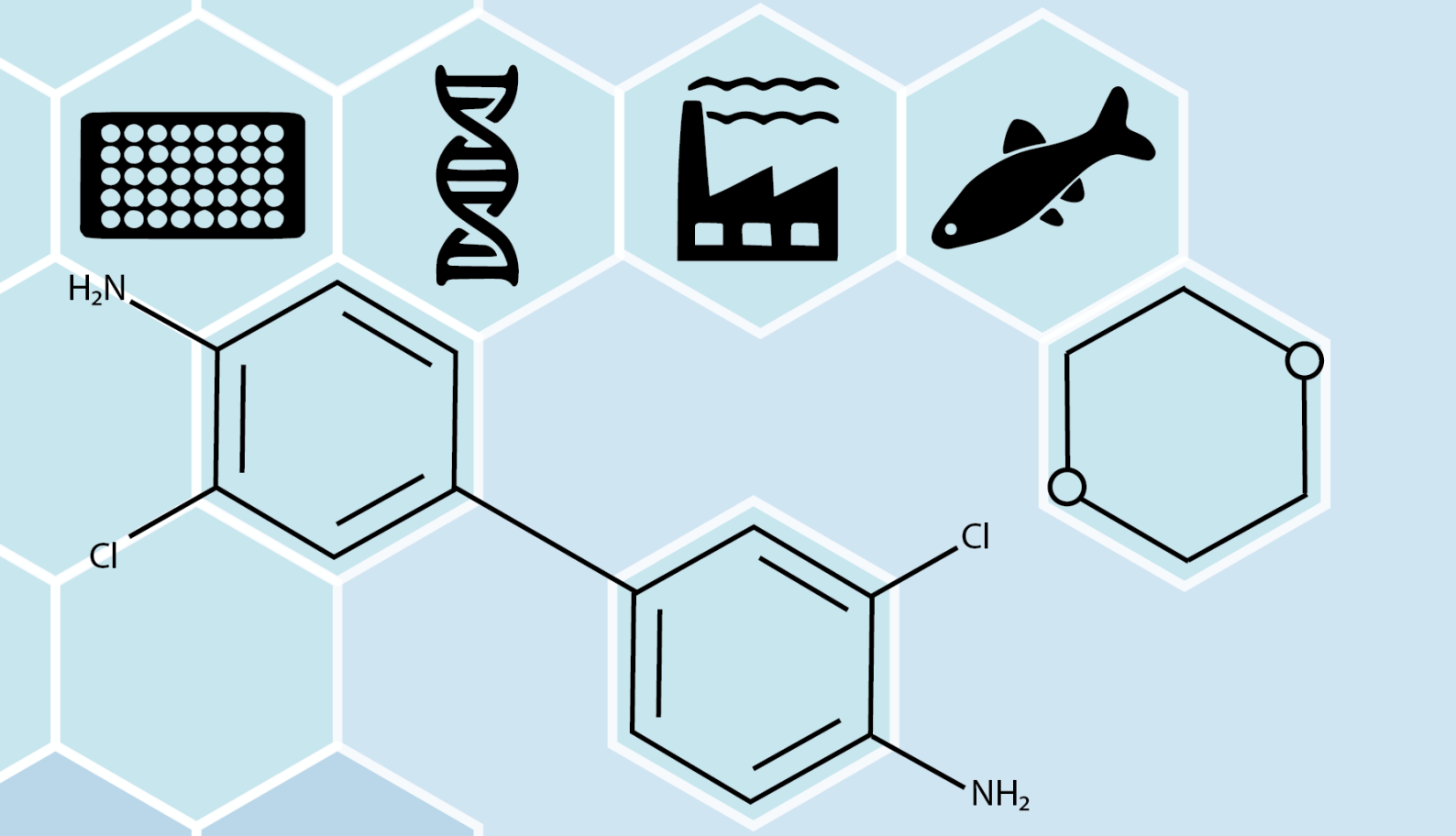
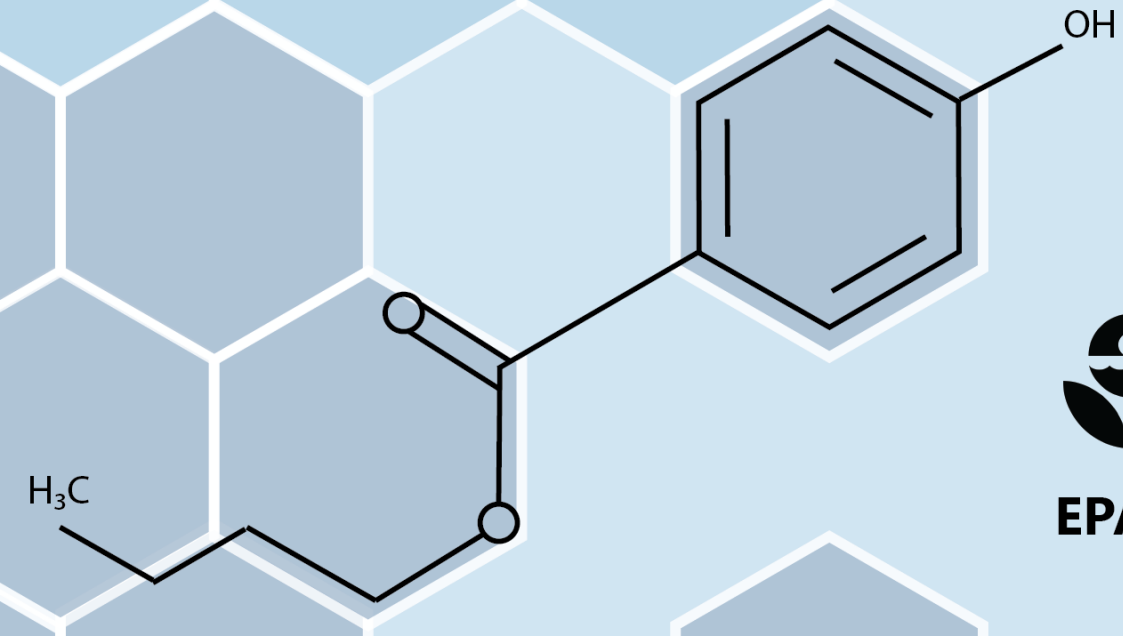


A Proof-of-Concept Study Integrating Publicly Available Information to Screen Candidates for Chemical Prioritization under TSCA



EPA

EPA/600/R-21-106

EPA Document# EPA/600/R-21-106
June 2021

United States Environmental Protection Agency
Office of Chemical Safety and Pollution Prevention
Office of Research and Development

A Proof-of-Concept Case Study Integrating Publicly Available Information to Screen Candidates for Chemical Prioritization under TSCA

June 2021

Table of Contents

Acknowledgements

This report was developed by the United States Environmental Protection Agency (U.S. EPA), Office of Chemical Safety and Pollution Prevention (OCSPP) and the Office of Research and Development (ORD).

Disclaimer

This document has been reviewed in accordance with the U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Suggested Citation

USEPA 2021. A Proof-of-Concept Case Study Integrating Publicly Available Information to Screen Candidates for Chemical Pre-Prioritization under TSCA. June 2021. U.S. Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention and Office of Research and Development, Washington, DC. EPA/600/R-21-106.

Reviewers

This document has been provided for review to EPA scientists, and peer reviewed by independent scientists external to EPA (listed below). A summary of comments and EPA's response to comments received from the independent external peer reviewers is provided through EPA's Science Inventory database of publicly released research products which is available by clicking here.

