Technical Support Document for the Final Fifth Contaminant Candidate List (CCL 5) - Chemical Contaminants

Office of Water (4607M)
EPA 815-R-22-002
October 2022
www.epa.gov/safewater

Table of Contents

Chapter 1 Introduction 1
Section 1.1 Background 1
Section 1.2 Overview of the CCL 5 Development Process 2
Chapter 2 Building the Universe 5
Section 2.1 Overview 5
Section 2.2 Assessing and Identifying Data Sources 5
Section 2.2.1 CCL 5 Primary Data Sources 7
Section 2.2.2 CCL 5 Supplemental Data Sources 10
Section 2.3 Developing a Pre-Universe 10
Section 2.3.1 Overview 10
Section 2.3.2 Extracting Relevant Data Elements for Developing the Pre-Universe 11
Section 2.3.3 Assigning Unique Contaminant Identifiers 13
Section 2.3.4 Saving Extracted Metrics in a Simple Data Format 14
Section 2.4 Enhancing the Universe 14
Section 2.4.1 Overview 14
Section 2.4.2 Refining DTXSID Assignments 14
Section 2.4.3 Additional Data Accessed via the Comptox Chemicals Dashboard 16
Section 2.4.4 Creating a Uniform Universe File 16
Chapter 3 Screening Universe Chemicals to Select the PCCL 18
Section 3.1 Overview 18
Section 3.2 Establishing the Screening Data 19
Section 3.2.1 Incorporating Universe Data Elements 19
Section 3.2.2 Calculating Screening Hazard Quotients (sHQs) 23
Section 3.3 Developing a Scoring Rubric 25
Section 3.3.1 Determining Screening Tiers 25
Section 3.3.2 Determining Relative Point Assignments Within Each Screening Tier 27
Section 3.4 Final Point Assignments and Screening Scores 35
Section 3.5 Using the CCL 5 Screening System 37
Section 3.6 Consideration of Publicly Nominated Chemicals 38
Section 3.6.1 Soliciting Public Nominations 38
Section 3.6.2 Summary of Chemical Nominations 38
Section 3.6.3 Consideration of Publicly Nominated Chemicals for the PCCL 39

Section 3.7 Chemicals Excluded from the PCCL 40
Section 3.7.1 Regulatory Determination 4 Chemicals 40
Section 3.7.2 Canceled Pesticides 40
Section 3.8 Summary of the PCCL 5 43
Chapter 4 Classification of PCCL Chemicals to Select the CCL 44
Section 4.1 Overview 44
Section 4.2 Supplemental Data Collection 45
Section 4.2.1 Occurrence Data 45
Section 4.2.2 Health Effects Data 47
Section 4.3 Calculated Data Elements 48
Section 4.3.1 Health Reference Levels and CCL Screening Levels 48
Section 4.3.2 Final Hazard Quotients 50
Section 4.3.3 Attribute Scores 51
Section 4.4 Contaminant Information Sheets (CISs) 60
Section 4.5 Evaluation Teams Listing Decision Process 61
Section 4.5.1 Evaluation Teams 61
Section 4.5.2 Evaluator Training 61
Section 4.5.3 Independent Reviews 62
Section 4.5.4 Listing Decisions 63
Section 4.6 Logistic Regression Analysis 64
Section 4.6.1 Overview 64
Section 4.6.2 Logistic Regression Applied to Validate the Selection of the PCCL 65
Section 4.6.3 Post-Evaluation Analysis: Exploring Listing Decision Determinants 69
Section 4.7 Selecting CCL 5 Chemicals 75
Chapter 5 CCL 5 Data Availability Assessment 80
Section 5.1 Overview 80
Section 5.2 Data Availability for CCL 5 Chemicals 80
Section 5.2.1 Occurrence 83
Section 5.2.2 Health Effects 84
Section 5.2.3 Analytical Methods 84
Section 5.3 Data Availability for PCCL 5 Chemicals not on CCL 5 85
Chapter 6 Data Management and Quality Assurance 90
Section 6.1 Overview 90
Section 6.2 Quality Assurance of PCCL 5 Development 90
Section 6.2.1 Overview 90

Section 6.2.2 Reviewing Input Data....91
Section 6.2.3 Reviewing Pre-processing Code....91
Section 6.2.4 QA/QC Procedure for DTXSID Assignments....92
Section 6.2.5 QA/QC Procedure for Screening Code....93
Section 6.2.6 QA/QC Procedure for Outputs....94
Section 6.3 Quality Assurance of CIS Development....94
Section 6.3.1 Overview....94
Section 6.3.2 Preparing Health Effects Data for CISs....95
Section 6.3.3 Preparing Occurrence Data for CISs....95
Section 6.3.4 Data Management and QA/QC of CISs....98
Chapter 7 References....99
List of Appendices....104
Appendix A - Primary Data Source Descriptions....A-1
Appendix B - Supplemental Data Sources....B-1
Appendix C - Publicly Nominated Chemical Contaminants....C-1
Appendix D - PCCL Chemical Contaminants....D-1
Appendix E - Protocol for the Occurrence Literature Review....E-1
Appendix F - Protocol for the Rapid Systematic Health Effects Literature Review....F-1
Appendix G - Protocol to Derive Health Concentrations....G-1
Appendix H - Protocol to Select Water Concentrations Used in Calculating Final Hazard Quotients....H-1
Appendix I - Protocol to Determine Potency Attribute Scores....I-1
Appendix J - Protocol to Determine Severity Attribute Scores....J-1
Appendix K - Protocol to Determine Prevalence Attribute Scores....K-1
Appendix L - Protocol to Determine Magnitude Attribute Scores....L-1
Appendix M - Protocol to Determine Magnitude Attribute Scores from Persistence-Mobility....M-1
Appendix N - Data Management for CCL 5....N-1
Appendix O – CCL 4 Chemicals Not Listed on CCL 5....O-1
Appendix P - Group of 23 DBPs included on CCL 5....P-1

Equations, Figures, and Tables

Equation 1. Formula for Calculating Screening Hazard Quotients....24
Equation 2. Formula for Calculating Final Hazard Quotients....50
Equation 3. Simple Logistic Regression Model....64

Figure 1. CCL Development Framework....3
Figure 2. CCL 5 Development Framework Step 1 - Building the Universe....5

Figure 3. Data Source Assessment Process 7
Figure 4. Development Framework Step 2 – Screening 19
Figure 5. Empirical Histogram of Log Transformed Screening Hazard Quotients Calculated for the Screening Step 25
Figure 6. Empirical Distribution of Cancer Slope Factors in the CCL 5 Universe 27
Figure 7. Ambient and Finished Water Detection Rates for the CCL 5 Universe Chemicals 28
Figure 8. Total Screening Scores for the CCL 5 Universe Chemicals 36
Figure 9. Health Effects and Occurrence Scores for the CCL 5 Universe Chemicals 37
Figure 10. Development Framework Step 3 – Classification 45
Figure 11. Rounded Logarithmic Distributions of CSFs, RfDs, NOAELs and LOAELs for the CCL 5 Universe 53
Figure 12. Flow Diagram of the Three-Step Iterative Process 65
Figure 13. Results of the Bayesian Simple Logistic Model of Probability of Listing vs Screening Score 69
Figure 14. AUC-ROC Curve for Screening Scores as a Predictor of Listing Decisions 73
Figure 15. AUC-ROC Curve for Parsimonious Model as a Predictor of Listing Decisions 74

Table 1. CCL 5 Health Effects Primary Data Sources 8
Table 2. CCL 5 Occurrence Primary Data Sources 9
Table 3. Cancer Classification Numeric Conversions 17
Table 4. Data Elements Assigned Points in the CCL 5 Screening System 20
Table 5. Formulas for Calculating Health Screening Levels (HSLs) 24
Table 6. Health and Occurrence Tiers for Points Assignments 26
Table 7. Point Assignments for Health-Related Data Elements 31
Table 8. Point Assignments for Occurrence-Related Data Elements 33
Table 9. Summary of Persistence Ranking Score 40
Table 10. Canceled Pesticides Assessed for Exclusion from PCCL 5 41
Table 11 Accounting on the PCCL 5 43
Table 12. Median Logarithmic Distribution Values by Toxicity Value 54
Table 13. Potency Scoring Equations by Toxicity Value 54
Table 14. CCL 5 Severity Categories 55
Table 15. Relationship between Data Elements Used to Score Prevalence and Magnitude 56
Table 16. Data Elements Used to Score Persistence and Mobility 59
Table 17. Survey Listing Decision Outcomes 63
Table 18. Summary Statistics for the MCMC Sample 67
Table 19. Summary Statistics of Probabilities of Listing at Screening Scores 3310 and 9050 Calculated from the MCMC Sample. 67
Table 20. Descriptive Statistics by Listing Decision Outcome 70
Table 21. Simple Logistic Regression Results 71
Table 22. Multiple Logistic Regression: Parsimonious Model 74
Table 23. Chemical Contaminants on the CCL 5 77
Table 24. Unregulated DBPs in the DBP Group on the CCL 5 79
Table 25. Data Availability for CCL 5 Chemicals 80
Table 26. Data Availability for PCCL 5 Chemicals not on CCL 5 86

List of Abbreviations and Acronyms

ADAF	Age-Dependent Adjustment Factor
ATSDR	Agency for Toxic Substances and Disease Registry
CADW	Canadian Drinking Water Quality
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
CDPR	California Department of Pesticide Regulation
CDR	Chemical Data Reporting
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CFR	Code of Federal Regulations
CSF	Cancer Slope Factor
CWS	Community water system
CWSS	Community Water System Survey
DSSTox	Distributed Structure Searchable Toxicity Public Database Network
DTXSID	Distributed Structure-Searchable Toxicity Substance Identifier
DWC	Drinking Water Committee
EEC	Estimated environmental concentrations
EDWC	Estimated drinking water concentrations
EPA	United States Environmental Protection Agency
FDA	Food and Drug Administration
fHQ	Final Hazard Quotient
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FR	Federal Register
HSL	Health Screening Level
HRL	Health Reference Level
HSDB	Hazardous Substances Data Bank
IARC	International Agency for Research on Cancer
ICR	Information Collection Rule
InChi	International Chemical Identifier
IOC	Inorganic Compounds
IRIS	Integrated Risk Information System
LD_{50}s	Median Lethal Doses
LOD	Limit of Detection
LOAEL	Lowest Observed Adverse Effect Levels
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
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 Academy of Sciences' National Research Council
NPDWR	National Primary Drinking Water Regulation
NTP	National Toxicology Program
NWIS	National Water Information System
OPERA	OPEn structure–activity/property Relationship App
OPP	Office of Pesticide Programs
PAD	Population Adjusted Dose
PCCL 5	Preliminary Contaminant Candidate List 5
PDP	Pesticide Data Program
PECO	Population, Exposure, Control, and Outcome
PPRTVs	Provisional Peer-Reviewed Toxicity Values
PWS	Public Water System
QSAR	Qualitative Structure-Activity Relationship
RD 3	Regulatory Determination 3
RD 4	Regulatory Determination 4
RD 5	Regulatory Determination 5
RfD	References Dose
RSR	Rapid Systematic Review
SAB	Science Advisory Board
SDWA	Safe Drinking Water Act
sHQ	Screening Hazard Quotient
SYR 3	Six-Year Review 3
TRI	Toxics Release Inventory
TSCA	Toxic Substance 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
USDA	United State Department of Agriculture
USGS	United States Geological Survey

Chapter 1 Introduction

Section 1.1 Background

Section 1412(b)(1)(B)(i) of the Safe Drinking Water Act (SDWA), as amended in 1996, requires the United States Environmental Protection Agency (EPA) 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.
- Are known or anticipated to occur in public water systems (PWSs).
- May require regulation under the SDWA.

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

The SDWA directs the agency to consider health effects and occurrence information for the unregulated contaminants to identify those contaminants that present the greatest public health concern related to exposure from drinking water. In identifying these contaminants, the SDWA requires that, when developing the CCL, EPA considers the National Contaminant Occurrence Database established under Section 1445(g) of the SDWA and consults the scientific community including the Science Advisory Board (SAB). 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.

EPA interprets broadly the criterion that contaminants are known or anticipated to occur in PWSs. In evaluating this criterion, EPA considers not only PWS monitoring data but also data on concentrations in ambient surface and ground waters, releases to the environment (e.g., Toxics Release Inventory), and production. Though such data may not establish conclusively that contaminants are known to occur in PWSs, EPA considers these data sufficient to anticipate that contaminants may occur in PWSs. The agency also considers adverse health effects that may pose a greater risk to lifestages and other sensitive groups that represent a meaningful portion of the population. Adverse health effects associated with infants, children, pregnant women, the elderly, and individuals with a history of serious illness in particular are evaluated.

In a regulatory action separate from the CCL, SDWA Section 1412(b)(1)(B)(ii) directs EPA to make regulatory determinations on at least five of the contaminants from the CCL every five years. Section 1412(b)(1)(A) of the SDWA specifies that 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.
- 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 identifies contaminants that serve as a short list to be considered for research and data collection efforts, such as the Unregulated Contaminant Monitoring Rule (UCMR). Only after additional data and

information are collected are contaminants considered for regulatory determination and rulemaking under the SDWA.

Prior to CCL 5, EPA had completed four previous cycles of CCLs since 1996 that are briefly described as follows:

- 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.

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

- EPA published the Third Contaminant Candidate List (CCL 3) on October 8, 2009 (74 FR 51850, USEPA, 2009f). In developing the CCL 3, 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. The third CCL (CCL 3) contained 104 chemicals or chemical groups and 12 microbial contaminants. EPA carried forward CCL 3 contaminants (minus those with regulatory determinations) to the Draft CCL 4.

- EPA published the Fourth Contaminant Candidate List (CCL 4) on November 17, 2016 (81 FR 81099, USEPA, 2016a). 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, except for two.

Section 1.2 Overview of the CCL 5 Development Process

The methodology for developing the Fifth Contaminant Candidate List (CCL 5) is based on the existing, three-step framework used previously for the CCL 3 (USEPA, 2009a). CCL 4 was a carryover from the CCL 3 and followed the same framework (USEPA, 2016a). In developing CCL 5, updates were made to allow 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 comprises three steps:

1. Building the Universe
2. Screening
3. Classification

Step 1 includes compiling the CCL 5 Universe of potential drinking water contaminants. During this step, EPA identified the primary data sources for CCL 5. As directed by the SDWA, 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 5 Universe and publicly nominated chemicals to identify a subset of chemicals that merit further review due to their potential to occur in PWSs and thereby pose a public health concern. This subset of chemicals is called the Preliminary CCL 5 (PCCL 5). In this step, 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. EPA screened chemicals to the PCCL 5 by evaluating the health effects and occurrence information provided in the data sources used to compile the CCL 5 Universe. The screening procedure is designed to balance known and unknown information regarding toxicity, exposure, and risk by assigning higher value to data that are more indicative of an occurrence in finished drinking water and potential to cause health effects.

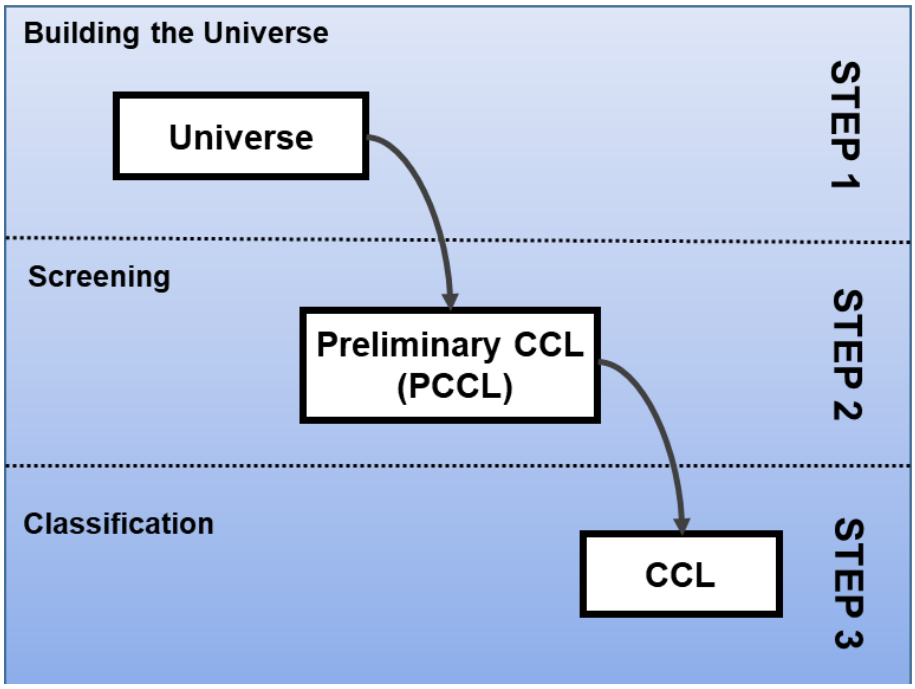


Figure 1. CCL Development Framework

Step 3 encompasses a structured approach for selecting CCL 5 from the PCCL 5. Following literature searches to collect supplemental data for the PCCL chemicals, the relevant data metrics for each chemical were summarized in a standardized document called a Contaminant Information Sheet (CIS). EPA scientists with a broad range of professional experience and relevant expertise evaluated chemicals for the CCL 5. These chemical evaluators used 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 5 chemicals should be listed on the CCL 5.

This Technical Support Document (TSD) describes in detail the process used to develop the CCL 5 for chemical contaminants and the updates made in response to expert input and recommendations from the SAB, NDWAC, NRC, 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 5 development process.
- Chapters 2, 3 and 4 describe in detail Steps 1, 2, and 3, respectively.
- Chapter 5 presents the data availability assessment of CCL 5 chemicals.
- Chapter 6 describes data management and quality assurance.

The companion documents to this chemicals TSD include the following:

- Technical Support Document for the Final Fifth Contaminant Candidate List (CCL 5) – Microbial Contaminants (USEPA, 2022a)
- Technical Support Document for the Final Fifth Contaminant Candidate List (CCL 5) – Contaminant Information Sheets (CISs), hereafter referred to as the CIS Technical Support Document (USEPA, 2022b)

All three technical support documents are accessible via the EPA docket (Docket ID No. EPA-HQ-OW-2018-0594) at https://www.regulations.gov.

Chapter 2 Building the Universe

Section 2.1 Overview

The purpose of Step 1 of the CCL 5 development process for chemical candidates is to construct a broad universe of potential drinking water chemical contaminants, as shown in blue in Figure 2. During this step, EPA compiled primary and supplemental data sources, identified 21,894 chemicals from primary data sources to form a CCL 5 Pre-Universe and then added supplemental data for pre-universe chemicals to create a CCL 5 Chemical Universe. For CCL 5, the agency retained all chemical contaminants identified in the pre-universe in the universe, which resulted in the largest and most data-rich CCL Universe generated to date.

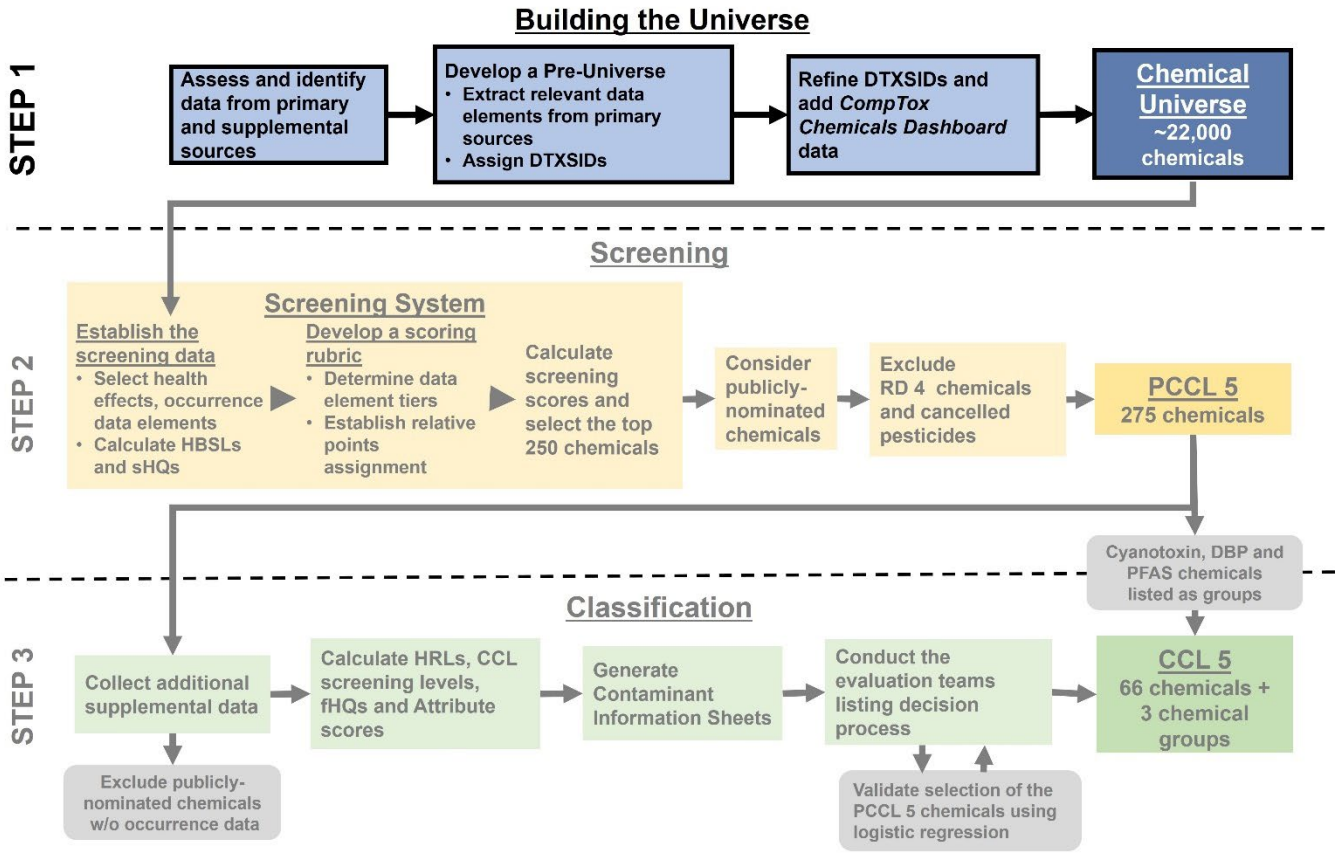


Figure 2. CCL 5 Development Framework Step 1 - Building the Universe

Section 2.2 Assessing and Identifying Data Sources

To initiate the CCL 5 development process, EPA compiled potential health effects and occurrence data sources that could be used to prioritize chemical contaminants for listing on the CCL 5. EPA compiled data sources identified from CCL 3 and CCL 4, along with data sources recommended by the CCL 5 EPA workgroup and subject matter experts. Information on how EPA addressed data sources provided through the public nomination process is described in Section 3.6.

As a result of this effort, EPA identified 134 potential data sources and further assessed their potential use for the CCL 5 development process. EPA accessed each potential data source online 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, EPA collects data on the volume of different chemicals produced in the U.S. under the Chemical Data Reporting (CDR) rule (USEPA, 2016b). This information can indicate potential occurrence of chemicals in the environment and therefore would be considered a relevant source of data for CCL 5 development. EPA's Integrated Risk Information System (IRIS) database would also be considered a relevant source of data, including 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.

- **Completeness:** 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, 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, EPA's Toxics Release Inventory (TRI) (USEPA, n.d.-b). Therefore, data from the more comprehensive source, TRI, were used while the other source was considered redundant and was not used. Note that multiple data sources may present values for the same data element; this does not make the data sources redundant. For example, 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.). EPA used both data sources and retained both RfC values (see the CIS for 1,4-dioxane in USEPA, 2022b).

- **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 5 development; the data are also accessible to the public (CDC, n.d.).

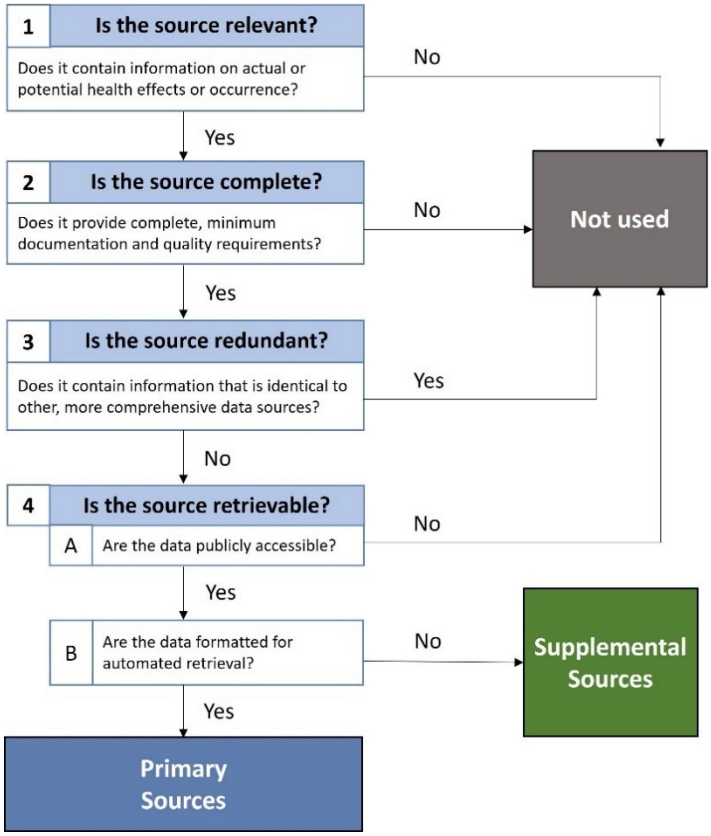


Figure 3. Data Source Assessment Process

These four assessment factors were used to evaluate data sources in 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 EPA's minimum guidelines for documentation and quality (NDWAC, 2004).

Data sources identified as relevant, complete, not redundant, and retrievable were considered primary data sources. Data sources that were not retrievable were set aside as supplemental sources. Twenty-one of the 134 potential data sources were excluded from further consideration in the CCL 5 process because they were not relevant or were incomplete or redundant, no longer existed, or had been combined with another data source. For example, the Distributed Structure Searchable Toxicity Public Database Network (DSSTox) was used in CCL 3, but it has since been incorporated into the CompTox Chemicals Dashboard, a supplemental data source for CCL 5 (see Section 2.4).

Section 2.2.1 CCL 5 Primary Data Sources

Out of the 134 potential sources of chemical data evaluated, 42 met all four assessment factors and therefore were considered primary data sources. One additional source, the Hazardous Substances Data Bank (HSDB) did not meet retrievability criteria because it lacked a format for automated retrieval but was still used as a primary 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 5 process. Therefore, additional effort was taken to extract these data, as was done with the CCL 3 process. EPA downloaded chemical data from these 43 primary data sources to a data directory to identify chemical contaminants for the pre-universe. These included 18 sources of health effects data listed in Table 1 and 25 sources of occurrence data listed in Table 2 and are described in Appendix A and in greater detail in Appendix N. EPA included health effects and occurrence data from primary data sources through December 2019. References for the primary data sources listed in Table 1 and Table 2 are provided in Appendix N.

Table 1. CCL 5 Health Effects Primary Data Sources

Data Source	Agency or Author[1]
Agency for Toxic Substances and Disease Registry (ATSDR) Minimal Risk Levels (MRLs)	Centers for Disease Control and Prevention (CDC)
Cancer Potency Data Bank	National Library of Medicine, U.S. Department of Health and Human Services (HHS)
Chemical Database	California Environmental Protection Agency (CalEPA) Office of Environmental Health Hazard Assessment (OEHHA)
Drinking Water Standards and Health Advisory Tables	EPA
Guidelines for Canadian Drinking Water Quality	Health Canada
Guidelines for Drinking-Water Quality	World Health Organization (WHO)
Hazardous Substances Data Bank	National Library of Medicine, HHS
Health-Based Screening Levels (HBSLs)	U.S. Geological Survey (USGS)
Human Health-Based Water Guidance Table	Minnesota Department of Health
Human Health Benchmarks for Pesticides	EPA
Integrated Risk Information System (IRIS)	EPA
International Agency for Research on Cancer Classifications	WHO
Maximum Recommended Daily Dose (MRDD) Database	U.S. Food and Drug Administration (FDA)
National Recommended Water Quality Criteria – Human Health Criteria	EPA
National Toxicology Program (NTP) Cancer Classifications	HHS
Provisional Peer-Reviewed Toxicity Values (PPRTVs)	EPA
Screening Levels for Pharmaceuticals	FDA Drugs@FDA database, National Institutes of Health (NIH) DailyMed Database
Toxicity Reference Database (ToxRefDB)	EPA

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

Table 2. CCL 5 Occurrence Primary Data Sources

Data Source	Agency or Author[1]
ATSDR Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) Substance Priority List	CDC
Chemical Data Reporting (CDR) Results	EPA
"Concentrations of prioritized pharmaceuticals in effluents from 50 large wastewater treatment plants in the US and implications for risk estimation"	Kostich et al. 2014
Disinfection By-product Information Collection Rule (DBP ICR)	EPA
"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
"Expanded target-chemical analysis reveals extensive mixed-organic-contaminant exposure in U.S. streams"	Bradley et al. 2017
Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) registered pesticides and pesticide ingredients	EPA
"Legacy and emerging perfluoroalkyl substances are important emerging water contaminants in the Cape Fear River Watershed of North Carolina"	Sun et al. 2016
National Health and Nutrition Examination Survey (NHANES)	CDC
National Inorganics and Radionuclides Survey (NIRS)	EPA
National Water Information System (NWIS)	Water Quality Portal, USGS
National Water-Quality Assessment (NAWQA)	Water Quality Portal, USGS
"Nationwide reconnaissance of contaminants of emerging concern in source and treated drinking waters of the United States"	Glassmeyer et al. 2017
"Nationwide reconnaissance of contaminants of emerging concern in source and treated drinking waters of the United States: Pharmaceuticals"	Furlong et al. 2017
Pesticide Data Program	U.S. Department of Agriculture (USDA)
Pesticide Use Estimates	USGS
"Pharmaceutical manufacturing facility discharges can substantially increase the pharmaceutical load to US wastewaters"	Scott et al. 2018
"Predicting variability of aquatic concentrations of human pharmaceuticals"	Kostich et al. 2010
"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
Surface Water Database (SURF)	California Department of Pesticide Regulation
"Suspect screening and non-targeted analysis of drinking water using point-of-use filters"	Newton et al. 2018
Toxics Release Inventory (TRI)	EPA
Unregulated Contaminant Monitoring Rule (UCMR) Cycles 1-3	EPA

Data Source	Agency or Author[1]
UCMR Cycle 4	EPA
Unregulated Contaminant Monitoring-State (UCM-State) Rounds 1 and 2	EPA

[1] References for the data sources listed in Table 2 are provided in Data Management Processing.

Section 2.2.2 CCL 5 Supplemental Data Sources

The use of primary data is critical to the entire CCL process, and it is often necessary to gather and extract additional data to further evaluate chemicals for listing on the CCL 5. As described in Section 2.2, EPA assessed data sources for potential use in the CCL 5 development process and set aside, as supplemental sources, 71 sources that met the relevance, completeness, and redundancy assessment factors but that were not retrievable because of their format. EPA also identified supplemental sources from data sources cited in public nominations (see Section 3.6) and conducted literature searches to identify further supplemental data on occurrence and health effects to aid in evaluating chemicals of interest (see Section 4.2). Though 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 detail and description to support CCL 5 listing decisions. See Appendix B for a complete list of supplemental data sources.

For health effects data, supplemental data sources were often closely related to a primary data source. For example, EPA's IRIS program provides an easily accessible and downloadable online database (https://cfpub.epa.gov/ncea/iris/search/index.cfm) that contains toxicity values for several hundred chemicals. The IRIS database met the four assessment factors to be a primary data source for CCL 5. Though 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 5 had to be manually extracted from IRIS assessments. Therefore, for certain chemicals, 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 Appendices F and G.

Supplemental occurrence data sources were also used to fill data gaps during the Step 3 classification process. For example, if primary data sources could not provide finished water data for a contaminant, EPA sought this information from a supplemental source identified through a literature search, from non-retrievable supplemental sources previously set aside, or from sources cited with public nominations. Many non-national scale studies on finished and ambient water were used to supplement the occurrence data from primary data sources (see Appendix B).

Section 2.3 Developing a Pre-Universe

Section 2.3.1 Overview

The pre-universe is a list of chemical contaminants identified through health and occurrence data extracted from primary data sources. Pre-universe development was conducted 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 below.

Generally, the pre-universe development involved pre-processing the CCL 5 primary data sources, which refers to the actions taken to identify chemicals from each source and transform data of various types, formats, and structures into a uniform and understandable format. Each data source used for CCL 5 has a unique data format and requires specific pre-processing steps to properly extract relevant data elements and create data entries. Additional information on pre-processing of primary data sources is provided in Appendix N. For CCL 5, data elements are defined as values or descriptors with unique meanings that characterize toxicological or occurrence information associated with chemical contaminants. Two examples of CCL 5 data elements are reference doses 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 74% finished water detection rate in public water systems for vanadium from UCMR 3, a nationally representative monitoring program.

EPA identified approximately 22,000 chemical contaminants, which formed the CCL 5 Pre-Universe, and created the CCL 5 Pre-Universe file for screening purposes (Step 2). The pre-universe file contained 41 types of data elements from the 43 primary data sources for a total of over 62,000 rows of individual data entries. See Table 3 in Section 3.2 for data elements extracted from primary data sources used in the screening step and Section N.5 of Appendix N for details about all 41 data elements extracted from primary data sources. In Step 2 (screening) and Step 3 (classification) of the CCL 5 development process, EPA extracted additional finished water and ambient water occurrence data elements from primary data sources.

The pre-universe was a starting point for chemical identification and data compilation, as was done during the CCL 3 process (USEPA, 2009a). The CCL 5 Pre-Universe file was later expanded to include additional data collected during the CCL 5 process, notably from supplemental sources compiled for Steps 2 and 3 of the CCL 5 process.

Section 2.3.2 Extracting Relevant Data Elements for Developing the Pre-Universe

Several relevant types of data elements were extracted for the development of the pre-universe. These categories include dose-response data, categorical toxicity data (e.g., cancer classifications), 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 median Lethal Doses (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.

Appendix N describes specifics about the pre-processing required to extract data elements from CCL 5 primary 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. 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, 2020), which include code and relevant notes.

Data sources may provide one or multiple data elements relevant to the CCL 5 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.

Following SAB recommendations for CCL 3, EPA prioritized extraction of the data elements most relevant to CCL 5 goals while developing the pre-universe (USEPA, 2009g). Therefore, EPA did not consider some data elements as relevant for CCL 5 because they are not necessarily directly implicated in health effects and/or occurrence in drinking water. For example, Furlong et al. (2017) provides concentration data in ambient and finished waters for pharmaceuticals and other contaminants of emerging concern, which were included in the pre-universe file. This study also provides chemical information, such as molecular weights, which were not included. Similarly, the Hazardous Substance Data Bank contains LD_{50} toxicity values, which were included in the pre-universe file, but it also contains EC_{50} (effective concentration) toxicity values, which were not included.

Some relevant data elements in primary data sources were not included in the pre-universe file because they were not needed for screening purposes (see Chapter 3), though they were appropriate for the classification process of CCL 5 (see Chapter 4). For example, EPA extracted additional data elements from primary data sources for chemicals in finished water and ambient water specifically for use in the classification process of CCL 5. In addition, if ambient or finished water concentration summary statistics were not readily available in the original data sources, summary statistics were calculated when possible and were considered part of the CCL 5 Pre-Universe. Specifics regarding data extraction and manipulation for these data elements are further described in Appendix N.

EPA updated how several data elements were treated in CCL 3 so they would be compatible with the CCL 5 screening process. For example, some primary CCL 5 occurrence data sources report non-detections for chemicals with water monitoring data. In CCL 3, however, finished and ambient water concentration summary statistics were based on analytical detections only and non-detections were not estimated or imputed (USEPA, 2009b). However, non-detections do not necessarily indicate that the chemical is absent and that the risk of exposure is zero, but rather indicate that the amount of chemical present is below a level that could be detected or quantified.

Therefore, recognizing the potential risk for exposure even when a chemical is reported as a non-detect, EPA adopted a more health protective approach to handle non-detections in ambient and finished water data in the screening stage of CCL 5. In this CCL cycle, EPA substituted maximum concentration values for chemicals with non-detects 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, in CCL 5, 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, EPA used half of the midpoint of the statistical range of the detection limits as the maximum concentration. For example, finished water monitoring data for propoxur provided by the U.S. Department of Agriculture Pesticide Data Program (USDA PDP) reported zero detects and a limit of detection (LOD) range of $6x10^{-6}$ µg/L – $4.13x10^{-4}$ µg/L. Therefore, EPA used $1.0175x10^{-4}$ 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-detects were handled for a specific data source are included in Appendix N.

One important difference in the health effects data elements used for the CCL 3 and CCL 5 processes is the inclusion of cancer slope factor (CSF) as a retrievable data element for CCL 5. During development of CCL 3, there was an insufficient number of CSF values in a retrievable format to be used for screening (USEPA, 2009b); however, they were used during the classification step of CCL 3. When primary data sources for the CCL 5 were collected, adding new retrievable toxicity data sources such as the Human Health Benchmarks for Pesticides resulted in 378 CSF values available in a retrievable format. Greater availability of CSFs meant it was possible to incorporate the CSF data element into the pre-universe file for screening purposes and use this data element in the classification step of the CCL 5 process.

Chemical contaminants with National Primary Drinking Water Regulations (NPDWRs) were also included in the pre-universe. However, these contaminants are already regulated; therefore, their inclusion in the CCL process is clearly unnecessary. Therefore, EPA extracted Maximum Contaminant Levels (MCLs) and Maximum Contaminant Level Goals (MCLGs) to easily identify regulated chemicals and their corresponding identifiers and remove them in the screening step (Step 2) of the CCL 5 development process. Regulated chemicals were not further considered for listing on the CCL 5.

Section 2.3.3 Assigning Unique Contaminant Identifiers

It is important that the data directory compiled for CCL 5 development correctly identifies health and occurrence information for specific chemicals across different sources, especially because the data sources may refer to chemicals using different identifiers. For example, the pharmaceutical gabapentin is referred to by four different identifiers across the CCL 5 primary data sources. To address this issue in the CCL 5 data directory, including the pre-universe file, EPA identifies chemicals by DTXSIDs (Distributed Structure-Searchable Toxicity Database Substance Identifiers) along with the original identifier provided by the data source.

EPA's DSSTox (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 EPA used to help identify chemicals and compile chemical-specific data for CCL 5. There are benefits of using DTXSIDs as the identifier system during the CCL 5 Pre-Universe and Universe development. First, DTXSIDs are curated by EPA to ensure that each DTXSID refers to one unique chemical or chemical group (Williams et al., 2017). Second, EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard) publishes mapping files that match DTXSIDs to other chemical identifiers, including chemical names, Chemical Abstracts Service (CAS) numbers, International Chemical Identifier (InChI) strings, and InChI keys. These mapping files allowed EPA to efficiently and accurately compile data provided by multiple data sources that used different chemical identifiers.

Some chemicals have no DTXSID on the CompTox Chemicals Dashboard website. In these cases, either "NA" or "NO_DTXSID" was temporarily entered in the ID field of the original source data. Further refinement of DTXSIDs occurred while building the universe and is discussed in Section 2.4.2.

Section 2.3.4 Saving Extracted Metrics in a Simple Data Format

At the beginning of the pre-universe development process, a "simple" data format was chosen so all primary and supplemental data could be easily combined and used in later steps of the CCL 5 development process. This simple format includes six critical pieces of information about each data entry required for the second step of the CCL 5 process:

- Chemical name or identifier as reported in the data source
- Chemical 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 N.

Section 2.4 Enhancing the Universe

Section 2.4.1 Overview

EPA used the pre-universe as a building block to prepare a universe of chemicals and related data elements that could be efficiently and effectively used during Steps 2 and 3 of the CCL 5 development process (Chapter 3 and Chapter 4). EPA refined DTXSIDs of chemicals identified in the pre-universe, added relevant supplemental data collected for pre-universe chemicals from EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard), and created a file to present data elements from different data sources in a uniform format. This universe file was used to screen chemicals for inclusion in the PCCL 5 and classify chemicals for inclusion in the CCL 5.

As mentioned in Section 2.1, the number of chemicals included in CCL 5 Pre-Universe and CCL 5 Universe was nearly the same; however, the amount of data associated with the CCL 5 Universe is far greater than that with the CCL 5 Pre-Universe.

An important difference in the Step 1 process for the CCL 3 and the CCL 5 was the use of selection criteria to narrow down the list of chemicals for inclusion in the CCL 3 Universe (USEPA, 2009a). In CCL 3, EPA reduced the number of unique substances identified from primary data sources from approximately 26,000 in the pre-universe to 6,003 in the universe based on availability of health effects and occurrence data (USEPA, 2009a). In CCL 5, EPA skipped this extra step and carried all chemicals identified in the pre-universe into the universe to undergo the Step 2 screening process (Chapter 3). With this improvement, EPA did not eliminate chemicals that could pose a public health risk through drinking water exposure but that are lacking either health or occurrence data, as was done in CCL 3. This modification to the CCL 3 development process resulted in the compilation of the most chemical- and data-rich CCL universe to date.

Section 2.4.2 Refining DTXSID Assignments

The CCL 5 data files identify chemicals by DTXSIDs, so that data entries associated with the occurrence or toxicity of a given chemical are assigned to the correct DTXSID. EPA further refined contaminant identifiers matched during the CCL 5 Pre-Universe development by grouping DTXSIDs for chemicals that would dissociate to the same compound in water (e.g., EPA assigned the same DTXSIDs

to lithium and lithium salts because they all form the lithium ion in water), rectifying incorrectly matched DTXSIDs during pre-universe development, and assigning unique DTXSIDs to contaminants without registered DTXSIDs from the CompTox Chemicals Dashboard.

EPA refined DTXSIDs manually when evidence suggested that certain chemicals should be grouped or distinguished from one another. EPA performed an extensive quality assurance (QA) of DTXSID assignments throughout the CCL 5 development process to catch incorrectly matched DTXSIDs (see Section 6.2 QA/QC of PCCL Development).

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 5 Universe, EPA considered all data relevant to lithium and lithium salts as one group and therefore grouped them under a single DTXSID.

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." Though this type of refinement may apply to many chemicals in the universe, it was not feasible for EPA to identify all instances, so efforts focused on identifying chemicals with ionized and/or salt forms (e.g., inorganic ions).

Another example of alterations to DTXSIDs was that EPA distinguished chemicals automatically matched to the same DTXSID which should have been considered unique substances for CCL 5 purposes. For example, entries described as "white phosphorous" and entries described as "phosphorous" were matched with the same DTXSIDs using the automated search tool in the CompTox Chemicals Dashboard. However, white phosphorous, an explosive compound used in munitions, has different chemical properties and toxicity than other forms of phosphorous that are ubiquitous in the environment. For the CCL 5 process, EPA matched data entries related to white phosphorous with a different DTXSID than is generally used to describe phosphorous compounds.

Another example of automatic matching of DTXSIDs using the CompTox Chemicals Dashboard that resulted in incorrect DTXSIDs was when the original source described the data entry with an abbreviation rather than the full chemical name. Some data entries labeled "DCPA" were matched to the DTXSID for dicalcium phosphate. Further investigation of the original source reports indicated that DCPA was meant to refer to dimethyl tetrachloroterephthalate, commonly known as "dacthal." In this case, EPA manually matched the DTXSID for dimethyl tetrachloroterephthalate to the DCPA entry.

EPA made additional efforts to assign correct DTXSIDs to data entries that did not have DTXSIDs. Some chemicals were not automatically linked to DTXSIDs because the synonym for the compound name was not included in the CompTox Chemicals Dashboard. An example of this is an entry for

"oestrogen," which is a British alternative spelling of estrogen. Other missing DTXSIDs could be attributed to misspellings or special characters in the original source files. Occasionally, the DTXSID for the entry had not been available at the time of pre-processing but was registered in the CompTox Chemicals Dashboard when developing the universe file. EPA manually matched the appropriate DTXSIDs in these cases.

If a DTXSID was not successfully matched to entries with missing DTXSIDs, EPA assigned a "NO_DTXSID" identifier to the entry. All chemicals with unique names were assigned a key of "NO_DTXSID" followed by a unique numeric string. Some manual correction of these NO_DTXSID assignments was needed to make sure entries describing the same chemical using different names were given the same NO_DTXSID assignment. For example, entries for "desulfinylfipronil amide" and "desulfinyl fipronil amide" were originally listed in the universe as distinct names because of how they were referenced in the primary data source, even though they clearly represent the same compound. Therefore, EPA assigned the unique numeric string of "437" to these chemicals, which resulted in a key of "NO_DTXSID437" for both entries.

Section 2.4.3 Additional Data Accessed via the CompTox Chemicals Dashboard

Due to advances in programming technologies and the enhanced capacity for systems to process large data sources, in this CCL cycle, EPA was able to download and append supplemental data from other relevant sources 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 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.

The CompTox Chemicals Dashboard was the only supplemental data source EPA relied on as a source of data elements for screening (Chapter 3), though only select data elements were used during this step. As described in Section 2.2.2, 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 N.

Section 2.4.4 Creating a Uniform Universe File

Several steps were required to ensure data elements from different sources 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 $(mg/kg/day)^{-1}$. To calculate distributions and compare the relative magnitude of data entries, all entries were also converted to a single numeric form. For example, EPA temporarily modified production data, which are reported as a range of pounds produced, to a single value (see Section 3.3.2 for additional details).

EPA also converted categorical cancer classifications to a numeric scheme (1, 2, or 3) according to the same methodology used for CCL 3 (USEPA, 2009b). In CCL 3, cancer classifications were distributed into numerical categories 1, 2, or 3 according to the designations provided in Table 3. 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, EPA retained the cancer classification from the source but created no new numeric data entry. For example, if a chemical has an EPA 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, EPA reverted back to the original cancer classification entries for the classification step of the CCL 5 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

With these modifications, EPA was able to compile and compare data from multiple sources for use during Steps 2 and 3 of the CCL 5 process (see Chapter 3 and Chapter 4). This is especially important for Step 2 of the CCL 5 process, which requires a uniform and comprehensive set of data elements to accurately screen the approximately 22,000 universe chemicals down to the PCCL 5.

