data collected by water resource management groups across the country. It includes details on collecting methods, type of contaminant found, and samples collected. This data source was used as a main data source for CCL 6. The data was categorized as a non-nationally representative ambient occurrence source.

Reference: STEWARDS/Agricultural Research Service (ARS). https://www.ars.usda.gov/anrds/ceap/ceap-home/. Accessed via the Water Quality Portal (WQP). URL: https://www.waterqualitydata.us/portal/. Accessed July 2023.

Data processing for screening:

- **Data download:** EPA downloaded the data from the STEWARDS/ARS source from the Water Quality Portal.
- **Data manipulation:** EPA wrote R code to remove unneeded columns and rename columns to match proper output format for maximum concentrations and percent detections.
- **Extracted data elements**: EPA wrote R code to extract the maximum concentrations and percent detections.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** Code was written to filter on detected data for summary statistics calculations. NAs were used to replace data where there is no minimum or calculable value.
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

Toxics Release Inventory (TRI) – Environmental Protection Agency

Data description: The Toxics Release Inventory (TRI) Explorer allows users to generate reports on releases, transfers, and waste managed that can be displayed by facility, chemical, geographic area, industry (NAICS code), or reporting years. The objective of TRI Explorer is to help communities identify facilities and chemical disposal or other release patterns that warrant further study and analysis. The TRI Explorer will generate Release Reports - including on- and off-site releases (i.e., off-site releases include transfers off-site to disposal and metals and metal compounds transferred to Publicly Owned Treatment Works (POTWs). This data source was used as a main data source for CCL 6.

Reference: USEPA. March 2023. Toxics Release Inventory (TRI) Program. https://www.epa.gov/toxics-release-inventory-tri-program. Accessed July 2023.

Data processing for screening:

- **Data download:** EPA generated a release chemical report, by selecting data from year 2021, "All of United States" for the Geographic Location field, "All chemicals" from the chemical field, "All Industries" from the industry field. EPA selected the 2021 National Analysis Dataset, released March 2023. EPA selected "Total On-site Disposal or Other Releases", "Total Off-site Disposal or Other Releases", "Total On-and Off-site Disposal or Other Releases", and "CAS Number" for report columns to include.
- **Data manipulation:** EPA wrote R code to clean data for processing. EPA produced a simple dataset where the total disposal (in lbs) for each chemical is reported using the "Total On-and Off-site Disposal or Other Releases" column.
- **Extracted data elements**: EPA wrote R code to extract the total pounds released in 2021 for each compound.

Data processing for classification:

- **Data download:** EPA downloaded the TRI Release Geography Reports associated with the 2021 release data used in the screening step in July 2023, from https://enviro.epa.gov/triexplorer/tri_release.geography. EPA selected the data option for total on- and off-site disposal and other releases.
- **Data manipulation:** EPA used the downloaded state release reports to manually count the number of states/territories from which a compound was reported released. If the reported release amount was 0 for total on- and off-site disposal or other releases for a given state or entity, the state was not counted.
- **Extracted data elements**: EPA extracted the total number of states from which a compound was released for the year 2021.

Unregulated Contaminant Monitoring Rule (UCMR) Cycles 1-4 – Environmental Protection Agency

Data description: These data represent all the Unregulated Contaminant Monitoring Rule (UCMR) sampling results from completed UCMR cycles. UCMR is a nationally representative survey of drinking water systems designed to provide a basis for future drinking water regulatory actions. UCMR 1 included monitoring for 26 contaminants between 2001 and 2003. UCMR 2 includes monitoring for 25 contaminants between 2008 and 2010. UCMR 3 included monitoring for 28 chemical contaminants and 11 microbes between 2013 and 2015. UCMR 4 included monitoring for 30 contaminants between 2018 and 2020. The EPA relied on the data processed and extracted during CCL 5 from this data source to support screening and classification for CCL 6. This information is described below. This data source was used as a main data source for CCL 6. The data was categorized as a nationally representative finished water occurrence source.

References:

USEPA. 1999. Revisions to the Unregulated Contaminant Monitoring Regulation for Public Water Systems; Final Rule. *Federal Register* 64(80): 50556.