Dr. Tara S. Barton-Maclaren, Health Canada

Dr. Weihsueh A. Chiu, Texas A&M University

Dr. Helen M. Goeden, Minnesota Department of Health

Dr. Kerry W. Nugent, Australian Industrial Chemicals Introduction Scheme (formerly NICNAS)

Dr. Edward J. Perkins, Department of Defense

List of Abbreviations and Acronyms

AD	Applicability Domain
ATSDR	Agency for Toxic Substances and Disease Registry (CDC)
BAF	Bioaccumulation Factor
BCF	Bioconcentration Factor
BER	Bioactivity-to-Exposure Ratio
CA	Chromosomal Aberration
CDC	Centers for Disease Control and Prevention
CMP	(Canadian) Chemicals Management Plan
CPCat	Chemical and Product Categories (database)
CPDat	Chemical and Product Database
CTS	Chemical Transformation Simulator (US EPA ORD)
DSSTox	Distributed Structure-Searchable Toxicity Database
DTXSIDs	DSSTox Substance Identifiers
ECHA	European Chemicals Agency
EcoSAR	Ecological Structure Activity Relationships
EFSA	European Food Safety Authority
EPA	US Environmental Protection Agency
ExpoCast	Exposure Forecasting
EUSES	European Union System for Evaluation of Substances
FDA	Food and Drug Administration
GHS	Globally Harmonized System of Classification and Labeling of Chemicals
HER	Human Hazard-to-Exposure Ratio
HESS	Hazard Evaluation Support System
HPVIS	High Production Volume Information System
HSDB	Hazardous Substances Data Bank
IAM	Information Availability Metric
IARC	International Agency for Research on Cancer
IG	Information Gathering
IRIS	Integrated Risk Information System
LLNA	Local Lymph Node Assay
MNT	Micronucleus Test
NAMs	New Approach Methods
NIOSH	National Institute for Occupational Safety and Health (CDC)
NTP	National Toxicology Program

OCSPP	Office of Chemical Safety and Pollution Prevention (OCSPP)
OECD	Organization for Economic Cooperation and Development
OPP	Office of Pesticide Programs (US EPA)
OPPT	Office of Pollution Prevention and Toxics (US EPA)
ORD	Office of Research and Development (US EPA)
PECs	Predicted Exposure Concentrations
PICS	Public Information Curation and Synthesis
PNECs	Predicted no Effect Concentrations
POD	Points-of-Departure
Pov	Overall Chemical Persistence
PPRTV	Provisional Peer-Reviewed Toxicity Values
QC	Quality Control
QA	Quality Assurance
QSAR	Quantitative Structure-Activity Relationship
ROC	Report on Carcinogenesis (NTP)
SCIL	Safer Chemical Ingredients List
SDM	Scientific Domain Metric
SEEM	Systematic Empirical Evaluation of Models
TER	TTC-to-Exposure Ratio
TEST	Toxicity Estimation Software Tool (US EPA)
TSCA	Toxic Substances Control Act
TSCA10	First Ten TSCA Work Plan Chemical Substances Selected for Evaluation
TSCA90	Chemical Substances from the 2014 Update to the TSCA Work Plan
POC238	238 Chemical Substances Selected for the POC Case Study
TTC	Threshold of Toxicological Concern
UVCBs	Unknown or Variable Composition, Complex Reaction Products and Biological Materials
WHO	World Health Organization
WOE	Weight-of-Evidence

1. Executive Summary

Regulatory agencies worldwide are looking to efficiently integrate information on chemical substances[1] in order to inform priorities for decisions and data requests. This document updates the US Environmental Protection Agency's (EPA) long-term strategy described in the Working Approach for Identifying Potential Candidate Chemicals for Prioritization[2] and presents the Public Information Curation and Synthesis (PICS) approach that integrates publicly-available hazard, exposure, persistence, and bioaccumulation information for chemical substances. The purpose of the PICS approach is to synthesize information from traditional and new approach methods (NAMs)[3] to understand the overall degree of potential concern as well as the relative coverage of potentially relevant human health and ecological toxicity and exposure information that could inform level of effort and resources that may be needed to evaluate that specific chemical substance. The PICS approach is based on two dimensions. The first dimension, Scientific Domain Metric (SDM), encompasses the synthesis of the traditional and NAM data to understand the overall degree of potential concern related to human health and the environment. The second dimension, Information Availability Metric (IAM), reflects the relative coverage of potentially relevant human health and ecological toxicity and exposure information that could inform level of effort and resources that may be needed to evaluate that specific chemical substance. The PICS approach is not designed to replace the prioritization process described in TSCA but aims to increase efficiency and focus expert review on chemical substances that may have a greater potential for designation as a high- or low priority candidate. A proof-of-concept case study was performed by applying the PICS approach to a subset of the TSCA active inventory. The results demonstrate that the approach discriminated between high- and low priority candidate chemical substances and identified potential information gaps. The PICS approach may be applied to large numbers of chemical substances and is an important tool for efficiently integrating and synthesizing large amounts of publicly available information, and aspects of the approach could be adapted and applied to other prioritization decision contexts.

2. Introduction

Regulatory agencies worldwide need to make decisions on chemical substances[4] based on a set of defined criteria on specific hazards of concern, exposure to specific populations, or

[1] Unless otherwise indicated, any references to "chemical" or "chemical substance" throughout this document means a "chemical substance" as defined in TSCA Section 3(2).
[2] https://www.epa.gov/sites/production/files/2018-09/documents/preprioritization_white_paper_9272018.pdf
[3] https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/alternative-test-methods-and-strategies-reduce
The term NAMs was recently introduced to cover any *in vitro, in silico,* or *in chemico* technique used to provide data or information for regulatory decision making.

persistence and bioaccumulation in the environment. There is a need for a consistent, timely and efficient approach to organize large numbers of chemical substances for further evaluation. This document describes an approach to integrate publicly available information on the more than 33,000 chemical substances on the non-confidential TSCA active inventory[5] to efficiently select chemical substances for expert review prior to prioritization. The information in this document expands on the long-term strategy previously described by the EPA[6], continuing with the development of the PICS approach for synthesizing information from traditional and NAMs in key scientific domains. These domains include human health hazard to exposure ratio (incorporating multiple specific toxicities), ecological hazard, carcinogenicity, genotoxicity, human exposure (general and susceptible populations), persistence/bioaccumulation, skin sensitization, skin irritation, and eye irritation. Of the seven domains used in the PICS approach, five were included in the previous Working Approach for Potential Candidates. The additional two domains (carcinogenicity; skin sensitization and skin/eye irritation) were included in the PICS approach based on their use in the 2014 TSCA Work Plan[7]. The detailed process was tested in a proof-of-concept case study. The PICS approach may help streamline the evaluation of chemical substances by transparently and reproducibly synthesizing available information and identify potential data gaps, and aspects of the approach could be adapted and applied to other prioritization decision contexts.

This document presents a proof-of-concept approach for EPA and the broader scientific community, and neither constitutes rulemaking by the EPA, nor can it be relied on to create a substantive or procedural right enforceable by any party in litigation with the United States. Non-mandatory language such as "should" provides recommendations and does not impose any legally binding requirements. Similarly, statements about what EPA expects or intends to do reflect general principles to guide EPA's activities and are not judgments or determinations as to what EPA will do in any particular case.

3. Background

3.1 Development and Implementation of New Approach Methods Under TSCA

Section 4 (h) of the Toxic Substances Control Act (TSCA), as amended by the Frank R. Lautenberg Chemical Safety for the 21st Century Act (P.L. 114-182), requires EPA to develop a Strategic Plan to promote the development and implementation of alternative test methods and strategies to reduce, refine or replace vertebrate animal testing and provide information of

[5] This list can be found at https://comptox.epa.gov/dashboard/chemical_lists/TSCA_ACTIVE_NCTI_0320
[6] https://www.epa.gov/sites/production/files/2018-09/documents/preprioritization_white_paper_9272018.pdf
[7] TSCA Work Plan Methods Document 2012 (https://www.epa.gov/sites/production/files/2014-03/documents/work_plan_methods_document_web_final.pdf).

equivalent or better scientific quality and relevance for assessing risks of injury to health or the environment of chemical substances or mixtures. EPA's *Strategic Plan to Promote the Development and Implementation of Alternative Test Methods Within the TSCA Program*[8] was released on June 22, 2018 and outlines the EPA's plan to reduce the use of vertebrates for chemical substances regulated under TSCA. As part of this Strategic Plan, EPA describes incremental steps for the development and integration of NAMs that are appropriate and fit-for-purpose for making TSCA-related decisions (e.g., identifying candidates for prioritization, prioritization, risk evaluations for new and existing chemical substances and other risk-based decisions). This multi-year strategic plan includes criteria for determination of what would be considered NAMs by the EPA and how they may be applied for evaluation of human health hazard, ecological hazard and exposure. In addition to this, EPA has developed a NAMs workplan[9] for reducing use of animals in chemical testing in order to prioritize Agency efforts and resources toward activities that aim to reduce the use of animal testing while continuing to protect human health and the environment. This workplan expands EPA's discussion of the development and use of NAMs for support of regulatory decision-making beyond TSCA and focuses on mechanisms for building confidence in the implementation of NAMs.