Chapter 3 Screening Universe Chemicals to Select the PCCL

Section 3.1 Overview

The purpose of Step 2 of the CCL 5 development process was to screen universe chemicals for inclusion on the PCCL 5 for further evaluation. The PCCL 5 comprised the top scoring universe chemicals that were advanced for further evaluation and publicly nominated chemicals. Certain top scoring chemicals and publicly nominated chemicals were not included on the PCCL 5 because they had ongoing agency actions under the Regulatory Determination 4 (RD 4) process or were cancelled pesticides the agency determined did not warrant further evaluation, as described further in Section 3.7.

In this step, 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 5, EPA modified the CCL 3 screening process to accommodate new data types and sources that have since become available but maintained the same screening framework based on the chemical's toxicity and occurrence properties (USEPA, 2009b). Similar to CCL 3, the CCL 5 screening process requires limited 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 5 screening system included the following actions, described in detail in this chapter:

1. Determine the data elements to be used for screening.
2. Determine health screening levels and calculate screening hazard quotients.
3. Establish a scoring rubric for the relative point assignment across health effects and occurrence data elements.
4. Assign points to the data elements available for each chemical and calculate a screening score.
5. Select chemicals based on screening scores for inclusion on the PCCL 5.

The CCL 5 screening process relies on a transparent and reproducible scoring rubric and point-based screening system implemented using the R programming language (R Core Team, 2020). 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 for the CCL 5 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. 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 point assignment and screening processes are further described in Section 3.2 and Section 3.3.

[1] Though screening scores were used to prioritize chemicals for inclusion on the PCCL 5, these scores do not reflect EPA's regulatory priorities for particular chemicals. The screening points system was designed to reflect the likelihood of a chemical being listed on the CCL 5, but the screening score itself did not influence the decisions 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 5.

EPA applied the point-based screening system across all chemical contaminants in the CCL 5 Universe to determine which of the approximately 22,000 universe chemicals warranted further consideration during the time- and resource-intensive classification process (see Chapter 4). Section 3.5 discusses the use of the CCL 5 screening system for this purpose. Figure 4 illustrates the screening process, indicated by yellow.

EPA also evaluated publicly nominated chemicals for inclusion on the PCCL 5, as discussed in Section 3.6. Finally, EPA excluded from PCCL 5 chemicals that did not warrant further evaluation, as discussed in Section 3.7. Section 3.8 contains a summary of the PCCL 5.

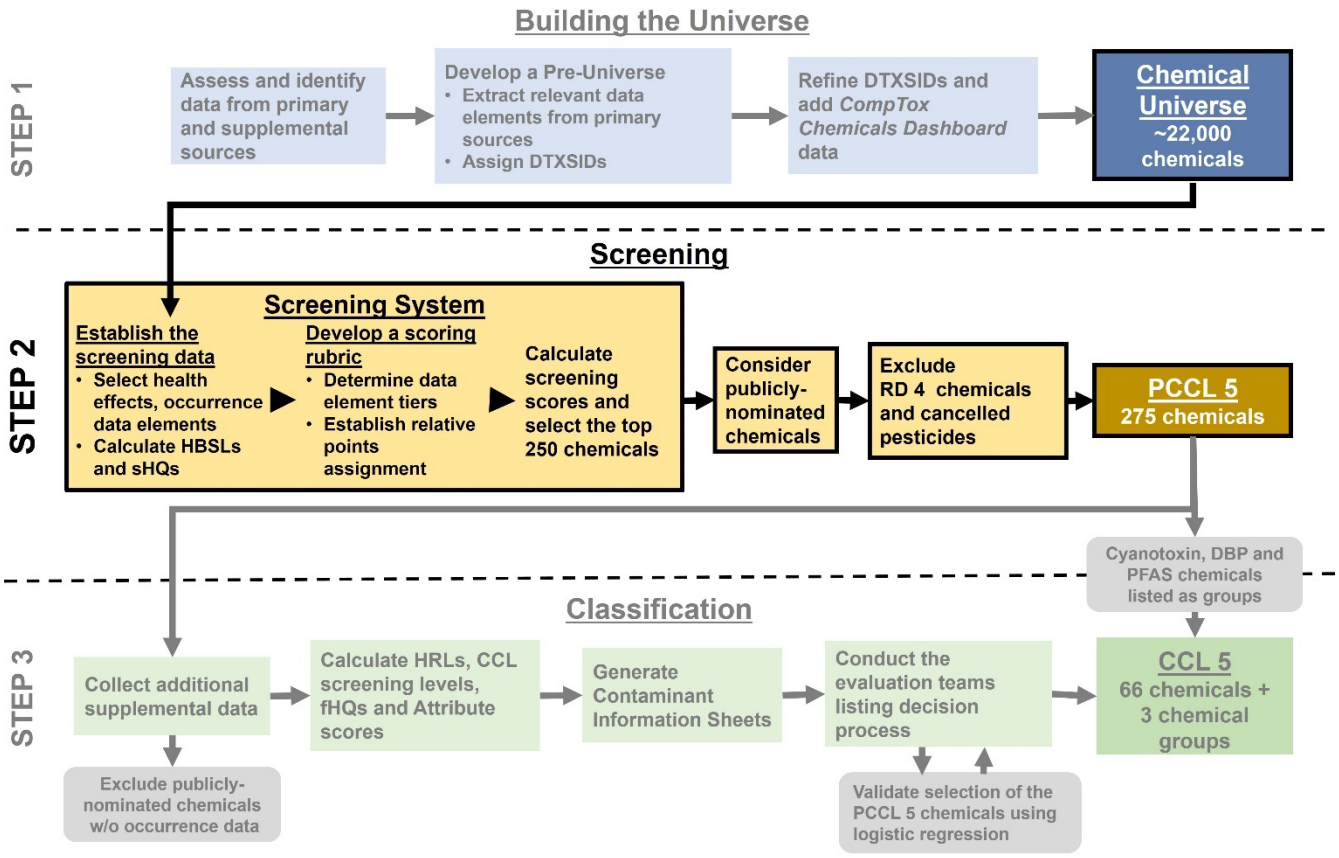


Figure 4. Development Framework Step 2 – Screening

Section 3.2 Establishing the Screening Data

Section 3.2.1 Incorporating Universe Data Elements

EPA designed the CCL 5 screening process to systematically consider the health effects and occurrence data from the CCL 5 Universe file and advance chemicals for further evaluation using consistent and transparent methods. During the CCL 5 Universe development process, EPA compiled 68 different data elements to consider for point assignment or as additional information for individual chemicals. Of these 68 data elements, EPA assigned points to 22 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 32 data elements not assigned points are included in Section N.5 of Appendix N.

Many of the data elements assigned points in CCL 5 are the same used in the CCL 3 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, Chemical Data Reporting (CDR) production volume and Toxics Release Inventory (TRI) chemical release data, and others.

There are also new data elements related to health and occurrence endpoints that EPA included in the CCL 5 screening process that were not available in a retrievable format or not used in previous CCL cycles. For example, EPA assigned health effects screening points to new assessment methods (sometimes referred to as NAMs) such as the percentage of active assays found in EPA's ToxCast *in vitro* screening. Similarly, EPA assigned occurrence points to lists of chemicals detected in human blood, serum, or urine as part of the CDC's NHANES biomonitoring program, in addition to points for contaminants with ambient and finished water percentage detection rates that were provided by nationally and non-nationally representative studies or surveys.

Table 4. Data Elements Assigned Points in the CCL 5 Screening System

Data Element	Description
Health Effects	
Acute benchmark	Short-term health-based concentration in water -- e.g., 10-day Health Advisories, acute or short-term guidance values from the Minnesota Department of Health, and acute Human Health Benchmark for Pesticides
Acute reference dose	Reference dose from a study with an acute exposure duration -- e.g., acute-duration MRLs, and acute population-adjusted doses from the Human Health Benchmarks for Pesticides
Androgen receptor chemicals	The list of chemicals identified by Kleinstreuer et al. (2017) and used to identify references with *in vitro* androgen receptor binding (downloaded from EPA's CompTox Chemicals Dashboard)
Cancer slope factor	Cancer risk per unit dose
Chronic benchmark	Chronic health-based concentration in water -- e.g., Lifetime Health Advisories, 10^{-6} cancer risk concentrations, chronic Human Health Benchmarks for Pesticides, and drinking water guidelines from WHO and Health Canada
Chronic LOAEL	Lowest Observed Adverse Effect Level from a study with a chronic exposure duration, a two-generation study, or a developmental toxicity study
Chronic NOAEL	No Observed Adverse Effect Level from a study with a chronic exposure duration, a two-generation study, or a developmental toxicity study
Developmental neurotoxins	This is a list of chemicals with data demonstrating effects on neurodevelopment, described in Table 1 of Mundy et al. (2015) (downloaded from EPA's CompTox Chemicals Dashboard)
Developmental neurotoxins (*in vivo)*	This is a list of chemicals documented to trigger developmental neurotoxicity (DNT) in at least two different laboratories, described in Table 5 of Aschner et al. (2017) (downloaded from EPA's CompTox Chemicals Dashboard)

Data Element	Description
Human neurotoxicants	A set of industrial chemicals that cause neurotoxicity identified by Grandjean and Landrigan (2006) (downloaded from EPA's CompTox Chemicals Dashboard)
LD_{50}	The lethal dose for 50% of the tested animals after a specified exposure duration
Mined literature for neurotoxins	List of chemicals associated with neurotoxicity compiled through automated literature mining of PubMed using Medical Subject Headings (MeSH) terms and associating these with single chemical substances (downloaded from EPA's CompTox Chemicals Dashboard)
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 from a PubMed search (downloaded from EPA's CompTox Chemicals Dashboard)
Reference dose	Reference dose from a study with a chronic exposure duration, a two-generation study, or a developmental toxicity study – e.g., chronic MRLs and chronic population-adjusted doses from Human Health Benchmarks for Pesticides
Subchronic benchmark	Benchmarks for a subchronic exposure duration.
Subchronic LOAEL	Lowest Observed Adverse Effect Level from a study with a subchronic exposure duration
Subchronic NOAEL	No Observed Adverse Effect Level from a study with a subchronic exposure duration
Subchronic reference dose	Reference dose from a study with a subchronic exposure duration -- e.g., intermediate-duration MRLs
TD_{50}	Dose associated with 50% of animals developing tumors, compiled by the Cancer Potency Data Bank
ToxCast assay percent active	Percent of active ToxCast *in vitro* assays tested (downloaded from EPA's CompTox Chemicals Dashboard)
Occurrence	
Biodegradation half-life – OPERA model	The predicted biodegradation half-life in days, according to the OPERA model (downloaded from EPA's CompTox Chemicals Dashboard)
Blood concentrations	90th percentile concentration in human blood, according to NHANES biomonitoring data
National ambient water detection rates	Detection rates in ambient water from nationally representative surveys – e.g., USGS Water Quality Portal National Ambient Water Quality Assessment (NAWQA)
National finished water detection rates	Detection rates in finished water from nationally representative monitoring programs – (e.g., UCMR 1-4) and National Inorganics and Radionuclides Survey (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)

Data Element	Description
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 EPA's Chemical Data Reporting (CDR)
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[1] 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] EPA's method for assigning maximum concentration values to non-detected chemicals in the screening step of CCL 5 is described in Chapter 2 and Appendix N.

Some data elements in the universe file were not assigned points for CCL 5 screening purposes. In general, 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 a large number of chemicals.
- Data element was not considered highly relevant to hazards associated with drinking water.
- Data element required chemical-specific data manipulation (e.g., unit conversions requiring chemical 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 N.5 of Appendix N 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 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 EPA toxicity values like oral RfDs (reported in units of mg/kg/day). However, EPA did include MADLs as a supplementary source of health effects information on the Chemical Information Sheets (CISs; see Chapter 4).

Furthermore, physical and chemical properties estimated by the EPA QSAR models TEST and OPERA, as well as toxicity values based on inhalation data, were not considered for point assignments. Though these data provide context to occurrence or health effects information, they are not considered directly relevant to potential hazards due to drinking water exposure. For example, 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 finished 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 along with the others for point assignment. Though data elements meeting the exclusion criteria described above were not assigned points in the CCL 5 screening system, these data elements were considered supplementary material and, along with MADLs, were provided to chemical evaluators during the classification process (see Chapter 4).

For certain data elements, points were not assigned because EPA decided to assign points to another equivalent data element or another data element describing similar data. In the cancer classifications, 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, EPA prevented chemicals from multiple sets of points for the same information.

For occurrence monitoring data in finished and ambient waters, 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.

Section 3.2.2 Calculating Screening Hazard Quotients (sHQs)

During the CCL 3 process, 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, 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 not expected to result in adverse health outcomes over a lifetime of exposure, and the 90^{th} percentile concentration of the chemical in ambient or finished water (USEPA, 2009c).

For CCL 5 chemicals that had the necessary health effects and occurrence information, EPA calculated a "screening hazard quotient" (sHQ), which 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. EPA used the sHQ during the screening phase of CCL 5 in the same way it used the HRL/concentration ratio during the classification phase of CCL 3.

To calculate the sHQ, EPA derived an element called the health screening level (HSL) to compare against the drinking water occurrence data for each chemical to inform whether a chemical has the potential to occur in finished drinking water at concentrations relevant to adverse health effects. A CCL 5 HSL is a calculated concentration of a chemical in drinking water derived from chronic toxicity values identified from primary data sources. Note that these HSLs are different metrics than the CCL 5 HRLs and CCL screening levels introduced in Chapter 4. HSLs were used in CCL 5 for initial coarse screening purposes only and were replaced by HRLs and CCL screening levels, which underwent manual review and expert discussion during their derivation, for classification.

HSLs were calculated according to the equations in Table 5, assuming a drinking water intake (DWI) of 33.8 ml/kg-day and 20% relative source contribution (RSC) (USEPA, 2019; USEPA, 2000b). When toxicity values such as NOAELs and LOAELs were available, the same default uncertainty factors (UFs) were applied as were used in CCL 3 (1,000x for NOAELs and 3,000x for LOAELs). If multiple types of toxicity values were available for a chemical, EPA calculated corresponding HSLs using each type of toxicity value and the most health protective HSL was used to compare against finished water concentrations. In CCL 5, EPA compiled all HSLs calculated for each chemical and denoted the most health protective HSL along with the corresponding source and data element information for future use.

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	$HSL = \frac{\left(\frac{CSF}{1x10^{-6}}\right)}{DWI}$
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, EPA calculated the screening hazard quotient for a chemical by dividing the maximum finished water concentration by the HSL (Equation 1). EPA chose maximum concentrations of a chemical in finished water for use only 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 295 of the universe chemicals. The logarithmic distribution of sHQs calculated for the screening step of CCL 5 is shown in Figure 5. It should be noted that the sHQ differs from the final hazard quotient (fHQ) calculated in the classification step of the CCL 5 process (see Chapter 4).

EPA incorporated the sHQ as a data element in the universe file and assigned points in the same way as other CCL 5 data elements. The process for distributing and applying screening points to each type of data element is described in Section 3.3.

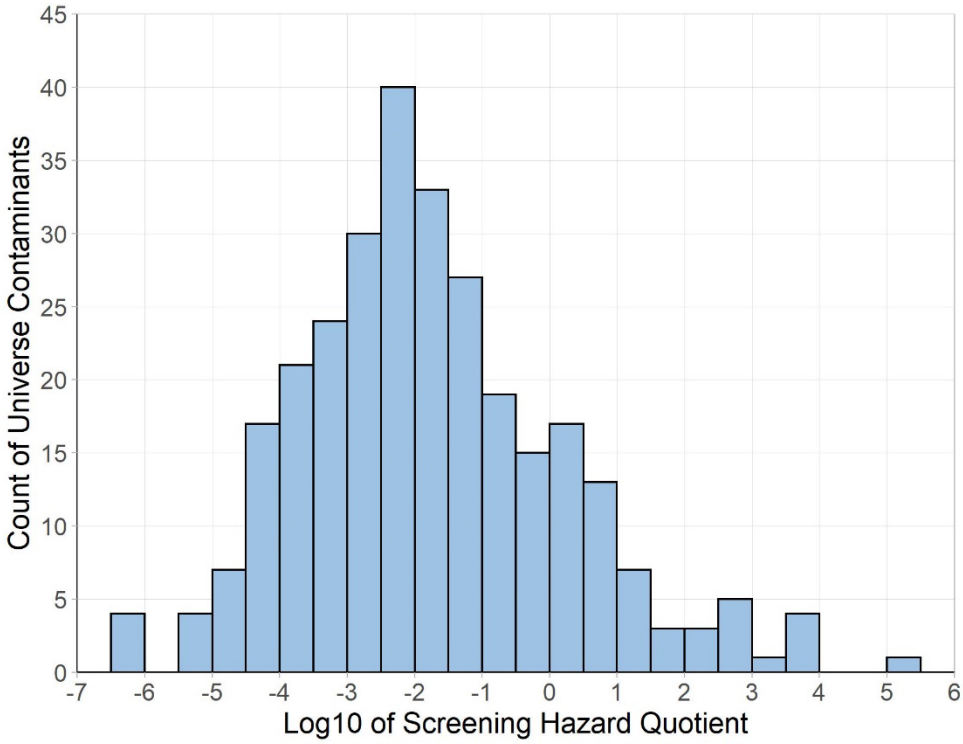


Figure 5. Empirical Histogram of Log Transformed Screening Hazard Quotients Calculated for the Screening Step

Section 3.3 Developing a Scoring Rubric

Section 3.3.1 Determining Screening Tiers

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 indicating a relatively indirect 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 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 screening hazard quotient (sHQ; see Section 3.2), 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 health screening level 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 decisions. Similarly, the lowest health tier (Health Tier 5) includes the percent of *in vitro* active results from EPA ToxCast screening and LD_{50}. These data elements may give an indication of relative toxicity but do not provide the information needed to derive toxicity values such as RfD or CSF, which are necessary for assessing 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, mined literature for neurotoxins, human neurotoxicants, developmental neurotoxins, developmental neurotoxins (*in vivo*), androgen receptor chemicals
Tier 5	TD_{50}, LD_{50}, ToxCast assay percent active, Number of PubMed articles
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 and of 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.

Section 3.3.2 Determining Relative Point Assignments Within Each Screening Tier

EPA analyzed the chemical-specific data for each data element and plotted distributions to ensure the data contained no obvious irregularities. 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). An example of the distribution of CSFs with the calculated quantiles represented with red 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 others.

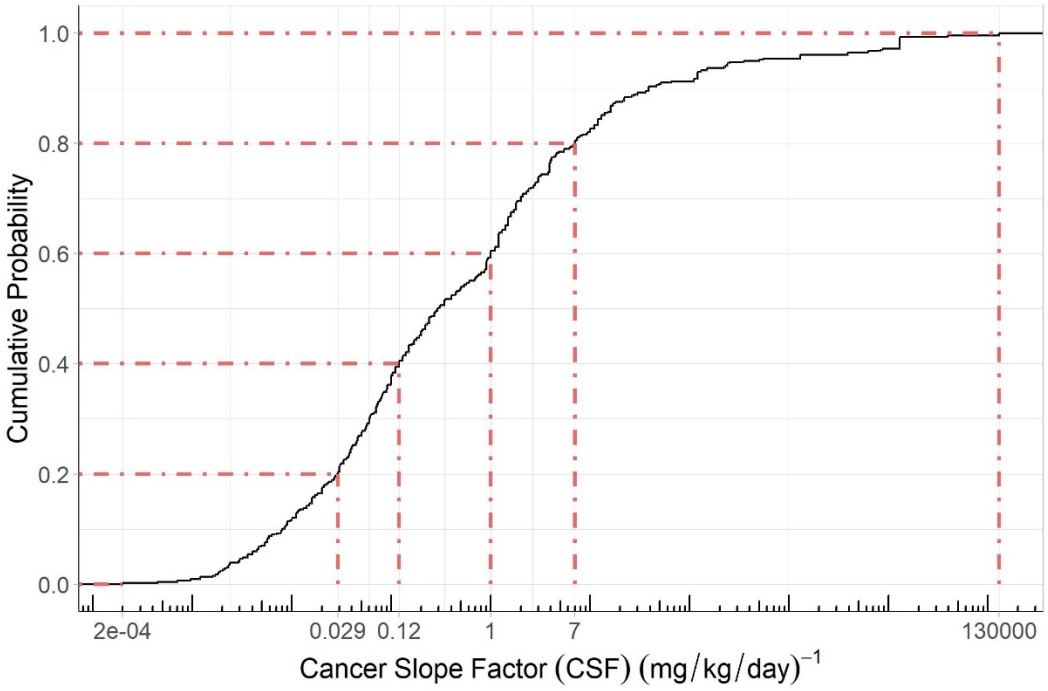


Figure 6. Empirical Distribution of Cancer Slope Factors in the CCL 5 Universe

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.

EPA retained all data associated with chemicals that are regulated with NPDWRs, i.e., when establishing the relative point assignments for each data element. Though points were assigned to regulated chemicals, they were not considered further in the CCL 5 process.

EPA evaluated the distribution of calculated sHQs for point assignments. The distribution of sHQ values is highly skewed with a median value of 0.01. 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. Similarly, an sHQ greater than 1 indicates the finished water concentration exceeds the HSL and, therefore, the chemical

may pose a greater potential hazard for public health. However, an sHQ less than 1 does not necessarily indicate a harmful effect is unlikely to occur.

Therefore, instead of limiting sHQ point assignments to chemicals with sHQs of 1 or higher, EPA assigned points to sHQ values that are equal to or exceed the median (0.01) or the top 50% of the sHQ values. EPA divided sHQ values equal to or greater than 0.01 into five categories based on orders of magnitude, or 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, EPA used this protocol if a data element value fell on a category boundary.

EPA evaluated the distributions of detection rates in ambient and finished water for point assignments. The distributions are highly skewed, likely 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, EPA developed points categories based on 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).

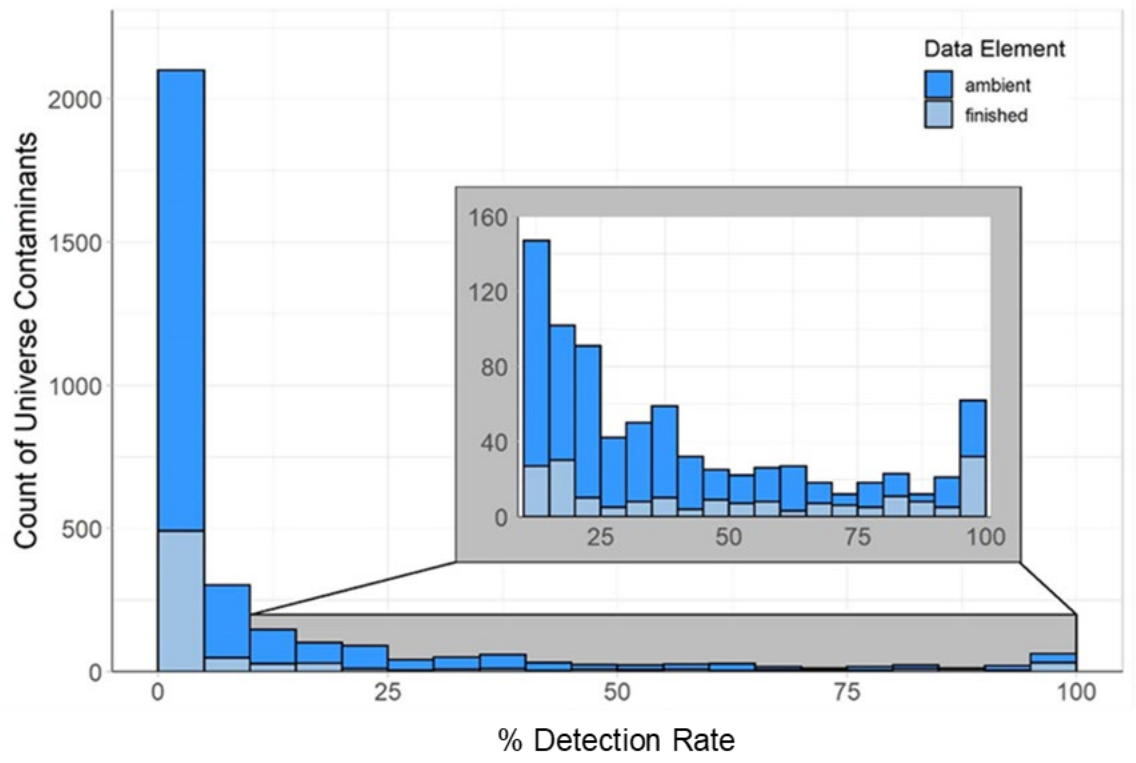


Figure 7. Ambient and Finished Water Detection Rates for the CCL 5 Universe Chemicals

EPA evaluated the distribution of chemical production volume data from the 2016 Chemical Data Reporting (CDR) for point assignments. Production volume data required special data processing steps to be incorporated into the CCL 5 screening rubric because 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 rather than as a numeric sum or singular value. For the screening step only, 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.

EPA's method of converting production volumes to single numerical values was driven by how the data was originally reported by CDR. 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 was the appropriate approach.

The lowest two production volumes available for chemicals in the CCL 5 Universe were "< 25000 lbs" (i.e., less than 25,000 lbs.) and "25000 - 100000 lbs" (i.e., between 25,000 lbs. and 100,000 lbs.). EPA determined that these two production volumes should not be assigned the same numerical value (25,000 lbs.) because CDR reports the production volumes as two distinct categories. Therefore, EPA temporarily substituted "12,500 lbs" for "<25000 lbs" (i.e., ½ of 25,000 lbs.) and "25,000 lbs" for "25000 - 100000 lbs" (i.e., the minimum value of the range). For all production volumes expressed as a range, the minimum value of the range was substituted for calculating point assignments.

Similarly, the highest two production volume categories available for chemicals in the CCL 5 Universe were "> 200,000,000,000 lbs" (i.e., greater than 200 billion lbs.) and "190,000,000,000 – 200,000,000,000 lbs" (i.e., between 190 billion lbs. and 200 billion lbs.). For CCL 5, EPA determined that these two production volume categories should not be assigned the same numerical value (200 billion lbs.) because CDR reports the production volumes as two distinct categories. Therefore, EPA substituted "190,000,000,000 lbs" for the "190,000,000,000 – 200,000,000,000 lbs" category and "200,000,000,000 lbs" to the "> 200,000,000,000 lbs" category specifically for calculating quantiles and relative point assignments. See Table 8 for points categories and point assignments for the chemical production volume data element.

EPA analyzed the distribution of predicted biodegradation half-life in the OPERA model from the CompTox Chemistry Dashboard (see Appendix N for additional information) for point assignments to incorporate physio-chemical considerations into the screening system. EPA established one point category for this data element (Table 8). For chemicals with biodegredation half-life prediction values shorter than 3.5 days, or below the 20th percentile, 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.

EPA analyzed the distribution of the number of PubMed articles data element provided by the CompTox Chemistry Dashboard (see Appendix N 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 5 Universe. 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.

The point categories determined in this stage are similar to those used in the CCL 3 criteria to screen the health effects and occurrence data for universe chemicals (USEPA, 2009b). 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.

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	3.00E+03	1.00E-01	1.00E-01	3.00E-02	3.00E-02	9.00E-03	9.00E-03	1.00E-03	1.00E-03	7.00E-10
Cancer Slope Factors	values - $(mg/kg/day)^{-1}$	2.00E-04	2.90E-02	2.90E-02	1.20E-01	1.20E-01	1	1	7	7	1.30E+05
Chronic Benchmarks	values - mg/L	2.50E+02	2.00E-01	2.00E-01	3.20E-02	3.20E-02	5.00E-03	5.00E-03	5.00E-04	5.00E-04	5.00E-12
Health Effects Tier 2	Points Assigned		150		250		350		450		550
Chronic NOAELs	values - mg/kg/day	4500	77.34	77.34	25	25	10	10	2.5	2.5	0.037
Chronic LOAELs	values - mg/kg/day	11270	257	257	100	100	33.9	33.9	8.7	8.7	0.002
Health Effects Tier 3	Points Assigned		100		200		300		400		500
Numeric Cancer Classifications	See Table 4	NA		NA			3		2		1
Subchronic RfDs	values - mg/kg/day	3.00E+03	6.00E-01	6.00E-01	1.00E-01	1.00E-01	1.00E-02	1.00E-02	2.00E-03	2.00E-03	5.00E-06
Subchronic Benchmarks	values - mg/L	5.00E+01	4.20E-01	4.20E-01	1.00E-01	1.00E-01	3.00E-02	3.00E-02	7.00E-03	7.00E-03	2.00E-07
Health Effects Tier 4	Points Assigned		50		100		150		200		250
Acute Benchmarks	values - mg/L	1.00E-07	3.00E-02	3.00E-02	2.00E-01	2.00E-01	1.00E+00	1.00E+00	4.00E+00	4.00E+00	1.00E+02
Acute RfDs	values - mg/kg/day	6.3	0.58	0.58	0.15	0.15	0.05	0.05	0.01	0.01	0.00002

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
Subchronic LOAELs	values - mg/kg/day	10635	263	263	80	80	30	30	6.7	6.7	0.0025
Subchronic NOAELs	values - mg/kg/day	5414	79	79	21.2	21.2	7.1	7.1	2.2	2.2	0.004
MRDDs	values - mg/kg/day	9.99E+02	2.50E+01	2.50E+01	6.67	6.67	2	2	3.33E-01	3.33E-01	1.00E-05
Mined Literature for Neurotoxins	presence on list	Yes		NA		NA		NA		NA	
Human Neurotoxicants	presence on list	Yes		NA		NA		NA		NA	
Developmental Neurotoxins	presence on list	Yes		NA		NA		NA		NA	
Developmental Neurotoxins (*in vivo*)	presence on list	Yes		NA		NA		NA		NA	
Androgen Receptor Chemicals	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	1.11E+08	1.56E+03	1.56E+03	3.60E+02	3.60E+02	9.72E+01	9.72E+01	1.92E+01	1.92E+01	1.21E-05
LD_{50}s	values -- mg/kg	4.39E+06	4.16E+03	4.16E+03	1.70E+03	1.70E+03	6.15E+02	6.15E+02	1.40E+02	1.40E+02	3.00E-04
ToxCast Assay Percent Active	values -- percent	>0	0.8	0.8	2.06	2.06	4.75	4.75	15.164	15.164	73.83
PubMed Articles	number of articles	50^{th}-90^{th} percentiles (81-3482 articles)		90^{th} percentile (>3482 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 water detection rates	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
Occurrence Tier 5	Points Assigned	50		100		150		200		250	
Chemical release information	values - lbs/year	>0	1.51E+01	1.51E+01	2.84E+03	2.84E+03	5.02E+04	5.02E+04	6.94E+05	6.94E+05	7.30E+08
Estimated Pesticide Application Rate	values - kg/year	>0	1.73E+02	1.73E+02	9.08E+03	9.08E+03	4.69E+04	4.69E+04	2.68E+05	2.68E+05	1.32E+08
Chemical production information	values- lbs/year	1.25E+04	2.50E+04	2.50E+04	1.00E+05	1.00E+05	1.00E+06	1.00E+06	1.00E+07	1.00E+07	2.00E+11
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 (<3.524106)									

[1] If a data element value falls on a category boundary, screening points are assigned according to the higher category.
[2] EPA assigned maximum concentration values to non-detected chemicals in the screening step of CCL 5. See Chapter 2 and Appendix N for additional information.

If multiple data entries for a single data element exist for a given chemical (e.g., a chemical has two different RfDs or two different non-nationally representative finished water detection rates available from different data sources), EPA assigned points using the data entry with the value that represents the maximum possible exposure or toxicity. Examples 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, EPA chose these values for each data element for several reasons:

- This is the most conservative and health-protective approach.
- With over 20,000 chemicals in the universe, it is not feasible to conduct a systematic review 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.

For example, when assigning occurrence screening points, EPA used the partial occurrence dataset from the UCMR 4 prior to the completion of all sampling and reporting activities. It is important to use more recent occurrence data in the screening process to ensure that new and potentially relevant information is not disregarded and that potentially hazardous chemicals are not discounted before the two teams of chemical evaluators can further investigate and review each chemical during the classification step (Chapter 4).

Section 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 occurrence, the maximum possible health effects and occurrence screening points that a chemical would accumulate were 6,200 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", is 14,050. The maximum screening score that an unregulated chemical in the CCL 5 Universe accumulated was 9,050 points. A histogram of screening scores for all chemicals in the CCL 5 Universe is shown in Figure 8.

EPA examined final point assignments and screening scores to ensure it considers 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 allows inclusion of a chemical with limited health effects data, but high occurrence, on the PCCL 5.

Propazine, for example, earned only 1,300 of 6,200 possible points for health effects data but was included in the PCCL because it earned a significant number of points (4,000 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 in the PCCL. For example, thiram earned only 600 points from occurrence data but 3,020 points from health effects data.

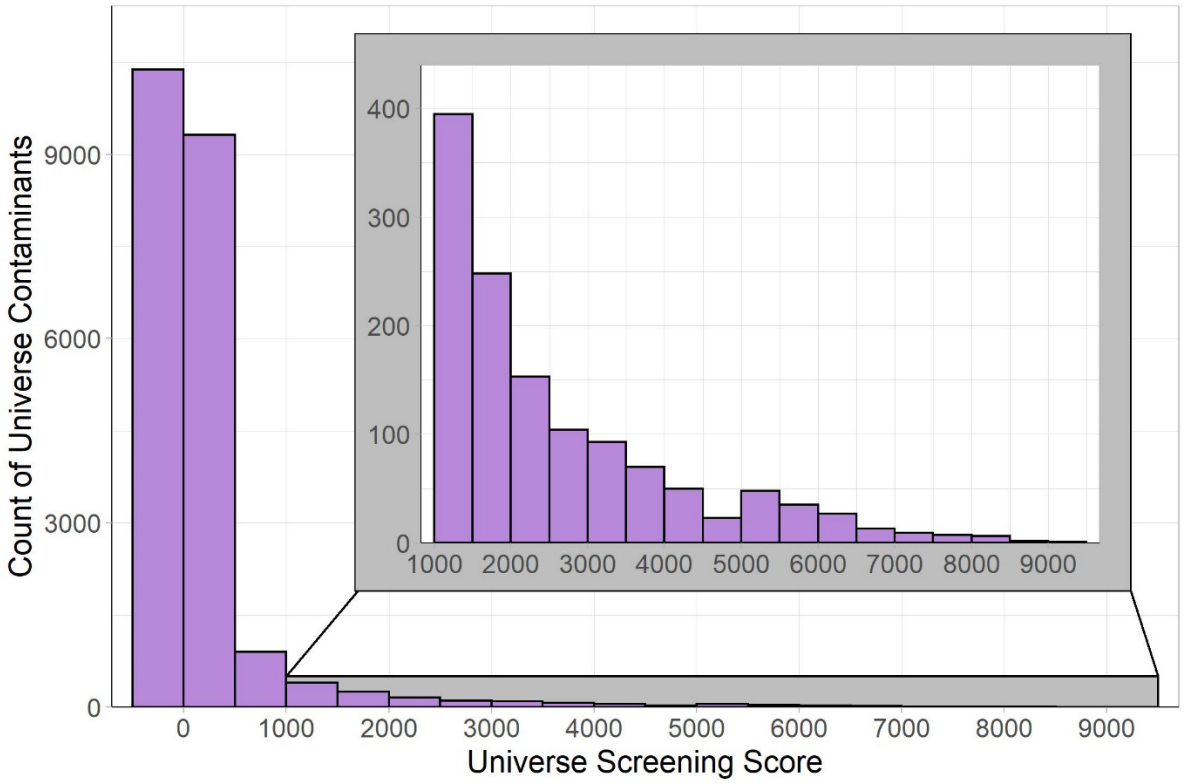


Figure 8. Total Screening Scores for the CCL 5 Universe Chemicals

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 (blue), 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 (red).

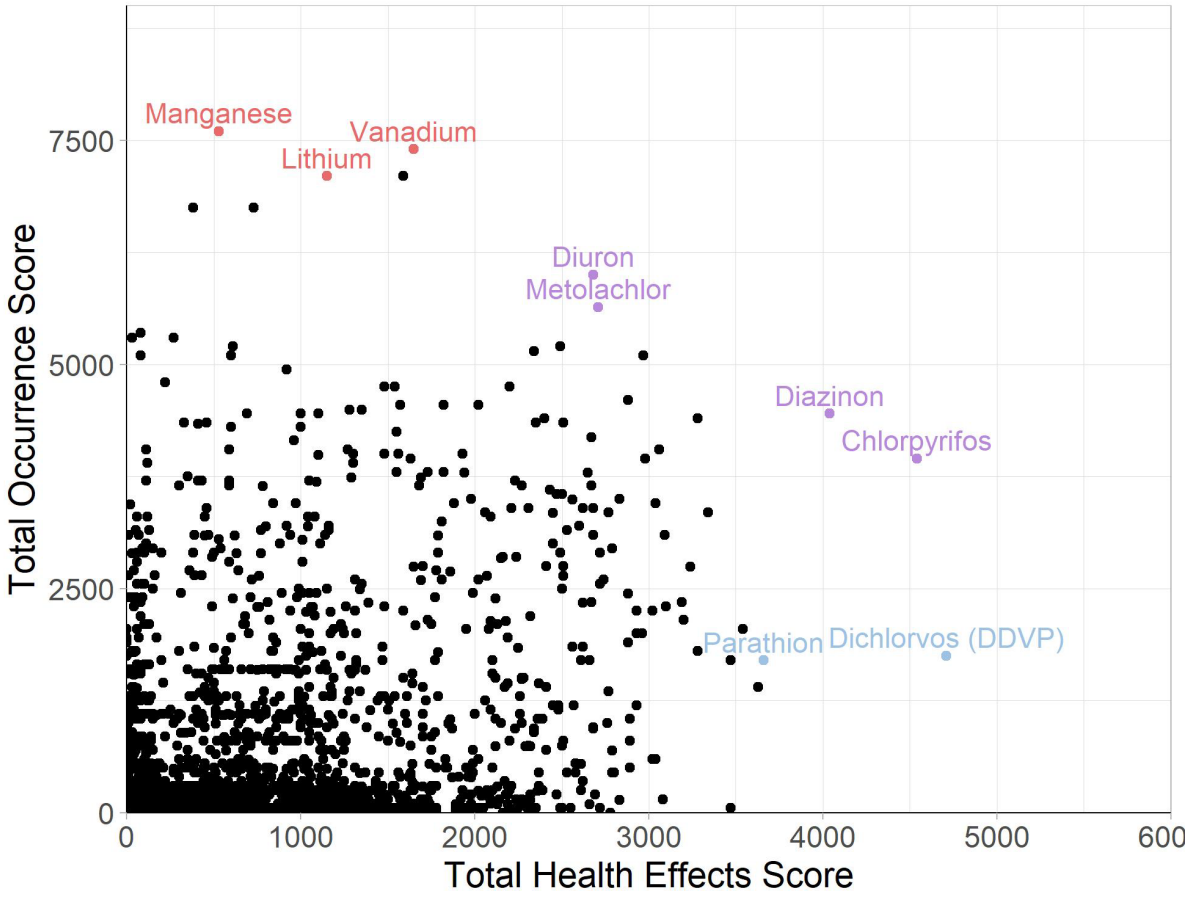


Figure 9. Health Effects and Occurrence Scores for the CCL 5 Universe Chemicals

Though screening scores were used to prioritize chemicals for inclusion on the PCCL, these scores do not reflect EPA's regulatory priorities for particular chemicals. The screening points system was designed to reflect the likelihood of a chemical being listed on the CCL 5, but the screening score itself did not influence the decisions 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 5. The points system and outcomes are solely a tool for CCL 5 screening purposes and statistical analyses (see Section 4.5) and should not be used or described in other contexts.

Section 3.5 Using the CCL 5 Screening System

The CCL 5 process has identified the broadest and most comprehensive universe of health and occurrence information to date. EPA used the screening system to take advantage of this information and identified the 250 top-scoring chemicals, with, theoretically, the potential for the greatest public health concern, for inclusion in the PCCL 5. Selecting a finite number of chemicals also allowed EPA to consider the resource requirements of compiling additional information, developing CISs and conducting evaluation teams' review during the classification step (see Chapter 4).

These 250 highest-scoring chemicals represent approximately the top 1% of chemicals in the CCL 5 Universe. By limiting evaluations to the top 250, all chemicals scoring at or above 3,320 points were

advanced for further consideration for the CCL 5. The highest score accumulated by a chemical in the CCL 5 Universe was 9,050, as mentioned above. Note that three chemicals (2,4-Dinitrophenol, Phosmet, and 4-Androstene-3,17-dione) had the same screening score of 3,320; therefore, a total of 252 chemicals were elevated for further consideration and potential inclusion on the PCCL 5. In this document, these 252 chemicals are referred to as the "top 250".

Section 3.6 Consideration of Publicly Nominated Chemicals

Section 3.6.1 Soliciting Public Nominations

On October 5, 2018, EPA published a request for public nominations of unregulated chemical and microbial contaminants to be considered for possible inclusion on the CCL 5 (83 FR 50364, USEPA, 2018). In accordance with the SDWA, which directs EPA to consider health effects and occurrence information when deciding whether to place contaminants on the CCL, EPA asked that nominations include responses to the following questions:

1. What is the contaminant's name, CAS registry number, and/or common synonym (if applicable)? Please do not nominate a contaminant that is already subject to a national primary drinking water regulation.
2. What are the data that you believe support the conclusion that the 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 or measured occurrence in sources of drinking water or provide information that shows the contaminant is released in the environment or is manufactured in large quantities and has a potential for contaminating sources of drinking water. Please provide the source of this information with complete citations for published information (i.e., author(s), title, journal, and date) or contact information for the primary investigator.
3. What are the data that you believe support the conclusion that the contaminant may require regulation? 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, or others). Please provide the source of this information with complete citations for published information (i.e., author(s), title, journal, and date) or contact information for the primary investigator.

Nominations were received via the EPA docket (Docket ID No. EPA-HQ-OW-2018-0594) on the Federal eRulemaking Portal (http://www.regulations.gov) and were also accepted by mail or hand delivery. EPA compiled and reviewed the information to identify the contaminants nominated and any supporting data submitted that could supplement data gathered by EPA to inform selection of the CCL 5.

Section 3.6.2 Summary of Chemical Nominations

EPA received public nominations for 73 unique chemicals, including chemicals used in commerce, pesticides, disinfection byproducts, pharmaceuticals, naturally occurring elements, and biological toxins. Chemicals nominated for consideration for the CCL 5 are shown in Appendix C.

In addition to individually nominated chemicals, EPA also received 7 nominations for chemical groups, including brominated haloacetic acids known as "HAA6Br," cyanotoxins, GenX chemicals

(hexafluoropropylene oxide dimer acid (HFPO-DA) and its ammonium salt), all the perfluoroalkyl and polyfluoroalkyl substances (PFAS) approved by the EPA Method 537.1, all PFAS more broadly, and the top 200 prescribed drugs of 2016 and their parents and metabolites. A public commenter also proposed that all CCL 4 contaminants be retained on the CCL 5, though EPA ultimately chose not to adopt this specific nomination proposal and, instead, subjected CCL 4 chemicals to the established screening and classification steps for listing on CCL 5; a summary table of CCL 4 chemicals that did not qualify for CCL 5 is provided in Appendix O. Perfluorononanoic acid (PFNA), perfluorooctane sulfonic acid (PFOS), and perfluorooctanoic acid (PFOA) received the most chemical nominations, each nominated by three organizations or individuals. Publicly nominated microbes are discussed in the Technical Support Document for the Final Fifth Candidate List (CCL 5) – Microbial Contaminants (USEPA, 2022a).

All public nominations can be viewed in the EPA docket (Docket ID No. EPA-HQ-OW-2018-0594) at https://www.regulations.gov.

Section 3.6.3 Consideration of Publicly Nominated Chemicals for the PCCL

EPA reviewed the publicly nominated chemical contaminants and identified the chemicals that were not already included in the top 250 (see Section 3.5) and not subject to proposed or promulgated NPDWRs and therefore needed to be considered for further analysis. Though nominated, EPA has since announced Final Regulatory Determinations for PFOA and PFOS (86 FR 12272, USEPA, 2021b) and decided not to consider these chemicals under CCL 5. EPA also did not add publicly nominated groups like "the top 200 most prescribed drugs in 2016 and their parents and metabolites" to the PCCL 5 because health effects and occurrence data must be linked to specific individual contaminants to be evaluated. However, individual chemicals in a nominated group were listed on the PCCL 5 if they were also nominated individually (e.g., morphine, part of "the top 200 most prescribed drugs in 2016") or if they were part of the CCL 5 Universe and included in the top 250 chemicals (e.g., 17-alpha ethynyl estradiol, part of "the top 200 most prescribed drugs in 2016"), as described in Section 3.5.

Of the 73 publicly nominated chemicals, 19 were already part of the CCL 5 Universe and included in the top 250 (see Section 3.5). Two nominated chemicals—ammonium perfluoro-2-methyl-3-oxahexanoate and perfluoro-2-methyl-3-oxahexanoic acid—are the ammonium salt and acid, respectively, of "Gen-X" (hexafluoropropylene oxide dimer acid, HFPO-DA). Both dissociate to form the same ion in water. Therefore, EPA included only the ammonium salt on the PCCL 5. EPA added the remaining 53 publicly nominated chemicals to the 252 highest-scoring chemicals to arrive at a total of 305 chemicals on the PCCL 5 (see Appendix D). Certain chemicals are then excluded from the PCCL 5, as described in Section 3.7.

For publicly nominated chemicals not in the CCL 5 Universe and added to the PCCL 5, further data collection was required so they could be evaluated for listing on the CCL 5. EPA assessed data sources cited with public nominations using the assessment factors described in Section 2.2 and extracted health effects and occurrence data from sources that were relevant, complete, and not redundant. Supplemental data sources were then used to fill any data gaps for particular chemical contaminants during Step 3 of the CCL 5 process (see Chapter 4). EPA also conducted literature searches to identify additional health effects and occurrence data, as described in Section 4.2. A complete list of supplemental sources can be found in Appendix B.

Thirteen of the publicly nominated chemicals did not have available water occurrence data, even after a literature search was conducted, and therefore were not evaluated by chemical evaluators for listing on the CCL 5; these are described further in Section 4.2.1.1.