USEPA. 2007. Unregulated Contaminant Monitoring Regulation (UCMR) for Public Water Systems Revisions. *Federal Register* 72(2): 367.

USEPA. 2012. Revisions to the Unregulated Contaminant Monitoring Regulation (UCMR 3) for Public Water Systems. *Federal Register* 77(85): 26071.

USEPA. 2016. Revisions to the Unregulated Contaminant Monitoring Rule (UCMR 4) for Public Water Systems and Announcement of Public Meeting; Final Rule. *Federal Register.* 81(244): 92666.

USEPA. 2022. The Fourth Unregulated Contaminant Monitoring Rule (UCMR 4): Data Summary, January 2022. Office of Water. EPA 815-S-22-001.

Data processing for screening:

- **Data download:** The results of UCMR 1-4 were downloaded from the following EPA website: https://www.epa.gov/dwucmr/occurrence-data-unregulated-contaminant-monitoring-rule.
- **Data manipulation:** If there were zero detections for a contaminant, half of the MRL was substituted for the maximum concentration. DTXSIDs were added. Data manipulations were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract maximum concentrations and percent of sites with detections in public water systems.

Data processing for classification:

- **Data download:** The same file used in the screening step was used to extract data for the classification step.
- **Data manipulation:** Data manipulation was minimal and limited to adding DTXSIDs. Concentration summary statistics were based on analytical detections only and maximum

concentrations for non-detected contaminants were not substituted for the classification step.

- **Extracted data elements**: EPA wrote R code to extract the minimum, median, 90th percentile, and maximum concentration of detections in public water systems in addition to method reporting levels (MRL), total number of sites, number of sites with detections, and percentage of sites with detections for each contaminant.

Unregulated Contaminant Monitoring Rule (UCMR), Cycle 5 – Environmental Protection Agency

Data Description: Similar in design to UCMR 4, UCMR 5 required all small and large PWSs (serving between 3,300 and 10,000 people and serving more than 10,001 people, respectively), plus a statistically representative national sample of 800 small PWSs (serving fewer than 3,300 people), were required to conduct assessment monitoring during a 12-month period between 2023 and 2025. UCMR 5 includes monitoring for 29 PFAS and lithium. Because the PFAS are included as a group on CCL 6, only lithium data was pulled from UCMR 5 for presentation on the Contaminant Information Sheet. These data are treated separately from the UCMR1-4 data because the monitoring dataset for UCMR 5 was not complete at the time of CCL 6 development. The UCMR 5 dataset used in CCL 6 are not final and are subject to change as updates become available. This data source was used as a main data source for CCL 6. No UCMR 5 data had been published as of July 2023, so the data source was not used for screening. The data was categorized as a nationally representative finished occurrence source.

Reference: USEPA. 2021. Revisions to the Unregulated Contaminant Monitoring Rule (UCMR 5) for Public Water Systems and Announcement of Public Meetings. Federal Register. Vol. 86, No. 245. p. 73131,

Data processing for screening:

- Data was not separately processed as a standalone simple file for screening.

Data processing for classification:

- **Data Download:** Data were downloaded from the Fifth Unregulated Contaminant Monitoring Rule Data Finder https://www.epa.gov/dwucmr/fifth-unregulated-contaminant-monitoring-rule-data-finder. Data was accessed July 10, 2025.
- **Data Manipulation:** Concentration summary statistics were based on analytical detections only.
- **Extracted Data Elements:** EPA extracted the minimum, median, 90th percentile and maximum concentrations of detections in public water systems, the total number of PWSs, and the number of systems with detections.

Unregulated Contaminant Monitoring – State (UCM-State) Rounds 1 and 2 – Environmental Protection Agency

Data description: The Unregulated Contaminant Monitoring (UCM) program was a drinking water monitoring effort that was a precursor to the Unregulated Contaminant Monitoring Rule (UCMR) program established in the 1996 amendments to the Safe Drinking Water Act (SDWA). Round 1 UCM data are from approximately 1988 to 1992 and were extracted from the Unregulated Contaminant Monitoring Information System (URCIS). The UCM-State Round 2 data are from 1993 to 1997 and were extracted from SDWIS.