3.2 Evaluating Existing Chemicals Under TSCA

Under Section 6(b) of TSCA, EPA is required to both prioritize and evaluate the risks of existing chemical substances. The law contains specific timetables, minimum chemical substance numbers, and general requirements for both prioritization and risk evaluation. Prioritization[10] is a 9- to 12-month public process in which chemical substances are designated as either high- or low priority for risk evaluation. A high priority chemical substance is one "that the Administrator concludes, without consideration of costs or other non-risk factors, may present an unreasonable risk of injury to health or the environment because of a potential hazard and a potential route of exposure under the conditions of use, including an unreasonable risk to potentially exposed or susceptible subpopulations identified as relevant by the Administrator." A low priority chemical substance is one that "the Administrator concludes, based on information sufficient to establish, without consideration of costs or other non-risk factors, that such chemical substance does not

[8] Alternative Test Methods and Strategies to Reduce Vertebrate Animal Testing in TSCA (https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/alternative-test-methods-and-strategies-reduce).
[9] More information can be found at https://www.epa.gov/chemical-research/epa-new-approach-methods-work-plan-reducing-use-animals-chemical-testing
[10] Final Rule, "Procedures for Prioritization of Chemicals for Risk Evaluation Under the Toxic Substances Control Act," available at https://www.regulations.gov/document?D=EPA-HQ-OPPT-2016-0636-0074.

meet the [High priority] standard." A designation of a chemical substance as low priority indicates that a risk evaluation is not warranted at that time.

TSCA requires that high priority chemical substances undergo risk evaluation to determine whether a chemical substance presents an unreasonable risk of injury to health or the environment, without consideration of costs or other non-risk factors, including an unreasonable risk to a potentially exposed or susceptible subpopulation[11] identified as relevant to the risk evaluation by the Administrator, under the conditions of use[12]. The risk evaluation must take no longer than three years with a possible six-month extension. If unreasonable risk is identified, EPA has two years with a possible extension of two additional years to finalize regulations so the chemical substance no longer presents such a risk.

On March 20, 2019, the EPA initiated the prioritization process for the first set of 20 high- and 20 low priority candidate chemical substances. The initiation of the prioritization process is followed by a 90-day public comment period for submitting relevant information. Upon completion of the public comment period, the EPA performs a screening review of the candidate chemical substances based on hazard and exposure potential, persistence and bioaccumulation, potentially exposed or susceptible subpopulations, storage near significant sources of drinking water, the conditions of use, and the volume of the chemical substance manufactured or processed. Based on the outcome of the screening review, the EPA will propose to designate a chemical substance as either a high priority or low priority chemical substance and release the information, analysis, and basis used to make the designation. The proposed designation will be followed by a second 90-day comment period prior to finalizing the designation.

3.3 Public Engagement

On December 11, 2017, EPA held a public meeting to gain input regarding identification of potential candidate chemical substances for prioritization. In preparation for this meeting, EPA published a discussion document including possible approaches to inform the dialogue at the

[11] "Potentially exposed or susceptible subpopulation," as defined in TSCA Section 3(12), means a group of individuals within the general population identified by the Administrator who, due to either greater susceptibility or greater exposure, may be at greater risk than the general population of adverse health effects from exposure to a chemical substance or mixture, such as infants, children, pregnant women, workers or the elderly (15 U.S.C. 2602).
[12] "Conditions of use" under TSCA means "the circumstances, as determined by the Administrator, under which a chemical substance is intended, known, or reasonably foreseen to be manufactured, processed, distributed in commerce, used or disposed of" (15 U.S.C. 2602). For purposes of prioritization, the Administrator may determine that certain uses fall outside the definition of "conditions of use".

meeting[13]. A Response to Comment document has been developed to address the comments[14]. EPA received 43 relevant comments in the docket associated with the public meeting[15]. There was consensus in the comments that EPA should proceed in a transparent manner with opportunities for public participation. However, there was no consensus around one or more of the proposed approaches the Agency presented. The most consistent support focused on use of the 2014 Update to the TSCA Work Plan[16] as the starting point for identifying high priority candidates. For low priority chemical substances, there was some support for using the chemical substances on the Safer Chemical Ingredients List (SCIL)[17], which was developed by EPA's Safer Choice Program, as a starting point. There was general support for the integration of NAMs for filling information gaps during the process to identify potential candidate chemical substances for prioritization; however, there was some concern regarding the readiness of these approaches for decision-making on prioritization for risk evaluation. There were opposing views regarding filling information gaps and EPA's authority to request submission of information, the use of voluntary submissions, when to request information, the quality of information, and how to use information from other jurisdictions (e.g., the European Union's Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH)).

On September 27, 2018, EPA released *A Working Approach for Identifying Potential Candidate Chemicals for Prioritization*[18] (to be called Working Approach for Potential Candidates throughout this document) that described both short- and long-term strategies for selecting candidate chemical substances for prioritization under TSCA. The long-term strategy in the Working Approach for Potential Candidates was adapted from the TSCA 2012 Work Plan process[19], but incorporated scientific advances in relevant fields, integration of NAMs, and modern information management technologies to integrate the large volume of information in an efficient, automatable and reproducible manner. The strategies presented in the Working Approach for Potential Candidates reflected public input received at the December 2017 meeting[20] and through

[13] Meeting materials for the December 11, 2017 Possible Approaches for Identifying Potential Candidates for Prioritization Public meeting can be found here: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/possible-approaches-identifying-potential-candidates .
[14] https://www.epa.gov/sites/production/files/2018-09/documents/publiccommentssummary_dec11_preprioritization_927.pdf
[15] The public comments received following the December 11, 2017 public meeting are available at www.regulations.gov in docket EPA-HQ-OPPT-2017-0586.
[16] https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/tsca-work-plan-chemical-assessments-2014-update
[17] This list can be found at https://comptox.epa.gov/dashboard/chemical_lists/SCILFULL
[18] https://www.epa.gov/sites/production/files/2018-09/documents/preprioritization_white_paper_9272018.pdf
[19] https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/tsca-work-plan-methods-document
[20] https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/meetings-and-webinars-amended-toxic-substances-control#12/11

the public docket for that meeting[21]. EPA accepted a second round of public comments on the proposed longer-term strategy. EPA received 26 unique comments in the docket[22]. Commenters also noted that data gaps do not necessarily equate to data needs, and that EPA should not prioritize solely based on information availability. Commenters included recommendations on specific topics, including susceptible and sensitive populations and increasing the use of exposure data to make this approach more risk-based. The comments also highlighted the need to clarify the purpose of the long-term strategy as a means for increasing efficiency of the expert review required for the selection of candidates for prioritization and not as a replacement for the formal prioritization and risk evaluation steps in the process. Finally, there appeared to be misconceptions about the difference between the bins outlined in the strategy and the chemical substance categories the Agency regularly uses in the TSCA New Chemicals Program to group chemical substances expected to show the same hazard characteristics. The present document and the PICS approach are intended to address these comments.

4. Public Information Curation and Synthesis (PICS) Approach

4.1 Overview of the PICS Approach

The PICS approach updates and expands on the long-term strategy described in the Working Approach for Potential Candidates document and integrates information from a variety of sources to better understand publicly available information for these chemical substances. The PICS approach synthesizes information from traditional methods and NAMs in key scientific domains including human health hazard to exposure ratio (incorporating multiple specific toxicities), ecological hazard, carcinogenicity, genotoxicity, exposure to susceptible populations, persistence/bioaccumulation, skin sensitization, and skin/eye irritation. For each scientific domain, a workflow was developed that specifies what information is utilized and the logic of how it is integrated. The methodology underlying the individual workflows are designed to incorporate scientific advances in each discipline and may differ from domain to domain. The domain-specific workflows are described in detail in the subsequent sections. Consistent with the *Strategic Plan to Reduce the Use of Vertebrate Animals in Chemical Testing*[23], the PICS approach integrates NAMs to fill gaps when traditional testing data are not available. In general, each workflow is based on previously accepted methods for prioritizing chemical substances under TSCA[24], with a focus on the use of data from study types for which there is traditionally the most confidence in the

[21] Further information can be found at docket EPA-HQ-OPPT-2017-0587.
[22] Further information can be found at docket EPA-HQ-OPPT-2018-0659.
[23] https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/strategic-plan-reduce-use-vertebrate-animals-chemical
[24] TSCA Work Plan Methods Document 2012 (https://www.epa.gov/sites/production/files/2014-03/documents/work_plan_methods_document_web_final.pdf).

regulatory toxicology community (e.g., *in vivo*), followed by those with decreasing confidence depending on the context for use (e.g., *in vitro, in silico*). Unless otherwise described, the domain-specific workflows generally utilize conservative assumptions to reduce the potential for false negatives at the initial screening stage. This document also presents potential options for future work to improve the approach, as well as caveats and limitations.