Section 3.7 Chemicals Excluded from the PCCL

Section 3.7.1 Regulatory Determination

In March 2021, under the fourth Regulatory Determination (RD 4) process, EPA made final regulatory determinations for eight chemicals: PFOS; PFOA; 1,1-dichloroethane; acetochlor; methyl bromide (bromomethane); metolachlor; nitrobenzene; and RDX (86 FR 12272, USEPA, 2021b). EPA also made a preliminary positive determination on strontium under the third Regulatory Determination (RD 3) process (79 FR 62715, USEPA, 2014). Therefore, EPA excluded these nine chemicals from the PCCL 5.

Section 3.7.2 Canceled Pesticides

The PCCL 5 contained 26 canceled pesticides. To exclude any canceled pesticides that are not persistent in the environment, EPA evaluated the persistence and occurrence of these canceled pesticides (e.g., biodegradation half-life, 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 EPA's 2012 TSCA Work Plan Chemicals: Methods document (USEPA, 2012), according to the pesticides' biodegradation half-life in air, water, soil and sediment.
2. End-of-use dates were used to determine when the canceled pesticides were last allowed to be used in the environment.
3. Occurrence monitoring data collected after the end-of-use dates were used to determine if a canceled pesticide had any detects and/or data spikes that would pose a public health concern.
4. Canceled pesticides that were assigned a persistence score of 3 were included in the PCCL 5.
5. Canceled pesticides that were assigned a score of 1 or 2 but had detects in drinking water were included in the PCCL 5, while those that had no or very few detects in ambient water were excluded from the PCCL 5.

Step 1. Canceled pesticides were assigned a persistence score based on their biodegradation half-life in the environment (see Table 9). If its biodegradation half-life was greater than 6 months, a canceled pesticide was assigned a persistence score of 3. If its half-life was greater than or equal to 2 months, a canceled pesticide was assigned a persistence score of 2. If its half-life was less than 2 months, then a canceled pesticide was assigned a persistence score of 1.

Table 9. Summary of Persistence Ranking Score

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

Step 2. End-of-use dates were used to determine when the canceled pesticides were last allowed to be used in the environment. For CCL 5, EPA did not use pesticide cancellation dates to assess their

persistence in the environment because, when a pesticide registration is canceled, EPA determines whether there is any significant potential risk associated with the use of the pesticide. If there is such concern, EPA generally makes a case-by-case determination about allowing continued distribution, sale, or use of existing stocks of the canceled pesticide (56 FR 29362, USEPA, 1991).

Step 3. EPA compared dates of occurrence monitoring data to end-of-use dates to determine if a canceled pesticide continued to have any detects and/or data spikes that would pose a public health concern. The data sources used for monitoring-include NAQWA, UCMR, UCM, NWIS, and SURF.

Step 4. EPA included canceled pesticides that were assigned a persistence score of 3 and showed detects in drinking water and/or ambient water in the PCCL 5.

Step 5. EPA evaluated canceled pesticides that received a persistence score of 1 or 2. If these canceled pesticides had detects in drinking water, it was included in the PCCL 5. If it had no or few detects in ambient water, it was excluded from the PCCL 5.

EPA assessed a total of 26 canceled pesticides for persistence. Four pesticides—dieldrin, aldrin, chlordecone (kepone), and ethion—were assigned a persistence score of 3 and showed detects in finished and/or ambient water and were included in the PCCL 5. Alpha-hexachlorocyclohexane was also included in the PCCL 5 because it showed drinking water detects in the UCMR 4 occurrence data (collected 2018-2019). This chemical is an organochloride, which is one of the isomers of hexachlorocyclohexane and a byproduct of the production of the canceled insecticide lindane.

The remaining 21 pesticides were excluded from the PCCL 5 because they were assigned a score of 1 or 2 and showed no detects in finished water or no to few 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 10 shows the canceled pesticides EPA assessed and ranked.

Table 10. Canceled Pesticides Assessed for Exclusion from PCCL 5

Chemical Name*	CASRN	DTXSID	Half-Life (days)	TSCA Persistence Score	Last End Use Date	Occurrence Monitoring Data
2,4,6-Trichlorophenol	88-06-2	DTXSID 5021386	9	1	10/10/1989	UCMR (2001-2003); NAWQA (2011)
3-Hydroxycarbofuran	16655-82-6	DTXSID 2037506	4	1	12/31/2009	NAWQA (2013-2017); SURF (1991-2011)
Aldrin*	309-00-2	DTXSID 8020040	329	3	5/15/1987	UCM (1993-1997); NAWQA (2002-2010)
alpha-Hexachlorocyclohexane*	319-84-6	DTXSID 2020684	19	1	10/1/2009	UCMR (2018-2019); NAWQA (2013-2017)
Azinphos-methyl	86-50-0	DTXSID 3020122	95	2	12/13/2013	NAWQA (2013-2017); SURF (1991-2017)
Benomyl	17804-35-2	DTXSID 5023900	5	1	12/31/2003	NAWQA (2015-2016); SURF (1992-2016)
Chlordecone (Kepone)*	143-50-0	DTXSID 1020770	914	3	4/4/1977	NWIS (2015)
Cyanazine	21725-46-2	DTXSID 1023990	5	1	12/31/2002	NAWQA (2013-2017); SURF (1993-2017)

Chemical Name*	CASRN	DTXSID	Half-Life (days)	TSCA Persistence Score	Last End Use Date	Occurrence Monitoring Data
Dacthal	1861-32-1	DTXSID 0024000	6	1	7/27/2005	NAWQA (2013-2017); SURF (1992-2017)
Dicofol	115-32-2	DTXSID 4020450	21	1	1/31/2013	SURF (2004-2017)
Dieldrin*	60-57-1	DTXSID 9020453	333	3	5/15/1987	UCM (1993-1997); NAWQA (2013-2017)
Disulfoton	298-040-4	DTXSID 0022018	143	2	12/31/2014	NAWQA (2013-2017); SURF (1991-2017)
Endosulfan	115-29-7	DTXSID 1020560	16	1	7/31/2016	SURF (1991-2017)
Endosulfan sulfate	1031-07-8	DTXSID 3037541	16	1	7/31/2016	NAWQA (2014-2017); SURF (1990-2017)
Ethion*	563-12-2	DTXSID 2024086	478	3	12/31/2004	NAWQA (2014-2017); SURF (1991-2002; 2007)
Fenamiphos	22224-92-6	DTXSID 3024102	5	1	10/6/2017	NAWQA (2013-2017)
Flusilazole	85509-19-9	DTXSID 3024235	201	2	12/31/2010	NAWQA (2013-2015)
Isofenphos	25311-71-1	DTXSID 8032417	3	1	1/26/2007	NAWQA (2014-2017); SURF (1991-1992; 2007)
Methamidophos	10265-92-6	DTXSID 6024177	5	1	12/31/2010	NAWQA (2013-2017); SURF (2005-2017)
Methidathion	950-37-8	DTXSID 5020819	141	2	12/30/2012	NAWQA (2013-2017); SURF (1991-2017)
Methyl parathion	298-00-0	DTXSID 1020855	5	1	12/31/2013	NAWQA (2013-2017); SURF (1991-2017)
Mevinphos	7786-34-7	DTXSID 2032683	4	1	7/1/1994	SURF (1992-2017)
Molinate	2212-67-1	DTXSID 6024206	4	1	8/31/2009	NAWQA (2013-2017); SURF (1991-2016)
p,p'-DDD	72-54-8	DTXSID 4020373	12	1	6/14/1972	NAWQA (2013-2015); SURF (1990-1995; 2007)
p,p'-DDT	50-29-3	DTXSID 4020375	20	1	6/14/1972	NAWQA (2013-2015); SURF (1990-1997; 2007)
Parathion	56-38-2	DTXSID 7021100	5	1	10/31/2003	NWIS (2008-2017)

Note: Asterisk (*) indicates canceled pesticides included on the PCCL5.

Section 3.8 Summary of the PCCL 5

The resulting PCCL 5 comprises a total of 275 chemicals. As shown in Table 11, this includes 252 of the highest scoring chemicals and 53 publicly nominated chemicals, from which 30 were excluded because they had ongoing agency actions under the Regulatory Determination 4 (RD 4) process or were cancelled pesticides the agency determined did not warrant further evaluation. The PCCL 5 also includes 23 DBPs, 7 cyanotoxins, and 18 PFAS chemicals. See Appendix D for all 275 chemicals on the PCCL 5.

Table 11 Accounting on the PCCL 5

Summation Steps	Number of Chemicals	Totals
Begin with the highest scoring chemicals (screened from Universe)	252	275 chemicals (PCCL 5)
(+) Add public nominated chemicals (those not already screened)	53	
(-) Exclude chemicals with regulatory determinations	9	
(-) Exclude canceled pesticides	21	
(-) Exclude DBPs (listed as a chemical group instead)	23	214 chemicals (PCCL 5 reviewed by the chemical evaluators)
(-) Exclude cyanotoxins (listed as a chemical group instead)	7	
(-) Exclude PFAS chemicals (listed as a chemical group instead)	18	
(-) Exclude publicly-nominated chemicals lacking occurrence data	13	

Chapter 4 Classification of PCCL Chemicals to Select the CCL

Section 4.1 Overview

The purpose of Step 3 of the CCL 5 development process was to narrow down the PCCL 5 chemicals to a CCL 5 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 5 chemical contaminants and reached a consensus on whether to recommend listing them on the CCL 5.

As was the case with past CCLs, the CCL 5 classification process adhered to principles that reflect the critical goals of the CCL:

- Classification must consider chemicals for listing based on a consideration of their potential for occurrence in 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 nature and spatial extent 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.

EPA's first task in this step involved the collection of additional health effects and occurrence information for the top-scoring and publicly nominated chemicals on the PCCL 5. EPA used supplemental sources that either were not identified during development of the universe or were not available in a retrievable format. EPA used this information to fill data gaps and calculate three types of data elements: health reference levels, final hazard quotients, and attribute scores (all referred to as calculated data elements). EPA then used these calculated data elements, along with relevant health effects and occurrence data metrics, to evaluate the contaminants on the PCCL 5 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 5 chemicals in Section 4.2, calculated data elements in Section 4.3, and the CISs in Section 4.4.

In the second task, 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 5 chemicals should or should not be listed on the CCL 5. A more detailed explanation of the team evaluation process is provided in Section 4.5.

Finally, to determine the number of chemicals to be reviewed by the evaluation teams and to assess the accuracy and performance of the screening scores and other relevant variables as a predictor of listing outcomes, EPA developed several logistic regression models. Further discussion on the logistic regression and its results is provided in Section 4.6.

Figure 10 illustrates the classification step for developing the CCL 5, shown in green:

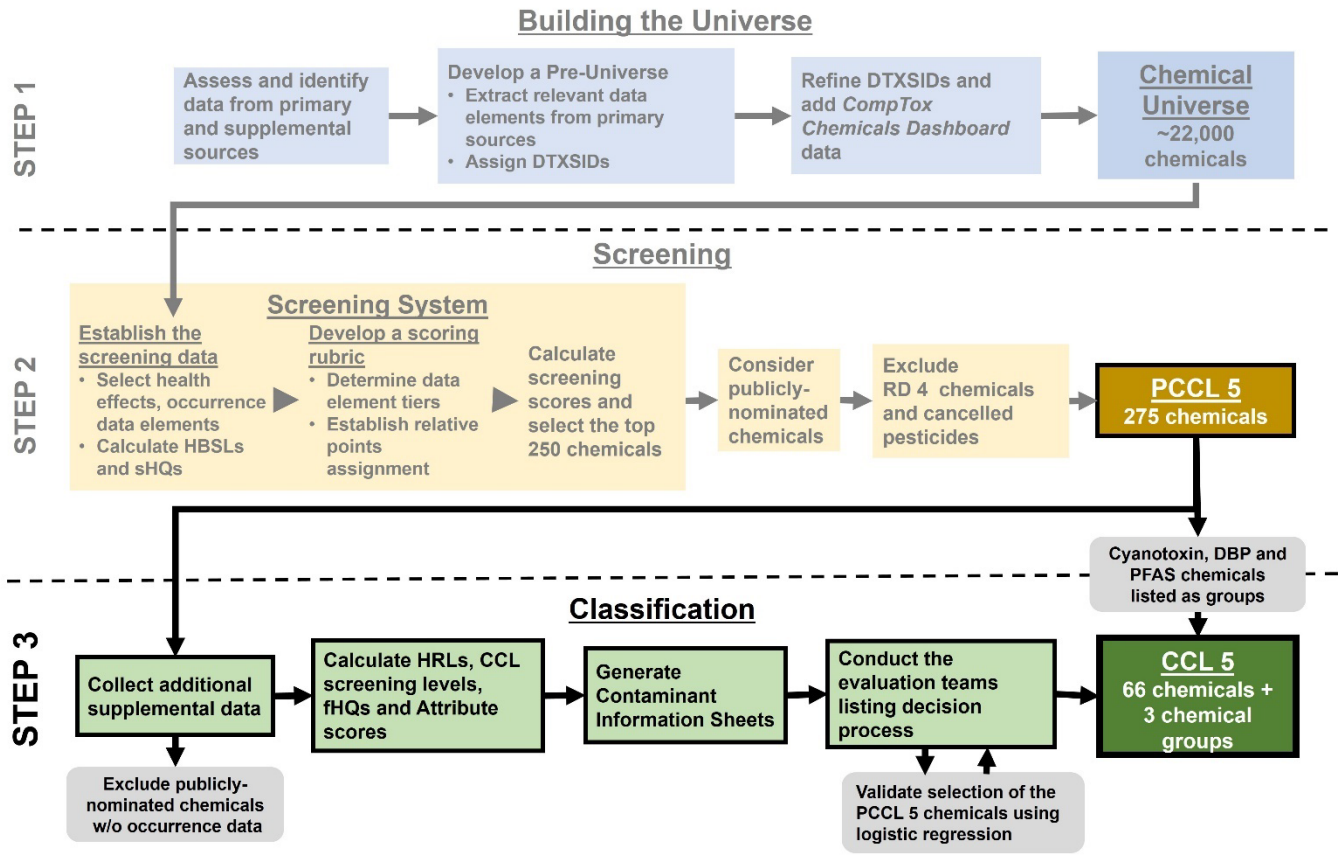


Figure 10. Development Framework Step 3 – Classification

Section 4.2 Supplemental Data Collection

Section 4.2.1 Occurrence Data

Section 4.2.1.1 Systematic Occurrence Literature Review

EPA's systematic literature reviews identified supplemental data to fill data gaps for PCCL 5 chemicals that required further evaluation. This included a search for additional peer-reviewed studies addressing the occurrence of chemicals in drinking water or ambient water. Targeted literature searches were conducted in 12 batches between March and June 2020 and covered studies published between 2010 and up to the time the specific literature search was completed. Many studies were highly localized in scope and evaluated as supplemental data only if other more comprehensive studies were not available.

This section summarizes the protocol used for conducting occurrence literature searches for CCL 5. For a full description of the occurrence literature search protocol and a list of supplemental occurrence literature used for CCL 5, see Appendix E.

EPA performed an internet search, primarily through Google Scholar, using the contaminant name and keywords such as "drinking water," "occurrence," and "occurrence in water." EPA also used Hazardous Substances Data Bank (HSDB) and EPA's abstract sifter to identify occurrence information to fill data gaps. EPA maintained a contaminant tracking list for all supplemental data sources identified.

EPA cross-checked data sources against the list of primary data sources identified during development of the CCL 5 Universe, described in Section 2.2.1 to avoid duplication of data. Some primary data sources were excluded from the occurrence literature review, with the exception of the HSDB, which was searched for available environmental fate and use data for the PCCL 5 chemicals.

EPA did not conduct occurrence literature searches for PCCL 5 chemicals that had nationally representative finished water data from UCMR 3 or UCMR 4. These chemicals were considered to already have the best available occurrence data to inform whether a contaminant was known to occur in PWSs and therefore no occurrence data were needed.

Thirteen of the publicly nominated chemicals added to the PCCL 5 did not have available water occurrence data, even after the systematic occurrence literature search was conducted (see Section 3.6.3). These chemicals were 1-phenylacetone, 3-monoacetylmorphine, 6-monoacetylmorphine, benzoic acid, benzoic acid glucuronide, hippuric acid, hydromorphone, hydromorphone-3-glucuronide, hydroxyamphetamide, isodrin, methamphetamine, morphine-6-glucuronide, and phenylpropanolamine. EPA discussed this group of chemicals with the two evaluation teams who decided not to examine them further for listing on the CCL 5. With no available data regarding measured occurrence in water and no relevant data provided by the nominators, the chemical evaluators agreed they could not determine the likelihood of these chemicals to present the greatest public health concern through drinking water exposure and therefore should not advance in the CCL 5 process. The 13 nominated chemicals with no water occurrence data were highlighted as having substantial data gaps (see Chapter 5). As a result, these chemicals were not evaluated for listing on the CCL 5.

Section 4.2.1.2 Estimated Occurrence Concentrations

EPA compiled estimated occurrence concentration data for pesticides on the PCCL 5 that lacked nationally representative finished and/or nationally representative ambient water data. These pesticides are registered through 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 environmental concentrations (EECs) or 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.

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), EPA selected the estimated surface water concentration 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, EPA selected the surface water-chronic-cancer estimate as the occurrence concentration for oxadiazon rather than estimated peak, acute, or chronic non-cancer concentrations.

For these pesticides, EPA compared modeled data from OPP with the health reference level. As part of the pesticide registration process, EPA calculates an EEC in water or EDWC depending on the year the last assessment was completed. The EEC and EDWC are derived from models that estimate the pesticide concentration in a reservoir used for drinking water. OPP used the PRZM-EXAMS model for surface water. Ground water concentrations were derived using the SCI-GROW regression model to represent exposure in shallow ground water. The modeled values allowed EPA to calculate the EEC or EDWC/HRL ratio for pesticides and/or their degradates.

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 final hazard quotient.

Section 4.2.1.3 State Drinking Water Compliance Monitoring Data and Six-Year Review 3
For the Third Six-Year Review (SYR 3), EPA requested, through an Information Collection Request (ICR), that primacy agencies voluntarily submit drinking water compliance monitoring data collected from 2006 through 2011 to EPA. Some primacy agencies submitted occurrence data for unregulated contaminants as well as regulated contaminants. EPA manually extracted occurrence data on PCCL 5 chemicals from the SYR 3 ICR data and supplemented these data by downloading additional publicly available monitoring data from state websites. The SYR 3 ICR data were included on the CISs. (Specific information on the SYR 3 ICR and state drinking water monitoring data used in CCL 5 can be found in Appendix N.)

Section 4.2.1.4 Community Water System Survey
EPA compiled additional occurrence data from the 2006 Community Water Systems Survey (CWSS) (USEPA, 2009d; 2009e). The 2006 CWSS gathered data on financial and operating characteristics from a sample of community water systems (CWSs) nationwide. Systems serving more than 500,000 people were included in the sample, and systems in that size category were surveyed about the concentrations of unregulated contaminants in their raw and finished water. EPA supplemented the CWSS by gathering additional information about contaminant occurrence from publicly available sources. EPA used the 2006 CWSS only as supplemental information and for illustrative purposes for CCL 5. The CWSS data were included on the CISs. (Specific information on CWSS data used in CCL 5 can be found in Appendix N.)

Section 4.2.2 Health Effects Data

Section 4.2.2.1 Rapid Systematic Literature Review
For chemicals with no available qualifying or non-qualifying health assessments, toxicity values identified through literature searches can be used to derive a CCL screening level (see Section 4.3.1). An RfD can be calculated by extracting NOAELs and LOAELs from peer-reviewed literature and dividing by the appropriate uncertainty factor. Subsequently, this RfD can be used for CCL screening level derivation.

As part of the classification step of CCL 5, EPA developed a rapid systematic review (RSR) protocol to identify supplemental health effects information for PCCL 5 chemicals identified during the CCL 5 screening process (see Chapter 3). 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 CCL 5 RSR protocol includes identification of health effects information (epidemiological and toxicological data as well as physiologically based pharmacokinetic models) and extraction of relevant data elements (e.g., NOAELs and LOAELs). Supplementary materials and literature search results for each chemical are accessible via the EPA docket (Docket ID No. EPA-HQ-OW-2018-0594).

The CCL 5 RSR protocol for identifying supplemental health effects data is designed to allow for screening and data synthesis for a large number of chemicals in a relatively short time frame and comprises the following:

- Targeted literature search
- Machine learning-based title-abstract screening to identify relevant literature
- Streamlined full text review and study quality evaluation of relevant literature
- Data extraction components of traditional systematic reviews

To increase efficiency and reduce redundancy of literature searches conducted by other offices and agencies, EPA did not conduct a health effects RSR for the following groups of PCCL 5 chemicals:

- Chemical pesticides registered under FIFRA which regularly undergo literature searches through OPP's registration review process
- FDA-registered pharmaceuticals for which EPA relied on lowest therapeutic doses extracted from FDA-approved labels
- Essential nutrients for which Institute of Medicine reports are regularly updated
- Chemicals currently prioritized by other agency processes (e.g., DBPs and PFAS)
- Nominated chemicals for which no water occurrence data were available (see discussion in Section 4.2.1.1)

Table F-1 in Appendix F lists the 53 PCCL 5 chemicals prioritized for the health effects RSR.

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 NOAEL and lowest LOAEL identified for each chemical (e.g., critical study, health effect endpoint) are populated on CISs (see Section 4.4) and used as important supplemental data to inform the chemical evaluators of potential health effects that can result from chronic oral exposure to chemical contaminants. The full health effects RSR protocol is available in Appendix F.

Section 4.3 Calculated Data Elements

Section 4.3.1 Health Reference Levels and CCL Screening Levels

Health reference levels (HRLs) and CCL screening levels, referred to collectively as health concentrations in this document, are non-regulatory health-based toxicity values and are expressed as concentrations of a chemical in drinking water that 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 levels in drinking water suggest a potential risk to human health. Both HRLs and CCL screening levels are expressed in µg/L.

HRLs are derived from toxicity values (e.g., RfDs, PADs, CSFs) extracted from qualifying health assessments. Qualifying health assessments are externally peer-reviewed, publicly available assessments published by EPA and other health agencies. These assessments generally follow methodologies consistent with EPA's current health guidelines and guidance documents (see Appendix G).

CCL screening levels are derived from toxicity values (e.g., RfD equivalents, CSF equivalents) extracted from non-qualifying health assessments. These publicly available assessments are published by health agencies to 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. CCL screening levels can also be derived from toxicity values such as NOAELs or LOAELs that are extracted from peer-reviewed studies identified through the CCL 5 RSR protocol (see Section 4.2.2.1). HRLs are preferentially derived over CCL screening levels.

EPA searched for all relevant health assessments for each PCCL 5 chemical identified for evaluation up until the evaluation team meetings (between March and July 2020; see Section 4.5). Appendix G describes the full protocol, briefly described below, for determining the assessment and data element most appropriate for deriving a health concentration for each chemical.

From each health assessment, EPA extracted toxicity values and other relevant data elements (e.g., cancer classifications) and compiled these in a single health effects data extraction spreadsheet. Generally, EPA relied on its most recently published health assessment as the source of toxicity values to derive the HRL. EPA relied on other sources if:

- No EPA health assessments were available for the chemical of interest.
- 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) to derive toxicity values.

For some chemicals of interest, no qualifying health assessments were available, so EPA relied on the most recently published non-qualifying health assessment to derive a CCL screening level. NOAELs and LOAELs extracted from peer-reviewed literature identified through the CCL 5 RSR process could be used as alternate toxicity values.

Appendix G also includes the procedure 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. EPA applied age-dependent adjustment factors (ADAFs) to chemicals identified as having a mutagenic mode of action to account for risks associated with early life exposure to mutagenic carcinogens. For non-carcinogens, the toxicity value (RfD or equivalent) was divided by an exposure factor (i.e., body weight-adjusted drinking water intake; USEPA, 2019) relevant to the target population and critical effect and multiplied by a 20% relative source contribution (USEPA, 2000b). Target populations considered for CCL 5 include sensitive populations such as bottle-fed infants, pregnant women, and lactating women. When selecting the relevant target population, EPA considered whether the critical effect had implications for a specific sensitive population. For example, if the critical effect was "altered neurodevelopment in infants," EPA would select bottle-fed infants as the relevant target population when deriving the health concentration. If a chemical has toxicity values based on both

cancer and non-cancer data, EPA selected the endpoint that resulted in the most health protective value as the final health concentration.

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 (see Section 4.5). EPA provides health concentrations derived from all available assessments in the health effects section of the CIS as an additional resource for the chemical evaluators. Health concentrations are reported in µg/L and can be directly compared with occurrence concentrations to assess whether concentrations in drinking water suggest a potential risk to human health.

Section 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 potency and the concentrations at which it may be found in water. To assess this relationship, EPA developed a metric called the final hazard quotient (fHQ). An fHQ is the ratio of a chemical's 90th percentile (of detections) water concentration to its health concentration (i.e., HRL or CCL screening level) at which no adverse effects are expected (as shown in Equation 2). When possible, this ratio was calculated for all PCCL 5 contaminants slated for further evaluation with empirical or modeled water data.

Equation 2. Formula for Calculating Final Hazard Quotients

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

The fHQ is an important benchmark that chemical evaluators can use to gauge the level of exposure concerns posed by each chemical in water. For the CCL 5, 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 100% of the health concentration value (high concern).

EPA selected the 90th percentile (of detections) water concentration as the point of comparison for the ratio, rather than the mean or median. 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 drinking water than others. For the CCL, if this concentration level was not available for a chemical, EPA used the next highest (i.e., 95th or 99th percentile) or the maximum reported value of detections.

EPA used a quality-based protocol (see Appendix H) to determine the data source for selecting the water concentration input across the different types of data available during the CCL 5 process. As in past iterations, EPA prioritized the use of nationally representative finished water data, choosing from the UCMR, UCM, NIRS, and DBP-ICR datasets first, if available.

For chemicals that lacked or had limited finished water data but had robust ambient water monitoring data such as NAWQA, EPA used the ambient water concentration to develop the ratio. For pesticides

with no measured nationally representative water data available, EPA used modeled water data developed by its Office of Pesticide Programs (OPP), when available, to calculate the fHQ. For contaminants with no water data (either empirical or modeled), the fHQ could not be calculated and the entry was left blank on the CIS.

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, EPA selected a CCL screening level derived from a non-qualifying assessment. If non-qualifying assessments were not available, EPA selected a CCL screening level derived from studies identified during the rapid systematic literature review. For contaminants with no toxicity values, the fHQ could not be calculated and the entry was left blank on the CIS document.

Section 4.3.3 Attribute Scores

Attribute scores are numeric values 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. To evaluate chemicals as potential CCL candidates, EPA needs to establish consistent comparative framework for the different types of data representing measures of the attributes.

During development of the CCL 3 and CCL 4 process, 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 dossiers of data and with variation in the types of data available for unregulated chemical contaminants. To enable comparisons, a scaling system that accepts a variety of input data yet provides a consistent comparative framework is needed.

In establishing the attribute scoring protocol for CCL 5, EPA chose to adhere 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 and prevalence and magnitude to describe occurrence during the development of CCL 3 (NRC, 2001). When occurrence data are not available, NRC also suggested that environmental fate properties (i.e., persistence and

mobility) could be used as surrogates to estimate potential for occurrence. As in the CCL 3, EPA agreed the recommended attributes were appropriate and consistent with data used in past decisions.

Each attribute as they relate to the CCL 5 data are described in the subsequent sections.

Section 4.3.3.1 Potency

The potency attribute score quantifies the potential for a chemical to cause adverse health effects based on the dose required to elicit the most sensitive adverse effect as identified in a single study or assessment. For CCL 5, the potency attribute score was quantified from the toxicity value (RfD, CSF, etc.) used to derive the health concentration (i.e., 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 adverse effects at lower doses).

The CCL 5 protocol for assigning potency scores is a modified version of the CCL 3 potency scoring protocol (USEPA, 2009c). Both methods require calibration of a set of toxicity values to normalize a scale with a range from 1 to 10. In CCL 3, EPA used a learning set of about 200 chemicals to calibrate this scale. In CCL 5, the potency score calibrations incorporated all available toxicity values from the universe—that is, a full range of potential potency (from low to high toxicity)—and established a scale to derive further scores.

EPA gathered CCL 5 Universe data and calibrated separate potency scoring scales for four types of toxicity values, including CSFs (and equivalents), RfDs (and equivalents), NOAELs, and LOAELs. Similar to the CCL 3 process, EPA plotted the logarithmic distribution of these toxicity values (rounded to the nearest integer) to assess the normality of the distributions and evaluate the possibility of developing a scale based on these measures. Distributions for each toxicity value type are shown in Figure 11.

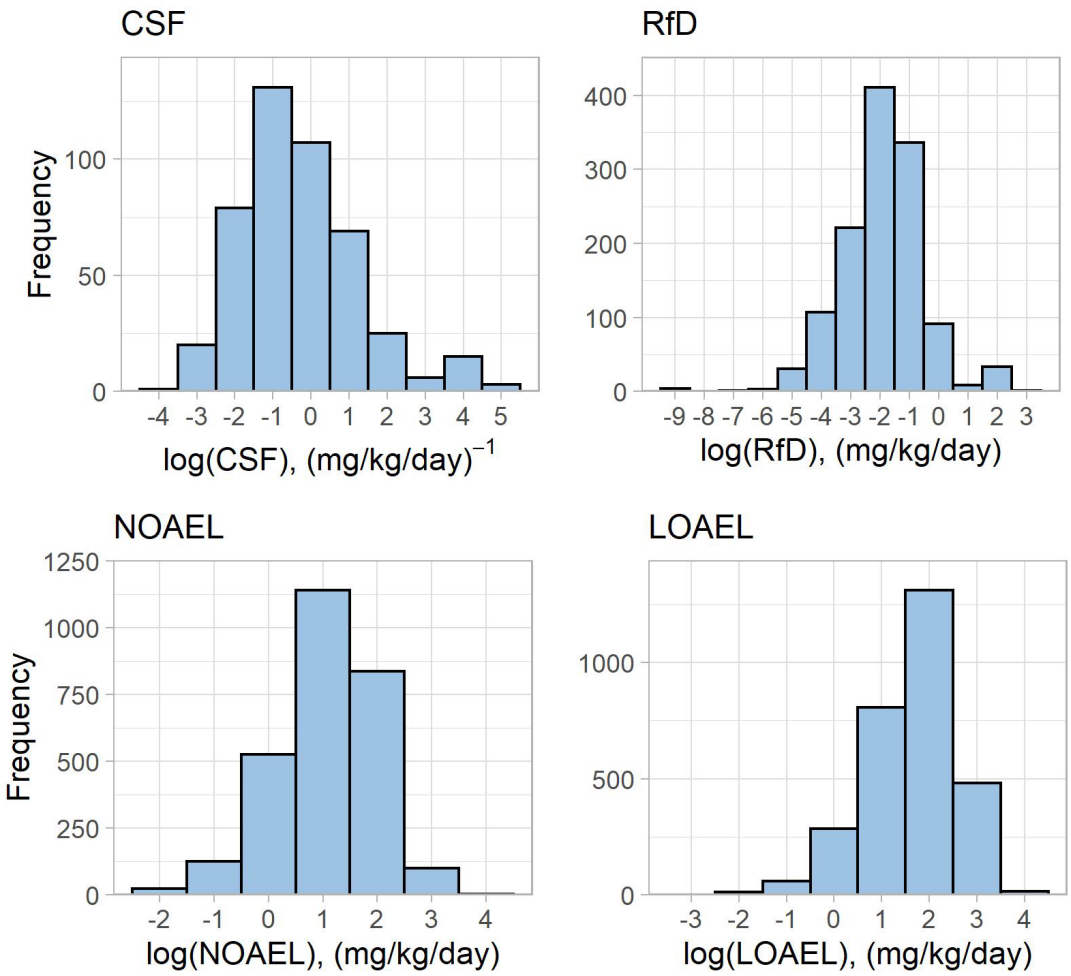


Figure 11. Rounded Logarithmic Distributions of CSFs, RfDs, NOAELs and LOAELs for the CCL 5 Universe

The logarithmic distributions for each type of toxicity value spread the chemical toxicity parameters across 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 distribution for the cancer slope factor is the most skewed of the four examined. This is similar to CCL 3 findings and indicates that chemicals in the universe with quantified CSFs are more likely to have larger CSFs (be more potent carcinogens) than have a lower CSF. As in CCL 3, no additional steps were taken to further normalize the CSF-based Potency scores across a range of 10. The values for each logarithmic distribution were normalized by binning the data into 10 levels with a center at level 5.

The logarithmic distribution 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 12. The distribution for each type of toxicity value is different, which necessitated different calibrations for each measure.

Table 12. Median Logarithmic Distribution Values by Toxicity Value

Toxicity Value	$\log_{10}$(median value)
RfD	-1.7827
CSF	-0.5302
NOAEL	1.1761
LOAEL	1.7324

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 13.

Table 13. Potency Scoring Equations by Toxicity Value

Toxicity Value	Potency Equation
RfD	Score = -1.7827 – $\log_{10}$(RfD) + 5
CSF	Score = -(-0.5302) + $\log_{10}$(CSF) + 5
NOAEL	Score = 1.1761 – $\log_{10}$(NOAEL) + 5
LOAEL	Score = 1.7324 – $\log_{10}$(LOAEL) + 5

As with the CCL 3 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. Note that due to differences in scale calibrations, potency scores derived for one type of toxicity value should be compared only to potency scores derived from that same type of toxicity value. Appendix I 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.

Section 4.3.3.2 Severity

The data source used to describe a chemical's potency is the same used to describe its severity. Severity is a descriptive measure of the adverse effect associated with the toxicity value (RfD, CSF, etc.) used to derive the potency score and health concentration (i.e., HRL or CCL screening level) for a specific chemical. Severity refers to the relative impact of an adverse physiological change caused by a chemical on the function or survival of a human or animal. CCL 5 severity categories correspond with the type of adverse outcome expected to occur at the LOAEL of a chemical.

Severity differs from the other attribute scores because it is a qualitative, not quantitative, chemical description. In previous CCL iterations, descriptions of severity were associated with a numerical scale. For CCL 5, EPA elected to simplify categorization of severity, given the potential range of effects and difficulty ascribing a quantitative level of adversity for effects, and retained categorical descriptions when referring to severity. The eight qualitative severity categories used in CCL 5 are listed in Table 14.

Table 14. CCL 5 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	Permanent developmental or gestational effects or effects that impact the ability of a population to reproduce
Carcinogen with linear mode of action	Effects resulting in a fatal disorder and 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 a fatal disorder and 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 a fatal disorder and any type of tumor confirmed to result from chemical exposure-induced mutagenicity
Reduced longevity	Effects resulting in premature mortality

Similar to CCL 3 and CCL 4, the CCL 5 severity category application requires scientific judgment. Appendix J 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.

Section 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 the occurrence of the chemical 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 element. Table 15 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 15. 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 at ambient water sites (nationally)	Median concentration of detections for a chemical at ambient water sites (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 drinking 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 5, data sources that could provide occurrence data ranged from direct measures of concentrations in water to annual measures of environmental release or production.

Section 4.3.3.4 Data Sources

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 1-4) datasets
- Unregulated Contaminant Monitoring-State Rounds 1 and 2 (UCM 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 Program (NAWQA)

Many chemicals evaluated through the CCL process did not have direct finished or ambient water measurements. To fill this gap, 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 2016
- 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

Section 4.3.3.5 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 or samples 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 K.

The CCL 5 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). EPA compiled a learning dataset of 207 chemicals to develop and calibrate scales for scoring the magnitude and prevalence attributes. 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. The analyses began with equal bin distributions, by equally dividing 100 percent of the sites with detections and 50 states with releases into 10 bins based on deciles. For prevalence, the bins provided a fairly good fit to the distribution but still required some adjustment because the equal bins tended to segregate chemicals by type. Chemicals with the highest percentage of detections scoring a 9 or 10 were naturally occurring inorganics. For example, in the NIRS for ground water, ions like sodium, calcium, and iron were all detected in ≥ 90% of the groundwater systems sampled.

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. 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 (IOCs).

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, 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.

Section 4.3.3.6 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 or samples with detections
3. Application of the chemical as a pesticide in pounds
4. Total releases of the chemical in pounds
5. Persistence-mobility data (see Section 4.3.3.7)

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

As with prevalence scoring, the CCL 5 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 4.3.3.3. In CCL 3, EPA explored a variety of potential scales that could be applied to the finished water concentration data. 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.

The first approach was to develop scales that used an array of compiled magnitude data and 10 bins with approximately equal numbers of contaminants in each, referred to as the equal number bins scale. Equal bins did not provide a good dispersion of scores. Accordingly, various log-scale options were explored. The magnitude data do not range across as many orders of magnitude as the potency RfD data, so various semi-logarithmic scales were evaluated to better represent distribution of values across the scale.

In evaluating and developing the calibration scale, water occurrence data presented a particular challenge because IOCs tended to skew the results. Many IOCs result from various anthropogenic processes, but most are of geologic origin as well and have relatively high measures for both prevalence and magnitude compared to most organic chemicals. For some of the semi-logarithmic magnitude scales, the only chemicals that could score high (e.g., a 10 or 9) would be IOCs. Such a scale would depress the score for organic chemicals. One approach that EPA evaluated was using different scales for IOCs and organic chemicals which, however, would make the scoring process overly complex. To keep the process straightforward and transparent EPA decided to use one scale for all water data. Scores were distributed across the range of values so organic contaminants as well as IOCs could receive high scores. 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).

Section 4.3.3.7 Persistence and 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 (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. 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. Therefore, 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 16. The measure of persistence reflects the time the chemical will remain unchanged in the environment. The first two measures of mobility represent the equilibrium ratio for the partitioning of the contaminant from one medium to another: K_{ow} (octanol: water) and K_H (air: water). K_{ow} is expressed as logs of the original measurements. For the third 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 16 are arranged in hierarchical order, with the most desirable at the top (i.e., the first data to be used if available).

Table 16. Data Elements Used to Score Persistence and Mobility

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

[1] The predicted biodegradation half-life, 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)

Section 4.3.3.8 Persistence and 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, 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 et al. (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 M for details).

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, EPA decided the somewhat conservative scores were acceptable as surrogate measures for magnitude when only persistence and mobility data were available for scoring.

Section 4.4 Contaminant Information Sheets (CISs)

EPA developed a CIS for each chemical on the PCCL 5 that was evaluated by the chemical evaluators to make listing recommendations for the CCL 5. Each CIS presents a contaminant's health and occurrence data gathered from primary and supplemental data sources along with health and occurrence statistical measures. CISs also include additional information about the contaminant, such as the identity of the contaminant and its usage, whether it was subject to past negative regulatory determinations, listed on past CCLs, and publicly nominated for the CCL 5. Due to the inclusion of more data in the CCL 5 process, CISs for the CCL 5 contain more information than those of past CCLs. An annotated CIS Key and the CISs for the CCL 5 can be found in the CIS Technical Support Document (USEPA, 2022b).

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), final hazard quotient (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 5. The first page also identifies whether the contaminant has been listed on the CCL 5; this information was added after the evaluation teams concluded their listing recommendations. 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 reference doses, cancer slope factors, and cancer classifications extracted from health assessments, and other health data from primary and supplemental data sources. The second page also summarizes results of the RSR of the health effects literature. 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 from primary data sources; and non-nationally representative finished and ambient water data from primary and supplemental sources. The third page also lists modeled environmental fate parameters for the contaminant. Data used to calculate statistical measures like attribute scores are highlighted.

Section 4.5 Evaluation Teams Listing Decision Process

Fourteen EPA scientists, referred to as chemical evaluators, reviewed the PCCL in batches to determine which chemicals should advance to listing on the CCL 5. Evaluation of each PCCL chemical involved the following:

- 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 and confidence rating in the underlying data for each chemical
- Team deliberations to reach a consensus following a facilitated discussion on whether or not to list each chemical, if needed

Section 4.5.1 Evaluation Teams

EPA divided the chemical evaluators into two seven-member teams to split the workload and expedite the listing decision process. The two teams had a similar composition of expertise and specialization. Participants included physical scientists, environmental engineers, toxicologists, program analysts, and environmental protection specialists from the Office of Water, Office of Research and Development, Office of Children's Health, and Office of Pesticide Programs. EPA also maintained a list of six 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 12 times between mid-March and early July 2020. Due to the COVID-19 pandemic, all team meetings were conducted virtually. The chemical evaluators discussed their independent reviews of each PCCL chemical in the batch to arrive at a consensus on whether or not to list it on the CCL. Batches ranged between seven to 30 chemicals, with a batch of 20 chemicals being the most common (i.e., 10 chemicals per evaluation team). Team meetings averaged between one and a half to two hours. When a team could not reach consensus, the chemical was tabled for a future meeting, allowing time to research additional information to help inform a final listing decision.

Section 4.5.2 Evaluator Training

Prior to beginning their reviews, the chemical evaluators participated in a training session to familiarize themselves with the background history of CCL, the SDWA requirements, and the listing approach to follow throughout the evaluations. The training introduced chemical evaluators to the process of taking

[2] Some chemical contaminants have five-page CISs.

chemicals from the universe through classification (i.e., steps 1 to 3) and their role in developing the CCL 5 chemicals.

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.

Section 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 received a link to a survey that asked for responses in three areas for each chemical in that batch:

- Provide a numeric listing decision 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 listing decisions in 1 to 3 sentences.

- Provide a rating of overall level of confidence for the data and information underlying the chemical:
  - Low – a score of 1
  - Medium – a score of 2
  - High – a score of 3

Based on the responses to question 1, EPA calculated the simple average for the list decisions across the individual chemical 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 decision. The thresholds for undertaking evaluation team discussion on a given chemical are shown in Table 17.

[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 17. Survey Listing Decision 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)
	1.0	2.0	3.0	4.0
Interpretation	Strong consensus	Weak consensus		Strong consensus
CCL 5 Outcome	Chemical not listed	Evaluation team discussion held to finalize the listing decision		Chemical listed

Section 4.5.4 Listing Decisions

After receiving and tallying the chemical evaluators' survey responses for a given batch, meeting coordinators prepared presentation slides for each chemical to support any necessary discussion based on the numerical average list decision. The presentation slides helped chemical evaluators understand the range in listing decisions and justifications for the current batch of chemicals and were used to guide discussions by the meeting facilitator during meetings. 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, range of individual list decisions, and general confidence in the underlying data. The facilitator then asked each evaluator to explain the listing decision and justification for the chemical, starting with evaluators who assigned the greatest listing certainty. 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 decision was considered final. If the consensus was weak, the outcome could be to go with the majority listing decision or table till a future team meeting pending further research.

Of the 275 PCCL 5 chemicals, the chemical evaluators reviewed 214 chemicals on PCCL 5; as shown previously in Section 3.8 under Table 11, the number 214 is derived by excluding 13 publicly nominated chemicals that lacked occurrence data in addition to the 23 DBPs, 7 cyanotoxins, and 18 PFAS chemicals that bypassed the evaluation process for listing as the three chemical groups.

Ultimately, 66 of the 214 evaluated PCCL 5 chemicals were recommended for placement on the CCL 5, shown in Section 4.7.

Section 4.6 Logistic Regression Analysis

Section 4.6.1 Overview

The PCCL 5 consists of 275 chemicals screened from the CCL 5 Universe by a new point-based screening process (Chapter 3). To select chemicals for the CCL 5, two teams of chemical evaluators reviewed 214 PCCL 5 chemicals (see Section 4.5). To ensure the efficacy of the screening process to a PCCL, EPA conducted statistical analyses and developed a logistic regression model to validate selection of the top 250 chemicals for the PCCL 5 while the evaluation team reviews were ongoing. Following conclusion of the reviews, EPA conducted further statistical analyses and logistic regression models to examine the efficacy of the screening process and to determine other factors associated with listing decisions. EPA developed simple (Section 4.6.2 and Section 4.6.3.1) and multiple (Section 4.6.3.2) logistic regression models for CCL 5.

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 or not a chemical is recommended for listing on the CCL 5. In CCL 5, the binary outcome of interest is the evaluation teams' list or not list decision. The independent variables or predictors that could influence listing decisions are screening scores, attribute scores, fHQs, etc. An example of a simple logistic regression model in the context of CCL 5 is modeling listing decision outcomes as a function of screening scores.

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 β_0 is an intercept term and β_1 is a slope coefficient. These concepts are further explained in the following sections.

EPA used data on the 214 PCCL 5 chemicals reviewed by the evaluation teams in the logistic regression models and additional statistical analyses described in the next two sections. Learning from the development and results of the CCL 3 prototype classification models (USEPA, 2009c), EPA assembled the chemicals' health effects and occurrence attribute scores. EPA also incorporated fHQs and input from the evaluation teams into the models. Listing decisions were coded as a binary variable (0 = not

list, 1 = list). For evaluated chemicals, screening scores ranged from 1900 to 9050. Scores for potency, magnitude, and prevalence attributes ranged from 1 to 10, where 1 represents the score for least potential for public health concern and 10 the score for greatest potential for public health concern. Severity categories were treated as a categorical variable with multiple levels (described in Section 4.3.3.2). Lastly, the fHQs were treated as a continuous variable with a range of 0.000009 to 8300.

The logistic regression classification models presented in this section were not used to categorize, prioritize, and/or classify PCCL 5 chemicals for inclusion on the CCL 5. EPA developed the statistical models to assess the screening and classification processes of CCL 5. The next sections describe the statistical analyses EPA conducted to investigate selection of the top 250 scoring chemicals for the PCCL 5, to determine the efficacy of the point-based screening process, and to discover if additional factors may have impacted listing decisions in the classification step of CCL 5.

Section 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 5. The screening framework was designed to rapidly prioritize the entirety of the CCL 5 Universe of chemicals while limiting manual review and human bias. With over 20,000 chemicals in the CCL 5 Universe, EPA used the screening scores to select and advance 250 chemicals for evaluation team review and potential inclusion on the PCCL 5 (Section 3.5).

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 5 by the evaluation teams. To investigate this relationship, EPA developed simple logistic regression models where screening scores were the sole predictor of listing decision outcomes. The goals of the simple logistic regression models were two-fold. First was 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. Second was to assess the accuracy and performance of the screening scores as a predictor of listing outcomes, as discussed in Section 4.6.3.2.

EPA developed a simple logistic regression model to provide 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 (9050) and lowest screening score (3310). The score of 3310 was the score of the CCL 5 Universe chemical listed directly below the cutoff score of 3320 for the top 250 chemicals.