UCM-State Round 1 monitoring initially involved 34 required volatile organic compounds (VOCs), 14 VOCs to be monitored at states' discretion, and two synthetic organic compounds (SOCs). Monitoring for unregulated compounds was to be conducted alongside monitoring for regulated compounds (USEPA, 1987). The final database for this round of monitoring included 62 regulated and unregulated contaminants (USEPA, 2001).

UCM-State Round 2 involved monitoring for 20 VOCs from the Round 1 required list and 14 VOCs from the Round 1 discretionary list, plus 13 SOCs and sulfate. The final database for this round of monitoring included 48 unregulated contaminants (USEPA, 2001).

There was no requirement that the monitoring data be reported to EPA and individual states maintained the data in different forms and formats. In the context of various initiatives and information collection requests (ICRs), many states voluntarily submitted the UCM data to EPA. EPA worked to assemble the state data into a composite dataset that would support national occurrence estimates. The UCM-State Round 1 database contains contaminant occurrence data from 38 states, Washington, D.C., and the U.S. Virgin Islands. The UCM-State Round 2 database contains data from 35 states and several tribes.

Processed versions of the data, called cross-sections, include the most complete and sound-quality state datasets and were constructed so that the data could be used to generate nationally representative summary statistics on contaminant occurrence. To develop the cross-sections, all states with monitoring data were first evaluated by their distribution across a range of pollution potential indicators and spatial/hydrogeologic diversity. A select group of states, representing a balanced distribution across these pollution potential measures and across the nation geographically, were then used to construct national cross-sections (one from Round 1 data and another from Round 2 data) that would provide reasonable representation of national occurrence. For more information on the construction of the UCM-State Round 1 and Round 2 cross-sections, see USEPA (2002). This data source was used as a main data source for CCL 6. The UCM-State Rounds 1 and 2 data were not used during screening, but were included on the Contaminant Information Sheets during classification. The data was categorized as a nationally representative finished occurrence source. The EPA relied on the data processed and extracted during CCL 5 from this data source to support screening and classification for CCL 6. This information is described below.

EPA considered finished drinking water maximum concentrations from all main data sources for calculating the screening hazard quotient in the screening step of CCL 6 (see section 3.2.2, of the main document) except UCM Program. Concerns about the age of the UCM data (data collection ranged from 1988-1997), high reporting limits, and the quality of the results contributed to EPA's decision to not consider this data source when calculating sHQs for CCL 6.

References:

USEPA. 2001. Occurrence of Unregulated Contaminants in Public Water Systems: An Initial Assessment. EPA 815-P-00-001. May 2001.

USEPA. 2002. Analysis of National Occurrence of the 1998 Contaminant Candidate List (CCL) Regulatory Determination Priority Contaminants in Public Water Systems. EPA 815-D-01-002. May 2002.

Data processing for screening:

- **Data download:** A Microsoft Access database containing the UCM data was downloaded from https://www.epa.gov/dwucmr/occurrence-data-unregulated-contaminant-monitoring-rule#12 on February 23, 2018. The cross-section files for UCM 1 and UCM 2 were used to extract data elements for screening step.
- **Data manipulation:** Data manipulation was minimal and limited to adding DTXSIDs.
- **Extracted data elements**: EPA wrote R code to extract the maximum concentrations of detections

Data processing for classification:

- **Data download:** The same data file used in screening was used for extracting data for the classification step.
- **Data manipulation:** Data manipulation was minimal and limited to adding DTXSIDs. **Extracted data elements**: EPA wrote R code to extract minimum, median, 90th percentile, and maximum concentration of detections in addition to total number of systems, number of systems with detections, percentage of systems with detections for each compound, total number of samples, number of samples with detections, and percentage of samples with detections for each contaminant.

"Urban Stormwater: An Overlooked Pathway of Extensive Mixed Contaminants to Surface and Groundwaters in the United States" – Masoner et al. 2019

Data description: Masoner et al., 2019 has data collected from 50 stormwater runoff events across 17 states between August 2016-December 2017. The study analyzed 438 organic chemicals (including pesticides, polyaromatic hydrocarbons, pharmaceuticals, and industrial and household chemicals) and 62 inorganic chemicals. This data source was used as a main data source for CCL 6. The data was categorized as a non-nationally representative ambient occurrence source.