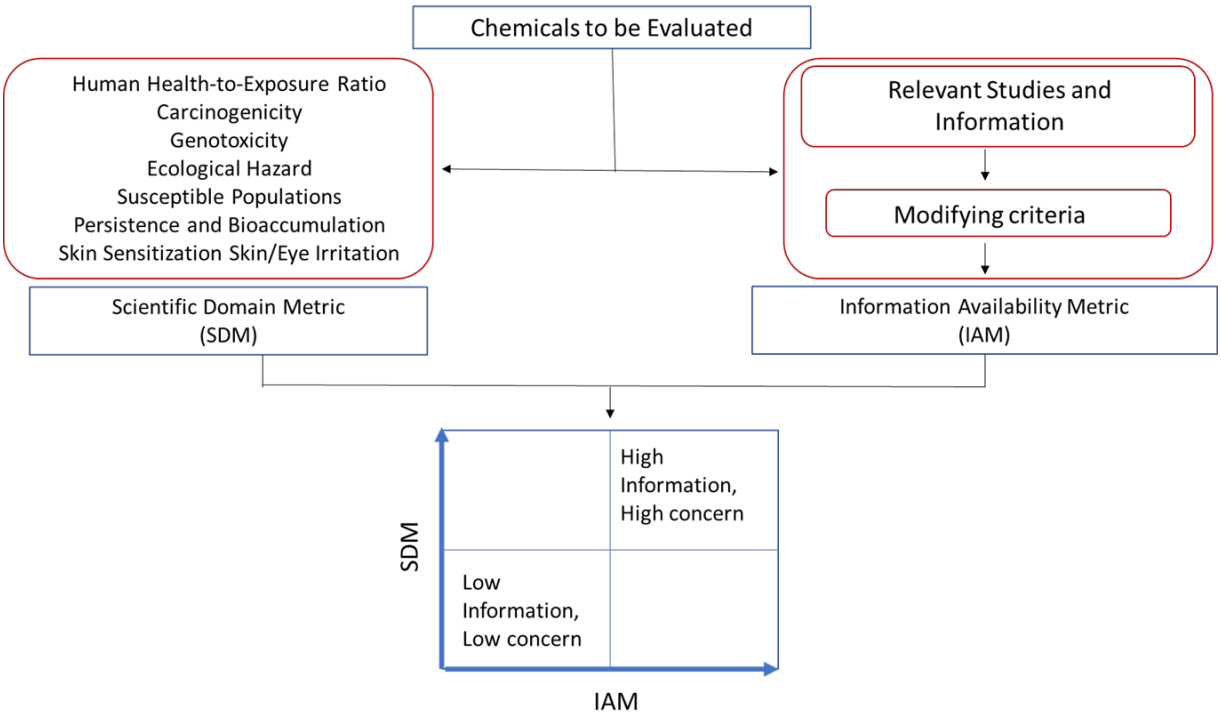


Figure 1. Schematic of the Public Information Curation and Synthesis (PICS) approach. The approach integrates publicly available information from seven scientific domains that represent human health and environmental hazard topics into a Scientific Domain Metric (SDM), and the amount and type of data in the Information Availability Metric (IAM). These two metrics are combined to give a visual display of the degree of potential concern and availability of publicly available information for the chemical substances assessed to inform future expert review of these chemical substances.

The PICS approach is based on two dimensions allowing visualization and separation of the chemical substances along each dimension (Figure 1). The first dimension reflects the overall degree of potential concern related to human health and the environment and is the integration of the individual results from the domain-specific workflows. In the PICS approach, this dimension is referred to as the Scientific Domain Metric (SDM).

The second dimension reflects the relative coverage of potentially relevant human health and ecological toxicity and exposure publicly available information that could inform level of effort and resources that may be needed to evaluate that specific chemical substance. This

dimension is referred to as the Information Availability Metric (IAM). The level of effort and resources is typically context specific and informed by expert judgment; however, an expert driven approach is not scalable to apply to the thousands of chemical substances on the TSCA active inventory at the initial screening stage. Therefore, a set of modifying criteria were used to inform the set of potentially relevant human health and ecological toxicity information. The modifying criteria were modeled after considerations used in the TSCA New Chemicals Program and include a combination of functional use considerations, environmental half-life, water solubility, molecular weight, and whether the chemical substance is a TSCA exempt polymer. The existence of an authoritative human health assessment would also contribute to this metric. In the PICS approach, the summary result from this dimension is referred to as the IAM.

The SDM and IAM are combined into a graphical representation of the PICS approach for the chemical substances on the TSCA active inventory. In response to public comments, the PICS approach moved away from the defined 'bins' of chemical substances that had been proposed in the Working Approach for Potential Candidates. The PICS approach does not determine what a result for a specific chemical substance represents, rather it provides a synthesis of the public information available for individual chemical substances.

4.2 What the PICS Approach is Intended to Accomplish

The current non-confidential, active TSCA inventory contains over 33,000 chemical substances[25] with varying amounts and types of available information. Historical approaches that search, compile, and manually evaluate relevant information are very time and resource intensive and are not be feasible for large number of chemical substances. As part of the development of a long-term strategy to inform selection of candidates for further review, an automated approach was developed that extracts, stores, and integrates publicly available information from traditional toxicology, exposure, and environmental fate-related studies, as well as NAMs. The approach relies on an information management and technology infrastructure to efficiently and transparently perform these functions and is one of possible tools that may inform candidate selection for prioritization of TSCA inventory chemical substances. A representation of the PICS approach within candidate identification is provided in Figure 2. The PICS approach is intended to accomplish the following aims:

- Understand the landscape of publicly available information on the over 33,000 data poor and data rich chemical substances on the TSCA active inventory and aid in identifying candidates for prioritization;
- Provide a transparent and reproducible process for integrating available information and identifying potential information gaps;

[25] https://www.epa.gov/tsca-inventory/tsca-inventory-notification-active-inactive-rule

- Increase efficiency and manage workload by focusing expert review on chemical substances that may have a greater potential for selection as high- or low priority candidates;
- Create a flexible and sustainable process that can adapt to scientific advances and continual generation of new scientific information; and
- Organize the process into modular workflows that can be readily adapted to address prioritization needs under other mandates.

4.3 What the PICS Approach is Not Intended to Accomplish

In order to manage expectations, it is also important to define what lies outside the domain of the PICS approach. The PICS approach is **not** intended to:

- Replace the formal TSCA prioritization or risk evaluation processes;
- Create a ranked list of chemical substances;
- Signal that the EPA has concerns with particular chemical substances or categories of chemical substances;
- Supplant expert judgment and review;
- Utilize confidential business information (CBI); or
- Incorporate systematic review of information to address study and data quality.