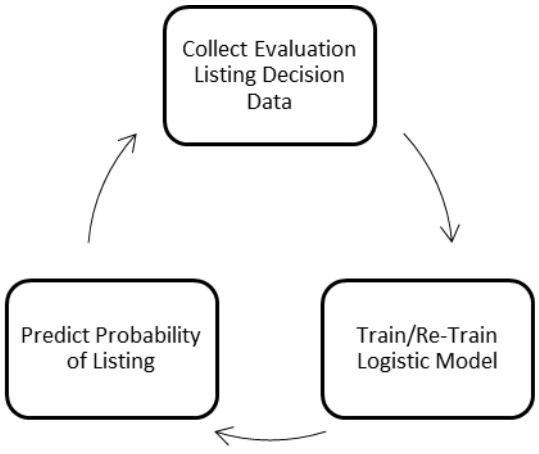


Figure 12. Flow Diagram of the Three-Step Iterative Process

To fit the logistic model according to screening scores and evaluation teams' listing decisions, 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.

Two teams evaluated PCCL 5 chemicals in 12 batches over several months. The iterative process began following completion of the sixth batch of chemical reviews and successively thereafter until all 214 chemicals were evaluated. The first six batches of chemical reviews provided 86 listing decisions and a starting point to begin model training. The screening scores for the first six batches ranged from 3480 to 9050, which represents a reasonable initial training dataset to obtain probabilities of listing at the screening score of 3310. Upon completion of each subsequent batch of chemical reviews (batches 7 to 12), the training dataset was updated with new listing decisions, the logistic model was re-trained, and the logistic model was used to predict listing probabilities. 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 12 batches of chemical reviews.

The chemicals' screening scores and evaluation teams' listing decisions were used to train the simple logistic regression model. Of the 214 chemicals evaluated by the evaluation teams, two publicly nominated chemicals, Heroin (DTXSID6046761) and Morphine-3-glucoronide (DTXSID80174157), lacked screening scores and were dropped from the model training. These chemicals were reviewed by the evaluation teams, but did not have data available in the universe for assignment of a screening score. Three publicly nominated chemicals, Morphine (DTXSID9023336), Gemfibrozil (DTXSID0020652), and Fluoxetine (DTXSID7023067), had screening scores below 3320. They were not included in the top 250 but were still reviewed by the evaluation teams and included in the final training dataset. The final training dataset consisted of 212 chemicals with screening scores ranging from 1900 to 9050 and listing decisions from the evaluation teams.

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. (2020) and Hoff (2009). The training dataset is well suited for Bayesian logistic regression due to 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 5 and, subsequently, which chemicals are candidates for the CCL 5.

An overview of the Bayesian simple logistic model developed for CCL 5 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. EPA assigned uniform prior distributions on β_0 and β_1. 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 18 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 18. Summary Statistics for the MCMC Sample

Parameter	Mean	2.5%	Median	97.5%
β_o (intercept)	-4.513	-6.018	-4.494	-3.11
β_1 (slope)	7.53E-4	4.793E-4	7.497E-4	0.001046

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 1.000753 (Table 18). Further discussion on statistical significance of screening scores as a predictor of listing outcomes is in Section 4.5.3.

After training the model, 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 (1900 to 9050). The logistic model was used to calculate probability of listing at screening scores using the parameter values from the posterior distribution. EPA focused on the probability of listing at the screening score of 9050, the score associated with highest scored chemical in the universe and on the PCCL 5, and the screening score of 3310. Table 19 contains summary statistics for the probabilities of listing at the screening scores of 3310 and 9050 calculated from the MCMC sample.

Table 19. Summary Statistics of Probabilities of Listing at Screening Scores 3310 and 9050 Calculated from the MCMC Sample.

Screening Score	Mean	5%	Median	95%
3310 (PCCL Rank #253)	0.12	0.075	0.12	0.18
9050 (PCCL Rank #1)	0.90	0.80	0.91	0.97

Figure 13 illustrates the results of the Bayesian simple logistic model where 212 listing decisions 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 (1900 to 9050). The range

[4] To perform the MCMC sampling, EPA used OpenBUGS (Bayesian inference Using Gibbs Sampling) version 3.2.3 rev. 1012 software (Lunn et al., 2009). Further analyses were conducted in R (R Core Team, 2020) in RStudio version 1.3.1056 using the CODA (Plummer et al., 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. 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.

of screening score values were discretized in evenly-spaced steps of 10 to create a 1-dimensional grid of values (1900, 1910, 1920…9050). 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 decisions are indicated by the red or light blue dots located where the listing probability is 1 (list) or 0 (not list). A small vertical offset was added to training dataset coordinates to enhance plot readability in Figure 13.

As indicated by Table 19 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 5, or the screening score equal to 9050, is 0.90. Conversely, the mean probability of listing at the screening score equal to 3310 is 0.12.

These results indicate the improved screening process achieved its intended goal to elevate chemicals for further review and inclusion on the CCL 5, based on data most relevant to drinking water exposure and potential for greatest health concern. With over 22,000 chemicals in the universe, EPA created a prioritization scheme that narrowed the focus of the evaluation teams' task of reviewing PCCL 5 chemicals for potential inclusion on the CCL 5. However, the probability of listing at the screening score 3310 is 0.12, which indicates the screening process may have missed advancing chemicals to the PCCL.

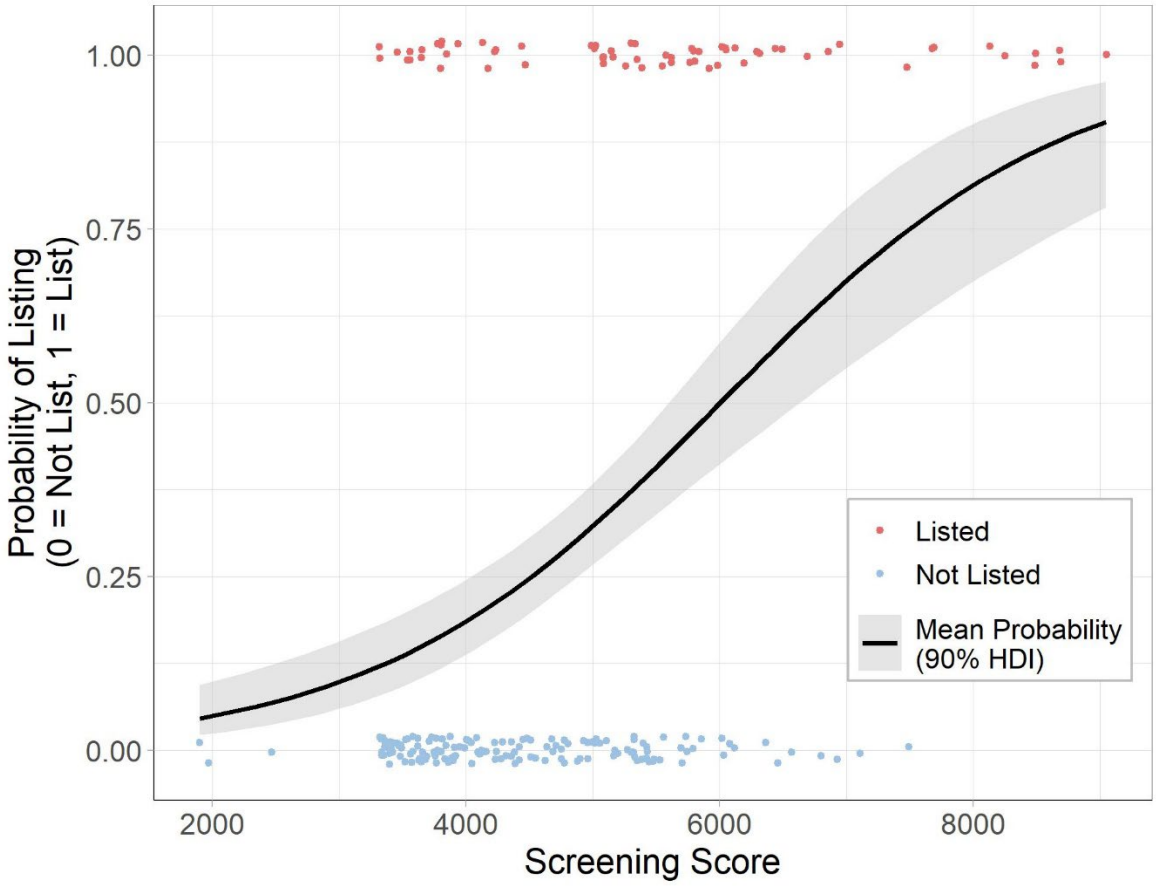


Figure 13. Results of the Bayesian Simple Logistic Model of Probability of Listing vs Screening Score

Though the screening scores incorporate health effects and occurrence data from primary 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 decisions, were unaccounted for in the simple logistic regression model. An example of this was observed in the listing decisions of the last three top scoring chemicals in the PCCL 5. 2,4-Dinitrophenol (DTXSID0020523), Phosmet (DTXSID5024261) and 4-Androstene-3,17-dione (DTXSID8024523) had the same screening score (3320); however, two of them were selected for inclusion on the CCL. It was evident that factors not captured by the screening scores were influencing the evaluation teams' listing decisions for these chemicals. 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 decisions. For this reason, there may be a disconnect between the screening scores and listing decision outcomes. Therefore, EPA conducted further statistical analyses to explore other factors not captured in the screening scores that may have influenced listing decision outcomes, as described in Section 4.6.3.

Section 4.6.3 Post-Evaluation Analysis: Exploring Listing Decision Determinants

As discussed in Section 4.6.2, a positive association was established between the screening scores and listing decisions. The higher a chemical's screening score, the higher its probability of being listed on

the CCL 5. However, EPA recognized that the screening scores may not be the only determining factor for listing decisions. Therefore, EPA explored other factors that may have impacted listing decisions and further evaluated how well the screening scores performed as a predictor of listing decisions.

Section 4.6.3.1 Exploratory Statistical Analysis

The first step of the analysis was to explore the dataset described in Section 4.5.1 through descriptive statistics. EPA calculated descriptive statistics for each variable stratified by listing decision (Table 20). This provided an early indication of which variables may be influential during the listing decision process and identification of any abnormalities in the data.

Table 20. Descriptive Statistics by Listing Decision Outcome

Variable	Not List	List
Potency	5.02 (1.30)	5.69 (1.50)
Prevalence	6.69 (3.38)	6.15 (3.49)
Magnitude	3.38 (2.34)	4.17 (2.14)
Final Hazard Quotient	69.9 (728)	12.5 (51.8)
Final Hazard Quotient (Deciles)	4.39 (2.65)	7.62 (1.97)
Screening Score	4433 (997)	5550 (1514)
Severity		
No adverse effects	7	1
Non-cancer effects	81	15
Reproductive and developmental effects	28	28
Carcinogen with linear MOA	14	20
Carcinogen with mutagenic MOA	0	1
Reduced longevity	1	1
Data unavailable	17	0

Mean (Standard Deviation) calculated for potency, prevalence, magnitude, fHQ, fHQ (Deciles), and screening scores. Frequency calculated for severity. EPA used the compareGroups package in R (Subirana et al., 2014) to calculate descriptive statistics.

As shown in Table 20, the average potency, magnitude, and screening scores were higher for chemicals that were listed compared to those not listed. However, the average prevalence score was higher for chemicals that were not listed. The average fHQ was found to be unexpectedly high for not-listed chemicals, and EPA determined a large outlier skewed the data. One chemical, 2,4-Dinitrotoluene, had an fHQ of 8,300 but was not listed. This resulted in not-listed chemicals having a higher average fHQ than listed chemicals (69.9 vs. 12.5). Further inspection of this chemical revealed that the water concentration used in the fHQ formula was based on one detect out of 3,873 samples from the UCMR 1 data. Therefore, the low occurrence of 2,4-Dinitrotoluene in national finished drinking water impacted the evaluation team's decision not to list this chemical on the CCL 5.

To alleviate the impact of outliers, EPA created a new fHQ variable, fHQ (Deciles). The fHQ values were normalized on a scale of 1 to 10 by dividing the values equally into 10 bins based on deciles where 10 is of greater concern and 1 is of lesser concern. As shown in Table 21, once the outliers were accounted for by the new fHQ variable, listed chemicals had higher average fHQ deciles (7.62) than not-listed chemicals (4.39). This adjustment also made the fHQs more suitable for further statistical modeling.

Not all severity categories were represented in the chemicals reviewed by the evaluation teams. For example, the severity category "Cosmetic effects" (Section 4.3.3.2) did not apply to any of the evaluated chemicals; therefore, this category is not represented in Table 21. Severity categories are descriptive measures, so EPA did not calculate mean and standard deviations for severity. Instead, EPA calculated frequencies for each severity category for list and no-list chemicals. Notable findings from the descriptive statistics included carcinogenic chemicals were more frequently listed than not listed, all chemicals that did not have data available to assign a severity category were not listed, and chemicals with non-cancer effects were not listed more frequently.

Descriptive statistics provide insight to underlying trends in the data. However, additional robust statistical tests are required to draw inferences about listing decisions. Therefore, EPA explored logistic regression models similar to those described in Section 4.6.2. EPA explored several simple logistic regression models to obtain odds ratios (OR) and establish statistical significance of the predictor variables. The value of an odds ratio indicates the strength and direction of the association between a dependent and independent variable (Porta, 2014). The results of the various simple logistic regression models are displayed in Table 21.

Table 21. Simple Logistic Regression Results

Variable	Odds Ratio [95% CI]	p-value
Potency	1.42 [1.13;1.77]	0.003†
Prevalence	0.96 [0.88;1.04]	0.306
Magnitude	1.15 [1.02;1.31]	0.019†
Final Hazard Quotient (Deciles)	1.66 [1.42;1.93]	<0.001†
Screening Scores	1.00 [1.00;1.00]	<0.001†

† Statistically significant at an alpha level of 0.05.

The results from the simple logistic regression models indicate that potency and magnitude are statistically significant predictors of listing decisions, but prevalence did not achieve statistical significance. The logistic regression model for screening scores displayed in Table 21 used a different method for parameter estimation compared to the Bayesian logistic regression model described in Section 4.6.2, but both models yielded very similar results. The model used in Table 21 shows that the screening scores are a statistically significant predictor of listing decisions while producing similar estimates (OR 1.0007, 95% CI: 1.0005, 1.0010). Once outliers were accounted for, the fHQ (deciles) variable achieved statistical significance and was shown to be the strongest individual predictor of listing decisions (OR 1.66, 95% CI: 1.42, 1.93). Because severity could not be treated as a continuous variable and the frequency of chemicals falling into several categories were too low to be amenable to modeling, it was not included in any of the logistic regression models.

Following the results of the simple logistic regression, EPA conducted further statistical analyses to assess if the multi-team approach affected the listing evaluations process (Section 4.5). The evaluation teams were modeled as a predictor of listing decisions where the odds of a chemical being listed were compared between each evaluation team. Initial results indicated that one team appeared to have higher odds of listing a chemical on the CCL 5. However, EPA recognized that the logistic models previously explored did not consider other important properties of the chemicals, such as chemical class.

Therefore, EPA conducted a confounding assessment to examine whether these observed differences in listing decisions between the teams could be due to such other factors as the class of chemicals each team evaluated. Confounding can be defined as the distortion of the true relationship between an independent and dependent variable by a third extraneous variable (Steenland & Savitz, 1998). EPA noted that one chemical class of pesticides, in particular, organophosphates, were assigned almost entirely to one evaluation team. In total, 19 organophosphates underwent evaluation and 17 were assigned to Team B. Of these 17, 13 were recommended to be listed, more than one-third of Team B's total. As a result, EPA hypothesized this might be a confounding factor for the association between the evaluation teams and listing decision outcomes.

Accordingly, EPA created a new variable for organophosphates so a confounding assessment could be conducted. To create the new variable, chemicals that were organophosphates were assigned a 1 and chemicals that were not organophosphates were assigned a 0. Overall, the results of the confounding assessment provided statistical evidence that a chemical being an organophosphate was a more important factor on listing decision outcomes than which team evaluated the chemical. Although these results suggested that the evaluation teams did not significantly affect the listing decision outcomes, the models employed were still relatively simple and straightforward. Therefore, EPA performed additional model diagnostics to further understand the determinants of listing decisions, as described in the next section.

Section 4.6.3.2 AUC-ROC as a Measure of Predictive Performance

One of the most widely used and important evaluation metrics to assess the performance of binary classification models is the area under the curve (AUC) receiver operating characteristics (ROC) curve. AUC-ROC curves have a wide array of applications and are proven useful tools in assessing and improving recreational water quality models (Holtschlag et al., 2008; Morrison et al., 2003).

AUC-ROC curves have a few important properties that allow them to assess the performance of classification models. First, they can measure the ability of an independent variable to correctly classify outcomes. In the context of the CCL 5, AUC-ROC curves can measure how well a given model correctly classifies chemicals as listed or not listed. Secondly, AUC-ROC curves can directly compare the discriminatory performance of multiple classification models that have different independent variables through a common AUC measurement (Holtschlag et al., 2008; Morrison et al., 2003). For the CCL 5, this allows for the direct comparison of the performances of simple logistic regression models and multivariable logistic models as predictors of listing decisions.

AUC-ROC curves use a straightforward scale to measure and compare the performance of classification models (Holtschlag et al., 2008; Tape, 2007):

- An AUC-ROC of 0.5-0.6 is considered very poor discriminatory performance
- An AUC-ROC of 0.6-0.7 is considered poor discriminatory performance
- AUC-ROC of 0.7-0.8 is considered good discriminatory performance
- An AUC-ROC of 0.8-0.9 is considered very good discriminatory performance
- An AUC-ROC of 0.9-1 is considered excellent discriminatory performance

Applied to the CCL 5, AUC-ROC curves compare the actual listing decisions made by the evaluation teams, to the predicted listing decisions made by a given model. In general, the more area that is under the ROC curve, the better the model is at discriminating between listed and not listed.

EPA applied these concepts to a simple logistic regression model with the screening scores as the sole predictor of listing decisions. AUC-ROC curves and estimates were obtained using the pROC package in R (Robin et al., 2011). The results displayed in Figure 14 indicated that as an individual variable, the screening scores were a moderate to good predictor of listing decisions (AUC = 0.72).

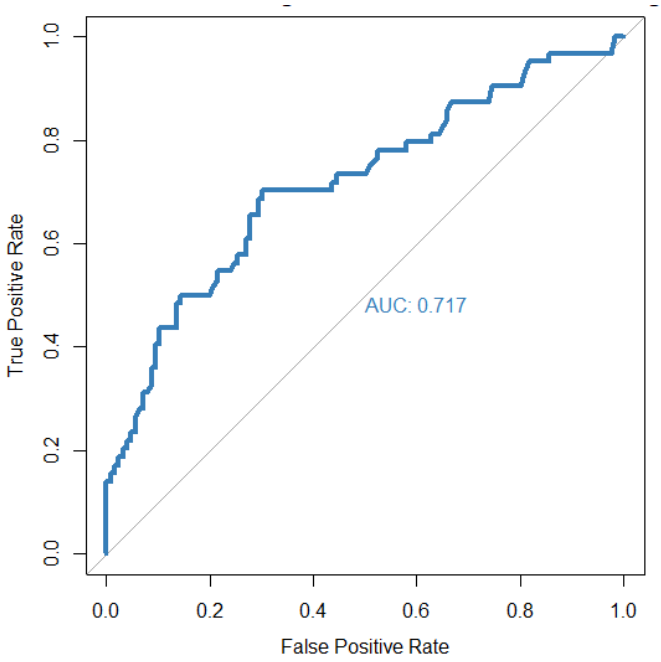


Figure 14. AUC-ROC Curve for Screening Scores as a Predictor of Listing Decisions

EPA then examined the performances of select multivariable logistic regression models. The first step was to examine the performance of a full logistic regression model, that is, a model that includes all possible independent variables as predictors.

In statistical modeling, the issue of over-fitting can be a concern when selecting a model. Any model can be made to fit a particular dataset very well by making the model more complex (this usually means estimating more model parameters). This addition of model complexity can come at the cost of a loss of general applicability.

Therefore, EPA conducted a model selection technique called backwards selection to arrive at a parsimonious model, that is, a model that has great predictive power while using a minimal number of predictors. Backwards selection based on p-values is a model selection technique that begins with all independent variables in the model and, at each step, the variable with the highest p-value is removed. In this analysis, the criterion to retain a variable in the model was a p-value below 0.05. Of the variables assessed for the CCL 5, prevalence, screening scores, fHQ (deciles), and organophosphates all met this criterion. The odds ratios, 95% confidence intervals, and p-values of the resulting parsimonious model are given in Table 22.

Table 22. Multiple Logistic Regression: Parsimonious Model

Variable	Odds Ratio	95% CI LL	95% CI UL	p-value
Prevalence	1.216	1.058	1.398	0.006
Screening Scores	1.001	1.000	1.001	0.001
Final Hazard Quotients (Deciles)	1.761	1.458	2.128	<0.001
Organophosphates	6.120	1.495	25.053	0.012

As shown in Figure 15, the parsimonious model was found to be a very good to excellent predictor of listing decisions (AUC = 0.89) while using a minimum number of predictors (prevalence, screening scores, fHQ (deciles), and organophosphates). Potency and magnitude were not selected as predictors for the final parsimonious model. This can be attributed to a statistical concept called multi-collinearity. Potency and magnitude are highly correlated with the fHQ (deciles) variable, which reduces their ability to achieve statistical significance when modeled together.

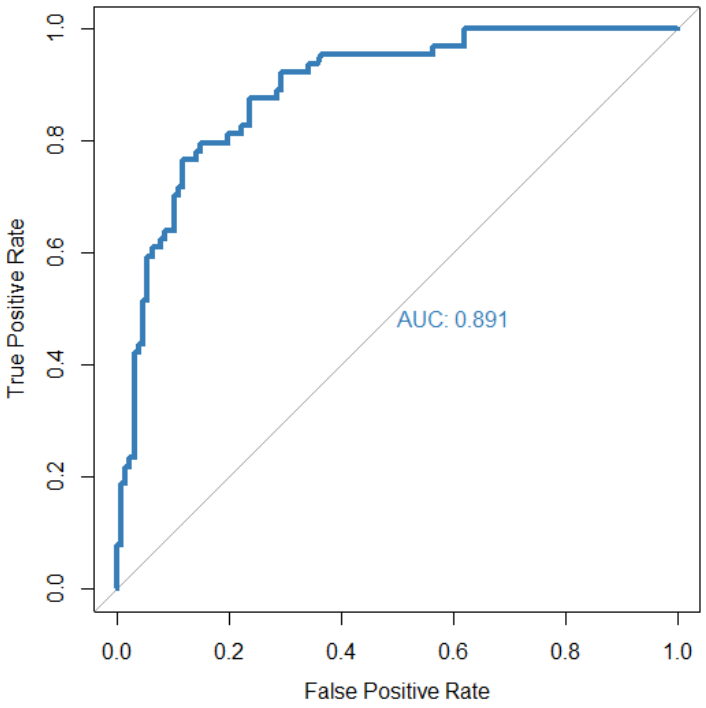


Figure 15. AUC-ROC Curve for Parsimonious Model as a Predictor of Listing Decisions

The various analyses described in this section revealed a few important findings about the CCL 5 screening and classification processes. Multiple statistical modeling techniques showed that the screening scores were a moderately good predictor of listing decisions. This finding lends confidence to the ability of screening scores to effectively prioritize the chemicals with the potential for the greatest public health concern.

In other words, the positive association observed with listing decisions suggests that the screening process was successful in providing a narrow, prioritized list of candidate contaminants for review by the chemical evaluators on the evaluation teams. The higher a chemical's screening score, the higher its odds were of the chemical being listed on the CCL 5.

EPA discovered that the screening scores were not the only determining factors for making listing decisions. Similar to the CCL3 classification algorithms, attribute scores and other chemical properties are major factors that influence listing decisions. The positive associations found between the listing decisions, the attribute scores, and the final hazard quotients (adjusted for outliers) suggest that EPA successfully developed scales and scoring mechanisms that normalize and accept a variety of input data.

The AUC-ROC analysis also led to the discovery of a parsimonious logistic regression model that was a very good to excellent predictor of listing decisions when comparing predicted to actual listing decisions. As a result, EPA developed a practical and effective tool that reasonably anticipates the ability of the human chemical evaluators to make decisions about listing chemical contaminants on the CCL 5. This opens the possibility for logistic regression-based decision support tools in future CCL iterations.

Section 4.7 Selecting CCL 5 Chemicals

The CCL 5 comprises 66 chemicals recommended by the evaluation teams, as described in Section 4.5, one group of cyanotoxins, one group of disinfection byproducts (DBPs), and one group of perfluoroalkyl and polyfluoroalkyl substances (PFAS). Table 23 presents chemical contaminants on the CCL 5.

Cyanotoxins, DBPs, and PFAS have been identified as agency priorities and contaminants of concern for drinking water under other EPA actions. Listing these three chemical groups on the CCL 5 does not necessarily mean EPA will make subsequent regulatory decisions for the entire group. Rather, EPA will evaluate scientific data on the listed groups, subgroups, and individual contaminants to inform any regulatory determinations for the group, subgroup, or individual contaminants in the group.

Addressing the public health concerns of cyanotoxins in drinking water remains a priority as specified in the 2015 Algal Toxin Risk Assessment and Management Strategic Plan for Drinking Water (USEPA, 2015). Cyanotoxins are toxins naturally produced and released by some species of cyanobacteria (previously known as blue-green algae), were listed on the CCL 3 and the CCL 4 as a group. EPA is listing a cyanotoxin group on the CCL 5, identical to the CCL 4 listing. The group of cyanotoxins includes, but is not limited to, anatoxin-a, cylindrospermopsin, microcystins, and saxitoxin. Cyanotoxins were monitored under the UCMR 4.

EPA is also listing 23 unregulated DBPs as a group on the CCL 5, as shown in Table 24. DBPs are formed when disinfectants react with naturally occurring materials in water. The DBPs within this group included both nominated and scored contaminants (see Appendix P). Under the Stage 2 Disinfectants and Disinfection Byproducts Rule, there are currently 11 regulated DBPs from three subgroups that include four trihalomethanes, five haloacetic acids, and two inorganic compounds (bromate and chlorite). Under the SYR 3, EPA identified 10 regulated DBPs (except for bromate) as candidates for revision (USEPA, 2017). For CCL 5, the group of unregulated DBPs includes both publicly nominated DBPs and DBPs in the top 250th scored chemicals that bypassed the evaluation teams' review due to other ongoing EPA actions. Listing these unregulated DBPs as a group on CCL 5 is consistent with EPA's decision identifying a number of microbial and disinfection byproduct (MDBP) drinking water regulations as candidates for revision in the SYR 3 of NPDWRs.

PFAS are a class of synthetic chemicals most 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 (AAAS, 2020; USEPA, 2018b). More than 4,000 PFAS have been manufactured and used globally since the 1940s (USEPA, 2019b), which would make listing PFAS individually on CCL 5 a challenge. EPA is listing PFAS as a group inclusive of any PFAS that fit the revised CCL 5 structural definition (except for PFOA and PFOS which are already in the regulatory development process). For the purpose of CCL 5, the structural

definition of per- and polyfluoroalkyl substances (PFAS) 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.

EPA is committed to addressing PFAS in drinking water and the environment in general. In October 2021, EPA announced a comprehensive PFAS Strategic Roadmap which outlined an integrated approach for tackling PFAS challenges in the environment. This strategic roadmap lays out EPA's whole-of-agency approach to addressing PFAS. The strategic roadmap builds on and accelerates implementation of policy actions identified in the Agency's 2019 action plan and commits to bolder new policies to safeguard public health, protect the environment, and hold polluters accountable. EPA's integrated approach to PFAS is focused on three central directives:

• Research. Invest in research, development, and innovation to increase understanding of PFAS exposures and toxicities, human health and ecological effects, and effective interventions that incorporate the best available science.

• Restrict. Pursue a comprehensive approach to proactively prevent PFAS from entering air, land, and water at levels that can adversely impact human health and the environment.

• Remediate. Broaden and accelerate the cleanup of PFAS contamination to protect human health and ecological systems.

EPA's approach is shaped by the unique challenges to addressing PFAS contamination. EPA cannot solve the problem of "forever chemicals" by tackling one route of exposure or one use at a time. EPA will continue to pursue a rigorous scientific agenda to better characterize toxicities, understand exposure pathways, and identify new methods to avert and remediate PFAS pollution. As EPA learns more about the family of PFAS chemicals, the Agency can do more to protect public health and the environment.

Table 23. Chemical Contaminants on the CCL 5

Chemical Name	CASRN[1]	DTXSID[2]
1,2,3-Trichloropropane	96-18-4	DTXSID9021390
1,4-Dioxane	123-91-1	DTXSID4020533
17-alpha ethynyl estradiol	57-63-6	DTXSID5020576
2,4-Dinitrophenol	51-28-5	DTXSID0020523
2-Aminotoluene	95-53-4	DTXSID1026164
2-Hydroxyatrazine	2163-68-0	DTXSID6037807
6-Chloro-1,3,5-triazine-2,4-diamine	3397-62-4	DTXSID1037806
Acephate	30560-19-1	DTXSID8023846
Acrolein	107-02-8	DTXSID5020023
alpha-Hexachlorocyclohexane	319-84-6	DTXSID2020684
Anthraquinone	84-65-1	DTXSID3020095
Bensulide	741-58-2	DTXSID9032329
Bisphenol A	80-05-7	DTXSID7020182
Boron	7440-42-8	DTXSID3023922
Bromoxynil	1689-84-5	DTXSID3022162
Carbaryl	63-25-2	DTXSID9020247
Carbendazim (MBC)	10605-21-7	DTXSID4024729
Chlordecone (Kepone)	143-50-0	DTXSID1020770
Chlorpyrifos	2921-88-2	DTXSID4020458
Cobalt	7440-48-4	DTXSID1031040
Cyanotoxins[3]	Multiple	Multiple
Deethylatrazine	6190-65-4	DTXSID5037494
Desisopropyl atrazine	1007-28-9	DTXSID0037495
Desvenlafaxine	93413-62-8	DTXSID40869118
Diazinon	333-41-5	DTXSID9020407
Dicrotophos	141-66-2	DTXSID9023914
Dieldrin	60-57-1	DTXSID9020453
Dimethoate	60-51-5	DTXSID7020479
Disinfection byproducts (DBPs)[4]	Multiple	Multiple
Diuron	330-54-1	DTXSID0020446
Ethalfluralin	55283-68-6	DTXSID8032386
Ethoprop	13194-48-4	DTXSID4032611
Fipronil	120068-37-3	DTXSID4034609
Fluconazole	86386-73-4	DTXSID3020627
Flufenacet	142459-58-3	DTXSID2032552
Fluometuron	2164-17-2	DTXSID8020628
Iprodione	36734-19-7	DTXSID3024154
Lithium	7439-93-2	DTXSID5036761
Malathion	121-75-5	DTXSID4020791
Manganese	7439-96-5	DTXSID2024169
Methomyl	16752-77-5	DTXSID1022267
Methyl tert-butyl ether (MTBE)	1634-04-4	DTXSID3020833
Methylmercury	22967-92-6	DTXSID9024198
Molybdenum	7439-98-7	DTXSID1024207

Chemical Name	CASRN[1]	DTXSID[2]
Nonylphenol	25154-52-3	DTXSID3021857
Norflurazon	27314-13-2	DTXSID8024234
Oxyfluorfen	42874-03-3	DTXSID7024241
Per- and polyfluoroalkyl substances (PFAS)[5]	Multiple	Multiple
Permethrin	52645-53-1	DTXSID8022292
Phorate	298-02-2	DTXSID4032459
Phosmet	732-11-6	DTXSID5024261
Phostebupirim	96182-53-5	DTXSID1032482
Profenofos	41198-08-7	DTXSID3032464
Propachlor	1918-16-7	DTXSID4024274
Propanil	709-98-8	DTXSID8022111
Propargite	2312-35-8	DTXSID4024276
Propazine	139-40-2	DTXSID3021196
Propoxur	114-26-1	DTXSID7021948
Quinoline	91-22-5	DTXSID1021798
Tebuconazole	107534-96-3	DTXSID9032113
Terbufos	13071-79-9	DTXSID2022254
Thiamethoxam	153719-23-4	DTXSID2034962
Tri-allate	2303-17-5	DTXSID5024344
Tribufos	78-48-8	DTXSID1024174
Tributyl phosphate	126-73-8	DTXSID3021986
Trimethylbenzene (1,2,4-)	95-63-6	DTXSID6021402
Tris(2-chloroethyl) phosphate (TCEP)	115-96-8	DTXSID5021411
Tungsten	7440-33-7	DTXSID8052481
Vanadium	7440-62-2	DTXSID2040282

[1] Chemical Abstracts Service Registry Number (CASRN) is a unique identifier assigned by the Chemical Abstracts Service (a division of the American Chemical Society) to every chemical substance (organic and inorganic compounds, polymers, elements, nuclear particles, etc.) in the open scientific literature. It contains up to 10 digits, seperated by hyphens into three number strings.

[2] Distributed Structure Searchable Toxicity Substance Identifiers (DTXSID) is a unique substance identifier used in EPA's CompTox Chemicals database, where a substance can be any single chemical, mixture or polymer.

[3] Toxins naturally produced and released by some species of cyanobacteria (previously known as "blue-green algae"). The group of cyanotoxins includes, but is not limited to: anatoxin-a, cylindrospermopsin, microcystins, and saxitoxin.

[4] This group includes 23 unregulated DBPs as shown in Table 24.

[5] For the purpose of CCL 5, the structural definition of per- and polyfluoroalkyl substances (PFAS) includes chemicals that contain at least one of these three structures (except for PFOA and PFOS which are already in the regulatory process):

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

Table 24. Unregulated DBPs in the DBP Group on the CCL 5

Chemical Name	CASRN	DTXSID
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		
Dichloroacetonitrile (DCAN)	3018-12-0	DTXSID3021562
Dibromoacetonitrile (DBAN)	3252-43-5	DTXSID3024940
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		
Chlorate	14866-68-3	DTXSID3073137
Formaldehyde	50-00-0	DTXSID7020637

Chapter 5 CCL 5 Data Availability Assessment

Section 5.1 Overview

CCL 5 development process included assessing the current availability of data for the chemical contaminants listed on CCL 5 and PCCL 5. In later steps, upon finalizing the CCL 5, 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.

Section 5.2 Data Availability for CCL 5 Chemicals

EPA provides the initial assessment of the current data availability of chemical contaminants on CCL 5 in Table 25. 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 5 contaminants and for further evaluation of contaminants under the Regulatory Determination 5 process (RD 5).

Contaminants in Group A have nationally representative finished drinking water data and qualifying health assessments. Contaminants in Group B have finished water data that are not nationally representative and qualifying health assessments. Contaminants in groups C, D, and E lack either a qualifying health assessment or finished water data and have more substantial data needs. EPA did not assess data availability for the cyanotoxins, DBPs, and PFAS groups because the availability of health effects and occurrence data varies with individual chemicals in each group. 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 may cause adverse health effects.

Table 25. Data Availability for CCL 5 Chemicals

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Is a Health Assessment Available?	Is an Analytical Method Available?
A. Contaminants with Nationally Representative Finished Water Occurrence Data and Qualifying Health Assessments					
1,2,3-Trichloropropane	96-18-4	DTXSID9021390	National finished water	Yes	Yes
1,4-Dioxane	123-91-1	DTXSID4020533	National finished water	Yes	Yes
2,4-Dinitrophenol	51-28-5	DTXSID0020523	National finished water	Yes	Yes
2-Aminotoluene	95-53-4	DTXSID1026164	National finished water	Yes	Yes
alpha-Hexachlorocyclohexane	319-84-6	DTXSID2020684	National finished water	Yes	Yes
Boron	7440-42-8	DTXSID3023922	National finished water	Yes	Yes
Carbaryl	63-25-2	DTXSID9020247	National finished water	Yes	Yes

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Is a Health Assessment Available?	Is an Analytical Method Available?
Chlorpyrifos	2921-88-2	DTXSID4020458	National finished water	Yes	Yes
Cobalt	7440-48-4	DTXSID1031040	National finished water	Yes	Yes
Dieldrin	60-57-1	DTXSID9020453	National finished water	Yes	Yes
Diuron	330-54-2	DTXSID0020446	National finished water	Yes	Yes
Ethoprop	13194-84-4	DTXSID4032611	National finished water	Yes	Yes
Lithium	7439-93-2	DTXSID5036761	National finished water	Yes	Yes
Manganese	7439-96-5	DTXSID2024169	National finished water	Yes	Yes
Molybdenum	7439-98-7	DTXSID1024207	National finished water	Yes	Yes
Oxyfluorfen	42874-03-3	DTXSID7024241	National finished water	Yes	Yes
Permethrin	52645-53-1	DTXSID8022292	National finished water	Yes	Yes
Profenofos	41198-08-7	DTXSID3032464	National finished water	Yes	Yes
Propachlor	1918-16-7	DTXSID4024274	National finished water	Yes	Yes
Quinoline	91-22-5	DTXSID1021798	National finished water	Yes	Yes
Tebuconazole	107534-96-3	DTXSID9032113	National finished water	Yes	Yes
Tribufos	78-48-8	DTXSID1024174	National finished water	Yes	Yes
Vanadium	7440-62-2	DTXSID2040282	National finished water	Yes	Yes
B. Contaminants with Non-Nationally Representative Finished Water Occurrence Data and Qualifying Health Assessments					
2-Hydroxyatrazine	2163-68-0	DTXSID6037807	Non-national finished water	Yes	No
Bromoxynil	1689-84-5	DTXSID3022162	Non-national finished water	Yes	No
Carbendazim (MBC)	10605-21-7	DTXSID4024729	Non-national finished water	Yes	No
Dicrotophos	141-66-2	DTXSID9023914	Non-national finished water	Yes	Yes
Ethalfluralin	55283-68	DTXSID8032386	Non-national finished water	Yes	No
Fipronil	120068-37-3	DTXSID4034609	Non-national finished water	Yes	No

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Is a Health Assessment Available?	Is an Analytical Method Available?
Fluometuron	2164-17-2	DTXSID8020628	Non-national finished water	Yes	Yes
Iprodione	36734-19-7	DTXSID3024154	Non-national finished water	Yes	No
Malathion	121-74-5	DTXSID4020791	Non-national finished water	Yes	Yes
Norflurazon	27314-13	DTXSID8024234	Non-national finished water	Yes	Yes
Phorate	298-02-2	DTXSID4032459	Non-national finished water	Yes	Yes
Phosmet	732-11-6	DTXSID5024261	Non-national finished water	Yes	No
Propanil	709-98-8	DTXSID8022111	Non-national finished water	Yes	Yes
Propargite	2312-35-8	DTXSID4024276	Non-national finished water	Yes	No
Propazine	139-40-2	DTXSID3021196	Non-national finished water	Yes	Yes
Propoxur	114-26-1	DTXSID7021948	Non-national finished water	Yes	Yes
Tebupirimfos	96182535	DTXSID1032482	Non-national finished water	Yes	No
Thiamethoxam	153719-23-4	DTXSID2034962	Non-national finished water	Yes	No
Tri-allate	2303-17-5	DTXSID5024344	Non-national finished water	Yes	No
C. Contaminants with Nationally Representative Finished Water Occurrence Data Lacking Qualifying Health Assessments					
17-alpha ethynyl estradiol	57-63-6	DTXSID5020576	National Finished Water	No	Yes
Methyl tert-butyl ether (MTBE)	1634-04-4	DTXSID3020833	National finished water	No	Yes
D. Contaminants with Qualifying Health Assessments Lacking Finished Water Occurrence Data					
6-Chloro-1,3,5-triazine-2,4-diamine	3397-62-4	DTXSID1037806	National ambient water	Yes	Yes
Acephate	30560-19-1	DTXSID8023846	National ambient water	Yes	Yes
Acrolein	107-02-8	DTXSID5020023	National ambient water	Yes	No
Anthraquinone	84-65-1	DTXSID3020095	National ambient water	Yes	No
Bensulide	741-58-2	DTXSID9032329	National ambient water	Yes	Yes
Bisphenol A	80-05-7	DTXSID7020182	National ambient water	Yes	No
Chlordecone (Kepone)	143-50-0	DTXSID1020770	Non-national ambient water	Yes	Yes
Deethylatrazine	6190-65-4	DTXSID5037494	National ambient water	Yes	No

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Is a Health Assessment Available?	Is an Analytical Method Available?
Desisopropyl atrazine	1007-28-9	DTXSID0037495	National ambient water	Yes	Yes
Diazinon	333-41-5	DTXSID9020407	National ambient water	Yes	Yes
Dimethoate	60-51-5	DTXSID7020479	National ambient water	Yes	Yes
Flufenacet (Thiaflumide)	142459-58-3	DTXSID2032552	National ambient water	Yes	No
Methomyl	16752-77-5	DTXSID1022267	National ambient water	Yes	Yes
Methylmercury	22967-92-6	DTXSID9024198	National ambient water	Yes	No
Terbufos	13071-79-9	DTXSID2022254	National ambient water	Yes	Yes
Tributyl phosphate	126-73-8	DTXSID3021986	National ambient water	Yes	No
Trimethylbenzene (1,2,4-)	95-63-6	DTXSID6021402	National ambient water	Yes	Yes
Tris(2-chloroethyl) phosphate (TCEP)	103476-24-0	DTXSID5021411	National ambient water	Yes	No
Tungsten	7440-33-7	DTXSID8052481	National ambient water	Yes	No
E. Contaminants Lacking Nationally Representative Finished Water Occurrence Data and Qualifying Health Assessments					
Desvenlafaxine	93413-62-8	DTXSID40869118	Non-national finished water	No	No
Fluconazole	86386-73-4	DTXSID3020627	Non-national finished water	No	No
Nonylphenol	104-40-5	DTXSID3021857	Non-national finished water	No	Yes

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.

The occurrence and health effects data used to categorize data availability can be found on the CISs. The following sections describe the types of data or information gaps listed in Table 25 and provide examples of contaminants that fall into each group.

Section 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, EPA is required to identify contaminants that were known or anticipated to occur in PWSs. EPA used nationally representative finished drinking water data as the best available occurrence information. However, in the absence of national representative finished water data, non-nationally representative finished drinking water occurrence data were also used. 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 25, EPA categorized occurrence data needs as follows:

- Finished drinking water occurrence data that are nationally representative. Data sources may include:
  - UCMRs (i.e., UCMR 1, UCMR 2, UCMR 3 and UCMR 4), the Unregulated Contaminant Monitoring – State (Round 1 and Round 2) and NIRS.
- Finished drinking water occurrence data that are not nationally representative. These data may include:
  - Finished water assessments by federal agencies (e.g., EPA, the US Department of Agriculture and 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 drinking 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).

Section 5.2.2 Health Effects

Under the regulatory determination process, EPA generally relies on externally 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 4) included EPA health assessments or peer-reviewed health assessments developed by other organizations such as the Agency for Toxic Substances and Disease Registry, World Health Organization, Health Canada, and the California EPA's Office of Environmental Health Hazard Assessment. The health assessment must have been peer-reviewed and must have used comparable methods, standards, and guidelines to an EPA health assessment.

For the CCL 5, as shown in Table 25, EPA categorized the health effects data availability in the following way:

- Health effects data are available. A peer-reviewed health assessment is available or is in the process of being revised.
- Health effects data currently not available. A peer-reviewed health assessment is not available or existing assessments do not include the derivation of toxicity values.

Section 5.2.3 Analytical Methods

To conduct nationally representative drinking water occurrence studies that could support a regulatory determination, 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 the purpose of CCL 5, EPA assessed the status of the development of analytical methods for drinking water.

- Though many methods for monitoring the CCL 5 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, as of September 2020, is presented in Table 25. For CCL 5 method availability assessment, EPA only considered EPA validated drinking water methods.

Section 5.3 Data Availability for PCCL 5 Chemicals not on CCL 5

To ensure that the evaluated chemicals most relevant to drinking water exposure were included on the CCL 5, EPA also assessed the data availability of PCCL 5 chemicals not included on CCL 5. 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.

For PCCL 5 chemicals not listed on CCL 5 with a health concentration (HRL or CCL screening level (CCL SL)) derived during the CCL 5 classification (see Chapter 4), EPA noted it in Table 26. For other chemicals, EPA provided the screening tier (see Chapter 3) for the best available health effects data. The health effects tiers established during the CCL 5 screening are:

- Tier 1 (T 1) health effects data including reference doses, cancer slope factors, 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 reference doses, and subchronic health-based concentrations
- Tier 4 (T 4) health effects data including acute RfDs, subchronic LOAELs, subchronic NOAELs, MRDDs, or a chemical is present on a list of known human neurotoxicants, and known neurodevelopmental disruptors
- Tier 5 (T 5) health effects data including TD50, LD50, ToxCast assay percent active, and number of PubMed articles

Contaminants were categorized into these three occurrence groups, as shown in Table 26:

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

The occurrence and health effects data listed in Table 26 are the best available data for that chemical contaminant.