Reference: Masoner, J. R., Kolpin, D. W., Cozzarelli, I. M., Barber, L. B., Burden, D. S., Foreman, W. T., Forshay, K. J., Furlong, E. T., Groves, J. F., Hladik, M. L., Hopton, M. E., Jaeschke, J. B., Keefe, S. H., Krabbenhoft, D. P., Lowrance, R., Romanok, K. M., Rus, D. L., Selbig, W. R., Williams, B. H., & Bradley, P. M. 2019. Urban Stormwater: An Overlooked Pathway of Extensive Mixed Contaminants to Surface and Groundwaters in the United States. Environmental science & technology, 53(17), 10070–10081. https://doi.org/10.1021/acs.est.9b02867

Data processing for screening:

- **Data download:** EPA downloaded the supporting information excel file from the reference link.
- **Data manipulation:** EPA wrote R code to combine Tables 9, 10, 18, 19 from Masoner 2019 Supplemental Information into one table and identify units. For contaminants with no detections, half the reporting level was substituted for the maximum concentration.
- **Extracted data elements**: EPA wrote R code to extract the maximum measured concentrations.

Data processing for classification:

- **Data download:** The input files used for pre-processing were used to extract data for classification.
- **Data manipulation:** Code was written to identify non-detections and non-samples within the data, including for entries that contained symbols or characters. Entries listed as "nd" or that contained the symbol "<" were treated as non-detections. Entries listed as either "not analyzed" or "not determined" were considered non-samples, as were entries that contained the character "V". Entries containing "E" followed by a number, were treated as just the number that followed "E" (for example E54.6 would be considered 54.6).
- **Extracted data elements:** EPA wrote R code to extract the minimum, median, and 90th percentile of detections in addition to total number of sites and the number of sites with detections.

Section M.3 Simple Data Format for the CCL 6

The simple data format is known as a two-dimensional flat file, which structures data that are stored as either a CSV or an Excel file. The simple data format is used to structure data extracted from main and select supplemental data sources for use in CCL 6.

An example of the simple data format is illustrated in Table 1. The simple data format consists of seven columns and each data entry in its own row:

- The first column, Name, provides the compound name as originally reported in the data source.
- The second column, CASRN, lists the CAS Registry Number, if used by the source as an identifier.
- The third column, Key, lists the DTXSIDs for the compounds.
- The fourth column, Value, lists values associated with the data entry.
- The fifth column, Unit, is the units for the value.
- The sixth column, Source, contains a shorthand indicator or acronym to describe the source of the data.
- The seventh and final column, Data Element, includes a shorthand code that describes the type of data element that the data entry is describing, such as "RfD"; data elements can refer to any of the value's data type, data group, measure, subset, and water type (e.g., ambient, finished, or wastewater effluent). For instance, a data element could represent the maximum concentration of a chemical in finished water or a RfD of any exposure duration.

Table M-1. Simple Data Format

Name	CASRN	Key	Value	Unit	Source	Data Element
chemical identifier reported by the data source	CASRN for the chemical	DTXSID number for the chemical	value associated with a specific data entry	units for the value	a code description for the data source	a code description for the data type (e.g., RfD, release)

Table M-2 provides an example of a data entry for a RfD from the Provisional Peer-Reviewed Toxicity Value (PPRTV) program for vanadium in the simple file format. The simple data format ensures the name of the chemical is always maintained as the identical name to the original data source. This allows traceability between processed data and the original source data. The simple data format also allows for the compilation of all available data into a single Pre-Universe file as described in Section 2.3 of the main document and is similarly used for much of the information considered and compiled for CCL 6.