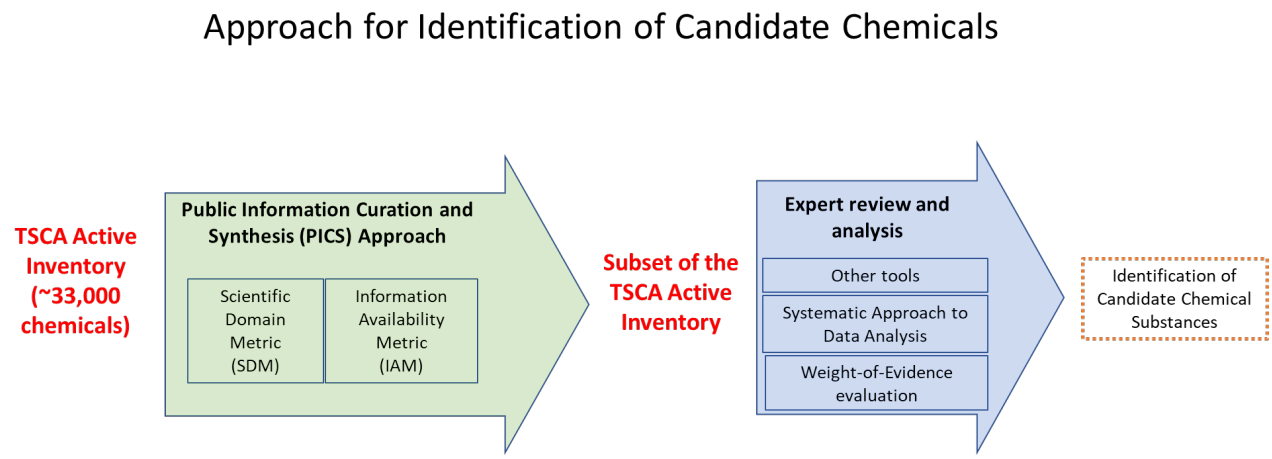


Figure 2. Schematic of the PICS Approach in Relation to Identifying High- and Low priority Candidate Chemical Substances. The PICS approach is a tool that can be used to inform identification of candidate chemical substances. The PICS approach combines results from domain-specific workflows and the relative coverage of potentially relevant human health and ecological toxicity information to identify a subset of the TSCA active inventory for additional expert review and analysis. Potential candidates identified using this approach combined with those from other tools (e.g., OncoLogic™) followed by expert review and weight-of-evidence (WOE) analysis is one approach that EPA can use to help select candidate chemical substances for prioritization.

5. A Proof-of-Concept Case Study

A subset of the TSCA active inventory was selected to test the PICS approach in a proof-of-concept (POC) case study. The subset of chemical substances was designed to evaluate the impact of various workflows across the scientific domains and gauge the impact of different modifying criteria on information availability. The chemical substances (described in further detail below) focused on including a broad range of chemical substances with varying levels of available information and included the initial proposed set of 20 high[26]- and 20 low priority[27] chemical substances. The PICS approach is intended to work broadly across the chemical substance landscape, but this case study was designed for particular application within the TSCA active inventory. Following the development of this approach, the PICS approach can also be applied to the broader TSCA active inventory or adapted to other decision contexts.

The POC case study was also used to develop standard operating procedures (SOPs) for quality control (QC) analysis of the large, Type 1 datasets[28] required to apply the approach to the TSCA active inventory. The SOPs are being implemented in an internal software system to assess, track and correct any discrepancies between data used as input and the source documents or files from which these data were obtained. The QC analysis was mainly focused on the data accuracy, and not determining the study or data quality. Further review of data and study quality would be performed by experts outside of the PICS approach prior to chemical substance selection. The QC software system being developed will also accommodate in-depth, expert review of studies for selected chemical substances and/or studies with summary values that trigger QC review.

5.1 Chemical Substance Selection, Curation and Quality Control

Chemical Structure and Identifier Mapping

Chemical substances on the non-confidential TSCA active inventory were downloaded from the EPA website[29]. Chemical substances without a CAS Registry Number were removed, and the remaining chemical substances mapped to Distributed Structure-Searchable Toxicity (DSSTox) substance identifiers (DTXSIDs) in the ChemReg chemical registration system[30], a

[26] This list can be found at https://comptox.epa.gov/dashboard/chemical_lists/TSCAHIGHPRI
[27] This list can be found at https://comptox.epa.gov/dashboard/chemical_lists/TSCALOWPRI
[28] Type 1 sources were defined as data sources storing reasonably available and relevant information that could be readily queried and extracted in a structured manner. This includes existing databases (and dashboards) that allow the user to sift through information using a graphical user-interface, a direct query such as Structured Query Language (SQL), or webservice Application Programming Interfaces (APIs). (https://www.epa.gov/sites/production/files/2018-09/documents/preprioritization_white_paper_9272018.pdf)
[29] https://www.epa.gov/tsca-inventory/how-access-tsca-inventory
[30] Grulke CM, Williams AJ, Thillanadarajah I, Richard AM. (2019). EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research. Computational Toxicology 12:100096.

database underpinning the CompTox Chemicals Dashboard[31,32]. Any chemical substances with conflicts between the TSCA identifiers and DSSTox records (e.g., discrepant CAS numbers or chemical substance names) were placed in a queue for mapping review by trained chemists. The mapped, non-confidential TSCA active inventory contained ~33,000 chemical substances and 25,275 with structural information visible via the CompTox Chemicals Dashboard[33].

Chemical Substance Selection

Chemical substances for the POC case study were selected from the mapped, non-confidential TSCA active inventory by the scientific experts designing the domain specific workflows and included:

- Initial proposed set of 20 high- and 20 low priority candidate chemical substances along with the initial first ten TSCA Work Plan chemical substances selected for evaluation in 2016 (TSCA10)[34];
- Chemical substances from the 2014 update to the TSCA Work Plan (TSCA90)[35];
- Chemical substances with well-studied effects in each of the scientific domains;
- A subset of chemical substances listed in the Food and Drug Administration's (FDA) Substances Added to Food inventory (formerly Everything Added to Foods in the United States list) and EPA's Safer Choice Safer Chemical Ingredients List (SCIL).

From these lists, a total of 238 chemical substances, called the POC238 (listed in Appendix A), was compiled for the development of workflows and metrics for each of the seven scientific domains[36]. The POC238 contains some chemical substances for which an expected biological response in one or more of the separate domains would serve as a reference for evaluation of how accurately the PICS approach identified potential hazards or environmental concerns. The POC238 was selected to span a range in the degree of potential concern and information availability; however, the overall information availability for the selected chemical substances was generally higher than for the overall TSCA active inventory (see Figure 14).

---

[31] Richard, AM (2004). DSSTOX website launch: Improving public access to databases for building structure-toxicity prediction models. Preclinica 2(2):103-108.

[32] Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, Patlewicz G, Shah I, Wambaugh JF, Judson RS, et al. (2017). The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. Journal of Cheminformatics 9(1):61.

[33] Curated list of non-confidential substances on the active TSCA inventory (https://comptox.epa.gov/dashboard/chemical_lists/TSCA_ACTIVE_NCTI_0320). The list contains 33,364 chemical substances as of March 2020.

[34] TSCA10 represents the first ten TSCA Work Plan chemical substances selected for risk evaluation in 2016.

[35] TSCA90 represents the TSCA Work Plan chemical substances from the 2014 update.

[36] The scientific domains include human hazard relative to exposure, ecological hazard, carcinogenicity, genotoxicity, exposure to susceptible populations, persistence and bioaccumulation, skin sensitization and skin/eye irritation

Chemical Substance Information Extraction and Quality Control

A key component of the PICS approach is the curation of data collected from "Type 1" data sources (as defined in the Working Approach for Potential Candidates). Type 1 data sources are publicly available and readily searchable, enabling data extraction in a structured form. Hazard, exposure, persistence, and bioaccumulation information was extracted from a range of Type 1 sources (listed in Appendix B). Curated traditional and NAM data were compiled and filtered prior to being analyzed for QC. More details on specific filtering of information sources for the individual workflows are described below.

QC analysis was performed on the data for the POC238 chemical substances to ensure the curation accuracy from primary published sources to database repository format, inform the development of formal quality assurance (QA) procedures, and obtain information on the scope and resources needed to perform QC for the entire active TSCA inventory. Specific approaches and considerations for the QC review are provided in Appendix C. The QC analysis focused on a determination of accuracy of extraction of the information from the Type 1 sources and did not examine or evaluate study conclusions. Additionally, no study quality considerations were evaluated during QC review. Reviewers did not perform a critical analysis of experimental design, statistical analyses, or data interpretation. Rather, reviewers compared the aggregated Type 1 data to primary and secondary sources. Reviewers flagged data that could not be confirmed to the primary source, even if the aggregated data matched the secondary source. However, certain secondary sources, such as the ECOTOX knowledgebase, the Integrated Risk Information System (IRIS), or chemical exposure data have existing QC or peer review processes. For these select databases, confirmation to secondary source was sufficient to pass QC review.