Table 26. Data Availability for PCCL 5 Chemicals not on CCL 5

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Best Available Health Effects Data
Group A. Contaminants with Nationally Representative Finished Water Data				
1,3-Butadiene	106-99-0	DTXSID3020203	Finished National	T 3
1,3-Dichloropropene	542-75-6	DTXSID1022057	Finished National	HRL
17-beta estradiol	50-28-2	DTXSID0020573	Finished National	CCL SL
1-Butanol	71-36-3	DTXSID1021740	Finished National	HRL
2,4-Dichlorophenol	120-83-2	DTXSID1020439	Finished National	HRL
2,4-Dinitrotoluene	121-14-2	DTXSID0020529	Finished National	HRL
2,6-Dinitrotoluene	606-20-2	DTXSID5020528	Finished National	HRL
4-Androstene-3,17-dione	63-05-8	DTXSID8024523	Finished National	CCL SL
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
Bromochloromethane	74-97-5	DTXSID4021503	Finished National	T 1
Calcium	7440-70-2	DTXSID9050484	Finished National	T 5
Chlorodifluoromethane	75-45-6	DTXSID6020301	Finished National	T 3
Chloromethane	74-87-3	DTXSID0021541	Finished National	T 3
EPTC	759-94-4	DTXSID1024091	Finished National	HRL
Linuron	330-55-2	DTXSID2024163	Finished National	HRL
Magnesium	7439-95-4	DTXSID0049658	Finished National	CCL SL
Metolachlor ESA	171118-09-5	DTXSID1037567	Finished National	CCL SL
Metolachlor OA	152019-73-3	DTXSID6037568	Finished National	CCL SL
p,p'-DDE	72-55-9	DTXSID9020374	Finished National	HRL
Phosphorus	7723-14-0	DTXSID1024382	Finished National	CCL SL
Potassium	7440-09-7	DTXSID9049748	Finished National	T 5
Prometon	1610-18-0	DTXSID6022341	Finished National	HRL
Silicon	7440-21-3	DTXSID0051441	Finished National	T 5
Sodium	7440-23-5	DTXSID1049774	Finished National	HRL
Terbacil	5902-51-2	DTXSID8024317	Finished National	HRL
Testosterone	58-22-0	DTXSID8022371	Finished National	CCL SL
Tin	7440-31-5	DTXSID1049801	Finished National	T 3
Group B. Contaminants with Non-Nationally Representative Finished Water Data				
2-(2-Methyl-4-chlorophenoxy)propionic acid	93-65-2	DTXSID9024194	Finished Non-National	HRL
2,4-Dichlorophenoxybutyric acid	94-82-6	DTXSID7024035	Finished Non-national	HRL
2-Methyl-4-chlorophenoxyacetic acid	94-74-6	DTXSID4024195	Finished Non-National	HRL
2-Methylnaphthalene	91-57-6	DTXSID4020878	Finished Non-National	HRL
Acetamiprid	135410-20-7	DTXSID0034300	Finished Non-National	HRL

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Best Available Health Effects Data
Acetophenone	98-86-2	DTXSID6021828	Finished Non-National	HRL
Acyclovir	59277-89-3	DTXSID1022556	Finished Non-National	CCL SL
Aldrin	309-00-2	DTXSID8020040	Finished Non-National	HRL
Ammonia	7664-41-7	DTXSID0023872	Finished Non-National	T 1
Atenolol	29122-68-7	DTXSID2022628	Finished Non-National	CCL SL
Azoxystrobin	131860-33-8	DTXSID0032520	Finished Non-National	HRL
Benfluralin	1861-40-1	DTXSID3023899	Finished Non-National	HRL
Bentazon	25057-89-0	DTXSID0023901	Finished Non-National	HRL
Benzophenone	119-61-9	DTXSID0021961	Finished Non-National	CCL SL
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
Bupropion	34911-55-2	DTXSID7022706	Finished Non-National	CCL SL
Caffeine	58-08-2	DTXSID0020232	Finished Non-National	T 3
Camphor	76-22-2	DTXSID5030955	Finished Non-National	T 5
Carbamazepine	298-46-4	DTXSID4022731	Finished Non-National	CCL SL
Carbon disulfide	75-15-0	DTXSID6023947	Finished Non-National	HRL
Chlorothalonil	1897-45-6	DTXSID0020319	Finished Non-National	HRL
Clomazone	81777-89-1	DTXSID1032355	Finished Non-National	HRL
Clopyralid	1702-17-6	DTXSID9029221	Finished Non-National	HRL
Clothianidin	210880-92-5	DTXSID2034465	Finished Non-National	HRL
Cotinine	486-56-6	DTXSID1047576	Finished Non-National	T5
Cycloate	1134-23-2	DTXSID6032356	Finished Non-National	HRL
Cyfluthrin	68359-37-5	DTXSID5035957	Finished Non-National	HRL
Cyhalothrin	68085-85-8	DTXSID6023997	Finished Non-National	HRL
Cypermethrin	52315-07-8	DTXSID1023998	Finished Non-National	HRL
Diazepam	439-14-5	DTXSID4020406	Finished Non-National	CCL SL
Dicamba	1918-00-9	DTXSID4024018	Finished Non-National	T 3
Dichlorvos	62-73-7	DTXSID5020449	Finished Non-National	HRL
Difenoconazole	119446-68-3	DTXSID4032372	Finished Non-National	HRL
Dimethenamid	87674-68-8	DTXSID4032376	Finished Non-National	HRL
Dimethenamid OXA	380412-59-9	DTXSID4037530	Finished Non-National	CCL SL
Esfenvalerate	66230-04-4	DTXSID4032667	Finished Non-National	HRL
Ethion	563-12-2	DTXSID2024086	Finished Non-National	HRL
Fenbuconazole	114369-43-6	DTXSID8032548	Finished Non-National	HRL
Fenitrothion	122-14-5	DTXSID4032613	Finished Non-National	HRL
Fenpropathrin	39515-41-8	DTXSID0024002	Finished Non-National	HRL
Fenthion	55-38-9	DTXSID8020620	Finished Non-National	HRL
Fexofenadine	83799-24-0	DTXSID00861411	Finished Non-National	CCL SL
Fluoranthene	206-44-0	DTXSID3024104	Finished Non-National	HRL
Fluoxetine	54910-89-3	DTXSID7023067	Finished Non-National	CCL SL
Galaxolide	1222-05-5	DTXSID8027373	Finished Non-National	CCL SL
Gemfibrozil	25812-30-0	DTXSID0020652	Finished Non-National	CCL SL
Hexazinone	51235-04-2	DTXSID4024145	Finished Non-National	HRL

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Best Available Health Effects Data
Imazapyr	81334-34-1	DTXSID8034665	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
Isophorone	78-59-1	DTXSID8020759	Finished Non-National	HRL
Isopropylbenzene	98-82-8	DTXSID1021827	Finished Non-National	HRL
Isoxaflutole	141112-29-0	DTXSID5034723	Finished Non-National	HRL
lambda-Cyhalothrin	91465-08-6	DTXSID7032559	Finished Non-National	HRL
Lidocaine	137-58-6	DTXSID1045166	Finished Non-National	CCL SL
Loratadine	79794-75-5	DTXSID2023224	Finished Non-National	CCL SL
Meprobamate	57-53-4	DTXSID3023261	Finished Non-National	CCL SL
Metalaxyl	57837-19-1	DTXSID6024175	Finished Non-National	HRL
Metformin	657-24-9	DTXSID2023270	Finished Non-National	CCL SL
Methocarbamol	532-03-6	DTXSID6023286	Finished Non-National	CCL SL
Methylbenzotriazole	29385-43-1	DTXSID0026171	Finished Non-National	CCL SL
Metoprolol	51384-51-1	DTXSID2023309	Finished Non-National	CCL SL
Metribuzin	21087-64-9	DTXSID6024204	Finished Non-National	HRL
Morphine	57-27-2	DTXSID9023336	Finished Non-National	CCL SL
Myclobutanil	88671-89-0	DTXSID8024315	Finished Non-National	HRL
N,N-Diethyl-m-toluamide	134-62-3	DTXSID2021995	Finished Non-National	T 4
Nicotine	54-11-5	DTXSID1020930	Finished Non-National	CCL SL
Oxadiazon	19666-30-9	DTXSID3024239	Finished Non-National	HRL
p-Cresol	106-44-5	DTXSID7021869	Finished Non-National	HRL
Pendimethalin	40487-42-1	DTXSID7024245	Finished Non-National	HRL
Phenanthrene	85-01-8	DTXSID6024254	Finished Non-National	T 3
Piperonyl butoxide	51-03-6	DTXSID1021166	Finished Non-National	HRL
Prometryn	7287-19-6	DTXSID4024272	Finished Non-National	HRL
Pronamide	23950-58-5	DTXSID2020420	Finished Non-National	HRL
Propiconazole	60207-90-1	DTXSID8024280	Finished Non-National	HRL
Prosulfuron	94125-34-5	DTXSID9034868	Finished Non-National	HRL
Pyrene	129-00-0	DTXSID3024289	Finished Non-National	HRL
Sitagliptin	486460-32-6	DTXSID70197572	Finished Non-National	CCL SL
Sulfamethoxazole	723-46-6	DTXSID8026064	Finished Non-National	CCL SL
Tamoxifen	10540-29-1	DTXSID1034187	Finished Non-National	CCL SL
Tebuthiuron	34014-18-1	DTXSID3024316	Finished Non-National	HRL
Tefluthrin	79538-32-2	DTXSID5032577	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
Triclopyr	55335-06-3	DTXSID0032497	Finished Non-National	HRL
Triclosan	3380-34-5	DTXSID5032498	Finished Non-National	HRL
Triethyl citrate	77-93-0	DTXSID0040701	Finished Non-National	CCL SL
Trifluralin	1582-09-8	DTXSID4021395	Finished Non-National	HRL
Tris(1,3-dichloro-2-propyl) phosphate	13674-87-8	DTXSID9026261	Finished Non-National	HRL

Chemical Name	CASRN	DTXSID	Best Available Occurrence Data	Best Available Health Effects Data
Tris(2-butoxylethyl) phosphate	78-51-3	DTXSID5021758	Finished Non-National	T 4
Verapamil	52-53-9	DTXSID9041152	Finished Non-National	CCL SL
Group C. Contaminants Lacking Finished Water Data				
1,1,2,2-Tetrachloroethane	79-34-5	DTXSID7021318	Ambient National	HRL
4-tert-Octylphenol	140-66-9	DTXSID9022360	Ambient National	CCL SL
Ametryn	834-12-8	DTXSID1023869	Ambient National	HRL
Butyl benzyl phthalate	85-68-7	DTXSID3020205	Ambient National	HRL
Cyprodinil	121552-61-2	DTXSID1032359	Ambient National	HRL
Diethyl phthalate	84-66-2	DTXSID7021780	Ambient National	HRL
Di-n-butyl phthalate	84-74-2	DTXSID2021781	Ambient National	HRL
Famoxadone	131807-57-3	DTXSID8034588	Ambient National	HRL
Heroin	561-27-3	DTXSID6046761		
Imazalil	35554-44-0	DTXSID8024151	Ambient National	HRL
Indoxacarb	173584-44-6	DTXSID1032690	Ambient National	HRL
Lactofen	77501-63-4	DTXSID7024160	Ambient National	HRL
Morphine-3-glucuronide	20290-09-9	DTXSID80174157		
Naled	300-76-5	DTXSID1024209	Ambient National	HRL
Naphthalene	91-20-3	DTXSID8020913	Ambient National	HRL
Phenol	108-95-2	DTXSID5021124	Ambient National	HRL
Pymetrozine	123312-89-0	DTXSID2032637	Ambient National	HRL
Pyraclostrobin	175013-18-0	DTXSID7032638	Ambient National	HRL
Pyridaben	96489-71-3	DTXSID5032573	Ambient National	HRL
Sulfentrazone	122836-35-5	DTXSID6032645	Ambient National	HRL
Sulfomethuron-methyl	74222-97-2	DTXSID0034936	Ambient National	HRL
Thiram	137-26-8	DTXSID5021332	Pesticide Application	HRL
Trifloxystrobin	141517-21-7	DTXSID4032580	Ambient National	HRL

Chapter 6 Data Management and Quality Assurance

Section 6.1 Overview

All steps of the CCL 5 development process underwent quality assurance/quality control (QA/QC) activities to ensure the integrity of the data and calculations used to generate CCL 5. The process consisted of two phases: QA/QC of the PCCL 5 Development (Section 6.2) and QA/QC of Contaminant Information Sheets (CISs) (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 5 Universe file, the screening code, classification data files, and the CISs were developed primarily using the R programming language. All code written to extract data from either primary 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. After building the CCL 5 Universe, EPA conducted input checks, such as verifying that the original source data matched the data contained in the CCL 5 Universe file. To check the accuracy of the screening code and ensure screening points were assigned correctly, EPA also conducted output checks. This entailed reviewing screening point assignments for a select sample of 20 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 two rounds of QA/QC in which data values on the CISs were spot-checked against the original data in the input files. Further details about QA/QC of the PCCL 5 development and CISs are described in the following sections.

Section 6.2 Quality Assurance of PCCL 5 Development

Section 6.2.1 Overview

For the PCCL development process, EPA wrote code using R (version 4.0.2) (R Core Team, 2020) and documented it using R Markdown (version 2.6) (Allaire et al., 2020). This allowed for transparent documentation and organization of the PCCL process and QA/QC activities. The EPA developed R Markdown files, which documented the PCCL 5 process, including the following:

- A series of individual R Markdown documents dedicated to pre-processing a primary data source (referred to as "pre-processing code" hereafter).
  - The goal of the pre-processing code was to extract and transform data relevant to screening from primary data sources to a simple data format. Details on the simple data format is described in Section 2.3.4 and Appendix N. The output of the pre-processing code are "simple" data files associated with each primary data source.
- Three separate R Markdown documents, which were used to develop the PCCL 5.
  - The first document, Making the Pre-Universe, was to aggregate the simple data files produced from the pre-processing code into the pre-universe file described in Section 2.3.
  - The second document, ID and Screen, was to manually correct DTXSIDs as necessary, assign unique internal-use NO_DTXSID identifiers for contaminants without existing DTXSIDs, and add 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 document, Screening (referred to as screening code hereafter), assessed the data in the universe file, 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 process.

Section 6.2.2 Reviewing Input Data

The first QA/QC activity was reviewing the input data used to build the pre-universe. EPA randomly sampled 300 data entries from the pre-universe file and checked them against the original data source. EPA ensured all primary data 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. EPA assessed the error rate using the beta distribution in R. Briefly, by checking N entries and finding K of them defective (errors), the estimated error rate would be K/N. An upper 92% confidence interval for the error rate estimate was calculated using the beta distribution: qbeta (0.92, K +0.5, N-K+0.5) where 0.5 are shape parameters for the beta distribution. For example, if 300 data entries (N = 300) are reviewed and zero errors are found (K = 0), the error rate is 0.5% and the null hypothesis could be rejected. If one error was found, the error rate would be 1.12% with 92% confidence, which is above the threshold or 1% or less error.

This QA/QC activity was intended to be a check of the input values, but it also captured any errors introduced in the pre-processing code. For example, a value could have been downloaded correctly from the original source but unintentionally corrupted when the data were written to a simple format file. The pre-universe file is an aggregate of the simple format files containing data elements from primary data sources relevant to the PCCL process. Therefore, checking random data entries in the pre-universe file also caught errors in the original source data and in the pre-processing code.

This QA/QC activity did not identify any errors in the original source data. However, the review identified one error introduced by the pre-processing code where data were being misclassified as a "factor" data type rather than a "numeric" data type. These data were used in the calculation of half of the method reporting limit (MRL) for maximum concentration for non-detects. The factor classification resulted in an incorrect calculation. EPA corrected the error in the code and ensured other pre-processing code documents did not include this error.

With one error identified in the 300 random samples of input data from the pre-universe file selected for review, the error rate for this QA/QC activity was > 1%. However, after correcting the one identified error, EPA moved forward with reviewing the pre-processing code and did not resample the pre-universe file for another round of input review. The reason was that the final QA/QC activity for the PCCL development included a review of the output from the screening code (Section 6.2.6). The input to the screening code was the universe file, so checking the output values would effectively repeat the input review process.

Section 6.2.3 Reviewing Pre-processing Code

The second QA/QC activity of the PCCL 5 development process was a systematic review of data processing for primary data sources used to generate the pre-universe file. The review was conducted by two team members who are proficient in R programming and code review by checking the functionality of the pre-processing code. Other members were responsible for reviewing the policy decisions

embedded in the data processing, such as how the data are labeled, how non-detections are treated, and other particularities specific to the given data source. EPA documented and addressed any coding errors identified by the QA/QC review team.

Section 6.2.4 QA/QC Procedure for DTXSID Assignments

The third QA/QC activity of PCCL 5 development was a review of DTXSID assignments to data entries in the pre-universe file. 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

The method used to assign DTXSIDs evolved over the course of the PCCL coding process. At the beginning, DTXSIDs were added to data in the pre-processing code using a mapping file downloaded from the CompTox Chemicals Dashboard, as described in Section 2.3.3. DTXSIDs were added when extracted data were written to a simple data format file. Any data entries that could not be automatically assigned a DTXSID using the mapping files were temporarily assigned a label of NA or NO_DTXSID. When compiling the simple data format files to form the pre-universe file, EPA manually reviewed and assigned DTXSIDs to data entries labeled with NA or NO_DTXSID. In some cases, no DTXSID existed for the data entry, but the data entry also represented a substance or data point relevant to the CCL. Examples include the number of biological specimens counted in a waterbody or mixtures of vapors that emerge from asphalt and street-paving activities. EPA characterized these data entries into one of three categories: not able to identify, not a chemical substance, or mixture of substances. Prior to finalizing the universe file, entries that fit into one of these categories were removed.

EPA changed the method of assigning DTXSIDs from using the mapping file, which was used in the pre-processing data, to using a more efficient batch search function in the CompTox Chemicals Dashboard. The Chemicals CompTox Dashboard is continuously being updated and refined. However, the mapping file is static (i.e., not updated over the course of the PCCL development process) and does not reflect subsequent updates to the CompTox Chemicals Dashboard. EPA determined that a more efficient approach would be to manually amend downloaded data with DTXSIDs using the batch search function, in which the original source data file was amended with DTXSIDs downloaded from the CompTox Chemicals Dashboard. Any entry without a DTXSID was designated with NO_DTXSID or NA. If the batch download resulted in no match, additional searching was performed to try to find the appropriate DTXSID. After the simple data files were compiled to form the pre-universe file, any data entries with no DTXSID were manually reviewed. If no DTXSID could be assigned, these entries remained in the pre-universe file and a unique internal-use DTXSID was assigned using the prefix NO_DTXSID followed by a unique number (further described in Section 2.4.2).

In September 2019, EPA reversed its initial decision to remove out of scope data entries from the pre-universe. As a result of QA checks, EPA determined that the method used to remove data entries from the pre-universe was not applied uniformly. Some entries that described mixtures (e.g., a data entry for xylenes could include mixtures of o, p, and m-xylene) had been automatically assigned DTXSIDs, so were not identified as pertaining to a mixture. This resulted in an uneven use of the rule for mixtures.

Because the PCCL development process was iterative (i.e., assigning DTXSIDs, manually reviewing data entries with missing DTXSIDs, assigning DTXSIDs, and assessing if data entries are out of scope over the course of a year), there was also concern that the standard for removing data entries evolved over time. Therefore, EPA decided to retain all data entries and pre-universe chemical contaminants in the universe to reduce the risk of removing data entries and chemicals that are relevant to the CCL.

The QA/QC of DTXSID assignments occurred in two phases. In the first phase, a random sample of 337 data entries was reviewed that were manually assigned DTXSIDs and previously identified as a mixture, not a chemical substance, or not able to identify. As a result of this review, EPA determined that these designations were poorly defined. Therefore, EPA took an updated random sample of 337 data entries of only the contaminants for which DTXSIDs were not automatically assigned and manually searched for the correct DTXSID.

EPA identified two types of errors in the QA/QC of DTXSID assignments. The first type of error is defined as an entry that was manually assigned an incorrect DTXSID. The second type of error is defined as a data entry that was previously assigned NO_DTXSID but for which a DTXSID was identified during the QA/QC. In the random sample used for QA (N=337), one data entry was assigned an incorrect DTXSID. EPA corrected this error in the R Markdown documents used to develop the PCCL. With one error associated with manual assignment of DTXSID numbers, the error rate for contaminants assigned a wrong DTXSID number is <1% with 92% confidence.

EPA identified 18 data entries that were previously assigned NO_DTXSID to have a DTXSID in the CompTox Chemicals Dashboard. Twelve of these were data entries associated with contaminants that could be confidently identified as a single chemical and relevant to the CCL process. The remaining six could be classified as mixtures, considered not a chemical substance, or represented a group of chemicals. Examples of these data entries include *Bacillus amyloliquifacien* (a bacteria) or metabisulfites (a group of compounds). Upon further investigation of the 12 data entries that previously did not have DTXSID numbers assigned and are relevant to the CCL process, EPA identified two entries, N-Acetyl-S-(3,4-dihydroxybutyl)-L-cysteine and N-Methyl-N-(3-oxopropyl)nitrous amide, that were assigned DTXSIDs on June 20, 2019, on the CompTox Chemicals Dashboard. The process of manually assigning DTXSID numbers to data entries in the universe occurred before June 20, 2019, and these contaminants may not yet have had a DTXSID assigned. With 12 errors associated with data entries relevant to the CCL, the error rate is 5.2% with 92% confidence.

Section 6.2.5 QA/QC Procedure for Screening Code

The fourth QA/QC activity was a detailed and rigorous review of all R code and R Markdown files written for the PCCL development process. The screening code review was conducted by an EPA reviewer who was not the primary code developer. 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

The EPA reviewer also reviewed the primary literature papers used to build the universe to ensure the data were being interpreted as intended and presented in the literature. No errors were identified as a result of the screening code review. The reviewer suggested several improvements to the coding style and efficiency of the scripts that had no impact on the code output. Subsequently, EPA incorporated improvements to the respective R Markdown files.

Section 6.2.6 QA/QC Procedure for Outputs

The final QA/QC activity of PCCL development was a review of output values of the screening code. This activity occurred in two phases. In the first phase, EPA checked a random sample of data entries in the Scored Universe file against the original source data to make sure data were not unintentionally altered. In the second phase, EPA reviewed data entries for 20 contaminants in the Scored Universe file to confirm all screening points were assigned correctly and summed to the expected screening score.

The process for checking the output values from the screening code against the original source data is effectively a repetition of the process used to check the input values. The QA/QC review team checked 337 data entries selected from the Scored Universe file against the original data to make sure that values and units were reported correctly. EPA identified zero errors in the output values as a result of this QA/QC review. This confirmed the results of the input check, which identified only one error in 300 samples of input data entries (Section 6.2.2). However, a QA/QC reviewer recommended changing the approach of imputing maximum concentration for non-detects in the USDA Pesticide Database Program (PDP) data. For some PDP compounds, limits of detections (LOD) are reported in the original source data as a range of values rather than a single value. In these cases, the average of the two values had been used to calculate the "maximum concentration" (half the average LOD). For CCL 5 EPA changed this approach and instead used half of the value of the midpoint between the minimum and maximum detection limits as the maximum concentration, as described in Section 2.3.2.

Twenty contaminants from the Scored Universe file were selected so that each data element involved in screening to a PCCL was represented in the review. For each compound, data entries that were assigned screening points were checked to ensure the assigned screening points matched the screening point assignment hierarchy (Chapter 3). EPA identified several issues in the review and determined these errors were systemic within the code, resulting in identical issues across several reviewed chemicals. Examples of errors include incorrect screening points being assigned for pesticide application rate data, environmental release data, and subchronic benchmarks. EPA corrected errors in the screening code R Markdown document, and screening points assignments were corrected across all chemical contaminants in the universe.

Section 6.3 Quality Assurance of CIS Development

Section 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 5, EPA identified and gathered data on the health effects and occurrence of each of the PCCL 5 chemicals evaluated then summarized this information on the CISs. This section also describes 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.

Section 6.3.2 Preparing Health Effects Data for CISs

Data extracted from health assessments were manually compiled in an Excel workbook. These data included reference doses, cancer slope factors, health endpoints, and information about the assessment (title, date of publication, citation). EPA used information extracted from the health assessments to identify the relevant target population and calculate health concentrations and attribute scores (severity and potency), all of which were included in the same Excel workbook. Two EPA staff members were responsible for ensuring accurate data extraction and calculations, one to perform the initial extractions and another to check for accuracy.

Preparation of the health effects data files for CIS development was performed using R scripts.[5] EPA spot-checked the scripts that read in the data, checked that formatting was consistent, and checked that there were no duplicates or contradictory data. The Excel workbook with data extracted from health assessments was reformatted for data placement onto the CIS Summary + Decision tab and Health Effects tab. Some of the health assessment data (e.g., RfD, CSF, and cancer classification data) were then converted using R into the simple data format as described in Appendix N.

EPA compiled chemical use information and CAS Registry numbers for the 214 chemicals reviewed by the two evaluation teams. These data were compiled in the simple format and reviewed by an EPA staff member. EPA selected 10 chemicals to check for the accuracy of CAS Registry numbers and chemical use information and did not identify any errors.

Other relevant health effects and summary data (e.g., previous CCL listing decisions, previous Regulatory Determination decisions, and the literature search summary) were also converted to simple format using an R script where necessary and were subsequently combined with the simple format health assessment data. 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.

Section 6.3.3 Preparing Occurrence Data for CISs

To extract and compile water occurrence data in support of the classification step (Chapter 4), EPA wrote code using R, referred to as R scripts, which was documented in R Markdown. EPA extracted occurrence related data elements, such as detection and concentration statistics (minimum, median, 90th percentile, maximum concentrations based on detects), and others for CIS development, occurrence attribute scoring, and fHQ calculations. This section describes the procedures EPA undertook to extract and gather occurrence data from primary data sources and supplemental occurrence data sources, such as data sources suggested through the CCL 5 public nominations process (Section 2.2.2 and Section 3.6) and data sources identified through literature searches (Section 4.2). This section also describes the QA/QC activities implemented during this process. See Appendix N for specific data processing information for primary data sources.

[5] Data restructuring and analysis were conducted using R version 3.6.2 (RStudio version 1.3.959). Every script was written in R Markdown (Allaire et al., 2020) 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. R script version control was maintained through a repository on GitLab, a web-based software development and IT operations lifecycle tool.

EPA developed a series of R scripts to extract and transform classification relevant data elements from eight occurrence related primary data sources (UCM Rounds 1 and 2, UCMR 1-3, Water Quality Portal (NWIS and NAWQA), NIRS, Toxics Release Inventory (TRI), USGS Pesticide Use Estimates, Furlong et al., 2017; and Glassmeyer et al., 2017). The R scripts were documented in R Markdown (Allaire et al., 2020).[6] 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 is described in Appendix N. Source and contaminant metadata such as water type and monitoring year ranges, were also included where possible.

The R scripts were reviewed by a QA/QC reviewer to ensure the functionality of the code. After the R scripts had passed code QA/QC, the simple files were combined into a single table using another R script and written to a CSV file. QA/QC lines of code checked the number of contaminants and unique DTXSIDs in each data source and were cross-referenced with the universe file. DTXSIDs were added to the data values based on DTXSID assignments in the universe file. QA/QC was performed by checking data points from the original source files against the produced simple data format file.

EPA developed an R program to standardize and automate data manipulation and extraction for the remaining 11 primary data sources (DBP ICR, CA SURF, USDA PDP, UCMR 4, Batt et al., 2016; Bradley et al., 2017; Bradley et al., 2018; Kostich et al., 2010; Kostich et al., 2014; Scott et al., 2018; and Sun et al., 2016), supplemental data sources (CWSS, SYR 3 ICR and State Drinking Water Data sets), and primary literature data sources identified in a targeted literature search. The occurrence data were reformatted into simple file format through R scripts.[7] The remainder of this section describes this process in detail.

Most occurrence data fit one of four general data structures, characterized as sample sites, samples, summarized by sites, and summarized by sources. The four data structures are short descriptions of how occurrence data were originally reported in a data source. If a data source did not fit one of these structures, a short data preparation script was written using R to convert it into one of those four data structures. This data preparation script was often required if the raw data included notes that affected the interpretation of the data by the R script, the source did not standardize site/sample names and contaminant names, or all the raw data were spread out across multiple source files. The data structure was noted in a Source Data Lookup Key, described below.

Supporting files were generated to allow automation of data manipulation, including matching input data and metadata where applicable and directing methods of re-structuring and data organization. These supporting files are referred to as lookup keys. Two lookup keys were developed in the process of producing the simple occurrence file to help automate the restructuring of all the varied occurrence data sources into a single simple data format:

[6] R scripts were written using R version 4.0.2 (R Core Team, 2020) in RStudio version 1.3.1056 using the tidyverse package library (Wickham et al., 2019).

[7] Data restructuring and analysis were conducted using R version 3.6.2 (RStudio version 1.3.959). Every 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. R script version control was maintained through a repository on GitLab, a web-based software development and IT operations lifecycle tool.

- The Source Data Lookup Key is an Excel file containing information on the occurrence data sources, their file locations, and the data structure of each of the sources. This key also contains source metadata information such as source names, citations, date range of monitoring, geographic range, and water type.
- The Contaminant DTXSID Lookup Key is an Excel file containing all contaminant names (e.g., synonyms) in all sources matched to their DTXSIDs, CAS Registry Numbers, and preferred names, thereby standardizing contaminant names and IDs. This key was checked and updated with each new additional data source to ensure complete coverage of all contaminant names. DTXSID and CAS Registry Numbers for contaminants were obtained using the batch search function in EPA's CompTox Chemistry Dashboard.

To convert the occurrence data into a simple file format, each was first filtered into one of three R scripts based on the data structure, as noted in the Source Data Lookup Key. In these three R scripts, all the data sources associated with those data structures are combined into a single data table under common column headers. QA lines of code in these scripts preview the data being generated and check that all data have contaminant names, an associated water type, and a source name. The product of these three scripts were also output to an Excel file and spot-checked for any anomalies by a QA reviewer.

After that checkpoint, all outputs from the three data structures were entered into a fourth R script, which bound the three intermediate outputs together, calculated concentration and detection statistics, and removed extraneous columns. The product of this fourth script was a clean, wide-format table containing data from all occurrence sources by contaminant, source, and water type. This table was then exported to an Excel file and spot-checked by a QA reviewer. The code was modified as new sources with slightly different formats and different data were added to allow it to more broadly accommodate diversity of format within each of the data structures.

Once all occurrence data were combined into a single table, two more R scripts were written to add DTXSIDs and re-structure the data into the simple data format. The first of these two R scripts assigned DTXSIDs to all data by matching up the contaminant names to DTXSIDs in the Contaminant 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 final R script restructured the clean, wide-format data with DTXSIDs into the simple format. In addition, it added in source metadata from the Source Data Lookup Key and contaminant metadata from the Contaminant DTXSID Lookup Key. Besides the final simple-format occurrence output, this script also output tables containing lists of all the unique sources and data elements for review and QA.

After all occurrence data from primary sources and supplemental sources were compiled in the simple file format, EPA used this information to calculate the occurrence attribute scores (prevalence and magnitude) and fHQs (Chapter 4). EPA compiled occurrence attribute scoring information and fHQs in Excel workbooks. Two EPA staff members were responsible for ensuring the accuracy of the values and calculations, one for manually assigning attribute scores and calculating fHQ values according to the attribute scoring and fHQ protocols (Appendix H), and the other for QA/QC of the values. EPA identified three errors during the QA/QC of occurrence attribute scores, such as a magnitude or prevalence score being assigned to the wrong occurrence data element. EPA identified five errors during the QA/QC of the fHQ values. An example of an error was incorrect rounding and significant figures in

the calculated fHQ value. EPA documented and corrected all errors identified during the QA/QC of occurrence attribute scores and fHQ values.

Section 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 created to assist with formatting all CIS data onto the Excel workbooks. One of these keys assisted with updating the universe file, while the other three assisted with data placement and formatting on the CISs. An R script pulled in all the re-formatted data source files and generated CISs. Using the openxlsx package (Schauberger & Walker, 2020) in R, CISs were created in Excel workbooks for each PCCL 5 chemical to be reviewed by the QA/QC 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, 2022b) accessible via the EPA docket (Docket ID No. EPA-HQ-OW-2018-0594)), and row heights and cell borders were added.

The CISs then underwent two rounds of QA/QC. During the first round, 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 had data, or whether the highlighting matched the scoring data correctly. If a new source had recently been identified and added from the occurrence literature search, the CISs for the chemicals in that source were checked to ensure the new source data formatted correctly through the entire process and printed correctly.

During the second round, the QA/QC reviewers performed spot-checks, conducted calculation cross-checks, looked for errors in rounding or significant figures, checked DTXSID hyperlinks, verified unit conversions, and scanned for missing data. These final 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 17-alpha-ethynyl estradiol had been calculated incorrectly by an order of magnitude due to version control issues. Once the second round of 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).

Chapter 7 References

Allaire, J.J., Y. Xie, J. McPherson, J. Luraschi, K. Ushey, A. Atkins, H. Wickham, J. Cheng, W. Chang, and R. Iannone. 2020. rmarkdown: Dynamic Documents for R. R package version 2.5, https://github.com/rstudio/rmarkdown.

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.

CalEPA. n.d. Office of Environmental Health Hazard Assessment (OEHHA). Chemicals. https://oehha.ca.gov/chemicals. Accessed May 2019.

CDC. n.d. Agency for Toxic Substances and Disease Registry (ATSDR). Minimal Risk Levels (MRLs) for Hazardous Substances. https://www.cdc.gov/TSP/MRLS/mrlsListing.aspx. Accessed April 10, 2018.

CDPR. n.d. Surface Water Database (SURF). https://www.cdpr.ca.gov/docs/emon/surfwtr/surfdata.htm. Accessed April 29, 2019.

Furlong, E.T., A.L. Batt, S.T. Glassmeyer, M.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.

Gelman, A., J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. 2020. Bayesian data analysis (3rd ed.). London: Chapman & Hall.

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.

HHS. n.d. National Institutes of Health (NIH). National Library of Medicine. Hazardous Substances Data Bank (HSDB). https://www.nlm.nih.gov/databases/download/hsdb.html. Accessed April 11, 2019.

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

Holtschlag, D. J., Shively, D., Whitman, R. L., Haack, S. K., & Fogarty, L. R. 2008. Environmental factors and flow paths related to Escherichia coli concentrations at two beaches on Lake St. Clair, Michigan, 2002–2005. Reston, VA: USGS.

Kleinbaum, D.G. and M. Klein. 2010. Logistic regression: A self-learning text (3rd ed.). New York, NY: Springer-Verlag.

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.

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.

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.

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

Lyman, W. J., W.F. Reehl, and D.H. Rosenblatt. 1990. Handbook of Chemical Property Estimation Methods, American Chemical Society, Washington, DC.

Morrison, A. M., K. Coughlin, J.P. Shine, B.A. Coull, and A. Rex. 2003. Receiver operating characteristic curve analysis of beach water quality indicator variables. Applied and Environmental Microbiology, 69(11), 6405.

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.

National Drinking Water Advisory Council (NDWAC). 2004. National Drinking Water Advisory Council Report on the CCL Classification Process to the U. S. Environmental Protection Agency, May 19, 2004.

National Research Council (NRC). 2001. Classifying Drinking Water Contaminants for Regulatory Consideration. National Academy Press, Washington DC.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC, R News, vol 6, 7-11.

Porta, M. 2014. A dictionary of epidemiology. 6th Edition. New York: Oxford University Press, 2014.

R Core Team 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Robin X., N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. Sanchez, and M. Müller. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 12, 77.

Schauberger P. and A. Walker. 2020. openxlsx: Read, Write and Edit xlsx Files. Retrieved from: https://CRAN.R-project.org/package=openxlsx.

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.

Steenland, K., Savitz, D.A. 1998. Topics in environmental epidemiology. Epidemiology, 9, 335. NY: Oxford university press.

Subirana, I., Sanz, H. & Vila, J. 2014. Building Bivariate Tables: The compareGroups Package for R. Journal of Statistical Software, 57(12), 1–16. https://www.jstatsoft.org/v57/i12/.

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.

Tape, T. G. 2007. The area under an ROC curve. University of Nebraska Medical Center. Retrieved from http://gim.unmc.edu/dxtests/roc3.htm. Accessed June 2020.

USEPA. 1991. Existing Stocks of Pesticides Products; Statement of Policy. Federal Register. Vol. 56, No. 123, p. 29362, June 26, 1991.

USEPA. 1998. Announcement of the Drinking Water Contaminant Candidate List; *Notice*. *Federal Register*. Vol. 63, No. 40. p. 10274, March 2, 1998.

USEPA. 2000a. Database of Database of Sources of Environmental Releases of Dioxin-Like Compounds in the United States. https://cfpub.epa.gov/ncea/dioxin/recordisplay.cfm?deid=20797. Accessed October 2018.

USEPA. 2000b. Methodology for Deriving Ambient Water Quality Criteria for the Protection of Human Health (2000). U.S. Environmental Protection Agency, Office of Water, Office of Science and

Technology, Washington, D.C. https://www.epa.gov/sites/production/files/2018-10/documents/methodology-wqc-protection-hh-2000.pdf.

USEPA. 2005. Notice - Drinking Water Contaminant Candidate List 2; Final Notice. *Federal Register*. Vol. 70, No. 36. p. 9071, February 24, 2005.

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

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

USEPA. 2009c. Final Contaminant Candidate List 3 Chemicals: Classification of the PCCL to CCL. EPA 815-R-09-008. August 2009. Washington, D.C.

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

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

USEPA. 2009f. Drinking Water Contaminant Candidate List 3 - Final. *Federal Register*. Vol. 74, No. 194. p. 51850, October 8, 2009.

USEPA. 2009g. SAB Advisory on EPA's Draft Third Drinking Water Contaminant Candidate List (CCL 3). U.S. Environmental Protection Agency, Office of the Administrator, Science Advisory Board. EPA-SAB-09-011. January 2009.

USEPA. 2012. TSCA Work Plan Chemicals: Methods Document, USEPS, Washington DC. February 2012. Available on the Internet at https://www.epa.gov/sites/production/files/2014-03/documents/work_plan_methods_document_web_final.pdf

USEPA. 2014. Announcement of Preliminary Regulatory Determination for Contaminants on the Third Drinking Water Contaminant Candidate List. *Federal Register*. Vol. 79, No. 202, p. 62716, October 20, 2014.

USEPA. 2016a. Drinking Water Contaminant Candidate List 4-Final. Federal Register. Vol. 81, No. 222. P. 81099, November 17, 2016.

USEPA. 2016b. Chemical Data Reporting (CDR) Results. https://www.epa.gov/chemical-data-reporting/2016-chemical-data-reporting-results#access. Accessed April 25, 2018.

USEPA. 2018. Request for Nominations of Drinking Water Contaminants for the Fifth Contaminant Candidate List. Federal Register Vol 83, No. 194, p. 50364, October 5, 2018.

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

USEPA. 2021a. Prepublication Notice. Proposed Rule. TSCA Section 8(a)(7) Reporting and Recordkeeping Requirements for Perfluoroalkyl and Polyfluoroalkyl Substances. June 10, 2021. EPA-HQ-OPPT-2020-0549.

USEPA. 2021b. Announcement of Final Regulatory Determinations for Contaminants on the Fourth Drinking Water Contaminant Candidate List. Federal Register. Vol. 86, No. 40, p. 12272, March 3, 2021.

USEPA. 2022a. Technical Support Document for the Final Fifth Candidate List (CCL 5) – Microbial Contaminants. EPA 815-R-21-004, October 2022.

USEPA. 2022b. Technical Support Document for the Final Fifth Contaminant Candidate List (CCL 5) - Contaminant Information Sheets. EPA 815-R-22-003, October 2022.

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

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

Wickham, H., M. Averick, J. Bryan, W. Chang, L.D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T.L. Pedersen, E. Miller, S.M. Bache, K. Müller, J. Ooms, D. Robinson, D.P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani. 2019. Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686. doi: 10.21105/joss.01686.

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.

List of Appendices

Appendix A - Primary Data Source Descriptions
Appendix B - Supplemental Data Sources
Appendix C - Publicly Nominated Chemical Contaminants
Appendix D - PCCL Chemical Contaminants
Appendix E - Protocol for the Occurrence Literature Review
Appendix F - Protocol for the Rapid Systematic Health Effects Literature Review
Appendix G - Protocol to Derive Health Concentrations
Appendix H - Protocol to Select Water Concentrations Used in Calculating Final Hazard Quotients
Appendix I - Protocol to Determine Potency Attribute Scores
Appendix J - Protocol to Determine Severity Attribute Scores
Appendix K - Protocol to Determine Prevalence Attribute Scores
Appendix L - Protocol to Determine Magnitude Attribute Scores
Appendix M - Protocol to Determine Magnitude Attribute Scores from Persistence-Mobility
Appendix N - Data Management for CCL 5
Appendix O - CCL 4 Chemicals Not Listed on CCL 5
Appendix P - Group of 23 DBPs included on CCL 5

Appendix A - Primary Data Source Descriptions

This appendix includes descriptions of the primary sources of health effects and occurrence data that form the CCL 5 Chemical Pre-Universe and Universe.

Primary 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.

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

The Cancer Potency Data Bank provides 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.

3. Chemical Database – California EPA (CalEPA) Office of Environmental Health Hazard Assessment (OEHHA)

CalEPA's Chemical Database provides health hazard information developed by the CalEPA OEHHA, including cancer potency data such as cancer slope factors

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

The DWSHA Tables provide EPA's drinking water regulations, health advisories, reference doses, and cancer risk values for drinking water contaminants. The tables are revised periodically. The 2018 edition was used in the development of the CCL 5.

5. Guidelines for Canadian Drinking Water Quality – Health Canada

The Guidelines for Canadian Drinking Water Quality provide health-based guidelines developed based on a systematic review 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.

6. Guidelines for Drinking-Water Quality – WHO

The WHO Guidelines for Drinking-Water Quality (2017, 4th ed.) provide guideline values for approximately 95 chemicals that, according to international risk assessments, show evidence of occurrence in drinking water and actual or potential health effects.

7. 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 did not meet retrievability criteria but was still used as a primary data source. The HSDB is a

data rich source, and the only source of LD_{50}s for the CCL 5 process. Therefore, additional effort was taken to extract this data.

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

USGS hosts a dataset of HBSLs for 808 contaminants. These 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). The HBSL list was revised in May 2018 to provide updated toxicity information and to make the data consistent with new EPA methods and exposure assumptions.

9. 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. The dataset contains acute, short-term, subchronic, chronic, and cancer health risk limits, health-based values, or risk assessment advice for 457 contaminants.

10. Human Health Benchmarks for Pesticides – EPA

EPA has developed human health benchmarks for 394 pesticides. These include benchmarks for acute and chronic exposures for the most sensitive populations (i.e., children and women of childbearing age) from exposure to pesticides that may be found in surface or ground water 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.

11. Integrated Risk Information System (IRIS) – EPA

EPA's IRIS contains toxicity data from assessments of 461 contaminants, including toxicity values (e.g., reference dose, oral slope factor) for health effects resulting from chronic exposure to chemicals.

12. 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 1,069 contaminants.

13. 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).

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

The Human Health Ambient Water Quality Criteria contain recommended water quality criteria for human health for 121 chemical 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.

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

The dataset is compiled from a list of 596 NTP peer-reviewed technical reports and includes cancer classifications for each contaminant based on short-term and long-term studies on rats and mice.

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

As part of EPA's Superfund and Resource Conservation & Recovery Act programs, PPRTVs including provisional reference doses, cancer slope factors, and cancer classifications are derived 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.

17. Screening Levels for Pharmaceutical Contaminants – FDA Drugs@FDA database, National Institutes of Health (NIH) DailyMED database

Screening Levels for pharmaceutical contaminants were calculated using human oral dosage and administration information obtained from two public access databases containing drug labels, the NIH DailyMED database and the Drugs@FDA database (FDA, 2018; NIH, 2018). The NIH DailyMed database contains over 122,000 publicly-available drug listings submitted as FDA-approved labels (NIH, 2018). Supplemental data for pharmaceuticals not available through the NIH DailyMED database was extracted from the Drugs@FDA database which includes information about most drug products approved since 1939 (FDA, 2018).

18. Toxicity Reference Database (ToxRefDB) – EPA

The ToxRefDB contains decades of results from approximately 5,900 in vivo animal toxicity studies on hundreds of chemicals, following strict guidelines set by EPA and NTP.

Primary 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 Byproducts Information Collection Rule (DBP ICR) – EPA

The Disinfection Byproducts 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 1,377 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. National Health and Nutrition Examination Survey (NHANES) Biospecimen Program – CDC

The CDC's "Fourth Report on Human Exposure to Environmental Chemicals, Updated Tables, January 2019" provides nationally representative, cumulative biomonitoring data for chemicals and metabolites measured in blood, serum, and urine samples from random subsamples collected in NHANES 1999–2000 through 2015-2016.

10. 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.

11. 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.

12. 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.

13. "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.

14. "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.

15. 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 2017.

16. 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 2016.

17. "Pharmaceutical manufacturing facility discharges can substantially increase the pharmaceutical load to US 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.

18. "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.

19. "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 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.

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

California's Department of Pesticide Regulation 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.

21. "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.

22. 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 2016 TRI dataset includes environmental release data on 503 on-site and off-site chemicals reported, disposed of or otherwise released in 2016.

23. Unregulated Contaminant Monitoring Rule (UCMR) Cycles 1-3 – EPA

Every five years, EPA develops a list of contaminants that PWSs must monitor as part of the UCMR program. 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 less than or equal to 10,000 people.

UCMR data are provided in EPA's National Contaminant Occurrence Database. This monitoring program provides a basis for future regulatory actions to protect public health.

24. UCMR Cycle 4 – EPA

UCMR 4 requires monitoring for 30 chemical contaminants between 2018 and 2020 using analytical methods developed by EPA and consensus organizations. Refer to description of UCMR 1-3 above for more information on data collection for the UCMR process.

25. 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.

References

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

FDA. 2018. Drugs @ FDA: FDA Approved Drug Products. https://www.accessdata.fda.gov/scripts/cder/daf/. Accessed October 2017.

NIH. 2018. DailyMed database. United States National Library of Medicine. 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 considered for filling data gaps in the CCL 5 process. This list includes supplemental data considered in CCL 3 and CCL 4 and data sources recommended by the CCL 5 EPA Workgroup and subject matter experts, cited in public nominations for the CCL 5, and identified through CCL 5 literature searches.