Table M-2. Example of a Data Entry for an RfD from EPA-PPRTV for Vanadium in the Simple File Format

Name	CASRN	Key	Value	Units	Source	Data Element
vanadium	7440-62-2	DTXSID2040282	7E-5	mg/kg/day	pprtv	rfd

Section M.4 EPA's CompTox Chemicals Dashboard Data Elements Used in CCL 6 and Descriptions

Data Element	Description
TEST Model Predictions	The Toxicity Estimation Software Tool (TEST) was developed by EPA to estimate toxicity and physical properties of chemicals. Additional information on the TEST model can be found in the following support document: "User's Guide for T.E.S.T. (version 4.2) (Toxicity Estimation Software Tool): A Program to Estimate Toxicity from Molecular Structure" (USEPA, 2016c). EPA included the following TEST predictions from the CompTox Chemicals Dashboard in the Universe: oral rat 50 percent lethal dose (LD_{50}), bioconcentration factor, developmental toxicity, Ames mutagenicity (mutagenicity), normal boiling point, water solubility, vapor pressure.
OPERA Model Predictions	The Open structure-activity Relationship App (OPERA) was developed by EPA and provides predictions for physicochemical properties, environmental fate parameters, and toxicity endpoints. More information on how the OPERA tool was developed can be found in Mansouri et al. (2016; 2018). EPA included the following OPERA predictions from the CompTox Chemicals Dashboard in the Universe: bioconcentration factor, biodegradation half-life, boiling point, Henry's law constant, octanol-water partition coefficient (Kow), organic carbon-water partition coefficient (Koc), vapor pressure, water solubility.
ExpoCast Exposure Predictions	This data element describes predicted daily exposure to a chemical in units of milligrams of a chemical per kilogram bodyweight per day. The value included for each chemical is a prediction of the median exposure level for the total population. Further information about the types of models used by the ExpoCast program for exposure predictions can be found in Wambaugh et al. (2014) and Ring et al. (2019).
ENDOCRINE: endocrine disruptor chemicals	This data element is the second and final list of chemicals identified under Tier 1 screening of the Endocrine Disruptor Screening Program (EDSP). The screening program was developed to determine whether certain substances have potential endocrine disrupting effects or may interact with the endocrine system.
ToxCast Assay Hit Count	This element reports the number of total *in vitro* assays tested under the ToxCast or Tox21 *in vitro* screening program, and the number of assays with the result of "active" for specific chemicals. Details on which assays were active and the associated AC50's can be found on the CompTox Chemicals Dashboard website, but this information is not available for download in a "retrievable" form. The ToxCast Assay Hit Count reports results as a fraction and a percent.
Number of PubMed Articles	This element includes the number of PubMed records associated with the given chemical structure. The value gives a sense of the amount of literature available that may not be "retrievable" for the Universe.
ANDROGEN: androgen receptor chemicals	This element is a list of chemicals used to find literature with *in vitro* androgen receptor binding data. This reference material was used to help develop a computational model for androgen receptor activity. More information on this model can be found in Kleinstreuer et al. (2017).
NEURO: Chemicals triggering developmental neurotoxicity *in vivo*	This element is a list of compounds documented to trigger developmental neurotoxicity in animal models in at least two different laboratories. The details describing the parameters for inclusion in this list are described in Table 5 of Aschner et al. (2017). This data element was combined with the other NEURO lists to form the neurotoxin/neurotoxicant data element for CCL 6.

Data Element	Description
NEURO: Human Neurotoxicants	This element is a list of 201 industrial chemicals compiled by Grandjean and Landrigan (2006) which are known to be neurotoxic to humans. This data element was combined with the other NEURO lists to form the neurotoxin/neurotoxicant data element for CCL 6.
NEURO: Chemicals demonstrating effects on neurodevelopment	This element is a list of compounds with data demonstrating effects on neurodevelopment. Mundy et al. (2015) performed a literature review of peer-reviewed studies and regulatory documents with the goal of evaluating the available evidence for chemicals that have been reported to alter brain development in animal tests or humans. The evidence found is described in Table 1 of Mundy et al. (2015). This data element was combined with the other NEURO lists to form the neurotoxin/neurotoxicant data element for CCL 6.
NEURO: Neurotoxicants from PubMed	This element is a list of chemicals thought to be neurotoxic, determined through automated literature mining of PubMed. The list was compiled using Medical Subject Headings (MeSH) search terms and associations of these with single chemical substances (when possible). In total, 4,528 chemicals were identified; this list contains 1,243 chemicals associated with 5 or more literature references, all of which have been registered in the CompTox Chemicals Dashboard. This data element was combined with the other NEURO lists to form the neurotoxin/neurotoxicant data element for CCL 6.