Over 25,000 total records were identified, with nearly 17,000 data points (68%) associated with primary sources. For this effort, a data point was deemed 'reviewed' if it matched the number in the authoritative secondary source; primary source review was not required. As an example, a point of departure (POD) extracted from the Integrated Risk Information System (IRIS) would have been confirmed against the on-line IRIS database, but not tracked back to the source paper for the POD. POD matching required that the chemical identity, POD value, and relevant metadata (e.g., units, exposure route, species) were consistent. The case study developed methods for data aggregation, curation, and evaluation, as well as QA recommendations to efficiently review Type 1 datasets.

5.2 Scientific Domain Metric Assessment

A comprehensive analysis of the publicly available information for the POC238 chemical substances was performed following data curation and QC. The overall SDM is determined by

summing the results from the individual scientific domain workflows described below for the following domains: (1) human health hazard relative to exposure; (2) ecological hazard; (3) carcinogenicity; (4) genotoxicity; (5) susceptible populations; (6) persistence and bioaccumulation; and (7) skin sensitization and skin/eye irritation (Figure 3). Of the seven domains used in the PICS approach, five were included in the previous Working Approach. The additional two domains (carcinogenicity; skin sensitization and skin/eye irritation) were included in the PICS approach based on their use in the 2014 TSCA Work Plan. Each of these workflows represent a mechanism for making a determination of potential concern for a compound in each domain based on publicly available data. These domains were selected based on their importance to understanding human health and ecological hazard, human exposure (including susceptible populations), past use in TSCA prioritization activities[37], and/or the statutory language in the Frank R. Lautenberg Chemical Safety for the 21st Century Act (P.L. 114-182)[38].

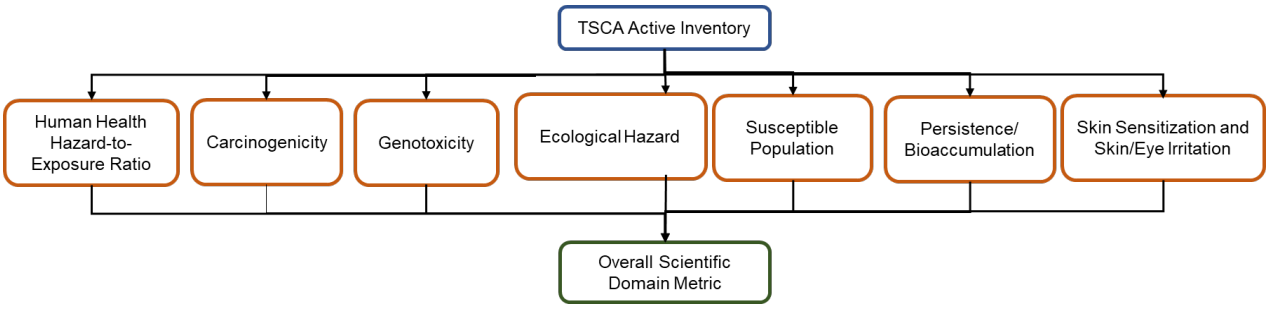


Figure 3. The seven scientific domains used to evaluate the degree of potential concern related to human health and the environment for each chemical substance. The overall SDM is the sum of the individual workflows within each domain.

Human Health Hazard-to-Exposure Ratio Domain

The identification of 2014 TSCA Work Plan chemical substances included consideration of human health hazard as well as information on exposure potential (TSCA 2012)[39]. As outlined in the Working Approach for Potential Candidates, the workflow described for this domain proposes the use of ratios of hazardous effect dose-response (e.g., point-of-departure) information to

[37] TSCA Work Plan Methods Document 2012 (https://www.epa.gov/sites/production/files/2014-03/documents/work_plan_methods_document_web_final.pdf).
[38] https://www.govinfo.gov/app/details/PLAW-114publ182 https://www.govinfo.gov/app/details/PLAW-114publ182
[39] TSCA Work Plan Methods Document 2012 (https://www.epa.gov/sites/production/files/2014-03/documents/work_plan_methods_document_web_final.pdf).

exposure predictions. As point-of-departure doses for hazardous effects approach exposure predictions, a greater degree of potential concern may be indicated, whereas doses for hazard and exposure separated by many orders of magnitude may suggest a lower degree of potential concern.

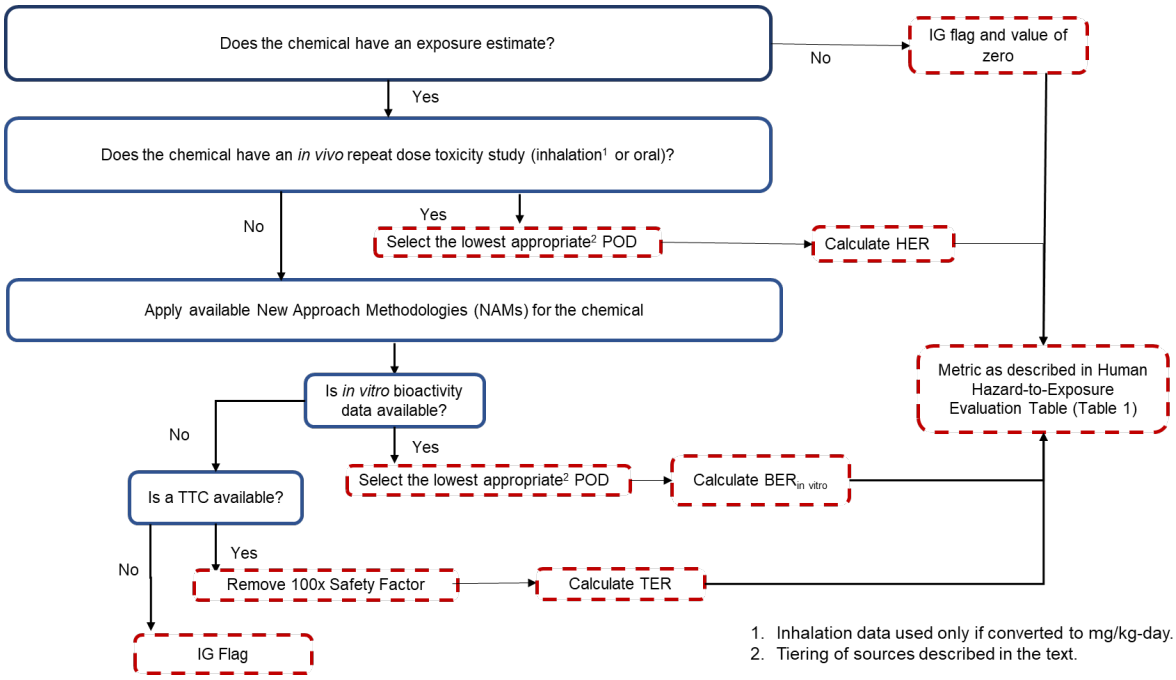


Figure 4. Workflow associated with the human health hazard-to-exposure ratio domain. Inhalation data used only if converted to mg/kg-day; tiering of sources described in the text. HER, hazard-to-exposure ratio calculated based on a point-of-departure from an *in vivo* repeat dose toxicity study divided by the median ExpoCast exposure estimate; BER, bioactivity-to-exposure ratio calculated based on the *in vitro-to-in vivo* extrapolation (IVIVE)-adjusted bioactivity estimates divided by the median ExpoCast exposure estimate; TER, threshold of toxicological concern-to-exposure ratio calculated based on the TTC divided by the median ExpoCast exposure estimate; POD, point-of-departure; and IG Flag, information gathering flag.