1. "1,4-dioxane monitoring in the Cape Fear River basin of North Carolina: An ongoing screening, source identification, and abatement verification study" – North Carolina Division of Water Resources 2017[1]
2. "An introduction to joint research by the USEPA and USGS on contaminants of emerging concern in source and treated drinking waters of the United States" – Kolpin et al. 2017
3. "Anthropogenic organic compounds in source water of nine community water systems that withdraw from streams, 2002-05" – Kingsbury et al. 2008
4. "Anthropogenic organic compounds in source water of selected community water systems that use groundwater, 2002-05" – Hopple et al. 2009
5. "A survey of occurrence and risk assessment of pharmaceutical substances in the Great Lakes Basin" – Uslu et al. 2013
6. Australian Drinking Water Guidelines – Australian Government National Health and Medical Research Council
7. "Human health screening and public health significance of contaminants of concern detected in public water supplies" – Benson et al. 2017
8. "Hormones and pharmaceuticals in groundwater used as a source of drinking water across the United States" – Bexfield et al. 2019
9. California Stream Quality Assessment (CSQA) – USGS
10. Chemicals of High Concern – Maine Department of Environmental Protection
11. Chemicals of High Concern – Minnesota Department of Health
12. Chemicals of High Concern to Children Reporting List – Washington State Department of Ecology
13. Chlorpyrifos Refined Drinking Water Risk Assessment for Registration Review – EPA[1]
14. Community Rolling Action Plan (CoRAP) – European Chemicals Agency (ECHA)
15. "Comparing the toxic potency in vivo of long-chain perfluoroalkyl acids and fluorinated alternatives" – Gomis et al. 2018[1]
16. CompTox Chemicals Dashboard – EPA
17. "Concentrations of glyphosate and atrazine compounds in 100 Midwest United States streams in 2013" – Mahler et al. 2016
18. "Concentrations of hormones, pharmaceuticals and other micropollutants in groundwater affected by septic systems in New England and New York" – Phillips et al. 2015[1]
19. "Contaminants of emerging concern in ambient groundwater in urbanized areas of Minnesota, 2009–12" – Erickson et al. 2014
20. Cumulative Estimated Daily Intake (CEDI) database – U.S. Food and Drug Administration (FDA)
21. "Cyanotoxins in US Drinking Water: Occurrence, Case Studies and State Approaches to Regulation" – AWWA

22. "Cytotoxicity of novel fluorinated alternatives to long-chain perfluoroalkyl substances to human liver cell line and their binding capacity to human liver fatty acid binding protein" – Sheng et al. 2017[1]
23. DailyMed database – U.S. National Library of Medicine
24. "Design and methods of the Midwest Stream Quality Assessment (MSQA), 2013" – Garrett et al. 2017
25. "Design and methods of the Southeast Stream Quality Assessment (SESQA), 2014" – Journey et al. 2015
26. "Detection of poly- and perfluoroalkyl substances (PFASs) in U.S. drinking water linked to industrial sites, military fire training areas, and wastewater treatment plants" – Hu et al. 2016[1]
27. "Developmental neurotoxicity of industrial chemicals" – Grandjean & Landrigan 2006
28. Dieldrin and Drinking Water – Minnesota Department of Health 2016
29. Dietary Reference Intake documents – National Academy of Medicine
30. Drinking Water & Groundwater Quality Standards/Advisory Levels – Wisconsin Department of Natural Resources
31. Drugs @ FDA: FDA Approved Drug Products – FDA
32. Electronic Data Transfer Library – California Water Boards Division of Drinking Water[1]
33. Environmental Hazard Evaluation (EHE) and Environmental Action Levels (EALs) – State of Hawaii Department of Health
34. Existing Substances Regulation (ESR) – ECHA
35. EXTOXNET Pesticide Information Profiles – Cooperative effort of University of California-Davis, Oregon State University, Michigan State University, Cornell University, and University of Idaho
36. "Factors affecting water quality in selected carbonate aquifers in the United States, 1993–2005" – Lindsey et al. 2008
37. "Formation and Occurrence of N-Chloro-2,2-dichloroacetamide, a Previously Overlooked Nitrogenous Disinfection Byproduct in Chlorinated Drinking Waters" – Yu & Reckhow 2017
38. Generally Regarded as Safe (GRAS) Notice Inventory – FDA
39. "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
40. Guidelines for Drinking-Water Quality documents – WHO
41. Health Advisory supporting documents – EPA Office of Water (OW)[1]
42. Health Canada Drinking Water Guidelines support documents
43. Health Effects Support Documents (HESDs) – EPA OW[1]
44. Human and Environmental Risk Assessment on ingredients of household cleaning products (HERA) – International Association for Soaps, Detergents and Maintenance Products (AISE) & European Chemical Industry Council (Cefic)
45. "Human health risk assessment of pharmaceuticals in water: An uncertainty analysis for meprobamate, carbamazepine, and phenytoin" – Kumar & Xagoraraki 2010
46. Human Health Risk Assessments – EPA Office of Pesticide Programs (OPP)
47. Indirect Additives Database – FDA
48. Initial Environmental Risk Assessment of Chemicals – Japan Ministry of Environment

49. Integrated Risk Information System (IRIS) Chemical Assessment Summaries – EPA[1]
50. IRIS Toxicological Reviews – EPA
51. Joint Meeting on Pesticide Residues (JMPR) Acceptable Daily Intakes (ADIs) – World Health Organization (WHO) & Food and Agriculture Organization of the United Nations (FAO)
52. "Key scientific issues in developing drinking water guidelines for perfluoroalkyl acids: Contaminants of emerging concern" – Post, Gleason, & Cooper 2017[1]
53. Literature Search for Supplemental Water Occurrence Data for Pharmaceuticals, Personal Care Products and Other Contaminants – EPA OW
54. Minnesota Department of Health Toxicological Summaries[1] – Minnesota Department of Health
55. National Aquatic Resource Surveys (NARS) – EPA
56. National Pesticide Use Database – NCFAP
57. National Toxicology Program (NTP) studies – U.S. Department of Health and Human Services (HHS)
58. NTP Report on Carcinogens: Monograph on Haloacetic Acids Found as Water Disinfection By-Products – HHS 2018[1]
59. NTP Technical Report on the Toxicology and Carcinogenesis Studies of Sodium Dichromate Dihydrate – HHS 2008[1]
60. Occupational Safety and Health Administration (OSHA) Permissible Exposure Limits (PELS) – National Institute for Occupational Safety and Health (NIOSH)
61. "Occurrence and Distribution of Iron, Manganese, and Selected Trace Elements in Ground Water in the Glacial Aquifer System of the Northern United States" – Groschen et al. 2009
62. "Occurrence and in vitro bioactivity of estrogen, androgen, and glucocorticoid compounds in a nationwide screen of United States stream waters" – Conley et al. 2017
63. 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.
64. "Occurrence of neonicotinoid insecticides in finished drinking water and fate during drinking water treatment" – Klarich et al. 2017
65. "Occurrence, sources and fate of pharmaceuticals and personal care products in the groundwater: A review" – Sui et al. 2015[1]
66. "Oral chromium exposure and toxicity" – Sun, Brocato, & Costa 2015[1]
67. Peer-reviewed studies identified through the health effects rapid systematic literature review (see Section 4.2.2 and Appendix F for more details and the spreadsheet titled "CCL5 Rapid Systematic Literature Review Results" for a full list of references)
68. "Perfluorinated compounds in the Cape Fear drainage basin in North Carolina" – Nakayama et al. 2007[1]
69. "Periphyton (1993-2011) and water quality (2014) data for ET&C article entitled Spatial and Temporal Variation in Microcystins Occurrence in Wadeable Streams in the Southeastern USA" – Loftin et al. 2016
70. "Pesticides in polar organic chemical integrative samplers (POCIS) for 97 Midwest U.S. streams, 2013" – Alvarez et al. 2016
71. Pesticide National Synthesis Project – USGS

72. Pesticide Residue Monitoring Program – FDA
73. Pesticide Toxicity Profile series – University of Florida
74. "Pharmaceutical contaminant concentration and watershed geospatial land-use/land-cover data for small wadeable streams in the Piedmont ecoregion of the USA assessed during the Southeastern Region Stream Quality Assessment during April through June 2014" – Bradley et al. 2016
75. "Polyfluoroalkyl Chemicals in the U.S. Population: Data from the National Health and Nutrition Examination Survey (NHANES) 2003–2004 and Comparisons with NHANES 1999–2000" – Calafat et al. 2007[1]
76. "Potential toxicity of complex mixtures in surface waters from a nationwide survey of United States Streams: Identifying in vitro bioactivities and causative chemicals" – Blackwell et al. 2019
77. Provisional Peer-Reviewed Toxicity Value (PPRTV) support documents – EPA
78. Public Health Goal support documents – California Environmental Protection Agency (CalEPA) Office of Environmental Health Hazard Assessment (OEHHA)[1]
79. "Quality of source water from public-supply wells in the United States, 1993–2007" – Toccalino et al. 2010
80. "Radionuclide and Pesticide data for sediment age and source analysis in the Midwest Stream-Quality Assessment Region (2013-2014)" – Gellis et al. 2016
81. "Reconnaissance of land-use sources of pesticides in drinking water, McKenzie River, Oregon" – Kelly, Anderson, & Morgenstern 2012
82. References cited in Table 1 of "Human health screening and public health significance of contaminants of concern detected in public water supplies" – Benson et al. 2017
83. Regional Stream Quality Assessment (RSQA) – USGS
84. Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) Registration Dossiers – ECHA
85. Reregistration Eligibility Decision (RED) documents – EPA OPP
86. Risk Assessment Information System (RAIS) – U.S. Department of Energy, Office of Environmental Management, Oak Ridge Operations Office
87. Risk-based screening values for soil and groundwater cleanup sites – Alabama Department of Environmental Management
88. "Risks to aquatic organisms posed by human pharmaceutical use" – Kostich & Lazorchak 2008
89. Six Year Review (SYR) 3 State Data on Unregulated Contaminants
90. State drinking water monitoring data for unregulated contaminants/contaminants of emerging concern (CECs) that are accessible online
91. State of California Chemicals Known to the State to Cause Cancer or Reproductive Toxicity – CalEPA
92. Steroidal hormones and other endocrine active compounds in shallow groundwater in nonagricultural areas of Minnesota—Study design, methods, and data, 2009–10" – Erickson 2012
93. "Source attribution of poly-and perfluoroalkyl substances (PFASs) in surface waters from Rhode Island and New York metropolitan area" –Zhang et al. 2016[1]
94. Southeast Stream Quality Assessment (SESQA) – USGS
95. Substances Added to Food inventory – FDA

96. Substances Registry Services (SRS) – EPA
97. Tap Water Database – Environmental Working Group (EWG)
98. "The methamphetamine problem in the United States" – Gonzalez, Mooney, & Rawson 2010[1]
99. "The quality of our nation's waters – Quality of water from domestic wells in principal aquifers of the United States, 1991-2004" – DeSimone, Hamilton, & Gilliom 2009
100. "The quality of our nation's waters—Water quality in principal aquifers of the United States, 1991–2010" – DeSimone, McMahon, & Rosen 2014
101. Toxicological Profiles – Centers for Disease Control and Prevention (CDC), Agency for Toxic Substances and Disease Registry (ATSDR)[1]
102. Toxic Substances Control Act (TSCA) Chemical Substance Inventory – EPA
103. TSCA Risk Evaluations and other technical support documents – EPA
104. TOXNET – NLM (includes the following supplemental sources: International Toxicity Estimates for Risk (ITER) Database, Drugs and Lactation Database [LactMed], and Chemical Carcinogenesis Research Information System [CCRIS])
105. "Trace elements and radon in groundwater across the United States: U.S. Geological Survey Scientific Investigations Report 2011-5059" – Ayotte et al. 2011
106. "Trace levels of dieldrin and bromacil in two Oahu Water Systems" – State of Hawaii Department of Health 2015
107. USGS/CA Groundwater Ambient Monitoring and Assessment (GAMA) Program – USGS
108. USGS/NAWQA Data Series 997 and associated fact sheets – USGS
109. Village Creek Dieldrin Screening: Final Report – EPA Region 4 2015
110. Workplace Environmental Exposure Levels (WEEL) Guides – Occupational Alliance for Risk Science (OARS)
111. "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

[1]These data sources were cited in public nominations.

Appendix C - Publicly Nominated Chemical Contaminants

Chemical Name	CASRN	DTXSID Number
1,1-Dichloroethane	75-34-3	DTXSID1020437
1,4-Dioxane	123-91-1	DTXSID4020533
1-Phenylacetone[1]	103-79-7	DTXSID1059280
2-(N-Methylperfluorooctane sulfonamido)acetic acid (Me-PFOSA-AcOH)	2355-31-9	DTXSID10624392
2-(N-Ethyl perfluorooctane sulfonamido) acetic acid (Et-PFOSA-AcOH)	2991-50-6	DTXSID5062760
2-[(8-Chloro-1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8-Hexadecafluorooctyl)oxy]-1,1,2,2-tetrafluoroethane-1-sulfonic acid (11Cl-PF3OUdS)	763051-92-9	DTXSID40892507
3-Hydroxycarbofuran	16655-82-6	DTXSID2037506
3-Monoacetylmorphine[1]	29593-26-8	DTXSID30183774
4,8-Dioxa-3H-perfluorononanoic acid (ADONA)	919005-14-4	DTXSID40881350
6-Monoacetylmorphine[1]	2784-73-8	DTXSID60182154
Ammonium perfluoro-2-methyl-3-oxahexanoate	62037-80-3	DTXSID40108559
Anatoxin A	64285-06-9	DTXSID50867064
Azinphos-methyl	86-50-0	DTXSID3020122
Benzoic acid[1]	65-85-0	DTXSID6020143
Benzoic acid glucuronide[1]	19237-53-7	DTXSID90940901
Bromochloroacetic acid (BCAA)	5589-96-8	DTXSID4024642
Bromochloroiodomethane (BCIM)	34970-00-8	DTXSID9021502
Bromodichloroacetic acid (BDCAA)	71133-14-7	DTXSID4024644
Bromodichloronitromethane (BDCNM)	918-01-4	DTXSID4021509
Bromodiiodomethane (BDIM)	557-95-9	DTXSID70204235
Chlorate	14866-68-3	DTXSID3073137
Chloro-diiodo-methane (CDIM)	638-73-3	DTXSID20213251
Chloropicrin (trichloro-nitromethane; TCNM)	76-06-2	DTXSID0020315
Chlorpyrifos	2921-88-2	DTXSID4020458
Cylindrospermopsin	143545-90-8	DTXSID2031083

Chemical Name	CASRN	DTXSID Number
Dibromochloroacetic acid (DBCAA)	5278-95-5	DTXSID3031151
Dibromochloronitromethane (DBCNM)	1184-89-0	DTXSID00152114
Dibromoiodomethane (DBIM)	593-94-2	DTXSID60208040
Dichloroiodomethane (DCIM)	594-04-7	DTXSID7021570
Fluoxetine	5491-89-3	DTXSID7023067
Gemfibrozil	25812-30-0	DTXSID0020652
Heroin	561-27-3	DTXSID6046761
Hippuric acid[1]	495-69-2	DTXSID9046073
Hydromorphone[1]	466-99-9	DTXSID8023133
Hydromorphone-3-glucuronide[1]	No CASRN	NO_DTXSID
Hydroxyamphetamide[1]	103-86-6	DTXSID3023134
Isodrin (Pholedrine, 4-Hydroxymethamphetamine)[1]	465-73-6	DTXSID7042065
Manganese	7439–96–5	DTXSID2024169
Methamphetamine[1]	537-46-2	DTXSID8037128
Microcystin LA	96180-79-9	DTXSID3031656
Microcystin LR	101043-37-2	DTXSID3031654
Microcystin LW	No CASRN	DTXSID70891285
Microcystin RR	111755-37-4	DTXSID40880085
Microcystin YR	101064-48-6	DTXSID00880086
Molybdenum	7439–98–7	DTXSID1024207
Morphine	57-27-2	DTXSID9023336
Morphine-3-glucuronide	20290-09-9	DTXSID80174157
Morphine-6-glucuronide[1]	20290-10-2	DTXSID40174158
N-Nitrosodiethylamine (NDEA)	55-18-5	DTXSID2021028
N-Nitrosodimethylamine (NDMA)	62-75-9	DTXSID7021029
N-Nitroso-di-n-propylamine (NDPA)	621-64-7	DTXSID6021032
N-Nitrosodiphenylamine (NDPhA)	86-30-6	DTXSID6021030
N-Nitrosopyrrolidine (NPYR)	930-55-2	DTXSID8021062
Perfluoro(2-((6-chlorohexyl)oxy)ethanesulfonic acid) (9Cl-PF3ONS)	756426-58-1	DTXSID80892506

Chemical Name	CASRN	DTXSID Number
Perfluoro-2-methyl-3-oxahexanoic acid	13252-13-6	DTXSID70880215
Perfluorobutane sulfonic acid (PFBS)	375-73-5	DTXSID5030030
Perfluorobutyric acid (PFBA)	375-22-4	DTXSID4059916
Perfluorodecanoic acid (PFDeA/PFDA)	335-76-2	DTXSID3031860
Perfluorododecanoic acid (PFDoA)	307-55-1	DTXSID8031861
Perfluoroheptanoic acid (PFHpA)	375-85-9	DTXSID1037303
Perfluorohexane sulfonic acid (PFHxS)	355-46-4	DTXSID7040150
Perfluorohexanoic acid (PFHxA)	307-24-4	DTXSID3031862
Perfluorononononanoic acid (PFNA)	375-95-1	DTXSID8031863
Perfluorooctanesulfonamide (PFOSA)	754-91-6	DTXSID3038939
Perfluorooctane sulfonic acid (PFOS)	1763-23-1	DTXSID3031864
Perfluorooctanoic acid (PFOA)	335-67-1	DTXSID8031865
Perfluorotetradecanoic acid (PFTA)[2]	376-06-7	DTXSID3059921
Perfluorotridecanoic acid (PFTrDA)[2]	72629-94-8	DTXSID90868151
Perfluoroundecanoic acid (PFUA/PFUnA)	2058-94-8	DTXSID8047553
Phenylpropanolamine[1]	37577-28-9	DTXSID4023466
Strontium	7440-24-6	DTXSID3024312
Tribromoacetic acid (TBAA)	75-96-7	DTXSID6021668
Triiodomethane (TIM)	75-47-8	DTXSID4020743

[1]Thirteen nominated chemicals did not have available water occurrence data, even after a systematic literature search was conducted, and therefore were not evaluated for listing on CCL 5. See Section 4.2.1.1 for more information.[2]Other acronyms that may be used: Perfluorotetradecanoic acid (PFTetDA) and Perfluorotridecanoic acid (PFTriDA).

Appendix D - PCCL Chemical Contaminants

Chemical Name *	CASRN	DTXSID	Screening score
1,1,2,2-Tetrachloroethane	79-34-5	DTXSID7021318	3440
1,2,3-Trichloropropane	96-18-4	DTXSID9021390	6690
1,2,4-Trimethylbenzene	95-63-6	DTXSID6021402	3560
1,3-Butadiene	106-99-0	DTXSID3020203	4420
1,3-Dichloropropene	542-75-6	DTXSID1022057	5170
1,4-Dioxane *	123-91-1	DTXSID4020533	7690
17-alpha ethynyl estradiol	57-63-6	DTXSID5020576	5620
17-beta-Estradiol	50-28-2	DTXSID0020573	6120
1-Butanol	71-36-3	DTXSID1021740	3390
1-O-Benzoylhexopyranuronic acid *	19237-53-7	DTXSID90940901	NA
1-Phenylacetone *	103-79-7	DTXSID1059280	100
2-(2-Methyl-4-chlorophenoxy)propionic acid (MCPP)	93-65-2	DTXSID9024194	5710
2-(N-Ethylperfluorooctanesulfonamido)acetic acid (Et-PFOSA-AcOH) *	2991-50-6	DTXSID5062760	0
2-(N-Methylperfluorooctanesulfonamido)acetic acid (Me-PFOSA-AcOH) *	2355-31-9	DTXSID10624392	150
2,4-Dichlorophenol	120-83-2	DTXSID1020439	3840
2,4-Dichlorophenoxybutyric acid	94-82-6	DTXSID7024035	3770
2,4-Dinitrophenol	51-28-5	DTXSID0020523	3320
2,4-Dinitrotoluene	121-14-2	DTXSID0020529	6020
2,6-Dinitrotoluene	606-20-2	DTXSID5020528	4960
2-[(8-Chloro-1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8-hexadecafluorooctyl)oxy]-1,1,2,2-tetrafluoroethane-1-sulfonic acid (11Cl-PF3OUdS) *	763051-92-9	DTXSID40892507	NA
2-Hydroxyatrazine	2163-68-0	DTXSID6037807	6950
2-Methyl-4-chlorophenoxyacetic acid (MCPA)	94-74-6	DTXSID4024195	7110
2-Methylnaphthalene	91-57-6	DTXSID4020878	3900
3-Monoacetylmorphine *	29593-26-8	DTXSID30183774	NA
4,8-Dioxa-3H-perfluorononanoic acid (ADONA) *	919005-14-4	DTXSID40881350	NA
4-Androstene-3,17-dione	63-05-8	DTXSID8024523	3320
4-tert-Octylphenol	140-66-9	DTXSID9022360	3380
6-Chloro-1,3,5-triazine-2,4-diamine	3397-62-4	DTXSID1037806	6050
6-O-Monoacetylmorphine *	2784-73-8	DTXSID60182154	NA
Acephate	30560-19-1	DTXSID8023846	5260
Acetamiprid	135410-20-7	DTXSID0034300	4000
Acetochlor ethanesulfonic acid (ESA)	187022-11-3	DTXSID6037483	4810

Chemical Name *	CASRN	DTXSID	Screening score
Acetochlor oxanilic acid (OA)	194992-44-4	DTXSID1037484	3990
Acetophenone	98-86-2	DTXSID6021828	3340
Acrolein	107-02-8	DTXSID5020023	3780
Acyclovir	59277-89-3	DTXSID1022556	4040
Alachlor ethanesulfonic acid (ESA)	142363-53-9	DTXSID6037485	5700
Alachlor oxanilic acid (OA)	171262-17-2	DTXSID1037486	4900
Aldrin	309-00-2	DTXSID8020040	6080
alpha-Hexachlorocyclohexane	319-84-6	DTXSID2020684	5350
Ametryn	834-12-8	DTXSID1023869	3580
Ammonia	7664-41-7	DTXSID0023872	4100
Anatoxin-a *	64285-06-9	DTXSID50867064	1230
Anthraquinone	84-65-1	DTXSID3020095	3850
Atenolol	29122-68-7	DTXSID2022628	3660
Azoxystrobin	131860-33-8	DTXSID0032520	5560
Benfluralin	1861-40-1	DTXSID3023899	3780
Bensulide	741-58-2	DTXSID9032329	3810
Bentazon	25057-89-0	DTXSID0023901	6030
Benzoic acid *	65-85-0	DTXSID6020143	1390
Benzophenone	119-61-9	DTXSID0021961	5030
Bifenthrin	82657-04-3	DTXSID9020160	5270
Bisphenol A	80-05-7	DTXSID7020182	5580
Boron	7440-42-8	DTXSID3023922	5810
Boscalid	188425-85-6	DTXSID6034392	5480
Bromacil	314-40-9	DTXSID4022020	4390
Bromochloroacetic Acid (BCAA) *	5589-96-8	DTXSID4024642	550
Bromodichloroacetic acid *	71133-14-7	DTXSID4024644	580
Bromodichloronitromethane *	918-01-4	DTXSID4021509	NA
Bromodiiodomethane *	557-95-9	DTXSID70204235	NA
Bromoxynil	1689-84-5	DTXSID3022162	5160
Bupropion	34911-55-2	DTXSID7022706	3520
Butyl benzyl phthalate	85-68-7	DTXSID3020205	4550
Caffeine	58-08-2	DTXSID0020232	4780
Calcium	7440-70-2	DTXSID9050484	5330
Camphor	76-22-2	DTXSID5030955	3420
Carbamazepine	298-46-4	DTXSID4022731	4380
Carbaryl	63-25-2	DTXSID9020247	5920
Carbendazim (MBC)	10605-21-7	DTXSID4024729	4440
Carbon disulfide	75-15-0	DTXSID6023947	5300
Chlorate *	14866-68-3	DTXSID3073137	5570

Chemical Name *	CASRN	DTXSID	Screening score
Chlordecone (Kepone)	143-50-0	DTXSID1020770	4130
Chlorodiiodomethane *	638-73-3	DTXSID20213251	2000
Chloromethane (Methyl chloride)	74-87-3	DTXSID0021541	4290
Chloropicrin *	76-06-2	DTXSID0020315	5320
Chlorothalonil	1897-45-6	DTXSID0020319	5350
Chlorpyrifos *	2921-88-2	DTXSID4020458	8490
Clomazone	81777-89-1	DTXSID1032355	3360
Clopyralid	1702-17-6	DTXSID9029221	4360
Clothianidin	210880-92-5	DTXSID2034465	5410
Cobalt	7440-48-4	DTXSID1031040	8690
Cotinine	486-56-6	DTXSID1047576	3460
Cycloate	1134-23-2	DTXSID6032356	4090
Cyfluthrin	68359-37-5	DTXSID5035957	3810
Cyhalothrin	68085-85-8	DTXSID6023997	3520
Cylindrospermopsin *	143545-90-8	DTXSID2031083	2260
Cypermethrin	52315-07-8	DTXSID1023998	4960
Cyprodinil	121552-61-2	DTXSID1032359	4710
Desethylatrazine	6190-65-4	DTXSID5037494	5770
Desisopropyl atrazine	1007-28-9	DTXSID0037495	5840
Desvenlafaxine	93413-62-8	DTXSID40869118	3650
Diazepam	439-14-5	DTXSID4020406	3730
Diazinon	333-41-5	DTXSID9020407	8490
Dibromoacetonitrile (DBAN)	3252-43-5	DTXSID3024940	4120
Dibromochloroacetic acid (DBCAA) *	5278-95-5	DTXSID3031151	50
Dibromochloronitromethane *	1184-89-0	DTXSID00152114	NA
Dibromoiodomethane *	593-94-2	DTXSID60208040	500
Dicamba	1918-00-9	DTXSID4024018	4280
Dichloroacetonitrile (DCAN)	3018-12-0	DTXSID3021562	4290
dichloroiodomethane *	594-04-7	DTXSID7021570	2400
Dichlorvos (DDVP)	62-73-7	DTXSID5020449	6460
Dicrotophos	141-66-2	DTXSID9023914	5020
Dieldrin	60-57-1	DTXSID9020453	7680
Diethyl phthalate	84-66-2	DTXSID7021780	3340
Difenoconazole	119446-68-3	DTXSID4032372	4230
Dimethenamid	87674-68-8	DTXSID4032376	5730
Dimethenamid Oxanilic acid degradate (OXA)	380412-59-9	DTXSID4037530	3540
Dimethoate	60-51-5	DTXSID7020479	6020
Di-n-butyl phthalate	84-74-2	DTXSID2021781	4240
Diuron	330-54-1	DTXSID0020446	8680

Chemical Name *	CASRN	DTXSID	Screening score
EPTC (Ethyl dipropylthiocarbamate)	759-94-4	DTXSID1024091	5740
Esfenvalerate	66230-04-4	DTXSID4032667	3760
Ethalfluralin	55283-68-6	DTXSID8032386	4230
Ethion	563-12-2	DTXSID2024086	4360
Ethoprop	13194-48-4	DTXSID4032611	6490
Famoxadone	131807-57-3	DTXSID8034588	3650
Fenbuconazole	114369-43-6	DTXSID8032548	3630
Fenitrothion	122-14-5	DTXSID4032613	4120
Fenpropathrin	39515-41-8	DTXSID0024002	3560
Fenthion	55-38-9	DTXSID8020620	3690
Fexofenadine	83799-24-0	DTXSID00861411	3400
Fipronil	120068-37-3	DTXSID4034609	6190
Fluconazole	86386-73-4	DTXSID3020627	4240
Flufenacet (Thiaflumide)	142459-58-3	DTXSID2032552	3940
Fluometuron	2164-17-2	DTXSID8020628	4170
Fluoranthene	206-44-0	DTXSID3024104	3910
Fluoxetine *	54910-89-3	DTXSID7023067	2470
Formaldehyde	50-00-0	DTXSID7020637	4920
Galaxolide (HHCB)	1222-05-5	DTXSID8027373	3810
Gemfibrozil *	25812-30-0	DTXSID0020652	1970
Halon 1011 (bromochloromethane)	74-97-5	DTXSID4021503	4640
HCFC-22 (Chlorodifluoromethane)	75-45-6	DTXSID6020301	3950
Heroin *	561-27-3	DTXSID6046761	NA
Hexazinone	51235-04-2	DTXSID4024145	5330
Hippuric acid *	495-69-2	DTXSID9046073	NA
Hydromorphone *	466-99-9	DTXSID8023133	860
Hydromorphone-3-glucuronide *	40505-76-8	NO_DTXSID	NA
Hydroxyamphetamide *	103-86-6	DTXSID3023134	150
Imazalil	35554-44-0	DTXSID8024151	4510
Imazapyr	81334-34-1	DTXSID8034665	3400
Imazaquin	81335-37-7	DTXSID3024152	3350
Imazethapyr	81335-77-5	DTXSID3024287	4230
Imidacloprid	138261-41-3	DTXSID5032442	5530
Indoxacarb	173584-44-6	DTXSID1032690	3770
Iprodione	36734-19-7	DTXSID3024154	6050
Isodrin *	465-73-6	DTXSID7042065	290
Isophorone	78-59-1	DTXSID8020759	4750
Isopropylbenzene (Cumene)	98-82-8	DTXSID1021827	3330
Isoxaflutole	141112-29-0	DTXSID5034723	3360

Chemical Name *	CASRN	DTXSID	Screening score
Lactofen	77501-63-4	DTXSID7024160	3680
lambda-Cyhalothrin	91465-08-6	DTXSID7032559	4780
Lidocaine	137-58-6	DTXSID1045166	3710
Linuron	330-55-2	DTXSID2024163	5450
Lithium	7439-93-2	DTXSID5036761	8250
Loratadine	79794-75-5	DTXSID2023224	4050
Magnesium	7439-95-4	DTXSID0049658	5430
Malathion	121-75-5	DTXSID4020791	6120
Manganese *	7439-96-5	DTXSID2024169	8130
Meprobamate	57-53-4	DTXSID3023261	3570
Metalaxyl	57837-19-1	DTXSID6024175	5060
Metformin	657-24-9	DTXSID2023270	4110
Methamphetamine *	537-46-2	DTXSID8037128	70
Bromochloroiodomethane (BCIM) *	34970-00-8	DTXSID9021502	1800
Triiodomethane *	75-47-8	DTXSID4020743	290
Methocarbamol	532-03-6	DTXSID6023286	3490
Methomyl	16752-77-5	DTXSID1022267	3800
Methyl mercury	22967-92-6	DTXSID9024198	3540
Methyl tert-butyl ether (MTBE)	1634-04-4	DTXSID3020833	6290
Methylbenzotriazole	29385-43-1	DTXSID0026171	4020
Metolachlor ethanesulfonic acid (ESA)	171118-09-5	DTXSID1037567	4680
Metolachlor oxanilic acid (OA)	152019-73-3	DTXSID6037568	4750
Metoprolol	51384-51-1	DTXSID2023309	4420
Metribuzin	21087-64-9	DTXSID6024204	6930
Microcystin LA *	96180-79-9	DTXSID3031656	-10
Microcystin LW *	157622-02-1	DTXSID70891285	0
Microcystin RR *	111755-37-4	DTXSID40880085	-10
Microcystin YR *	101064-48-6	DTXSID00880086	0
Microcystin-LR *	101043-37-2	DTXSID3031654	3750
Molybdenum *	7439-98-7	DTXSID1024207	7480
Morphine *	57-27-2	DTXSID9023336	1900
Morphine 6-glucuronide *	20290-10-2	DTXSID40174158	NA
Morphine-3-Glucuronide *	20290-09-9	DTXSID80174157	NA
Myclobutanil	88671-89-0	DTXSID8024315	4510
N,N-Diethyl-m-toluamide (DEET)	134-62-3	DTXSID2021995	5430
Naled	300-76-5	DTXSID1024209	3630
Naphthalene	91-20-3	DTXSID8020913	4930
Nicotine	54-11-5	DTXSID1020930	5860
N-Nitrosodiethylamine (NDEA) *	55-18-5	DTXSID2021028	4110

Chemical Name *	CASRN	DTXSID	Screening score
N-nitrosodimethylamine (NDMA) *	62-75-9	DTXSID7021029	6330
N-Nitrosodi-n-butylamine	924-16-3	DTXSID2021026	3490
N-Nitroso-di-n-propylamine (NDPA) *	621-64-7	DTXSID6021032	3250
N-Nitrosodiphenylamine (NDPhA) *	86-30-6	DTXSID6021030	1720
N-nitrosopyrrolidine (NPYR) *	930-55-2	DTXSID8021062	3500
Nonylphenol	25154-52-3	DTXSID3021857	5550
Norflurazon	27314-13-2	DTXSID8024234	5390
o-Toluidine	95-53-4	DTXSID1026164	3560
Oxadiazon	19666-30-9	DTXSID3024239	4620
Oxyfluorfen	42874-03-3	DTXSID7024241	6320
p,p'-DDE	72-55-9	DTXSID9020374	7490
p-Cresol	106-44-5	DTXSID7021869	5110
Pendimethalin	40487-42-1	DTXSID7024245	4450
Perfluoro(2-((6-chlorohexyl)oxy)ethanesulfonic acid) (9Cl-PF3ONS) *	756426-58-1	DTXSID80892506	NA
Perfluorobutanesulfonic acid (PFBS) *	375-73-5	DTXSID5030030	5930
Perfluorobutanoic acid (PFBA) *	375-22-4	DTXSID4059916	4310
Perfluorodecanoic acid (PFDeA/PFDA) *	335-76-2	DTXSID3031860	2650
Perfluorododecanoic acid (PFDoA) *	307-55-1	DTXSID8031861	2400
Perfluoroheptanoic acid (PFHpA) *	375-85-9	DTXSID1037303	3200
Perfluorohexanesulfonic acid (PFHxS) *	355-46-4	DTXSID7040150	5450
Perfluorohexanoic acid (PFHxA) *	307-24-4	DTXSID3031862	2450
Perfluorononononanoic acid (PFNA) *	375-95-1	DTXSID8031863	5140
Perfluorooctanesulfonamide (PFOSA) *	754-91-6	DTXSID3038939	170
Perfluorotetradecanoic acid (PFTA) *	376-06-7	DTXSID3059921	700
Perfluorotridecanoic acid (PFTrDA) *	72629-94-8	DTXSID90868151	1100
Perfluoroundecanoic acid (PFUA/PFUnA) *	2058-94-8	DTXSID8047553	2640
Permethrin	52645-53-1	DTXSID8022292	6440
PFPrOPrA / Perfluoro-2-methyl-3-oxahexanoic acid *	No CASRN	DTXSID40108559 / DTXSID70880215	0 / NA
Phenanthrene	85-01-8	DTXSID6024254	4130
Phenol	108-95-2	DTXSID5021124	3880
Phenylpropanolamine *	14838-15-4	DTXSID4023466	210
Phorate	298-02-2	DTXSID4032459	5620
Phosmet	732-11-6	DTXSID5024261	3320
Phosphorus	7723-14-0	DTXSID1024382	5020
Phostebupirim (Tebupirimphos)	96182-53-5	DTXSID1032482	5080
Piperonyl butoxide	51-03-6	DTXSID1021166	4690
Potassium	7440-09-7	DTXSID9049748	5180
Profenofos	41198-08-7	DTXSID3032464	5980

Chemical Name *	CASRN	DTXSID	Screening score
Prometon	1610-18-0	DTXSID6022341	6570
Prometryn	7287-19-6	DTXSID4024272	5330
Pronamide	23950-58-5	DTXSID2020420	5320
Propachlor	1918-16-7	DTXSID4024274	5150
Propanil	709-98-8	DTXSID8022111	4990
Propargite	2312-35-8	DTXSID4024276	5090
Propazine	139-40-2	DTXSID3021196	5300
Propiconazole	60207-90-1	DTXSID8024280	5790
Propoxur	114-26-1	DTXSID7021948	3650
Prosulfuron	94125-34-5	DTXSID9034868	3620
Pymetrozine	123312-89-0	DTXSID2032637	3480
Pyraclostrobin	175013-18-0	DTXSID7032638	5000
Pyrene	129-00-0	DTXSID3024289	3910
Pyridaben	96489-71-3	DTXSID5032573	3760
Quinoline	91-22-5	DTXSID1021798	3460
Silicon	7440-21-3	DTXSID0051441	4160
Sitagliptin	486460-32-6	DTXSID70197572	3580
Sodium	7440-23-5	DTXSID1049774	5430
Sulfamethoxazole	723-46-6	DTXSID8026064	3830
Sulfentrazone	122836-35-5	DTXSID6032645	4480
Sulfometuron methyl	74222-97-2	DTXSID0034936	3490
Tamoxifen	10540-29-1	DTXSID1034187	3410
Tebuconazole	107534-96-3	DTXSID9032113	5090
Tebuthiuron	34014-18-1	DTXSID3024316	5200
Tefluthrin	79538-32-2	DTXSID5032577	3410
Terbacil	5902-51-2	DTXSID8024317	3880
Terbufos	13071-79-9	DTXSID2022254	5010
Testosterone	58-22-0	DTXSID8022371	3920
Tetraconazole	112281-77-3	DTXSID8034956	5390
Thiabendazole	148-79-8	DTXSID0021337	4320
Thiamethoxam	153719-23-4	DTXSID2034962	4470
Thiobencarb	28249-77-6	DTXSID6024337	4880
Thiram	137-26-8	DTXSID5021332	3620
Tin	7440-31-5	DTXSID1049801	3860
Triallate	2303-17-5	DTXSID5024344	5340
Tribromoacetic acid (TBAA) *	75-96-7	DTXSID6021668	100
Tribufos	78-48-8	DTXSID1024174	5780
Tributyl phosphate	126-73-8	DTXSID3021986	5800
Triclopyr	55335-06-3	DTXSID0032497	6800

Chemical Name *	CASRN	DTXSID	Screening score
Triclosan	3380-34-5	DTXSID5032498	5480
Triethyl citrate	77-93-0	DTXSID0040701	3360
Trifloxystrobin	141517-21-7	DTXSID4032580	3470
Trifluralin	1582-09-8	DTXSID4021395	5400
Tris(1,3-dichloro-2-propyl) phosphate (TDCP)	13674-87-8	DTXSID9026261	6370
Tris(2-butozylethyl) phosphate (TBEP)	78-51-3	DTXSID5021758	3750
Tris(2-chloroethyl) phosphate (TCEP)	115-96-8	DTXSID5021411	6860
Tungsten	7440-33-7	DTXSID8052481	3810
Vanadium	7440-62-2	DTXSID2040282	9050
Verapamil	52-53-9	DTXSID9041152	3340

Note: Asterisk (*) indicates publicly nominated chemical contaminants. Screening scores of "NA" indicate publicly nominated chemical contaminants that were not identified from primary data sources and therefore had no available data in the CCL 5 Universe. See Sections 3.3.2 and 3.4 of the main document for a description of the point assignment process and calculation of screening scores for each chemical.

Appendix E - Protocol for the Occurrence Literature Review

The goal of occurrence literature searches was to identify state data, guidance from other government agencies, and peer-reviewed studies that would fill occurrence data gaps and aid the evaluations of PCCL 5 chemicals that required further evaluation in the classification step. EPA conducted a targeted literature search for occurrence data based on the type of data already available for a PCCL 5 chemical in the universe. This Appendix describes the protocol developed by EPA for conducting these targeted occurrence literature searches.

- For chemicals having national finished water data from primary data sources such as UCMR, UCM, or NIRS:
    - EPA did not conduct occurrence literature searches for chemicals which had UCMR 3 or UCMR 4 data.
    - EPA conducted occurrence literature search for finished water data collected in the last 10 years for chemicals that had no UCMR 3 or UCMR 4 even though UCMR 2, UCMR 1, UCM (Round 1 and/or 2), or NIRS data were available.
- For chemicals having national ambient water data occurrence data as the best available occurrence data, EPA conducted an occurrence literature search for non-national finished water data collected within the last 10 years.
- For chemicals having pesticide application data as the best available occurrence data, EPA conducted an occurrence literature search for non-national finished water and non-national ambient water data, both collected within the last 10 years.
- For chemicals having chemical production volume data as the best available occurrence data in the CCL 5 Universe, EPA conducted a literature search for non-national finished water and non-national ambient water data, both collected within the last 10 years.
- For chemicals with release data as the best available occurrence, EPA conducted a literature search for non-national finished water and non-national ambient water data both collected within the last 10 years.

All literatures searches were repeated by a quality control reviewer to ensure all relevant primary literature was identified. Also, the results of the literature searches were reviewed to assure relevance. EPA used Google Scholar, HSDB, as well as EPA abstract sifter. Keywords included “drinking water,” “occurrence,” and “occurrence in water.”

References

The following 12 supplemental occurrence data sources containing contaminant ambient or finished water data were identified as part of the targeted literature reviews for PCCL 5 chemicals that required further evaluation.

Arnold, TL, DeSimone, L.A., Bexfield, L.M., Lindsey, B.D., Barlow, J.R., Kulongoski, J.T., Musgrove, MaryLynn, Kingsbury, J.A., and Belitz, Kenneth. 2016. 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, 56 p., http://dx.doi.org/10.3133/ds997.

Bexfield, L.M., P.L. Toccalino, K. Belitz, W.T. Foreman, and E.T. Furlong. 2019. Hormones and pharmaceuticals in groundwater used as a source of drinking water across the United States. Environmental Science & Technology. 53: 2950-2960.

EPA Region 4. 2015. Village Creek Dieldrin Screening. https://www.birminghamal.gov/wp-content/uploads/2017/08/15-0308-Village-Creek-Dieldrin-Screening-Final-Report-v081015.pdf

Kumar, A. and Xagoraraki, I. 2010. Human health risk assessment of pharmaceuticals in water: An uncertainty analysis for meprobamate, carbamazepine, and phenytoin. Regulatory Toxicology and Pharmacology. 57(2-3): 146-156.

Padhye, L.P., Yao, H., Kung'u, F.T., and Huang, C.H. 2014. Year-long evaluation on the occurrence and fate of pharmaceuticals, personal care products, and endocrine disrupting chemicals in an urban drinking water treatment plant. Water Research. 51: 266-276.

Klarich, K.L., Pflug, N.C., DeWald, E.M., Hladik, M.L., Kolpin, D.W., Cwiertny, D.M. and LeFevre, G.H. 2017. Occurrence of neonicotinoid insecticides in finished drinking water and fate during drinking water treatment. Environmental Science & Technology Letters, 4(5): 168-173.

Minnesota Department of Health. October 2016. Dieldrin and Drinking Water. https://www.health.state.mn.us/communities/environment/risk/docs/guidance/gw/dieldrininfo.pdf

USGS. n.d. Southeast Stream Quality Assessment (SESQA). https://webapps.usgs.gov/rsqa/#!/region/SESQA

State of Hawaii Department of Health. 2015. Trace levels of dieldrin and bromacil in two Oahu Water Systems. https://health.hawaii.gov/news/files/2013/05/TRACE-LEVELS-OF-DIELDRIN-AND-BROMACIL-DETECTED-IN-TWO-OAHU-WATER-SYSTEMS.pdf

USGS. 2009. 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 appendixes.

USGS. Reconnaissance of Land-Use Sources of Pesticides in Drinking Water, McKenzie River, Oregon. U.S. Geological Survey Scientific Investigations Report 2012–5091, 46 p. plus appendixes.

Uslu, M., Jasim, S., Arvai, A., Bewtra, J. and Biswas, N. 2013. A Survey of Occurrence and Risk Assessment of Pharmaceutical Substances in the Great Lakes Basin. Ozone: Science & Engineering. 35(4): 249-26

Appendix F - Protocol for the Rapid Systematic Health Effects Literature Review

The focus of the CCL 5 health effects rapid systematic review (RSR) was on identifying animal toxicity studies with dose-response data relevant to chronic oral exposure to chemical contaminants. This RSR for supplemental health effects information was divided into four steps:

- Step 1: Literature identification
- Step 2: Title-abstract screening
- Step 3: Full text review and study quality evaluation
- Step 4: 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, or 4. The following protocol outlines the identification of supplemental health effects information as part of the classification process of CCL 5. Refer to Section 4.2.2 of the text for additional information.

Step 1: Literature identification

a) Health Assessment Identification

A key element of the RSR process is to leverage toxicity information for PCCL chemicals that was derived from previously published health or hazard assessments. To achieve this, EPA started by conducting literature searches to identify the most recently published assessments that provide information on health effects resulting from oral exposure routes for each chemical. Assessments used to inform the RSR protocol for PCCL chemicals included:

- Agency for Toxic Substances and Disease Registry – Toxicological Profiles
- California EPA Office of Environmental Health Hazard Assessment – Public Health Goals
- EPA Office of Water – Drinking Water Health Advisories or Health Effects Support Documents
- Health Canada – Guidelines for Canadian Drinking Water Quality
- Integrated Risk Information System – Chemical Assessment Summaries or Toxicological Reviews
- EPA Superfund Program - Provisional Peer-Reviewed Toxicity Values
- World Health Organization – Drinking Water Quality Guidelines

If a chemical had at least one of the assessments 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 relevant assessments listed above were not date limited. Relevant risk assessment documents and search date limits for each chemical that underwent the RSR process are provided in Table F-1.