Section M.5 Data Elements Not Assigned Screening Points

Data Element	Description
Acute LOAEL	Lowest Observed Adverse Effect Level in a study with an acute exposure duration
Acute NOAEL	No Observed Adverse Effect Level in a study with an acute study duration
Acute reference concentration	Concentration level at which no adverse effects are likely to occur after an acute inhalation exposure
Action level	EPA National Primary Drinking Water Regulations action level for lead and copper
Ames mutagenicity assay results – TEST model	Prediction of mutagenicity based on whether the chemical has tested positive for induction of revertant colony growth in any strain of *Salmonella typhimurium* (downloaded from EPA's CompTox Chemicals Dashboard)
Androgen receptor chemicals	Presence on the androgen receptor chemicals list
Bioconcentration factor – OPERA model	Predicted bioconcentration factor (ratio of concentration in fish tissue to concentration in surrounding water) from the OPERA model (downloaded from EPA's CompTox Chemicals Dashboard)
Bioconcentration factor – TEST model	Predicted bioconcentration factor (ratio of concentration in fish tissue to concentration in surrounding water) from the TEST Model (downloaded from EPA's CompTox Chemicals Dashboard)
Boiling point – OPERA model	Predicted normal boiling point in degrees Celsius from the OPERA Model (downloaded from EPA's CompTox Chemicals Dashboard)
Boiling point – TEST model	Predicted normal boiling point in degrees Celsius from the TEST Model (downloaded from EPA's CompTox Chemicals Dashboard)
Cancer classification	The cancer classification as designated by EPA, NTP, or IARC. EPA converted cancer classifications to a numerical form that directly translated to a screening point assignment. See Section 2.4.4 of the main document for information on this conversion.
Developmental Toxicity – TEST model	Prediction of whether a chemical is a potential developmental toxin (downloaded from EPA's CompTox Chemicals Dashboard)
Endocrine Disruptor Screening Program (EDSP) List 2 Chemicals	List of endocrine disruptor chemicals from the final EDSP List 2 (downloaded from EPA's CompTox Chemicals Dashboard)
ExpoCast exposure level prediction	Predicted daily exposure to a chemical based on the median exposure level for the total population. Further information about the types of models used by the ExpoCast program for exposure predictions can be found in Wambaugh et al. (2014) and Ring et al. (2019) (downloaded from EPA's CompTox Chemicals Dashboard)
Henry's Law Constant – OPERA model	Predicted Henry's Law constant from the OPERA Model (downloaded from EPA's CompTox Chemicals Dashboard)
Inhalation LOAEL	Lowest Observed Adverse Effect Level from a chronic inhalation study
Inhalation NOAEL	No Observed Adverse Effect Level from a chronic inhalation study
Inhalation Subchronic LOAEL	Lowest Observed Adverse Effect Level from a subchronic inhalation study
Inhalation Subchronic NOAEL	No Observed Adverse Effect Level from a subchronic inhalation study
Inhalation unit risk	Lifetime extra risk for a chronic inhalation exposure scenario resulting in carcinogenicity
Koc – OPERA model	Predicted organic carbon-water partition coefficient (Koc) from the OPERA Model (downloaded from EPA's CompTox Chemicals Dashboard)
K_{ow} – OPERA model	Predicted log octanol water partition coefficient ($\log(K_{ow})$) from the OPERA Model (downloaded from EPA's CompTox Chemicals Dashboard)
MADL	Maximum allowable dose level for reproductive toxicity from CalEPA OEHHA