The calculation of the human health hazard-to-exposure ratio (HER) domain metric is based on a workflow that incorporates a tiered selection of hazard information as well as exposure estimates from the EPA ExpoCast (Exposure Forecasting) modeling effort (Figure 4)[40]. The third generation ExpoCast Systematic Empirical Evaluation of Models (SEEM3) exposure model[41] is a meta-model for aggregate population median dose intake rate and incorporates twelve different

[40] https://www.epa.gov/sites/production/files/2014-12/documents/exposure_forecasting_factsheet.pdf
[41] Ring CL, Arnot JA, Bennett DH, Egeghy PP, Fantke P, Huang L, Isaacs KK, Jolliet O, Phillips KA, Price PS, Shin HM. (2018). Consensus modeling of median chemical intake for the US population based on predictions of exposure pathways. Environmental Science & Technology 53(2):719-32.

exposure predictors covering sources that are near[42]- and far-field[43]. Four distinct source-based exposure pathways were considered: non-pesticide dietary, consumer products, far-field chemical, and far-field industrial. Chemical substances with other exposure pathways are outside of the domain of the models and are noted with an information gathering (IG) flag. IG flags are used to bring attention to specific aspects of the workflow decisions that may impact the results and may denote whether the data falls within the applicability domain of the model. SEEM3 is calibrated using chemical substance intake rates from biomonitoring data from the Centers for Disease Control and Prevention (CDC) National Health and Nutrition Examination Survey; (NHANES)[44]; and are used in place of the SEEM3 predictions[45]. As further described in Ring et al. 2017[46], NHANES was used as a more comprehensive dataset which allowed for incorporation of interindividual variability, including across different demographics. PODs from dose-response curves from traditional *in vivo* toxicity studies are divided by the median ExpoCast intake rate estimate to provide a HER. The approach uses only POD values with units of mg/kg-bw/day from repeat dose studies (including multiple specific toxicities, e.g., reproductive toxicity). Therefore, the majority of included studies assessed the oral route of exposure, although other routes of exposure were included if the units had been converted appropriately (e.g., inhalation exposure concentration converted to an equivalent mg/kg-bw/day dose). The POD used for HER calculation was either the minimum of the set, or if a human health-relevant POD estimate from an authoritative regulatory agency was available (ATSDR, EFSA, EPA HEAST, EPA OPP, EPA IRIS or EPA PPRTV), it was used in the analysis. When *in vivo* studies are not available, *in vitro* bioactivity estimates from ToxCast are converted into an oral dose equivalent using high-throughput toxicokinetic (HTTK) approaches[47,48] (called the *in vitro*-to-*in vivo* extrapolation (IVIVE) POD) and divided by the ExpoCast exposure estimate to provide a bioactivity-to-

[42] Near-field represents exposures occurring proximal to use-field (e.g., sources inside the home, for example from consumer products).
[43] Far-field represents exposures occurring far from use or as a result of environmental emission (e.g., ambient sources outside the home, for example from industrial releases).
[44] Centers for Disease Control and Prevention. "Fourth report on human exposure to environmental chemical substances, updated tables." US Department of Health and Human Services, Centers for Disease Control and Prevention (2017).
[45] Ring CL, Arnot J, Bennett DH, Egeghy P, Fantke P, Huang L, Isaacs KK, Jolliet O, Phillips K, Price PS, Shin HM, Westgate JN, Setzer RW, Wambaugh JF. (2019). Consensus Modeling of Median Chemical Intake for the U.S. Population Based on Predictions of Exposure Pathways. Environmental Science and Technology 53(2):719–732.
[46] Ring CL, Pearce RG, Setzer RW, Wetmore BA, Wambaugh JF. (2017). Identifying populations sensitive to environmental chemical substances by simulating toxicokinetic variability. Environment International 106:105-118.
[47] Pearce RG, Setzer RW, Davis JL, Wambaugh JF. (2017). Evaluation and calibration of high-throughput predictions of chemical distribution to tissues. Journal of Pharmacokinetics and Pharmacodynamics 44(6):549-565.
[48] Pearce RG, Setzer RW, Strope CL, Wambaugh JF, Sipes NS. (2017). httk: R package for high-throughput toxicokinetics. Journal of Statistical Software 79(4):1.

exposure ratio (BER)[49,50]. Finally, when neither *in vivo* nor *in vitro* studies are available, the most relevant threshold of toxicological concern (TTC) value is assigned when appropriate and divided by the ExpoCast exposure estimate to provide a TTC-to-exposure ratio (TER)[51]. Note that TTC values are *in silico* NAMs derived using the Toxtree software application (Ideaconsult Ltd)[52] by calculating the lower 95th-percentile POD for each of the classes of chemical substances considered, and then applying a safety factor of 100. In the current application, this safety factor is removed because lack of *in vivo* data is accounted for separately in the IAM. From a practical standpoint, if this safety factor was left in place, a vast majority of chemical substances with only a TTC value would be designated as high concern, regardless of exposure level.

Human Health Hazard-to-Exposure Evaluation

A human HER domain metric is assigned in a tiered fashion based on the magnitude of the HER, BER, or TER value. The order of preference is HER > BER > TER (i.e., if the HER is available, it is used preferentially over BER and TER). For volatile substances, PODs from traditional *in vivo* repeat dose toxicity studies that have units converted to mg/kg-bw/day are utilized, followed by IVIVE POD estimates using *in vitro* bioactivity data from ToxCast and toxicokinetic estimates from HTTK. EPA does not initially incorporate TTC values for volatile substances since well-established TTC values for the inhalation route of exposure are not yet available.

For each chemical, each metric was assigned a value in the range from 1 to 4, to allow combining the metrics in a consistent way. This is adapted from the strategy used in the 2012 TSCA Work Plan Methodology[53]. Most of the metrics naturally fell into discrete categories from low concern=1 to high concern=4. For the HER and ecological hazard domains (see below) we converted the continuous value to this scale using Formula 1:

$$Domain\ metric = 4 - 3 \times \frac{log_{10}(\text{HER/BER/TER}) - log_{10}(\text{HER/BER/TER})_{min}}{log_{10}(\text{HER/BER/TER})_{max} - log_{10}(\text{HER/BER/TER})_{min}} \quad (1)$$

[49] Paul-Friedman K, Gagne M, Loo LH, Karamertzanis P, Netzeva T, Sobanski T, Franzosa JA, Richard AM, Lougee RR, Gissi A, Lee JJ, Angrish M, Dorne JL, Foster S, Raffaele K, Bahadori T, Gwinn MR, Lambert J, Whelan M, Rasenberg M, Barton-Maclaren T, Thomas RS. (2020). Utility of In Vitro Bioactivity as a Lower Bound Estimate of In Vivo Adverse Effect Levels and in Risk-Based Prioritization. Toxicological Sciences 173(1):202-225.
[50] Wetmore BA, Wambaugh JF, Allen B, Ferguson SS, Sochaski MA, Setzer RW, Houck KA, Strope CL, Cantwell K, Judson RS, LeCluyse E. (2015). Incorporating high-throughput exposure predictions with dosimetry-adjusted *in vitro* bioactivity to inform chemical toxicity testing. Toxicological Sciences 148(1):121-36.
[51] Patlewicz G, Wambaugh JF, Felter SP, Simon TW, Becker RA. (2018). Utilizing Threshold of Toxicological Concern (TTC) with high throughput exposure predictions (HTE) as a risk-based prioritization approach for thousands of chemicals. Computational Toxicology 7:58-79.
[52] Patlewicz G, Jeliazkova N, Safford RJ, Worth AP, Aleksiev B. (2008). An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. SAR and QSAR in Environmental Research 19(5-6):495-524.
[53] TSCA Work Plan Chemicals: Methods Document, https://www.epa.gov/sites/production/files/2014-03/documents/work_plan_methods_document_web_final.pdf

Note that the maximum and minimum values are taken across all chemicals with HER values so that the domain metric is scaled relative to HER values. The first term ($\log_{10}$ (HER/BER/TER)) uses an HER for a chemical, if available, followed by BER and TER. This sets the domain metric of the chemical with the lowest HER (highest concern) to a value of 4 and sets the domain metric of the chemicals with the highest HER (lowest concern) to 1 (Table 1). The minimum and maximum HER values will be somewhat sensitive to the set of chemicals included, but these values are taken from the 2838 out of the TSCA active inventory with value for either HER, BER or TER. The largest HER is 5.03×10^{14} (1,2,5,6,9,10-hexabromocyclododecane) and the smallest value is 0.89 (ethenylsilanetriyl triacetate). A value of 0 is given in the absence of information and the substance is flagged for future information gathering. The *in vivo* hazard data is derived from the EPA ToxValDB database. The *in vitro* ToxCast data is obtained from the EPA invitroDB database. These datasets as well as the toxicokinetic data parameters are publicly available through the EPA CompTox Chemicals Dashboard[54].