Table F-1. Date Limitations for CCL 5 RSR Chemicals

Chemical	Search Date	Search Date Limit	Most Recent Assessment
1,1,2,2-Tetrachloroethane	4/7/2020	9/1/2009	IRIS, September 2010
1,2,4-Trimethylbenzene	4/6/2020	9/1/2015	IRIS, September 2016
1,3-Butadiene	2/13/2020	11/1/2001	IRIS, November 2002
1,4-Dioxane[c]	NA	NA	NA
1-Butanol	3/16/2020	3/1/1986	IRIS, March 1987
2,4-Dichlorophenol	3/25/2020	7/1/2006	PPRTV, July 2007
2,4-Dinitrophenol	4/14/2020	3/1/2010	ATSDR, March 2011
2,4-Dinitrotoluene	12/17/2019	2/1/2015	ATSDR, February 2016
2,6-Dinitrotoluene	2/13/2020	2/1/2015	ATSDR, February 2016
2-Aminotoluene	4/6/2020	12/1/2011	PPRTV, December 2012
2-Methylnaphthalene	3/25/2020	9/1/2006	PPRTV, September 2007
4-Androstene-3,17-dione	4/14/2020	no date limit	none
4-tert-Octylphenol	4/7/2020	no date limit	none
Acetophenone	4/14/2020	6/1/2010	PPRTV, June 2011
Ammonia	3/25/2020	9/1/2015	IRIS, September 2016
Anthraquinone	1/14/2020	2/1/2010	PPRTV, February 2011
Benzophenone	2/7/2020	no date limit	none
Bisphenol A	12/17/2019	9/1/1987	IRIS, September 1988
Boron	2/13/2020	11/1/2009	ATSDR, November 2010
Bromochloromethane	4/14/2020	9/1/2008	PPRTV, September 2009
Butyl benzyl phthalate	12/17/2019	10/1/2001	PPRTV, October 2002
Camphor	4/7/2020	no date limit	none
Carbon disulfide	2/13/2020	8/1/1995	ATSDR, August 1996
Chlorodifluoromethane	3/25/2020	8/1/2002	IRIS, August 2003
Cobalt	10/22/2019	8/1/2007	PPRTV, August 2008
Cotinine	4/6/2020	no date limit	none
Diethyl phthalate	4/14/2020	6/1/1994	ASTDR, June 1995
Di-n-butyl phthalate	12/17/2019	9/1/2000	ATSDR, September 2001
Fluoranthene	3/13/2020	12/1/2011	PPRTV, December 2012
Galaxolide	4/6/2020	no date limit	none
Isophorone	2/13/2020	7/1/2017	ATSDR, July 2018
Isopropylbenzene (cumene)	4/14/2020	9/1/2001	IRIS, September 2002
Lithium	10/21/2019	6/1/2007	PPRTV, June 2008
Manganese	10/25/2019	5/1/2018	HC, May 2019
Methylbenzotriazole	3/13/2020	no date limit	none
Methylmercury	1/15/2020	3/1/2012	ATSDR, March 2013
Methyl tert-butyl ether	1/14/2020	1/1/2005	HC, January 2006
Molybdenum	2/13/2020	4/1/2016	ATSDR, April 2017
Nonylphenol	2/7/2020	no date limit[a]	none[b]
p-Cresol	12/17/2019	9/1/2009	PPRTV, September 2010

Chemical	Search Date	Search Date Limit	Most Recent Assessment
Phenanthrene	3/13/2020	3/1/2008	PPRTV, March 2009
p,p'-Dichlorodiphenyldichloroethylene	10/22/2019	9/1/2016	PPRTV, September 2017
Pyrene	3/25/2020	9/1/2006	PPRTV, September 2007
Quinoline	3/16/2020	7/1/2005	IRIS, July 2006
Silicon	2/13/2020	no date limit	none
Tin	3/25/2020	8/1/2004	ATSDR, August 2005
Tributyl phosphate	12/17/2019	9/1/2011	ATSDR, September 2012
Triethyl citrate	4/14/2020	no date limit	none
Tris(1,3-dichloro-2-propyl) phosphate	2/13/2020	9/1/2011	ATSDR, September 2012
Tris(2-butoxyethyl) phosphate	10/22/2019	9/1/2011	ATSDR, September 2012
Tris(2-chloroethyl) phosphate	2/7/2020	9/1/2011	ATSDR, September 2012
Tungsten	1/14/2020	9/1/2014	PPRTV, September 2015
Vanadium	10/22/2019	9/1/2011	ATSDR, September 2012

PPRTV = Provisional Peer-Reviewed Toxicity Values; ASTDR = Agency for Toxic Substances and Disease Registry; IRIS = Integrated Risk Information System; HC = Health Canada; NA = not applicable.
[a]no date limit = search date was open ended.
[b]none = no previous assessment identified.
[c] A literature search was conducted as part of a separate EPA and CalEPA joint effort. The search was date limited from 2009 to 4/12/2019 or 4/15/2019, depending on the database. Title-abstract and full text screening were completed with PECO criteria very similar to the CCL 5 PECO statement. Thus, EPA used the results of this literature search and screen and began the review efforts for 1,4-dioxane at the study quality stage.

b) Peer-Reviewed Study Identification

The next portion of the literature identification step included searches for peer-reviewed human and animal studies related to chronic oral exposure to the PCCL 5 chemicals of interest. To ensure that all relevant literature for each chemical was captured, EPA first curated a list of search synonyms for each chemical using two databases: the CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard) and ChemID*Plus* (https://chem.nlm.nih.gov/chemidplus/). The Chemicals Dashboard was searched using DSSToxIDs previously assigned for each chemical (see Chapter 2). The active CASRN retrieved from this DSSToxID search was then used to search the ChemID*Plus* 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) were included in the search string. Duplicate and ambiguous synonyms were 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 Lithium is provided in Table F-2. Duplicate references across the two databases were removed.

Table F-2. Example PubMed and Web of Science Search Strings for Lithium

Date of Search: 10/21/2019; Date Limit: 6/01/2007 (most recent assessment: PPRTV, June 2008) Language = English Number of results = 5,127 Database = PubMed	
Set	**Search Strategy**
Set 1 (Synonyms)	("DTXSID5036761"[tiab] OR "7439-93-2"[rn] OR "Lithium"[mh] OR "Lithium"[tiab] OR "Lithium metal"[tiab] OR "Lithium atom"[tiab] OR "Lithium element"[tiab] OR "UN 1415"[tiab] OR "Lithium, elemental"[tiab] OR "EC 231-102-5"[tiab] OR "EINECS 231-102-5"[tiab] OR "HSDB 647"[tiab] OR "Lithium, metallic"[tiab] OR "UNII-9FN79X2M3F"[tiab] OR "UN1415"[tiab])
Set 2 (Tox Filter)	AND (Tox[sb] OR "Toxicol Sci"[TA])
Limit: Language	AND (English[lang])
Date of Search: 10/21/2019; Date Limit: 6/01/2007 (most recent assessment: PPRTV, June 2008) Language = English All terms searched in Topic (Title, Abstract, and Keywords) Number of results = 536 Database = Web of Science	
Set	**Search Strategy**
Set 1 (Synonyms)	("DTXSID5036761" OR "7439-93-2" OR "Lithium" OR "Lithium metal" OR "Lithium atom" OR "Lithium element" OR "UN 1415" OR "Lithium, elemental" OR "EC 231-102-5" OR "EINECS 231-102-5" OR "HSDB 647" OR "Lithium, metallic" OR "UNII-9FN79X2M3F" OR "UN1415")
Set 2 (Tox filter)	AND (("adverse effects" AND ("Amino Acids, Peptides, and Proteins " OR "Biological Factors " OR "Biomedical Materials" OR "Dental Materials" OR Carbohydrates OR "Chemical Actions" OR "Chemical Uses" OR "Complex Mixtures" OR "drug therapy" OR "Environment Health" OR "Public Health" OR Enzymes OR Coenzymes OR food OR

	beverages OR Hormones OR "Hormone Substitutes" OR "Hormone Antagonists" OR "Heterocyclic Compounds" OR "household products" OR Lipids OR "Macromolecular Substances" OR "Nucleic Acids" OR Nucleotides OR Nucleosides "Pharmaceutical Preparations" OR Phytochemicals OR "Polycyclic Compounds" OR radiotherapy)) OR (("chemically induced" OR "chemical induced") AND ("Animal Diseases" OR "Cardiovascular Diseases" OR "Congenital Diseases" OR "Congenital Abnormalities" OR "Hereditary Diseases" OR "Hereditary Abnormalities" OR "Neonatal Diseases" OR "Neonatal Abnormalities" OR "Digestive System Diseases" OR "Disorders of Environmental Origin" OR "Environmental Disorders" OR "Endocrine System Diseases" OR "Eye Diseases" OR "Urogenital Diseases" OR "Pregnancy Complications" OR "Hemic Diseases" OR "Lymphatic Diseases" OR "Immune System Diseases" OR "Immune Diseases" OR "mental disorders" OR "Musculoskeletal Diseases" OR "Neoplasms" OR "Cancer" OR "Nervous System Diseases" OR "Nutritional Diseases" OR "Metabolic Diseases" OR "Otorhinolaryngologic Diseases" OR "Pathological Conditions" OR "Pathological Signs" OR "Pathological Symptoms" OR "Respiratory Tract Diseases" OR "Stomatognathic Diseases" OR "Skin Diseases" OR "Connective Tissue Diseases" OR "Liver injury")) OR (("drug effects" OR "drug induced") AND ("birth weight" OR "Genetic Phenomena" OR "Integumentary System Physiological Phenomena" OR "Ocular Physiological Phenomena" OR "Reproductive Physiological Phenomena" OR "Urinary Physiological Phenomena" OR "liver injury")) OR "drug-induced abnormalities" OR "occupational accidents" OR "adverse drug reaction reporting systems" OR "Drug-Induced Akathisia" OR "biohazard release" OR "chemical burns" OR carcinogen* OR Carcinogenesis OR cardiotox* OR Cardiotoxicity OR "chemical hazard release" OR "chemical terrorism" OR "Chemically-Induced Disorders" OR "chemical induced disorders" OR "Colony Collapse" OR "Drug Interactions" OR "Drug Recalls" OR "Drug-Induced Dyskinesia" OR ecotox* OR Ecotoxicology OR "Environmental Health" OR "environmental illness" OR "environmental monitoring" OR "environmental pollutants" OR "environmental pollution" OR "Environmental Restoration" OR "Environmental Remediation" OR "Fetal Alcohol Spectrum" OR "forensic toxicology" OR "hazardous substances" OR hepatotox* OR immunotox* OR "Metabolic Inactivation" OR "LC50" OR "Material Safety Data Sheets" OR mutagen* OR mutagenesis OR nephrotox* OR neurotox* OR noxae OR "occupational diseases" OR "persian gulf syndrome" OR Pesticides OR poison* OR poisoning OR "substance-induced psychoses" OR terata* OR terato* OR Teratogenesis OR "Toxic Actions" OR toxic OR "toxicity tests" OR Toxicokinetics OR "Toxicological Phenomena" OR toxicology OR toxif* OR toxig* OR "Toxin-Antitoxin Systems")

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 applied to the identified literature in order 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.

Step 2: Title-abstract screening

EPA defined population, exposure, control, and outcome (PECO) criteria to determine relevance to animal hazard for the title-abstract screening (Step 2) and full text reviews (Step 3). Table F-3 presents the CCL 5 PECO statement outlining inclusion criteria for animal hazard studies. 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. Studies solely describing human health effects due to chemical exposure are not amenable to the RSR process due to the complexity of epidemiological data and the level of effort required to extract

relevant results. Therefore, further descriptions of health effect data derived exclusively from human studies are not included here.

Table F-3. Animal Hazard PECO Statement for Rapid Systematic Review Screening

PECO Element	Evidence
Populations	**Animal:** 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.
	In vitro/cell toxicity studies or *in silico*/modeling toxicity studies should be tagged as "supplemental".
Exposures	**Relevant Chemical Forms:** [a]
	Animal: Controlled exposure to the chemical of interest via oral routes. Any exposure length is acceptable for reproductive or developmental exposure. 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 exposure levels. Studies involving exposure to mixtures will be included only if animals are exposed exclusively to the relevant chemical at 2 exposure levels.
	Acute exposure (<28 days), alternative exposure routes, (e.g., inhalation, dermal, injection or unknown/multiple routes), single dose groups, and exposure to mixtures will be tagged as "supplemental".
Comparators	**Animal:** A concurrent control group exposed to vehicle-only treatment or an untreated control.
Outcomes	All health outcomes (both cancer and noncancer), including clinical chemistry endpoints. Meta-analysis presenting new hazard findings from a compilation of existing literature should be included. Studies evaluating hazard in animals with a gene knock-out should be included with an additional supplemental tag. Studies evaluating changes in organ morphology, even if the study design is targeted at evaluating protective effects, should be included.
	Studies containing only mechanistic data should be excluded and will be tagged as "supplemental." Studies evaluating hazards or mechanisms in disease models (e.g. mice pretreated to induce diabetes or mania) should be excluded and tagged as "supplemental."

[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 and assessed any studies with an "unsure" tag. 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. Tag categories and their descriptions are provided in Table F-4.

Table F-4. Tags for Rapid Systematic Review Title-Abstract Screening

Category	Description
Included	
Animal Hazard	Reference meets animal PECO criteria in Table F-3.
Unsure	Full text review is required to determine whether a reference is relevant.
Excluded	
Human Hazard	Reference does not meet animal PECO or supplemental criteria.
Supplemental	
Supplemental	Add this tag if the study contains any of the following types of information: - **Alternative Exposure Route/Duration/Levels:** non-PECO exposure route (e.g., inhalation, dermal, injection), duration < 28 days, or single exposure level. - **Mixture:** target chemical administered as a mixture. - **Alternative Species:** non-PECO vertebrate. - **Mechanistic**: data on mode of action (e.g. oxidative stress, genotoxicity, DNA/RNA/protein inductions, bioinformatics). - **In Vitro**: exposure occurred *in vitro* (cells, tissues, biochemical reactions). - **Toxicokinetics (TK)**: includes TK or physiologically-based pharmacokinetic (PBPK) models; data on mammalian absorption, distribution, metabolism, or excretion (ADME). - **Exposure Only:** contains only data on human exposure (e.g. biological matrices, predicted or occupational exposure) or measures target chemical in relevant human exposure matrices (e.g. food, drinking water, air). - **Case-Report**: Study design that reports data for a small number of individuals without a comparison group (human only). - **Secondary Data Source**: Secondary data source (e.g. reviews, commentaries, editorials) with hazard data for humans or other mammals.
Notes	
Unsure	If full text review is required, tag as "Include – Unsure" and add a note explaining the reason for uncertainty.
Agency Assessment	If the literature is a relevant agency assessment, include and tag to the appropriate evidence stream (human, animal). In the notes section of SWIFT, indicate "Agency Assessment" so that it can undergo further review.
Abstract Only	If you encounter a relevant reference that is clearly an abstract only, exclude it. In the notes section of SWIFT, indicate "Abstract Only" to track the justification.
Non-English Language	If you encounter a reference that is not in English, exclude it. In the notes section of SWIFT, indicate "Non-English Language" to track the justification.

Reviewers prioritized studies during title-abstract screening using the SWIFT-Active machine learning tool (https://www.sciome.come/swift-activescreener/). This tool uses initial title-abstract screening tag results for each chemical as a training set to develop algorithms that predict the number of relevant studies in the entire pool of references for that chemical (Howard et al., 2020). References most likely to be relevant to PECO criteria are prioritized and provided to screeners for review first. This allows for the review of a fraction of the references from the entire literature search for a given chemical. For this RSR, only references that meet 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,

- SWIFT-Active predicted that >80% of relevant references were identified and the last 300 references screened were not relevant, or
- all references were screened.

In special cases, screening was stopped prior to meeting one of these conditions. For example, 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 to be amenable to the screening, review, and extraction steps of the RSR protocol. Similarly, the review for pyrene was temporarily halted because the original reference list was found to contain a high number of benzo(a)pyrene studies indicating the literature search had inadvertently captured an off-topic chemical exposure. In the latter case, reviewers employed a Keyword Analysis Tool (KAT), a tool developed by ICF International Inc., used when an off-topic term skews literature search results. The KAT allowed for the removal of the approximately 2,000 references that were identified to only contain terms for benzo(a)pyrene and screeners were able to resume title-abstract screening for relevant pyrene references.

Step 3: Full text review and study quality evaluation

a) 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://hawcproject.org/portal/). EPA completed full text reviews concurrently with the streamlined study quality evaluation described in Step 3b. References identified as "include", or relevant, during title-abstract screening were subject to a full text review comprised of a primary review and a secondary quality control review by a senior staff member. The animal hazard PECO criteria (Table F-3) were again used to confirm reference relevancy. In the full text review stage, EPA also reviewed studies identified during the title-abstract screen as "supplemental" and tagged accordingly to catalogue potentially useful information. EPA did not evaluate studies identified as only supplemental past the full text review phase. EPA conducted study quality evaluations in HAWC for each reference determined to meet the animal hazard PECO criteria at the full text review step.

b) Study quality evaluation

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. 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 F-5.

Table F-5. CCL 5 Study Quality Metrics and Overall Score Descriptions

Metric 1 – Exposure
References should be evaluated for the following components of exposure characterization, chemical administration, and exposure timing:
Exposure Characterization • Source and/or CAS RN of the administered chemical were reported. • Purity of the chemical was 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.
Exposure 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 were well reported, sensitive, and appropriately applied. • Outcomes were assessed consistently across exposure groups. • Assessors were blinded to exposure status for subjective outcomes.
Results Presentation • Number of animals used was presented for each exposure group and was 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 were presented with means and a measure of variance for continuous endpoints. For dichotomous endpoints, incidence was reported for each exposure group. • Results were presented separately for sex and age (if relevant).
Note: Relevant outcomes are listed in Table X-3, which include clinical chemistry and histopathology endpoints. Do not score a study based on mechanistic outcomes.
Metric 3 – Confounding
References should be evaluated for the following potential sources of confounding factors:
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 indicate 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 Metric 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 in Table 2, which is not a simple sum of individual metric scores.

- A rating of **Good** should be used when the study fully reports all information requested in Metrics 1-3 and presents no concerns or uncertainties.
- A rating of **Adequate** 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 **Adequate**.
- A rating of **Deficient** 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 **Deficient** overall score indicates that caution should be used when considering data from that study.
- A score of **Critically Deficient** 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 **Critically Deficient**.

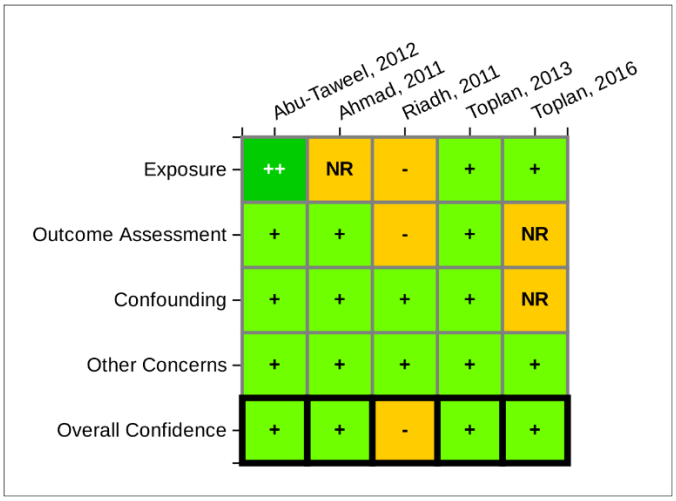
Metric Score	Description
Good	Direct evidence that all components of the criteria were met. No concerns likely to bias the results.
Adequate	Direct or indirect evidence that all components of the criteria were met. Minor concerns are unlikely to significantly bias the results.
Deficient	Evidence that some components of the criteria were not met. Concerns may significantly bias the results.
Not Reported	Information necessary to apply criteria is missing. Explain source of uncertainty in comment.
Critically Deficient	Evidence that components of the criteria were not addressed appropriately. Major concerns are likely to significantly bias the results.

Studies with an overall score of Good, Adequate, or Deficient proceeded to data extraction while studies with an overall score of Critically Deficient were removed from further review. A senior toxicologist reviewed and either confirmed or modified these scores prior to progression to the data extraction step.

Figure F-1 is an example of the output of the study quality assessment conducted in HAWC for lithium. In the case of lithium, five 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 domain (exposure, outcome assessment, confounding, other concerns) as well as the overall confidence for each included study using different colors to visually represent the quality of the chemical's evidence base.

Figure F-1. Example HAWC Study Quality Heatmap for Lithium



Step 4: Data extraction

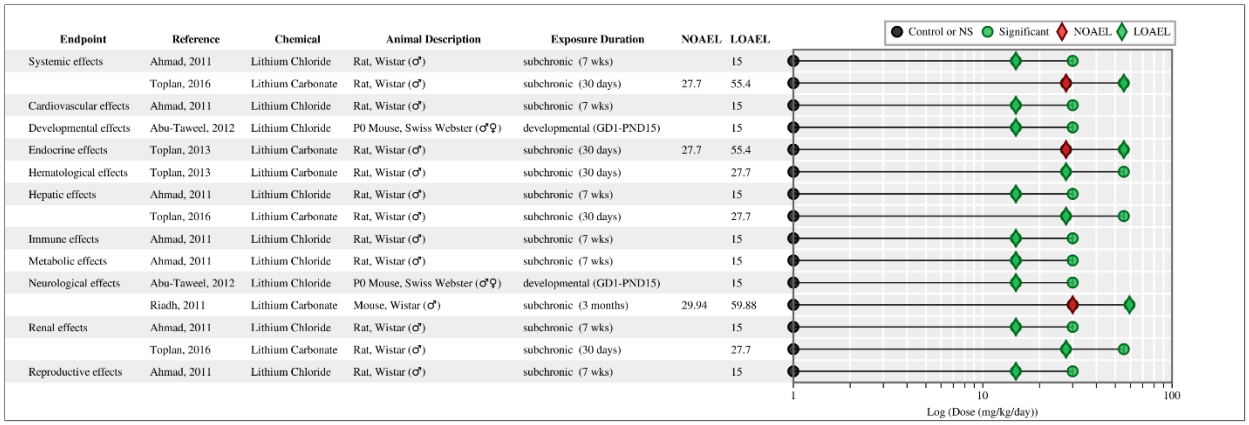
Studies that met the study quality metrics with an overall score of Good (high), Adequate (medium), or Deficient (low) proceeded to the data extraction step. EPA used HAWC to conduct a simple extraction of animal hazard data and capture the LOAEL and NOAEL at the "health outcome category" level. A senior 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 effect category were extracted collectively. EPA considered mechanistic data and outcomes as supplemental data and therefore did not complete data extraction for these endpoints. The possible health effect categories are listed below:

- Carcinogenicity
- Cardiovascular effects
- Dermal effects
- Developmental effects
- Endocrine effects
- Gastrointestinal effects
- Hematological effects
- Hepatic effects
- Immune effects
- Metabolic effects
- Musculoskeletal/ connective tissue effects
- Neurological effects
- Ocular effects
- Renal effects
- Reproductive effects
- Respiratory effects
- Systemic effects

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 effect category, the lowest LOAEL and NOAEL across endpoints were quantified. Data pivots were created in HAWC to summarize the findings across references for each chemical. See Figure F-2 for an

example pivot for Lithium; 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 non-significant response).

Figure F-2. Example HAWC Data Pivot for Lithium



Endpoint	Reference	Chemical	Animal Description	Exposure Duration	NOAEL	LOAEL
Systemic effects	Ahmad, 2011	Lithium Chloride	Rat, Wistar (♂)	subchronic (7 wks)		15
	Toplan, 2016	Lithium Carbonate	Rat, Wistar (♂)	subchronic (30 days)	27.7	55.4
Cardiovascular effects	Ahmad, 2011	Lithium Chloride	Rat, Wistar (♂)	subchronic (7 wks)		15
Developmental effects	Abu-Taweel, 2012	Lithium Chloride	P0 Mouse, Swiss Webster (♂♀)	developmental (GD1-PND15)		15
Endocrine effects	Toplan, 2013	Lithium Carbonate	Rat, Wistar (♂)	subchronic (30 days)	27.7	55.4
Hematological effects	Toplan, 2013	Lithium Carbonate	Rat, Wistar (♂)	subchronic (30 days)		27.7
Hepatic effects	Ahmad, 2011	Lithium Chloride	Rat, Wistar (♂)	subchronic (7 wks)		15
	Toplan, 2016	Lithium Carbonate	Rat, Wistar (♂)	subchronic (30 days)		27.7
Immune effects	Ahmad, 2011	Lithium Chloride	Rat, Wistar (♂)	subchronic (7 wks)		15
Metabolic effects	Ahmad, 2011	Lithium Chloride	Rat, Wistar (♂)	subchronic (7 wks)		15
Neurological effects	Abu-Taweel, 2012	Lithium Chloride	P0 Mouse, Swiss Webster (♂♀)	developmental (GD1-PND15)		15
	Riadh, 2011	Lithium Carbonate	Mouse, Wistar (♂)	subchronic (3 months)	29.94	59.88
Renal effects	Ahmad, 2011	Lithium Chloride	Rat, Wistar (♂)	subchronic (7 wks)		15
	Toplan, 2016	Lithium Carbonate	Rat, Wistar (♂)	subchronic (30 days)		27.7
Reproductive effects	Ahmad, 2011	Lithium Chloride	Rat, Wistar (♂)	subchronic (7 wks)		15

Data extracted from relevant studies are summarized on the health effects page of Contaminant Information Sheet (CIS) for each chemical (see Section 4.4 for further descriptions of CISs and the *Technical Support Document for the Final Fifth Contaminant Candidate List (CCL5) - Contaminant Information Sheets*). For further details related to RSR results for individual chemicals, refer to the file titled "CCL 5 Rapid Systematic Literature Review Results" provided in the CCL 5 docket (EPA-HQ-OW-2018-0594).

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/

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 G - Protocol to Derive Health Concentrations

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

Step 1: EPA identified relevant qualifying health assessments and selected the appropriate toxicity value for derivation of the Health Reference Level (HRL).

For CCL 5, 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 guidelines and methodologies are found in the 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* (2011 edition and individual chapter updates),
- *Recommended Use of Body Weight$^{3/4}$ as the Default Method in Derivation of the Oral Reference Dose* (USEPA, 2011)
- *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)

EPA considered the following health assessments as qualifying assessments:

- EPA Integrated Risk Information System (IRIS) Chemical Assessment Summaries and Toxicological Reviews
- EPA Office of Water Health Advisory (HA) documents and Health Effects Support Documents (HESDs)
- EPA Provisional Peer-Reviewed Toxicity Value (PPRTV) support documents
- EPA Toxic Substances Control Act (TSCA) Risk Evaluations and other technical support documents
- EPA Office of Pesticide Programs (OPP) Human Health Risk Assessments (HHRAs) and Reregistration Eligibility Decision (RED) documents
- 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

If available, websites for these types of health assessments are listed in the references section of this appendix.

If the contaminant was a currently registered pesticide or pesticide metabolite/degradate regulated under the Federal Insecticide, Rodenticide, and Fungicide Act (FIFRA), EPA identified the most recent publicly available EPA OPP health assessment and used the population adjusted dose (PAD) or CSF derived in that assessment to derive the HRL. If the chemical was a registered pesticide, but all uses in the United States have been canceled, EPA followed the procedure for TSCA persistence review (see Section 3.7.2) to determine if the chemical poses risk to human health through drinking water exposure. If the pesticide was deemed persistent, EPA followed the standard procedure for active-use pesticides and identified the most recent available OPP health assessment for data extraction. If the pesticide was not deemed persistent, it was not referred for review by an evaluation team and no further action was required. Pesticide metabolites or degradates were treated as pesticides only if an OPP assessment assigns or derives toxicity values for these chemicals. If not, EPA identified toxicity values from other sources for the derivation of health concentrations.

If a chemical had a single assessment that provides 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 assessments meeting the acceptance criteria, EPA derived the HRL from the most recent EPA assessment, unless an approved assessment from another source incorporated critical studies published after EPA's most recent assessment in their toxicity value derivations. If multiple assessments were available from EPA, or if there were multiple assessments presenting more current science than the most recent EPA assessment, EPA selected the most recent published assessment to derive the HRL.

If no qualifying health assessments were available for a chemical, EPA searched for relevant non-qualifying health assessments, as described below.

Step 2: If qualifying health assessments were not available, 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, 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, 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 and provides relevant health information but does not necessarily follow standard EPA methodologies and/or is not externally peer-reviewed by subject matter experts. EPA generally has not considered these assessments for regulatory purposes in the past, but recognizes that they 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, 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,

- Derived No Effect Levels (DNELs) from European Chemicals Agency (ECHA) Registration Dossiers,
- Tolerable Upper Intake (TUI) levels from the Institute of Medicine (IOM) Dietary Reference Intake documents, and
- Lowest Therapeutic Doses (LTDs) from FDA-approved pharmaceutical labels.

If available, websites for these types of health assessments are listed in the references section of this appendix.

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, EPA selected toxicity values from the most recent published assessment for derivation of the CCL Screening Level. If no non-qualifying health assessments were available, EPA referenced toxicity values extracted during the rapid systematic literature review (see Section 4.2.2). During this literature review, EPA identified No Observed Adverse Effect Levels (NOAELs) extracted from available PECO-relevant studies and noted the overall lowest NOAEL and its associated critical effect. If no NOAELs were identified, EPA noted the overall Lowest Observed Adverse Effect Level (LOAEL) and its associated critical effect. Similar to previous CCL protocols (USEPA, 2009), an uncertainty factor (UF) of 1,000 was applied to NOAELs and an UF of 3,000 was applied to LOAELs – these values were then used as surrogate RfDs for derivation of CCL Screening Levels.

If a chemical was used as an active ingredient in an FDA-approved pharmaceutical, EPA preferentially relied on a Lowest Therapeutic Dose (LTD) value extracted from an FDA-approved pharmaceutical label. In CCL 5, EPA considered LTDs as similar to lowest observed effect levels (LOELs and applied a standard uncertainty factor of 3,000 to this dose (UFs of 10x for intraspecies extrapolation, 10x for subchronic-to-chronic study extrapolation, 10x for extrapolation from LOEL to NOEL, and 3x for database deficiencies). The resulting values, referred to as "Screening Levels for Pharmaceuticals", are considered equivalent to an RfD and were used to derive CCL Screening Levels for all pharmaceutical chemicals.

Step 3: Derive the health concentration.

The process used to derive health concentrations was similar to the process the Agency uses to derive HRLs for Regulatory Determination (USEPA, 2020). 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 drinking water intake (DWI), relevant to the target population and critical effect (USEPA, 2019) and multiplying by a 20% relative source contribution to account for non-water sources of exposure (USEPA, 2000). Potentially sensitive subpopulations were considered when determining the target population. Relevant target populations and their corresponding exposure factors are presented in Table G-1.

Table G-1. Exposure Factors Used for Derivation of Health Concentrations

Target Population	DWI	Description of exposure metric	Citation
General Population	33.8 mL/kg-day	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
Bottle-fed infants	151 mL/kg-day	90th percentile combined direct and indirect drinking water consumption of community water, consumers-only, birth to <1 year, normalized by age range duration.	2019 Exposure Factors Handbook Chapter 3, Table 3-58, CSFII 1994-1996, 1998
Pregnant women	33.3 mL/kg-day	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 mL/kg-day	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
Women of childbearing age	35.4 mL/kg-day	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

DWI = drinking water intake; NHANES = national health and nutrition examination survey; CSFII = continuing survey of food intake by individuals

Table G-2 exhibits the formulae used to derive health concentrations from the various data elements. All health concentrations were converted to units of µg/L to compare with CCL 5 occurrence concentrations. If a chemical had no available qualifying or non-qualifying health assessments or studies identified through the rapid systematic review process, or if the available health assessments elect not to derive toxicity values, EPA did not derive a health concentration.

Table G-2. 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	$\boldsymbol{HRL} = \frac{1x10^{-6}}{CSF} * \sum_i \left(\frac{F_i * ADAF_i}{DWI_i}\right)$

HRL = Health Reference Level; RfD = reference dose; DWI = drinking water intake; RSC = relative source contribution; NOAEL = no observable adverse effect limit; LOAEL = no observable adverse effect limit; CSF = cancer slope factor; ADAF = age-dependent adjustment factor

Note: concentrations derived from NOAELs and LOAELs 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.

Toxicity values identified through this process were used to inform and derive several other health effects metrics including the potency attribute score (Section 4.3.3.1) and severity category (Section 4.3.3.2). For each PCCL 5 chemical, all health-related information was compiled and presented on the corresponding Contaminant Information Sheet (CIS). Adaptations of this compilation of health effects data, edited to fit this document, are presented in Table G-3 (non-cancer effects) and Table G-4 (cancer effects). Table G-3 depicts the derivation of an RfD-based (non-cancer) HRL for lithium while Table G-4 depicts the derivation of a CSF-based (cancer) HRL for oxadiazon. The health concentrations are presented in columns 10 and 11 of Table G-3 and Table G-4, respectively.

Table G-3. Example Health Assessment Data Compilation for Non-Cancer Effects of Lithium

1	2	3	4	5	6	7	8	9	10	11
Name	**DTXSID**	**Assessment source**	**Assessment title (date)**	**RfD (or equivalent) value**	**RfD critical study**	**Critical effect**	**Target population**	**Exposure factor (mL/kg-day)**	**HRL (µg/L, 1 sig. figure)**	**Notes on non-cancer HRL**
lithium	DTXSID 5036761	PPRTV	Provisional Peer Reviewed Toxicity Values for Lithium (2008)	2 µg/kg-day	Baldessarini and Tarazi, 2001	renal, neurologic and endocrine gland effects	general population	33.8	10	"The onset of impaired renal concentrating capacity typically is within the first 2 years of treatment. Although altered renal function appears to be reversible early in treatment, it may be progressive during the first decade of lithium treatment, leading to irreversible damage over time."

PPRTV = Provisional Peer-Reviewed Toxicity Values; RfD = reference dose; HRL = health reference level

Table G-4. Example Health Assessment Data Compilation for Cancer Effects of Oxadiazon

1	2	3	4	5	6	7	8	9	10	11	12
Name	**DTXSID**	**Assessment source**	**Assessment title (date)**	**CSF**	**CSF critical study**	**Tumor types or locations**	**Cancer classification**	**Target population**	**Exposure factor (mL/kg-day)**	**HRL (ug/L, 1 sig. figure)**	**Notes on cancer HRL**
oxadiazon	DTXSID 3024239	OPP	Human Health Scoping Document in Support of Registration Review (2014c)	0.0711 $(mg/kg\text{-}day)^{-1}$	Shirazu, 1987	increase in liver adenomas and/or carcinomas combined in males	L	general population	33.8	0.4	"A dose-related increase in transformation frequencies was observed in an *in vitro*… assay, but other assays for mutagenic or clastogenic potential were negative."

OPP = Office of Pesticide Programs; CSF = cancer slope factor; L cancer classification = likely to carcinogenic to humans; 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, 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 were subsequently used for derivation of the 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. 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/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://echa.europa.eu/information-on-chemicals/registered-substances.

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/guidelines-canadian-drinking-water-quality-summary-table.html.

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/production/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. 2005. Guidelines for Carcinogen Risk Assessment. Risk Assessment Forum. https://www3.epa.gov/ttn/atw/cancer_guidelines_final_3-25-05.pdf.

USEPA. 2005. Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens. Risk Assessment Forum. https://www.epa.gov/sites/production/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. 2008. Provisional Peer Reviewed Toxicity Values for Lithium. Office of Research and Development.

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

USEPA. 2011. Recommended Use of Body Weight$^{3/4}$ 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. 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 Advisory, Risk Assessment Forum. https://www.epa.gov/sites/production/files/2015-01/documents/ddef-final.pdf.

USEPA. 2014c. Oxadiazon. Human Health Assessment Scoping Document in Support of Registration Review. DP No. D420616. Office of Chemical Safety and Pollution Prevention.

USEPA. 2018. 2018 Edition of the Drinking Water Standards and Health Advisories Table. Office of Water. https://www.epa.gov/sites/production/files/2018-03/documents/dwtable2018.pdf.

USEPA. 2019. 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. 2020. Announcement of Preliminary Regulatory Determinations for Contaminants on the Fourth Drinking Water Contaminant Candidate List. 40 CFR Part 141, Volume 85, No. 47, pgs. 14098-14142. EPA–HQ–OW–2019–0583; FRL–10005–88– OW.

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

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/provisional-peer-reviewed-toxicity-values-pprtvs-assessments.

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/chemicals-undergoing-risk-evaluation-under-tsca.

World Health Organization (WHO). 2017. Guidelines for Drinking-Water Quality (4th Edition). https://www.who.int/water_sanitation_health/publications/drinking-water-quality-guidelines-4-including-1st-addendum/en/

Appendix H - Protocol to Select Water Concentrations Used in Calculating Final Hazard Quotients

A. Does the chemical have UCMR 1-4 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. the next highest percentile (95th or 99th)
        iii. maximum
    2. No –> move on to item B.

B. Does the chemical have UCM Round 1 or Round 2 data with greater than 0 detects?
    1. Yes –> select concentration in the following order, depending upon availability:
        i. 90th percentile
        ii. the next highest percentile (95th or 99th)
        iii. maximum

        (note: if a compound has both Round 1 *and* Round 2 data, choose the higher of the two reported concentrations)
    2. No –> move on to item C.

C. Does the chemical have NIRS data with greater than 0 detects?
    1. Yes –> select concentration in the following order, depending upon availability:
        i. 90th percentile
        ii. the next highest percentile (95th or 99th)
        iii. maximum
    2. No –> move on to item D.

D. Does the chemical have Disinfection Byproducts ICR finished concentration data with greater than 0 detects?
    1. Yes –> select concentration in the following order, depending upon availability:
        i. 90th percentile
        ii. the next highest percentile (95th or 99th)
        iii. maximum
    2. No –> move on to item E.

E. Does the chemical have NAWQA (total ambient water) concentration data with greater than 0 detects?
    1. Yes –> select concentration in the following order, depending upon availability:

i. 90th percentile
ii. the next highest percentile (95th or 99th)
iii. maximum

2. No –> move on to item F.

F. Is the chemical a pesticide with modeled concentration data from an OPP evaluation?
   1. Yes –> select the highest value from "Surface Water Chronic" or "Ground Water Chronic" modeled concentration in the following order, depending upon availability:
      i. 90th percentile
      ii. the next highest percentile (95th or 99th)
      iii. maximum
   2. No –> move on to item G.

G. 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. If 90th percentile is not available, choose the next highest percentile, then the maximum.
         ii. If there are multiple states with overlapping monitoring periods (data available from the same period), select the highest 90th percentile concentration value. If 90th percentile is not available, choose the next highest percentile, then the maximum.
      b. USDA PDP
         i. 90th percentile
         ii. the next highest percentile (95th or 99th)
         iii. maximum
      c. Individual studies that are primary data sources (Glassmeyer et al. 2017, Furlong et al. 2017, Bradley et al. 2018, Batt et al. 2016, and Sun et al. 2016) and results from the literature search
         i. If multiple studies provide concentration data for a compound, choose the highest 90th percentile concentration available. If 90th concentrations are not available, choose the next highest percentile, then the maximum.
      d. Community Water Systems survey

i. 90th percentile
ii. the next highest percentile (95th or 99th)
iii. maximum

2. No –> move on to item H

H. 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 ("total ambient water")
            i. 90th percentile
            ii. the next highest percentile (95th or 99th)
            iii. maximum
        b. CA Surf
            i. 90th percentile
            ii. the next highest percentile (95th or 99th)
            iii. maximum
        c. USDA PDP ("all ambient water" data)
            i. 90th percentile
            ii. the next highest percentile (95th or 99th)
            iii. maximum
        d. Individual studies that are primary data sources (e.g., Glassmeyer et al., 2017; Furlong et al., 2017; Bradley et al., 2017) and results from the literature search
            i. If multiple studies provide concentration data for a compound, choose the highest 90th percentile concentration available. If the 90th percentile concentrations are not available, choose the maximum value.

    2. No –> move on to item I.

I. Does the chemical have wastewater treatment plant (WWTP) effluent concentration data from primary 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. the next highest percentile (95th or 99th); select the highest next percentile if multiple studies are available.
        iii. maximum; select the highest maximum concentration if multiple studies are available.
    2. No –> move on to item J.

J. If none of the concentrations listed above are available, an fHQ is not calculated and the fHQ entry on the CIS is left blank.

Appendix I - Protocol to Determine Potency Attribute Scores

This appendix outlines the three-step protocol used to identify data and derive and select the final potency score for a chemical of interest. Table I-1, referenced throughout this protocol, provides an example of the information gathered during the identification of health effects data relevant to the potency attribute score for 2,6-dinitrotoluene (2,6-DNT).

Table I-1. Health Assessment Data Relevant to the Potency Score Extracted for 2,6-Dinitrotoluene (2,6-DNT)

Chemical	DTXSID	Assessment Source	Health Assessment (Date)	Toxicity Value	Critical Study	Potency Equation	Potency Score
2,6-DNT	DTXSID 5020528	PPRTV	Provisional Peer-Reviewed Toxicity Values for 2,6-Dinitrotoluene (2013)	0.0003 mg/kg-day (RfD)	Lee et al. (1976)	Score = -1.7827 – $\log_{10}$(RfD) + 5	7
2,6-DNT	DTXSID 5020528	OW	Drinking Water Health Advisory for 2,4-Dinitrotoluene and 2,6-Dinitrotoluene (2008)	0.001 mg/kg-day (RfD)	Lee et al. (1976)	Score = -1.7827 – $\log_{10}$(RfD) + 5	6
2,6-DNT	DTXSID 5020528	PPRTV	Provisional Peer-Reviewed Toxicity Values for 2,6-Dinitrotoluene (2013)	1.5 $(mg/kg\text{-}day)^{-1}$ (CSF)	Leonard et al. (1987)	Score = -(-0.5302) + $\log_{10}$(CSF) + 5	6
2,6-DNT	DTXSID 5020528	OW	Drinking Water Health Advisory for 2,4-Dinitrotoluene and 2,6-Dinitrotoluene (2008)	0.667 $(mg/kg\text{-}day)^{-1}$ (CSF)	Ellis et al. (1979); Lee et al. (1985)	Score = -(-0.5302) + $\log_{10}$(CSF) + 5	5
2,6-DNT	DTXSID 5020528	IRIS	Chemical Assessment Summary, 2,4-/2,6-Dinitrotoluene Mixture (1990)	0.68 $(mg/kg\text{-}day)^{-1}$ (CSF)	Ellis et al. (1979)	Score = -(-0.5302) + $\log_{10}$(CSF) + 5	5

2,6-DNT = 2,6-Dinitrotoluene; PPRTV = Provisional Peer-Reviewed Toxicity Values; OW = Office of Water; IRIS = Integrated Risk Information System; RfD = reference dose; CSF = cancer slope factor

Step 1: Identify toxicity values from available sources of health effects information.

If available, EPA first identified toxicity values (RfDs, CSFs, etc.) extracted from all published health assessments; when no health assessments existed, EPA extracted toxicity values (NOAELs or LOAELs) from studies identified through the rapid systematic literature review (see Section 4.2.2.1). EPA compiled toxicity values from these sources in a table along with other relevant health information. An adapted version of this table depicting the available health information with columns related to the potency of 2,6-dinitrotoluene (2,6-DNT) is provided above (Table I-1).

Step 2: Derive potency scores for each extracted toxicity value.

EPA derived a potency score for each toxicity value using the equations listed in Table 8 in Section 4.3.3.1 of this document. Each type of toxicity value has a different potency scoring equation based on the distribution and calibration of the available values; EPA selected the appropriate equation for derivation based on the type of toxicity value presented (i.e., EPA derived a potency score for an RfD using the equation calibrated for RfDs). EPA did not derive a potency score if toxicity values were not identified for the chemical of interest. Table I-1 lists the available toxicity values for 2,6-DNT, each associated potency equation, and the subsequently derived scores.

Step 3: Select the final potency score.

EPA selected the potency score that corresponded with the toxicity value used to derive the health concentration (i.e., HRL or CCL screening level) and severity category to list on the summary page of the Contaminant Information Sheet (CIS). Because this potency score is associated with the health concentration selected to derive the final hazard quotient (see Section 4.3.1 and 4.3.2, respectively), it is not necessarily the highest potency score available for a chemical. For instance, in the example of 2,6-DNT presented in Table I-1, the health concentration and potency score selected were based on a CSF extracted from the 2013 PPRTV assessment (highlighted in yellow).

Due to differences in scale calibrations, potency scores derived for one type of toxicity value should only be compared to potency scores derived from that same type of toxicity value. In the case of 2,6-DNT, although the potency score based on the RfD extracted from the same 2013 PPRTV assessment is higher than the potency score associated with the CSF and cancer effects, the potency score derived from the CSF was selected to list on the summary page of the CIS. This was because the health concentration derived from the CSF is more health protective (i.e. lower) than that derived from the RfD and because the health concentration from the CSF was further used to derive the final hazard quotient. However, EPA provides potency scores that correspond with toxicity values from additional assessments or toxicity value types in the health effects section of the CIS as an additional resource for the chemical evaluators.

References

USEPA. 1990. Chemical Assessment Summary, 2,4-/2,6-Dinitrotoluene mixture. National Center for Environmental Assessment, Integrated Risk Information System (IRIS).

USEPA. 2008. Drinking Water Health Advisory for 2,4-Dinitrotoluene and 2,6-Dinitrotoluene. Office of Water.

USEPA. 2013. Provisional Peer-Reviewed Toxicity Values for 2,6-Dinitrotoluene. Office of Research and Development.

Appendix J - Protocol to Determine Severity Attribute Scores

This appendix outlines the steps required to identify the appropriate severity category for a chemical of interest. Table J-1, referenced throughout the protocol, provides an example of the information gathered during the identification of health effects data relevant to the severity category for 2,6-dinitrotoluene (2,6-DNT).

Table J-1. Health assessment data extracted for 2,6-Dinitrotoluene (2,6-DNT)

Chemical	DTXSID	Assessment Source	Health Assessment (Date)	Toxicity Value	Critical Study	Critical Effect	Severity Categories	Final Severity
2,6-DNT	DTXSID 5020528	PPRTV	Provisional Peer-Reviewed Toxicity Values for 2,6-Dinitrotoluene (2013)	0.0003 mg/kg-day (RfD)	Lee et al. (1976)	Increased incidence of splenic extramedullary hematopoiesis	Non-cancer effects	Non-cancer effects
2,6-DNT	DTXSID 5020528	OW	Drinking Water Health Advisory for 2,4-Dinitrotoluene and 2,6-Dinitrotoluene (2008)	0.001 mg/kg-day (RfD)	Lee et al. (1976)	Neurotoxicity, Heinz bodies, bile duct hyperplasia, liver and kidney histopathology, and increased incidence of death	Non-cancer effects; Reduced longevity	Reduced longevity
2,6-DNT	DTXSID 5020528	PPRTV	Provisional Peer-Reviewed Toxicity Values for 2,6-Dinitrotoluene (2013)	1.5 $(mg/kg\text{-}day)^{-1}$ (CSF)	Leonard et al. (1987)	Hepatocellular carcinomas	Carcinogen with a linear MOA	Carcinogen with a linear MOA
2,6-DNT	DTXSID 5020528	OW	Drinking Water Health Advisory for 2,4-Dinitrotoluene and 2,6-Dinitrotoluene (2008)	0.667 $(mg/kg\text{-}day)^{-1}$ (CSF)	Ellis et al. (1979); Lee et al., (1985)	Hepatocellular carcinomas and neoplastic nodules; mammary gland adenomas, fibroadenomas, fibromas, and adenocarcinomas/carcinomas	Carcinogen with a linear MOA	Carcinogen with a linear MOA
2,6-DNT	DTXSID 5020528	IRIS	Chemical Assessment Summary, 2,4-/2,6-Dinitrotoluene Mixture (1990)	0.68 $(mg/kg\text{-}day)^{-1}$ (CSF)	Ellis et al. (1979)	Hepatocellular carcinomas and neoplastic nodules; mammary gland adenomas, fibroadenomas, fibromas, and adenocarcinomas/carcinomas	Carcinogen with a linear MOA	Carcinogen with a linear MOA

2,6-DNT = 2,6-Dinitrotoluene; PPRTV = Provisional Peer-Reviewed Toxicity Values; OW = Office of Water; IRIS = Integrated Risk Information System; RfD = reference dose; CSF = cancer slope factor; MOA = mode of action

Step 1: Identify the critical effect and corresponding toxicity value.

EPA first identified the critical effect corresponding with the toxicity value of interest (RfD, CSF, etc.) 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 4.3.3.1) for that chemical. EPA identified critical effects from all available sources of health effects information and compiled them in a health effects data table along with other relevant health information. An adapted version of this table depicting the available health information with columns related to the severity of an example chemical, 2,6-Dinitrotoluene (2,6-DNT), is provided above (Table J-1).