Data Element	Description
Maximum concentration in ambient water	Maximum concentration in ambient water from a given source such as Batt et al. (2016), Bradley et al. (2017), and others.
Maximum concentration in finished water	Maximum concentration in finished water from a given source such as UCMR 1-4, Glassmeyer et al. (2017), and others. This data element was used to calculate the screening hazard quotient (sHQ) which was assigned screening points.
Maximum concentration in groundwater	Maximum concentration of a chemical observed in ground water (USDA PDP)
Maximum contaminant level	EPA National Primary Drinking Water Regulations maximum contaminant level (MCL)
Maximum contaminant level goal	EPA National Primary Drinking Water Regulations maximum contaminant level goal (MCLG)
Non-targeted detection frequency	Number of samples (12) with detects in Newton et al. (2018) non-targeted study of Britta Water Filter extracts
Public Nomination	Indicates whether the contaminant was nominated via the public nominations process. See Section 3.6.2 of the main document for a summary of chemical nominations.
Rat LD_{50} - TEST model	Predicted oral rat LD_{50} (single dose of a chemical resulting in 50% mortality of the test species) from the TEST Model (downloaded from EPA's CompTox Chemicals Dashboard)
Reference concentration	Concentration level at which no adverse effects are likely to occur after chronic inhalation exposure
Subchronic reference concentration	Concentration level at which no adverse effects are likely to occur after subchronic inhalation exposure
ToxCast assay fraction	The fraction of active ToxCast *in vitro* assays tested over the total number of assays tested for a chemical (downloaded from EPA's CompTox Chemicals Dashboard)
Vapor pressure – OPERA model	Predicted vapor pressure in mmHg at 25˚ C from the OPERA QSAR Model (downloaded from EPA's CompTox Chemicals Dashboard)
Vapor pressure – TEST model	Predicted vapor pressure in mmHg at 25˚ C from the TEST Model (downloaded from EPA's CompTox Chemicals Dashboard)
Water solubility – OPERA model	Predicted water solubility in mol/L at 25˚C from the TEST Model (downloaded from EPA's CompTox Chemicals Dashboard)
Water solubility – TEST model	Predicted water solubility in mol/L at 25˚C from the TEST Model (downloaded from EPA's CompTox Chemicals Dashboard)

EDSP = Endocrine Disruptor Screening Program; IARC = International Agency for Research on Cancer; K_{ow} = Octanol-water Partition Coefficient; LD_{50} = Median Lethal Dose; LOAEL = Lowest Observed Adverse Effect Level; MADL = Maximum Allowable Dose Level; NOAEL = No Observed Adverse Effect Level; NTP = National Toxicology Program; OPERA = OPEn (q)saR App; QSAR = Quantitative Structure-Activity Relationship; sHQ = Screening Hazard Quotient; TEST = Toxicity Estimation Software Tool; USDA PDP = United States Department of Agriculture Pesticide Data Program

Section M.6 References

Aschner, M., S. Ceccatelli, M. Daneshian, E. Fritsche, N. Hasiwa, T. Hartung, H.T. Hogberg, M. Leist, A. Li, W.R. Mundy, S. Padilla, A.H. Piersma, A. Bal-Price, A. Seiler, R.H. Westerink, B. Zimmer and P.J. Lein. 2017. Reference compounds for alternative test methods to indicate developmental neurotoxicity (DNT) potential of chemicals: Example lists and criteria for their selection and use. ALTEX - Alternatives to animal experimentation. 34(1):49-74. doi: 10.14573/altex.1604201.

Batt, A.L., T.M. Kincaid, M.S. Kostich, J.M. Lazorchak and A.R. Olsen. 2016. Evaluating the extent of pharmaceuticals in surface waters of the United States using a national-scale rivers and streams assessment survey. Environmental Toxicology and Chemistry. 35(4):874-81. https://doi.org/ 10.1002/etc.3161.

Bradley, P.M., C.A. Journey, K.M. Romanok, L.B. Barber, H.T. Buxton, W.T. Foreman, E.T. Furlong, S.T. Glassmeyer, M.L. Hladik, L.R. Iwanowicz, D.K. Jones, D.W. Kolpin, K.M. Kuivila, K.A. Loftin, M.A. Mills, M.T. Meyer, J.L. Orlando, T.J. Reilly, K.L. Smalling, and D.L. Villeneuve. 2017. Expanded Target-Chemical Analysis Reveals Extensive Mixed-Organic-Contaminant Exposure in U.S. Streams. Environmental Science & Technology. 51(9): 4792–4802. https://doi.org/10.1021/acs.est.7b00012.