Step 2: Select the appropriate severity category for each critical effect.

Based on the critical effect related to the toxicity value, EPA selected the appropriate severity category. Table J-2 lists the eight possible severity categories. The severity categories selected for each chemical and critical effect were reviewed for accuracy and consistency by EPA experts from the Office of Water's Health and Ecological Criteria Division. If the assessment or study lists multiple critical effects associated with the LOAEL, EPA listed each applicable severity category. If there was no available toxicity value or a corresponding critical effect for a chemical, a severity category was not applied and the entry was left blank. Table J-1 lists the available toxicity values for 2,6-DNT and each associated severity category.

Table J-2. CCL 5 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

Generally, if a chemical is associated with effects unrelated to carcinogenicity, and has co-critical effects that correspond with several severity categories, EPA selected one category for that assessment based on the hierarchy of effects listed below:

reduced longevity > reproductive and developmental effects > non-cancer effects.

An example of this for 2,6-DNT is depicted in Table J-1. In this example, the 2008 Office of Water Health Advisory presents multiple non-cancer co-critical effects for 2,6-DNT. One co-critical effect includes "increased incidence of death" which corresponds with a severity category of "reduced longevity". In this case, while there are also critical effects identified by this assessment that would fall into the severity category of "non-cancer effects", EPA selected "reduced longevity" 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 severity categories. For these instances, the severity category that corresponds to the critical effect and associated toxicity value also used to derive the health concentration (see Section 4.3.1) was selected as the final severity category and listed on the summary page of the Contaminant Information Sheet (CIS).

Generally, the final severity category corresponds to the most protective health concentration. In the example of 2,6-DNT presented in Table J-1, the health concentration and potency score were based on the cancer slope factor from the 2013 Provisional Peer-Reviewed Toxicity Values health assessment (highlighted in yellow). Therefore, the severity category listed on the summary page of the CIS was "carcinogen with a linear mode of action". Other severity categories identified through this process are presented within the health effects section of the CIS as an additional resource for the chemical evaluators.

References

USEPA. 1990. Chemical Assessment Summary, 2,4-/2,6-Dinitrotoluene mixture. National Center for Environmental Assessment, Integrated Risk Information System (IRIS).

USEPA. 2008. Drinking Water Health Advisory for 2,4-Dinitrotoluene and 2,6-Dinitrotoluene. Office of Water.

USEPA. 2013. Provisional Peer-Reviewed Toxicity Values for 2,6-Dinitrotoluene. Office of Research and Development.

Appendix K - 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 K-1. Use the same type of data to score prevalence as for magnitude.

Table K-1. Hierarchy of Prevalence Data Elements

Rank	Prevalence Data Element	Type of Data
1	Percent of PWSs with detections	National scale / representative data (UCMR 1-4 has highest priority, then UCM State Rounds 1-2, then NIRS) from 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 EPA
5	Production volume in pounds per year	Chemical Data Reporting (CDR) data from 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 K-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 K-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 K-2. Prevalence Scoring Scales

Prevalence Score	1	2	3	4	5
	% Finished Water with Detections (PWSs)	% Ambient Water with Detections (Sites/Samples)	# States Reporting Pesticide in Use	# States Reporting TRI Total Releases	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 L - 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 L-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 M).

Table L-1. Hierarchy of Magnitude Data Elements

Rank	Prevalence Data Element	Type of Data
1	Median concentration of PWSs with detection	National scale / representative data (UCMR 1-4 has highest priority, then UCM State Rounds 1-2, then NIRS) from 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 EPA
5	Persistence-mobility	Empirical and modeled environmental fate data from 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 L-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 L-2 (e.g., prevalence is scored using production volume data), use the Persistence-Mobility scoring approach to develop a magnitude score (see Appendix M).

Table L-2. Magnitude Scoring Scales

Magnitude Score	1	2	3	4	5
	Finished Water Median Concentration of Detections(ug/L)	**Ambient Water Median Concentration of Detections(ug/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 M - 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 M-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	Log Octanol/Water Partitioning Coefficient (log K_{OW})	dimensionless	>4	1-4	<1
2	Henry's Law Coefficient (K_H)	dimensionless	>0.042	0.042 – 4.2x 10^{-6}	<4.2x10^{-6}
3	Solubility in water	µg/L	<1000	1000-1,000,000	>1,000,000

Appendix N - Data Management for CCL 5

Section N.1 Overview

EPA documented all processes related to data management and decision-making in developing the CCL 5. This appendix describes the data management, processing, and extraction steps performed for the primary and select supplemental data sources used in developing the CCL 5.

Section N.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 N.3 provides details about the simple data format EPA used to compile and structure data extracted for CCL 5.

Section N.4 provides the data elements and their descriptions extracted from EPA's CompTox Chemicals Dashboard.

Section N.5 provides a list of data elements of the CCL 5 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 EPA's exclusion criteria for data element point assignment.

Section N.6 provides references for sections N.3, N.4, and N.5.

Section N.2 Data Source Descriptions and Pre-Processing Specifics for Primary and Select Supplemental CCL 5 Data Sources

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

Data description: According to ATSDR, "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/mrls/index.html).

ATSDR develops MRLs for the oral and the inhalation route of exposure and for acute, intermediate, and chronic exposure durations. For pre-universe development, 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. This data source was used as a primary data source for CCL 5.

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 April 2018.

Data download: The data were copied and pasted from the table into an Excel spreadsheet. After CDC published additional MRLs for PFNA and PFHxS, the MRLs for these compounds were added to the original data.

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 (equivalent to acute reference doses (RfDs), subchronic RfDs, chronic RfDs, acute reference concentrations (RfCs), subchronic RfCs, and chronic RfCs). Oral data were used in the screening step; however, inhalation data were extracted for use in the classifications step for reference on the Contaminant Information Sheets (CISs).

ATSDR Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) Substance Priority List – 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. The SDWA requires that CERCLA priority substances be considered as part of the CCL development process. This data source was used as a primary data source for CCL 5. (Description adapted from ATSDR's Substance Priority List website.)

Reference: Centers for Disease Control and Prevention (CDC). 2017. 2017 ATSDR Substance Priority List. https://www.atsdr.cdc.gov/spl/index.html. Accessed March 2018.

Data download: EPA downloaded the 2017 Substance Priority List for use in CCL 5.

Data manipulation: Data manipulation was minimal and restricted to adding DTXSIDs.

Extracted data elements: EPA wrote R code to extract list-type data elements, which were assigned a value of 1 to indicate presence on the Substance Priority List.

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

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 primary data source for CCL 5.

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 download: The original NIH-ToxNet website EPA accessed to download the CPDB has since been retired. The CPDB data can now be accessed through this link: https://www.nlm.nih.gov/databases/download/cpdb.html.

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. EPA presented only the minimum and maximum TD_{50} values on CISs for chemicals with multiple entries.

Chemical 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, child-specific reference doses, Proposition 65 safe harbor numbers, soil-screening levels, and fish advisories. This data source was used as a primary data source for CCL 5.

Reference: California Environmental Protection Agency (CalEPA). n.d. Office of Environmental Health Hazard Assessment (OEHHA) Chemicals. https://oehha.ca.gov/chemicals. Accessed May 2019.

Data download: The option to export database as a comma separated values (CSV) file was selected.

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: Public health goals were extracted and treated as chronic duration benchmarks, oral slope factors were extracted and coded as cancer slope factors (CSFs), and notification levels were also coded as chronic benchmarks. Maximum allowable daily levels (MADLs) for chemicals causing reproductive toxicity, inhalation unit risks (IURs), and RfCs were extracted and included in the universe as a reference, but these were not used for screening.

Chemical Data Reporting (CDR) Results – EPA

Data description: These data represent production volume information collected by EPA under the Toxic Substances Control Act (TSCA). This data source was used as a primary data source for CCL 5.

Reference: USEPA. 2016. 2016 Chemical Data Reporting (CDR) Results. https://www.epa.gov/chemical-data-reporting/2016-chemical-data-reporting-results#access. Accessed April 2018.

Data download: EPA downloaded CDR's 2016 National Aggregate Production Volume dataset for use in CCL 5.

Data manipulation: Data manipulation was minimal and limited to adding DTXSIDs.

Extracted data elements: EPA wrote R code to extract the national aggregate production volume data. These data are reported in categories of production volume rather than a numeric sum of production volume (i.e., 1,000,000 - 10,000,000 lb or 1,000,000,000 - 5,000,000,000 lb).

Community Water System Survey (CWSS) – EPA

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. This data source was used as a supplemental data source for CCL 5.

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 detects, and percentage of samples with detects 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 – EPA

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 882,000 chemicals. This data source was used as a supplemental data source for CCL 5.

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 November 2018 from the following website address: https://comptox.epa.gov/dashboard/. Data were downloaded using the batch search tool for all unique DTXSIDs identified during pre-universe development. The batch search tool allows searches only for less than 5,000 unique identifiers at once. Multiple batches were required to search dashboard data for every chemical in the pre-universe.

Data manipulation: Results from the OPERA and TEST models, which were not deemed relevant to the CCL 5 goals, were removed. No other 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 5.

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

Data description: This is an EPA Office of Research and Development 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. This data source was used as a primary data source for CCL 5.

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.

Pre-processing steps for screening:

- **Data download:** EPA downloaded the publication and supplemental data file for use in CCL 5. 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. This data source was considered a non-nationally representative ambient water study for the screening step of CCL 5.

Pre-processing steps 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-detects were reclassified as concentrations below the method reporting level (MRL). 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 detects for each contaminant. This data source was considered a non-nationally representative occurrence study and the water data were categorized as wastewater effluent for the classification step of CCL 5.

Disinfection Byproducts Information Collection Rule (DBP ICR) – EPA

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 monitoring data 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 primary data source for CCL 5.

References: USEPA. 2000. ICR Auxiliary 1 Database. EPA 815-C-00-002.

Pre-processing steps for screening:

- **Data download:** EPA downloaded the DBP ICR Aux 1 Microsoft Access database on October 31, 2018 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 were converted to 0 (non-detects). Maximum concentration of detects for each contaminant was calculated and DTXSIDs were added. Data manipulation steps were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract the maximum concentration of detections. This data source was treated as a nationally representative finished water survey.

Pre-processing steps for classification:

- **Data download:** The DBP ICR Aux 1 database downloaded for screening was used for extracting data used in classification.
- **Data manipulation:** Three Excel worksheets (TUXANLYT, TUXDBP, and TUXSAMPLE) were extracted from the Microsoft Access database. All have different relevant data and are in different data structures, so worksheets were reformatted and combined into one table. Concentration data reported as -999 were converted to 0 (non-detects). Summary statistics of concentration data and detection information were calculated. 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 concentrations of detections in addition to total number of sites, number of sites with detections, and percentage of sites with detects for each contaminant. This data source was treated as a nationally representative finished water survey.

Drinking Water Standards and Health Advisory (DWSHA) Tables – EPA

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 primary data source for CCL 5.

Reference: USEPA. 2018. Edition of the Drinking Water Standards and Health Advisories Tables. https://www.epa.gov/sites/production/files/2018-03/documents/dwtable2018.pdf. Accessed November 2019.

Data download: The data in the PDF document 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. 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.

"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 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. This data source was used as a primary data source for CCL 5.

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.

Pre-processing steps 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. The data source was treated as a non-nationally representative ambient water study.

Pre-processing steps 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, 90^{th} 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. The data source was treated as a non-nationally representative ambient water study.

"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, 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. This data source was used as a primary data source for CCL 5.

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.

Pre-processing steps 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. The data source was treated as a non-nationally representative ambient water study.

Pre-processing steps 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 detects for each contaminant. The data source was treated as a non-nationally representative ambient water study.

Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) registered pesticides and pesticide ingredients – EPA

Data description: This list represents the active pesticide and pesticide ingredients currently registered by EPA in the U. S. The SDWA requires that registered pesticides be considered in CCL development. This data source was used as a primary data source for CCL 5.

Reference: USEPA. 2017. 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 download: The EPA's Pesticide Chemical Search Database contains links to regulatory documents for pesticides (https://iaspub.epa.gov/apex/pesticides/f?p=chemicalsearch). EPA accessed the list of compounds included in the Pesticide Chemical Search Database on October 19, 2018 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 2017.

Data manipulation: No data manipulation was necessary.

Extracted data elements: EPA wrote R code to extract list-type data elements, which were assigned a value of 1 to indicate that a pesticide or pesticide ingredient was registered on the FIFRA list.

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 allowable concentrations (MACs) for chemical and physical parameters in drinking water. This data source was used as a primary data source for CCL 5.

Reference: Health Canada (HC). n.d. Guidelines for Canadian Drinking Water Quality – Summary Table. https://www.canada.ca/en/health-canada/services/environmental-workplace-health/reports-publications/water-quality/guidelines-canadian-drinking-water-quality-summary-table.html. Accessed October 2018.

Data download: EPA copied and pasted Table 2 containing MACs into a CSV file.

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 MACs. MACs were considered chronic benchmarks.

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) was published in 2017. This data source was used as a primary data source for CCL 5.

Reference: World Health Organization (WHO). 2017. *Guidelines for drinking-water quality*. 4th edition, incorporating the 1st addendum. https://www.who.int/publications/i/item/9789241549950. Accessed October 2018.

Data download: EPA downloaded the PDF, accessed the table containing guideline values (Table A3.3), and copied and pasted the values into a CSV file.

Data manipulation: Data manipulation was minimal and restricted to altering the format of chemical identifiers (e.g., adding DTXSIDs).

Extracted data elements: EPA wrote R code to extract the guideline values. Guideline values were treated as chronic benchmarks.

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 primary source for CCL 5 as it is data-rich and the only source of LD_{50} for the CCL 5 process.

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. Accessed April 2019.

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.

Health-Based Screening Levels (HBSLs) – U.S. 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 primary data source for CCL 5.

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 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.

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 primary data source for CCL 5.

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 June 2018.

Data download: EPA copied and pasted the table of health-based guidance values into a CSV file.

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 also altered the format of chemical identifiers for each entry (e.g., added DTXSIDs). Data manipulation steps were conducted using R.

Extracted data elements: EPA wrote R code to extract the acute, subchronic and chronic benchmarks. Short-term and subchronic guidance values were considered subchronic benchmarks.

Human Health Benchmarks for Pesticides – EPA

Data description: The Human Health Benchmarks for Pesticides are published by EPA and were last updated in 2017. 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. 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 primary data source for CCL 5.

Reference: USEPA. n.d. Human Health Benchmarks for Pesticides. https://iaspub.epa.gov/apex/pesticides/f?p=HHBP:home:11786831942978. Accessed March 2018.

Data download: EPA copied and pasted HHBP data into a CSV file.

Data manipulation: EPA selected the 10^{-6} cancer risk level as the basis of the benchmarks to compare cancer risk concentrations across multiple sources. Other 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 acute and chronic benchmarks, acute and chronic population adjusted doses (treated as acute and chronic RfDs, respectively), and CSFs.

Integrated Risk Information System (IRIS) – EPA

Data description: EPA's Office of Research and Development 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 5 process. This data source was used as a primary data source for CCL 5.

Reference: USEPA. n.d. Integrated Risk Information System (IRIS). IRIS Advanced Search. https://cfpub.epa.gov/ncea/iris/search/index.cfm?keyword. Accessed May 2019.

Data download: EPA exported the complete IRIS database as an Excel file.

Data manipulation: EPA altered the format of chemical identifiers for each entry (e.g., added DTXSIDs) and converted cancer classifications from other 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.

Extracted data elements: EPA wrote R code to extract oral toxicity values that include RfDs, subchronic RfDs, and CSFs in addition to cancer classifications. Inhalation data including RfCs and inhalation unit risks (IURs) were also extracted but are not used in the screening step.

International Agency for Research on Cancer (IARC) Cancer 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 primary data source for CCL 5.

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 April 2018.

Data download: EPA downloaded the list of classifications (volumes 1-128) as a CSV file.

Data manipulation: EPA altered the format of chemical identifiers for each entry (e.g., added DTXSIDs) and 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. Data manipulation steps were conducted using R.

Extracted data elements: EPA wrote R code to extract the monograph conclusions (group 1, 2A, 2B, or 3), considered cancer classifications for screening purposes.

"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 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. This data source was used as a primary data source for CCL 5.

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.

Pre-processing steps 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. This data source was considered a non-nationally representative ambient water study for the screening step.

Pre-processing steps 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 detects, and percentage of sites with detects. This data source was considered a non-nationally representative ambient water study for the classification step.

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

Data description: The Food and Drug Administration 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 5 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 primary data source for CCL 5.

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 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, 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. In previous CCLs, MRDDs were considered equivalent to LOAELs. For CCL 5, the MRDDs are considered a distinct toxicity data type.

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

Data description: The Fourth Report of Human Exposure to Environmental Chemicals was published in 2019 by the Centers for Disease Control (CDC). 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 primary data source for CCL 5.

Reference: Centers for Disease Control and Prevention (CDC). 2019. Fourth Report on Human Exposure to Environmental Chemicals, Updated Tables. U.S. Department of Health and Human Services. https://www.cdc.gov/exposurereport/**.** Accessed February 2019.

Data download: EPA downloaded Volumes I and II of the Fourth Report for use in CCL 5. The January 2019 release of this report was the most recent version available for universe development.

Data manipulation: The report was exported into an Excel spreadsheet. The most recent year of results for each compound were copied to a separate data file. The date with the most recent data are variable from compound to compound depending on when the last year of biomonitoring for that analyte occurred. DTXSIDs were added. The table containing minimum reporting levels (MRLs) was amended to the table containing the analyte results.

Extracted data elements: EPA wrote R code to extract the 90th percentile concentrations for each compound in addition to the matrix in which the analyte was measured (blood, serum, and urine).

National Inorganics and Radionuclides Survey (NIRS) – EPA

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. This data source was used as a primary data source for CCL 5.

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.

Pre-processing steps 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. This data source was treated as a nationally representative finished water study.

Pre-processing steps 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 detects for each chemical. This data source was treated as a nationally representative finished water study.

National Primary Drinking Water Regulations – EPA

Data description: National Primary Drinking Water Regulations (NPDWRs) are legally enforceable primary standards and treatment techniques applicable to public water systems. 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 5.

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 April 2019.

Data download: NPDWRs were copied and pasted into a CSV file.

Data manipulation: Data manipulation was minimal and restricted to adding DTXSIDs.

Extracted data elements: EPA wrote R code to extract Maximum Contaminant Levels (MCLs) and Maximum Contaminant Level Goals (MCLGs).

National Recommended Water Quality Criteria - Human Health Criteria – EPA

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.

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 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.

National Toxicology Program (NTP) Cancer Classifications – HHS

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).

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 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.

"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 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. This data source was used as a primary data source for CCL 5.

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.

Pre-processing steps 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. Treated water data were considered finished water data, and source water data were considered ambient water data. This data source was considered a non-nationally representative occurrence study.

Pre-processing steps 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. Source water data were considered ambient water data, and treated water were considered finished water data. Quantitative detection rate data are relevant to the classification step and included on the Contaminant Information Sheets. Sampling year ranges for each study phase and reporting limits were also extracted. This data source was treated as a non-nationally representative occurrence study.

"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 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. This data source was used as a primary data source for CCL 5.

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.

Pre-processing steps 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 (the next data source below). 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. Source water data were treated as ambient water data. Qualitative detection frequencies were used in the screening step as these metrics are a more conservative estimate of detection than quantitative detection rates. This data source was treated as a non-nationally representative occurrence study.

Pre-processing steps 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. Source water data were treated as ambient water data. Quantitative detection rate data are relevant to the classification step and included on the Contaminant Information Sheets. Sampling year ranges for each study phase and reporting limits were also extracted. This data source was treated as a non-nationally representative occurrence study.

Pesticide Data Program (PDP) – 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 primary data source for CCL 5.

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.

Pre-processing for screening:

- **Data download:** EPA downloaded the most recent 10 years (2008-2017) of occurrence data on untreated water, finished water, and groundwater on May 29, 2019, from the website address: https://apps.ams.usda.gov/pdp. The summary of findings option was selected for the output report.
- **Data manipulation:** Percentage detection rates were calculated using fields for the number of samples analyzed and number of samples with detects. If a pesticide had no detections and a limit of detection (LOD) was reported, half of the LOD was replaced for the maximum concentration value. If a pesticide had no detections and a range of LODs were reported, the maximum concentration value was replaced by half of the midpoint of the LOD range. DTXSIDs were added. Data manipulation steps were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract maximum concentrations and percentage of detection data. Groundwater and untreated water are considered ambient water. Finished water samples are considered finished water data. This data source was considered a non-nationally representative occurrence water study.

Pre-processing for classification:

- **Data download:** EPA compiled all water data (untreated, finished and ground water) available from 2001 onward in January 2020 from the website address: https://apps.ams.usda.gov/pdp. The analytical results option was selected for the output report.

- **Data manipulation:** Summary concentrations based on analytical detections and percentage of site detection rates were calculated. DTXSIDs were added. Data manipulation steps were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract minimum, median, 90th percentile and maximum concentration of detections as well as total number of sites, number of sites with detections, and percentage of sites with detects for each contaminant in finished water, untreated water, ground water, and combined untreated and ground water. This dataset was considered a non-nationally representative occurrence study.

Pesticide Use Estimates – 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 primary data source for CCL 5.

References:

U.S. Geological Survey (USGS). n.d. National Water-Quality Assessment (NAWQA) Project: The Pesticide National Synthesis Project. https://water.usgs.gov/nawqa/pnsp/usage/maps/county-level/. Accessed February 2019.

Pre-processing steps for screening:

- **Data download:** EPA downloaded the "High Estimate Agricultural Pesticide Use by Crop Group 1992-2016" dataset. The dataset was converted to a CSV file.
- **Data manipulation:** EPA calculated the total application rates for each compound for each year that data were available using R. DTXSIDs were added.
- **Extracted data elements**: EPA wrote R code to extract the total application rate for the most recent year for each compound.

Pre-processing steps for classification:

- **Data download:** The same data file used in screening was used for extracting data for the classification step.
- **Data manipulation:** No additional data manipulation steps were required.
- **Extracted data elements**: EPA wrote R code to extract the total number of states the pesticide was used in and the most recent year reported associated with the total application rate that was calculated in the pre-processing steps for screening.

"Pharmaceutical manufacturing facility discharges can substantially increase the pharmaceutical load to US 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. This data source was used as a primary data source for CCL 5.

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.

Pre-processing steps for screening:

- **Data download:** EPA downloaded the publication and supplemental data file for use in CCL 5. 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. This study was treated as a non-nationally representative ambient water study in the screening step.

Pre-processing steps 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. This data source was treated as a non-nationally representative wastewater effluent study in the classification step.

"Predicting variability of aquatic concentrations of human pharmaceuticals" – Kostich et al. 2010

Data description: This is an EPA Office of Research and Development 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. This data source was used as a primary data source for CCL 5.

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.

Pre-processing steps 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.

Pre-processing steps 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 primary 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.

Provisional Peer-Reviewed Toxicity Values (PPRTVs) – EPA

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. PPRTVs include toxicity values 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 primary data source for CCL 5.

Reference: USEPA. n.d. Provisional Peer-Reviewed Toxicity Values. https://www.epa.gov/pprtv/provisional-peer-reviewed-toxicity-values-pprtvs-assessments. Accessed March 2019.

Data download: EPA exported PPRTV data as an Excel file from the PPRTV Library housed by Oak Ridge National Laboratory (https://hhpprtv.ornl.gov/quickview/pprtv_compare.php).

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 RfDs, subchronic RfDs, and CSFs were extracted in addition to cancer classifications. Inhalation data including RfCs, subchronic RfCs, and inhalation unit risks (IURs) were also extracted.

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. 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. This data source was used as a primary data source for CCL 5.

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.

Pre-processing steps 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. This data source was treated as a non-nationally representative finished water study.

Pre-processing steps 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 detects for each contaminant. This data source was treated as a non-nationally representative finished water study.

Screening Levels for Pharmaceutical Contaminants – FDA Drugs@FDA database, National Institutes of Health (NIH) DailyMED database

Data description: Screening levels for pharmaceuticals were calculated from human oral dosage and administration information obtained from public access databases containing FDA-approved drug labels (FDA, 2018; NIH, 2018). The lowest (total daily) therapeutic dose (LTD) to an adult patient population was utilized. LTDs are the minimum total daily dose (adjusted for adult body weight) at which a therapeutic effect is achieved and are more similar to a traditional point of departure (i.e., lowest observed effect level [LOAEL]) than the maximum recommended daily dose (MRDD), which was sometimes used as the POD for pharmaceuticals in previous CCL

efforts. Similar to past procedures, an uncertainty factor of 3,000 (10x for intraspecies extrapolation, 10x for subchronic-to-chronic study extrapolation, 10x for extrapolation from the LOEL to no observed effect level [NOEL], and 3x for database deficiencies) and exposure factors were applied to the LTD to derive screening levels for the general population and bottle-fed infants, in final units of µg/L. This data source was used as a primary data source for CCL 5.

Reference: US FDA. 2018. Drugs @ FDA: FDA Approved Drug Products. https://www.accessdata.fda.gov/scripts/cder/daf/. Accessed October 2017.

NIH (National Institutes of Health). 2018. DailyMed database. United States National Library of Medicine. https://dailymed.nlm.nih.gov/dailymed/. Accessed October 2017.

Data download: EPA retrieved FDA-approved labels from the websites listed above and copied and pasted relevant data into Excel files.

Data manipulation: Other than the calculations described above (application of uncertainty factors and exposure factors), data manipulation for this source was minimal and was limited to altering the format of chemical identifiers (e.g., adding DTXSIDs).

Extracted data elements: LTDs were extracted from FDA-approved labels, from which the screening level for each compound was calculated. Screening levels are considered chronic benchmarks for screening purposes.

State Drinking Water Monitoring Datasets and EPA's Third Six-Year Review – EPA

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 Third Six-Year Review (SYR 3) of drinking water regulations, some states submitted PWS occurrence data for unregulated contaminants along with the requested data on regulated contaminants. 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 SYR3 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.

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 5 and included on the Contaminant Information Sheets.

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/drinking-water. 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-data-base. 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.wi.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 detects for each PCCL 5 chemical.

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 detects 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 detects 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 detects for each PCCL 5 chemical.

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

Data description: California's Department of Pesticide Regulation (DPR) 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 primary data source for CCL 5.

Reference: California Department of Pesticide Regulation (DPR). n.d. Surface Water Database (SURF). https://www.cdpr.ca.gov/docs/emon/surfwtr/surfdata.htm. Accessed April 2019.

Pre-processing steps for screening:

- **Data download:** EPA downloaded the complete SURF database.
- **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 NWIS were removed. Later in the data collection process, EPA noticed some USGS data were not included in the NWIS dataset, so it conducted a second round of data processing and included these data in the SURF database. These data-processing steps resulted in two summary data files, which were subsequently combined using R. Maximum concentration and percentage detects were calculated for each contaminant.
- **Extracted data elements**: EPA wrote R code to extract the maximum concentration of detections and percentage of detection information. This data source was treated as a non-nationally representative ambient water study.

Pre-processing steps for classification:

- **Data download:** The data files downloaded for screening were used for extracting data used in classification.
- **Data manipulation:** The summary data files described in the pre-processing steps for screening above were used to extract data for classification. Summary statistics of concentration data and detection information were calculated. 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 detects for each contaminant. This data source was treated as a non-nationally representative ambient water study.

"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 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. This data source was considered as a primary data source for CCL 5 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 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.

Toxicity Reference Database (ToxRefDB) – EPA

Data description: The Toxicity Reference Database (ToxRefDB) contains the results of thousands of *in vivo* animal toxicity studies conducted over the last 30 years. This database was compiled by EPA and released in 2014. The purpose of the database is to describe dose-response animal toxicity data with a standardized vocabulary so that the results are accessible and searchable. This data source was used as a primary data source for CCL 5.

Reference: USEPA. n.d. Exploring ToxCast Data: Downloadable Data. Animal Toxicity Studies: Effects and Endpoints. Toxicity Reference Database. https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data. Accessed July 2018.

Data download: EPA used the Download Animal Toxicity Data link from the website listed above to access the zip file of ToxRefDB data and downloaded nel_lel_noael_loael summary and study_tg_effect_endpoint.

Data manipulation: Data manipulation for this source was minimal and limited to altering the format of chemical identifiers (e.g., adding DTXSIDs).

Extracted data elements: Studies in ToxRefDB are coded and categorized by study type. For the screening step, subacute studies (SAC) are considered acute NOAELs or LOAELs, subchronic studies (SUB) are considered subchronic NOAELs or LOAELs, and chronic (CHR), multigenerational reproductive (MGR), prenatal development (DEV), and reproductive/fertility (REP) studies are considered chronic NOAELs or LOAELs. Both oral and inhalation studies were extracted, though only oral studies were used for screening purposes. For the purpose of screening from the universe to the PCCL, studies marked as having a usability of 1/2 or 3 (guideline acceptable or non-guideline acceptable, respectively) were extracted. Studies marked as having a usability of 4, 5, or 6 (unacceptable, incomplete/deficient report, or not evaluated, respectively) were not included.

Toxics Release Inventory (TRI) – EPA

Data description: The Toxics Release Inventory (TRI) Program was developed by EPA as part of the Emergency Planning and Community Right-to-Know Act to inform citizens of chemical releases from industrial facilities. TRI tracks the industrial management of toxic chemicals that may cause harm to human health and the environment. A release refers to emitting a compound to the air, discharging the compound to water, or placing a compound in a landfill. The TRI includes a summary of release reports for each calendar year and totals the pounds-per-year of each compound released. This data source was used as a primary data source for CCL 5.

Reference: USEPA. n.d. Toxics Release Inventory (TRI) Program. https://www.epa.gov/toxics-release-inventory-tri-program. Accessed April 2018 and January 2020.

Pre-processing steps for screening:

- **Data download:** EPA downloaded the 2016 data from the TRI Explorer Release Reports on April 24, 2018, from https://iaspub.epa.gov/triexplorer/tri_release.chemical. The data option for total on and off-site disposal and other releases was selected. As of March 2021, this website has been updated, and TRI Explorer Release Reports can now be accessed at: https://enviro.epa.gov/triexplorer/tri_release.chemical.
- **Data manipulation:** Data manipulation was minimal and restricted to adding DTXSIDs.
- **Extracted data elements**: EPA wrote R code to extract the total pounds released in 2016 for each compound.

Pre-processing steps for classification:

- **Data download:** EPA downloaded the TRI Release Geography Reports associated with the 2016 release data used in the screening step on January 3, 2020, from https://iaspub.epa.gov/triexplorer/tri_release.geography. EPA selected the data option for total on- and off-site disposal and other releases. As of March 2021, the original website has been updated and TRI Explorer and geography reports can now be accessed at: https://enviro.epa.gov/triexplorer/tri_release.geography.
- **Data manipulation:** EPA used the downloaded state release reports to manually count the number of states 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 2016.

Unregulated Contaminant Monitoring (UCM) Program – EPA

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. Round 1 UCM data are from approximately 1988 to 1992 and were extracted from the Unregulated Contaminant Monitoring Information System (URCIS). The UCM Round 2 data are from 1993 to 1997 and were extracted from SDWIS.

UCM 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 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, 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 Round 1 database contains contaminant occurrence data from 38 states, Washington, D.C., and the U.S. Virgin Islands. The UCM 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 Round 1 and Round 2 cross-sections, see USEPA (2002). This data source was used as a primary data source for CCL 5.

EPA considered finished drinking water maximum concentrations from all primary data sources for calculating the screening hazard quotient in the screening step of CCL 5 (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 5.

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.

Pre-processing steps 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

Pre-processing steps 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 detects for each compound, total number of samples, number of samples with detections, and percentage of samples with detections for each contaminant. This data source was treated as a nationally representative finished water data source for classification for CCL 5 and included on the Contaminant Information Sheets.

Unregulated Contaminant Monitoring Rule (UCMR) Cycles 1-3 – EPA

Data description: These data represent all the Unregulated Contaminant Monitoring Rule (UCMR) sampling results from completed UCMR cycles. UCMR is 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 including monitoring for 25 contaminants between 2008 and 2010. UCMR3 included monitoring for 28 chemical contaminants and 11 microbes between 2013 and 2015. This data source was used as a primary data source for CCL 5.

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.

Pre-processing steps for screening:

- **Data download:** The results of UCMR 1-3 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. This data source was treated as a nationally representative finished water survey for the screening step.

Pre-processing steps 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 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, percentage of sites with detects for each contaminant, total number of samples, number of samples with detections, and percentage of samples with detections for each contaminant. This data source was treated as a nationally representative finished water survey.

Unregulated Contaminant Monitoring Rule (UCMR), Cycle 4 – EPA

Data description: Similar in design to UCMR 1, 2 and 3, UCMR 4 required surface water systems to monitor quarterly and groundwater systems to monitor semiannually to capture seasonal variability. See USEPA (2016) for more information on the UCMR 4 study design and data analysis, including a complete list of analytes. For UCMR 4, all large and very large PWSs (serving between 10,001 and 100,000 people and serving more than 100,000 people, respectively), plus a statistically representative national sample of 800 small PWSs (serving 10,000 people or fewer), were required to conduct assessment monitoring during a 12-month period between January 2018 and December 2020. These data are treated separately from the UCMR1-3 data because the monitoring dataset for UCMR 4 was not complete at the time of CCL 5 development. The UCMR 4 dataset used in CCL 5 are not final and are subject to change as updates become available. This data source was used as a primary data source for CCL 5.

References:

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. 2019. The Fourth Unregulated Contaminant Monitoring Rule (UCMR 4): Data Summary, October 2019. Office of Water. EPA 815-S-19-005.

USEPA. 2020. The Fourth Unregulated Contaminant Monitoring Rule (UCMR 4): Data Summary, January 2020. Office of Water. EPA 815-S-20-001.

Pre-processing steps for screening:

- **Data download:** The fifth National Contaminant Occurrence Database (NCOD) release of UCMR 4 results received as of October 2019.
- **Data manipulation:** If there were zero detections for a contaminant, half 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 concentration and percent detection of drinking water systems were extracted. This data source was considered a nationally representative finished water occurrence survey for the screening step.

Pre-processing steps for classification:

- **Data download:** The sixth NCOD release of UCMR 4 analytical results received as of December 2019.
- **Data manipulation:** Concentration summary statistics based on analytical detections and detection rate information were calculated. DTXSIDs were added. Data manipulation steps were conducted using R.
- **Extracted data elements**: EPA wrote R code to extract 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 detects for each contaminant. This data source was considered a nationally representative finished water occurrence survey for the classification step.

National Water Information System (NWIS) and National Ambient Water Quality Assessment (NAWQA) Programs – Water Quality Portal (WQP), USGS

Data description: The Water Quality Portal is a collaborative tool sponsored by EPA, USGS, and the National Water Quality Monitoring Council (NWQMC) that allows access to water quality data collected by state, tribal, local and federal agencies. The Water Quality Portal is used to access the USGS National Water Information Services (NWIS) database. The NWIS relational database houses every piece of data that USGS collects, including information like gauge heights and compound concentration data and results from the National Water-Quality Assessment (NAWQA) program. The goals of the NAWQA program include assessing the condition of the nation's streams, rivers, and groundwater and identifying how those conditions are changing over time. The NAWQA program is designed to be statistically representative of water conditions in the nation. NAWQA data are considered nationally representative, whereas NWIS results are not expected to be statistically representative of the U.S. These data sources were used as primary data sources for CCL 5.

Reference:

United States Geological Survey (USGS). n.d. National Water-Quality Assessment (NAWQA) Program. Accessed via the Water Quality Portal (WQP). URL: https://www.waterqualitydata.us/portal/. Accessed January 2018.

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 January 2018.

Pre-processing steps for screening:

- **Data download:** In the Water Qualtiy Portal, EPA downloaded all data from the NAWQA monitoring program from 1991 through 2017. The results in the NWIS database that were not associated with the NAWQA program were downloaded for samples collected from 2008 through 2017. Raw data were download using REST API and saved into a SQL Server database. Data excluded from the analysis include non-water data, data from media other than ground water or surface water (e.g., leachate, etc.), and data with non-standard units of measure.
- **Data manipulation:** Raw data were stored in a SQL Server database and prepared for analysis (e.g., concentrations are converted to common units) and summarized using R.

Combined surface water and ground water data were summarized and output in a CSV file. DTXSIDs were added.

- **Extracted data elements**: EPA wrote R code to extract maximum concentration and percent detection in study sites were extracted. Combined surface water and ground water data were categorized as ambient water for the screening step. Ambient water data from the NAWQA program (non-NWIS) were considered nationally representative and water data from the NWIS database (non-NAWQA) were considered non-nationally representative.

Pre-processing steps for classification:

- **Data download:** The same data used in the screening step were used to extract data for the classification step.
- **Data manipulation:** Raw data stored in the SQL Server database prepared for the screening step were used to prepare data used in the classification step. Summary statistics and detection information were calculated for combined surface water and ground water samples, and for surface water samples and ground water samples separately. Data were output to a CSV file and DTXSIDs were added.
- **Extracted data elements:** EPA wrote R code to extract minimum, median, 90^{th} percentile and maximum concentration of detections as well as total number of sites, number of sites with detections, and percentage of sites with detects for each contaminant in surface water, ground water, and combined surface water and ground water. Combined surface water and groundwater data from the NAWQA program (non-NWIS) were considered nationally representative and data from the NWIS database (non-NAWQA) were considered non-nationally representative.

Section N.3 Simple Data Format for the CCL 5

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 primary and supplemental data sources for use in CCL 5.

An example of the simple data format is illustrated in Table N-1. The simple data format consists of six columns and each data entry in its own row:

- The first column, Name, provides the compound name as originally reported in the data source. Some sources only report CAS Registry Numbers (CASRN) or PubChem Compound IDs (CID) as identifiers—in this case the CASRN or CID is listed in the Name column.
- The second column, Key, lists the DTXSIDs for the compounds.
- The third column, Value, lists values associated with the data entry.
- The fourth column, Unit, is the units for the value.
- The fifth column, Source, contains a shorthand indicator or acronym to describe the source of the data.
- The sixth and final column, Data Element, includes a shorthand code that describes the type of data element that the data entry is describing, such as an LD50; 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 an LD_{50}.

Table N-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 5.

Table N-1. Example of the Simple Data Format

Name	Key	Value	Unit	Source	Data Element
chemical identifier reported by the data source	DTXSID number or unique identifier for compounds which a DTXSID could not be identified (NO_DTXSIDXXXX)	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 (i.e., RfD, release)

Table N-2. Example of a Data Entry for an RfD from EPA-PPRTV for Vanadium in the Simple File Format

Name	Key	Value	Units	Source	Data Element
vanadium	DTXSID2040282	7E-5	mg/kg/day	pprtv	rfd

Section N.4 EPA's CompTox Chemicals Dashboard Data Elements Used in CCL 5 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, 2016). EPA included the following TEST predictions from the CompTox Chemicals Dashboard in the universe: oral rat 50 percent lethal dose (LD50), 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, 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. 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).
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.
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).

Data Element	Description
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.

Section N.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	Acute reference concentration (inhalation exposures)
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)
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 designated by EPA, NTP, or IARC. EPA converted cancer classifications to a numerical form which were assigned screening points. 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 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 unit risk	Unit risk for a chronic inhalation exposure scenario resulting in carcinogenicity
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
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

Data Element	Description
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} from the TEST Model (downloaded from EPA's CompTox Chemicals Dashboard)
Reference concentration	Chronic reference concentration – inhalation exposure
Subchronic reference concentration	Reference concentration based on an inhalation study with a subchronic exposure duration
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 N.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. 2016. User's Guide for T.E.S.T. (version 4.2) (Toxicity Estimation Software Tool). Office of Research and Development.

Appendix O – CCL 4 Chemicals Not Listed on CCL 5

Chemical Name	CASRN	DTXSID
A. CCL 4 Chemicals Below the Top 250 Screening Points Threshold for Inclusion on PCCL 5		
1,1,1,2-Tetrachloroethane	630-20-6	DTXSID2021317
17alpha-estradiol	57-91-0	DTXSID8022377
2-Methoxyethanol	109-86-4	DTXSID5024182
2-Propen-1-ol	107-18-6	DTXSID8020044
4,4'-Methylenedianiline	101-77-9	DTXSID6022422
Acetaldehyde	75-07-0	DTXSID5039224
Acetamide	60-35-5	DTXSID7020005
Aniline	62-53-3	DTXSID8020090
Benzyl chloride	100-44-7	DTXSID0020153
Butylated hydroxyanisole	25013-16-5	DTXSID7020215
Captan	133-06-2	DTXSID9020243
Clethodim	110429-62-4	DTXSID3034458
Cumene hydroperoxide	80-15-9	DTXSID3024869
Dimethipin	55290-64-7	DTXSID0024052
Equilenin	517-09-9	DTXSID2052156
Equilin	474-86-2	DTXSID7047433
Erythromycin	114-07-8	DTXSID4022991
Estriol	50-27-1	DTXSID9022366
Estrone	53-16-7	DTXSID4022367
Ethylene glycol	107-21-1	DTXSID8020597
Ethylene oxide	75-21-8	DTXSID0020600
Ethylene thiourea	96-45-7	DTXSID5020601
Formaldehyde	50-00-0	DTXSID7020637
Germanium	7440-56-4	DTXSID8052483
Hexane	110-54-3	DTXSID0021917
Hydrazine	302-01-2	DTXSID3020702
Mestranol	72-33-3	DTXSID0020814
Methanol	67-56-1	DTXSID2021731
Nitroglycerin	55-63-0	DTXSID1021407
N-Methyl-2-pyrrolidone	872-50-4	DTXSID6020856
Norethindrone (19-Norethisterone)	68-22-4	DTXSID9023380
n-Propylbenzene	103-65-1	DTXSID3042219
Oxirane, methyl	75-56-9	DTXSID5021207
Oxydemeton-methyl	301-12-2	DTXSID8025541
sec-Butylbenzene	135-98-8	DTXSID2022333
Tebufenozide	112410-23-8	DTXSID4034948
Tellurium	13494-80-9	DTXSID9032119
Thiodicarb	59669-26-0	DTXSID0032578
Thiophanate-methyl	23564-05-8	DTXSID1024338
Toluene diisocyanate	26471-62-5	DTXSID0024341

Chemical Name	CASRN	DTXSID
Triethylamine	121-44-8	DTXSID3024366
Triphenyltin hydroxide (TPTH)	76-87-9	DTXSID1021409
Urethane	51-79-6	DTXSID9021427
Vinclozolin	50471-44-8	DTXSID4022361
Ziram	137-30-4	DTXSID0021464
B. CCL 4 Chemicals Screened to PCCL 5, "Not List" Decisions by the Chemical Evaluators		
1,3-Butadiene	106-99-0	DTXSID3020203
1-Butanol	71-36-3	DTXSID1021740
Acetochlor ethanesulfonic acid (ESA)	187022-11-3	DTXSID6037483
Acetochlor oxanilic acid (OA)	184992-44-4	DTXSID1037484
Alachlor ethanesulfonic acid (ESA)	142363-53-9	DTXSID6037485
Alachlor oxanilic acid (OA)	171262-17-2	DTXSID1037486
Chloromethane (Methyl chloride)	74-87-3	DTXSID0021541
Estradiol (17-beta estradiol)	50-28-2	DTXSID0020573
HCFC-22	75-45-6	DTXSID6020301
Halon 1011 (bromochloromethane)	74-97-5	DTXSID4021503
Metolachlor ethanesulfonic acid (ESA)	171118-09-5	DTXSID1037567
Metolachlor oxanilic acid (OA)	152019-73-3	DTXSID6037568
C. CCL 4 Chemicals Screened to PCCL 5, Removed for Regulatory Determination 4 Status		
1,1-Dichloroethane	75-34-3	DTXSID1020437
Acetochlor	34256-82-1	DTXSID8023848
Methyl bromide (bromomethane)	74-83-9	DTXSID8020832
Metolachlor	51218-45-2	DTXSID4022448
Nitrobenzene	98-95-3	DTXSID3020964
Perfluorooctanesulfonic acid (PFOS)	1763-23-1	DTXSID3031864
Perfluorooctanoic acid (PFOA)	335-67-1	DTXSID8031865
RDX (Hexahydro-1,3,5-trinitro-1,3,5-triazine)	121-82-4	DTXSID9024142
D. CCL 4 Chemicals Screened to PCCL 5, Removed as Cancelled Not Persistent Pesticides		
3-Hydroxycarbofuran	16655-82-6	DTXSID2037506
Methamidophos	10265-92-6	DTXSID6024177

Appendix P - Group of 23 DBPs included on CCL 5

Chemical Name	DTXSID	Nominated	Top 250
Haloacetic Acids			
Bromochloroacetic acid (BCAA)	DTXSID4024642	X	
Bromodichloroacetic acid (BDCAA)	DTXSID4024644	X	
Dibromochloroacetic acid (DBCAA)	DTXSID3031151	X	
Tribromoacetic acid (TBAA)	DTXSID6021668	X	
Haloacetonitriles			
Dichloroacetonitrile (DCAN)	DTXSID3021562		X
Dibromoacetonitrile (DBAN)	DTXSID3024940		X
Halonitromethanes			
Bromodichloronitromethane (BDCNM)	DTXSID4021509	X	
Chloropicrin (trichloronitromethane, TCNM)	DTXSID0020315	X	X
Dibromochloronitromethane (DBCNM)	DTXSID00152114	X	
Iodinated Trihalomethanes			
Bromochloroiodomethane (BCIM)	DTXSID4021503	X	
Bromodiiodomethane (BDIM)	DTXSID70204235	X	
Chlorodiiodomethane (CDIM)	DTXSID20213251	X	
Dibromoiodomethane (DBIM)	DTXSID60208040	X	
Dichloroiodomethane (DCIM)	DTXSID7021570	X	
Iodoform (triiodomethane, TIM)	DTXSID4020743	X	
Nitrosamines			
Nitrosodibutylamine (NDBA)	DTXSID2021026		X
N-Nitrosodiethylamine (NDEA)	DTXSID2021028	X	X
N-Nitrosodimethylamine (NDMA)	DTXSID7021029	X	X
N-Nitrosodi-n-propylamine (NDPA)	DTXSID6021032	X	
N-Nitrosodiphenylamine (NDPhA)	DTXSID6021030	X	
Nitrosopyrrolidine (NPYR)	DTXSID8021062	X	X
Others			
Chlorate	DTXSID3073137	X	X
Formaldehyde	DTXSID7020637		X