Bradley, P.M., D.W. Kolpin, K.M. Romanok, K.L. Smalling, M.J. Focazio, J.B. Brown, M.C. Cardon, K.D. Carpenter, S.R. Corsi, L.A. DeCicco, J.E. Dietze, N. Evans, E.T. Furlong, C.E. Givens, J.L. Gray, D.W. Griffin, C.P. Higgins, M.L. Hladik, L.R. Iwanowicz, C.A. Journey, K.M. Kuivila, J.R. Masoner, C.A. McDonough, M.T. Meyer, J.L. Orlando, M.J. Strynar, C.P. Weis, and V.W. Wilson. 2018. Reconnaissance of mixed organic and inorganic chemicals in private and public supply tapwaters at selected residential and workplace sites in the United States. Environmental Science & Technology. 52, 23:13972–13985. https://doi.org/10.1021/acs.est.8b04622.

Glassmeyer, S.T., E.T. Furlong, D.W. Kolpin, A.L. Batt, R. Benson, J.S. Boone, O. Conerly, M.J. Donohue, D.N. King, M.S. Kostich, H.E. Mash, S.L. Pfaller, K.M. Schenck, J.E. Simmons, E.A. Varughese, S.J. Vesper, E.N. Villegas, and V.W. Wilson. 2017. Nationwide reconnaissance of contaminants of emerging concern in source and treated drinking waters of the United States. Science of The Total Environment. 581-582: 909-922. https://doi.org/10.1016/j.scitotenv.2016.12.004.

Grandjean, P. and P.J. Landrigan. 2006. Developmental neurotoxicity of industrial chemicals. The Lancet. 368(9553): 2167-2178. doi: 10.1016/S0140-6736(06)69665-7.

Kleinstreuer, N.C., P. Ceger, E.D. Watt, M. Martin, K. Houck, P. Browne, R.S. Thomas, Casey, W.M., Dix, D.J., Allen, D., Sakamuru, S., Xia, M., Huang, R. and Judson, R. 2017. Development and validation of a computational model for androgen receptor activity. Chemical Research in Toxicology. 30(4):946-964. https://doi.org/10.1021/acs.chemrestox.6b00347.

Mansouri, K., C.M. Grulke, A.M. Richard, R.S. Judson, and A.J. Williams. 2016. An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling. SAR QSAR Environ Res. 27(11): 939-965. doi:10.1080/ 1062936X.2016.1253611.

Mansouri, K., C.M. Grulke, R.S. Judson, and A.J. Williams. 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. Journal of Cheminformatics. 10(1): 10. doi:10.1186/s13321-018-0263-1.

Mundy, W.R., S. Padilla, J.M. Breier, K.M. Crofton, M.E. Gilbert, D.W. Herr, K.F. Jensen, N.M. Radio, K.C. Raffaele, K. Schumacher, T.J. Shafer, and J. Crowden. 2015. Expanding the test set: Chemicals with potential to disrupt mammalian brain development. Neurotoxicology and Teratology. 52A: 25-35. doi: 10.1016/j.ntt.2015.10.001.

Newton, S.R., R.L. McMahen, J.R. Sobus, K. Mansouri, A.J. Williams, A.D. McEachran and M.J. Strynar. 2018. Suspect screening and non-targeted analysis of drinking water using point-of-use filters. Environmental Pollution. 234: 297-306. https://doi.org/10.1016/j.envpol.2017.11.033.

Ring, C.L., J.A. Arnot, D.H. Bennett, P.P. Egeghy, P. Fantke, L. Huang, K.K. Isaacs, O. Jolliet, K.A. Phillips, P.S. Price, H. Shin, J.N. Westgate, R.W. Setzer, and J.F. Wambaugh. 2019. Consensus modeling of median chemical intake for the U.S. population based on predictions of exposure pathways. Environmental Science & Technology. 53(2): 719-732. doi: 10.1021/acs.est.8b04056.

Wambaugh, J.F., A. Wang, K.L. Dionisio, A. Frame, P. Egeghy, R. Judson, and R.F. Setzer. 2014. High throughput heuristics for prioritizing human exposure to environmental chemicals. Environmental Science & Technology. 48(21): 12760-12767. doi: 10.1021/es503583j.

USEPA. 1996. Guidelines for Reproductive Toxicity Risk Assessment. https://www.epa.gov/sites/default/files/2014-11/documents/guidelines_repro_toxicity.pdf

USEPA. 2016. User's Guide for T.E.S.T. (version 4.2) (Toxicity Estimation Software Tool). Office of Research and Development