Table 1. Criteria used to calculate the human hazard to exposure ratio domain metric

Metric	HER, BER, or TER value[1]
0	No available data (hazard or exposure)
1	Result is on a continuum based on Formula 1, i.e., 1 = highest HER, BER, TER (lowest concern); 4 = lowest HER, BER, TER (highest concern)
2	
3	
4	

Information Gathering (IG) Flags: Note concerning key study types with no *in vivo* data (repeat dose, reproductive, developmental); secondary source data; predicted data; lack of exposure data

[1]HER, hazard-to-exposure ratio calculated based on in vivo repeat dose toxicity studies divided by the median ExpoCast exposure estimate; BER, bioactivity-to-exposure ratio calculated based on IVIVE bioactivity estimates divided by the median ExpoCast exposure estimates; TER, TTC-to-exposure ratio calculated based on the TTC divided by the median ExpoCast exposure estimate.

Limitations and Longer-term Options

When only data from acute *in vivo* studies was available, the data were not considered sufficient for calculation of the HER, which uses hazard information from *in vivo* repeat dose

[54] https://comptox.epa.gov/dashboard

studies, including studies for specific endpoints including reproductive and developmental toxicology. (Note though that the presence or absence of acute data is included in the IAM, described below.) Ongoing research will be needed to determine how to utilize acute toxicity information in the absence of repeat dose toxicity information to estimate a POD for HER calculation. In the current implementation, POD values with typical inhalation units (mg/m^3 or ppm) have been excluded. Converting these inhalation values to the oral equivalent dose value requires, at a minimum, knowing whether the chemical substance has local or systemic effects. This information is not typically captured in the current Type 1 information sources and will require either manual curation of the relevant studies, or development of a semi-automated approach to select the appropriate exposure effect class. Similarly, conversion of dermal exposure was not addressed for this case study. Future efforts could incorporate data from these additional routes of exposure.

Calculation of the BER is influenced by selection of a minimum *in vitro* potency value from high-throughput bioactivity screening data and the HTTK approaches used to derive an administered equivalent dose. Although this is an area of ongoing research, current evidence supports that this global bioactivity approach is conservative[55] and that further efforts to refine may provide additional pathway specific PODs (increase relevance). In the POC, features of the concentration-response curves fit to the ToxCast high-throughput bioactivity data have been used to identify a minimum potency value showing bioactivity. In future iterations, we propose leveraging ongoing research on how to best identify the minimum credible *in vitro* potency values from ToxCast and other sources of high-throughput bioactivity data (e.g. high-throughput transcriptomics), as well as ongoing and iterative improvements in HTTK modeling. Selected choices in HTTK modeling approaches can also include a consideration of interindividual toxicokinetic variability, or not, depending on the scenario; in a conservative approach to an initial screening of substances, use of estimated parameters for toxicokinetically-susceptible individuals to derive administered equivalent doses may be informative.

Additionally, QSAR models were considered to estimate *in vivo* PODs as a fourth level in the hazard estimation process (*in vivo*>IVIVE>QSAR>TTC), but at the time of development, a valid QSAR model was not available. Finally, the current TTC values are limited to oral exposures. We are reviewing the latest research efforts related to the use of TTC for other routes of exposure, and any future improvements to this approach may expand the domain of applicability of the TTC

[55] Paul Friedman K, Gagne M, Loo LH, Karamertzanis P, Netzeva T, Sobanski T, Franzosa JA, Richard AM, Lougee RR, Gissi A, Lee JJ, Angrish M, Dorne JL, Foster S, Raffaele K, Bahadori T, Gwinn MR, Lambert J, Whelan M, Rasenberg M, Barton-Maclaren T, Thomas RS. (2020). Utility of In Vitro Bioactivity as a Lower Bound Estimate of In Vivo Adverse Effect Levels and in Risk-Based Prioritization. Toxicological Sciences 173(1):202-225.

to incorporate these updates such as in Nelms and Patlewicz (2020)[56]. This may also help to address limitations of this approach related to potential screening of compounds for which *in vitro* assays (the basis of BER) or TTC (the basis of TER) do not perform well.

Carcinogenicity Domain

The probable or known carcinogenicity of a chemical substance was considered in selecting the 2014 TSCA Work Plan chemical substances[57]. Carcinogenicity was not included as a separate domain in the previous Working Approach for Potential Candidates, due to limited availability of Type 1 carcinogenicity data sources. In the Working Approach for Potential Candidates, the genotoxicity domain was considered as a surrogate for carcinogenicity. In the PICS approach, carcinogenicity and genotoxicity are included as separate domains due to the fact that carcinogenicity may be associated with non-genotoxic as well as genotoxic mechanisms.

The ability of an agent to cause cancer in humans is typically assessed using a weight-of-evidence approach, considering exposure, epidemiology, animal cancer data, and mechanistic data, including genotoxicity and pharmacokinetic/pharmacodynamic information. Major national and international organizations convene expert panels to perform these evaluations, resulting in authoritative assessments of the potential of agents to induce cancer in humans (e.g., IARC, IRIS). EPA has its own guidelines for cancer that consider mechanistic data as an important component of carcinogenicity[58]. In the absence of such evaluations for human cancer, the ability of the agent to cause cancer in *in vivo* rodent models is an indication of the potential of an agent to be carcinogenic to humans. Rodent chronic bioassays include the standard protocols established by organizations such as the National Toxicology Program (NTP)[59], as well as more generalized guidance from institutions such as the Organization for Economic Cooperation and Development (OECD)[60]. These data have been compiled in ToxValDB. The presence of lesions believed to have resulted from carcinogenesis were used in a binary fashion in the scoring process. That is, the potency of the carcinogen is not reflected in the evaluation process, just the evidence of carcinogenicity (yes or no) (Figure 5).

---

[56] Nelms MD, Patlewicz G. (2020) Derivation of new Thresholds of Toxicological Concern values for exposure via inhalation for environmentally-relevant chemicals. Frontiers in Toxicology 2:5.
[57] https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/tsca-work-plan-methods-document
[58] https://www.epa.gov/sites/production/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf
https://www.epa.gov/sites/production/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf
[59] NTP Toxicology/Carcinogenicity Study Overview. https://ntp.niehs.nih.gov/testing/types/cartox/index.html
[60] OECD Draft guidance http://www.oecd.org/chemicalsafety/testing/44960015.pdf

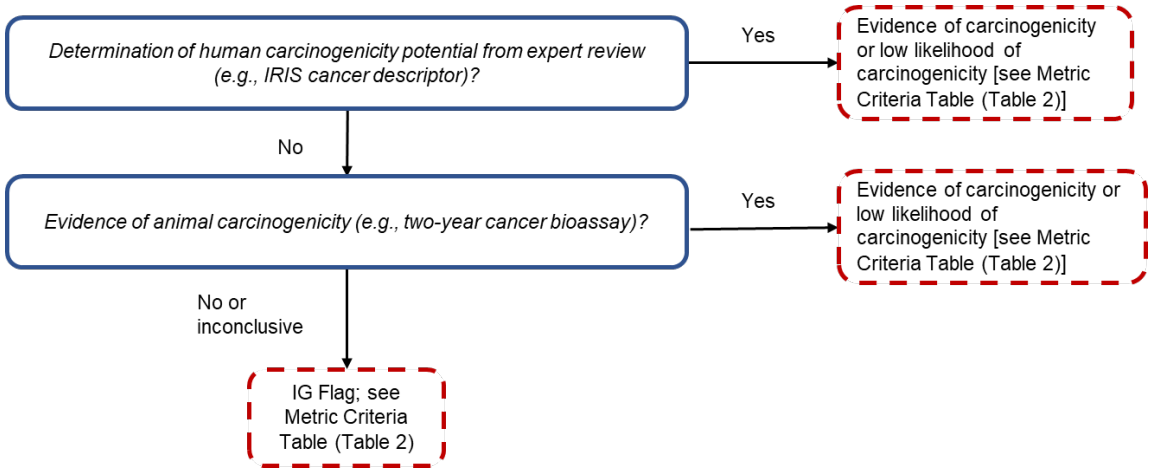


Figure 5. Tiered evaluation process associated with the carcinogenicity domain. The workflow begins with the determination of human carcinogenicity from an authoritative source and ends at one of the red, dashed line boxes. Blue, solid line boxes represent intermediate decision points. IG flag = information gathering flag.

Carcinogenicity Evaluation

The carcinogenicity domain metric is determined from a two-tiered evaluation workflow of the available carcinogenicity data in humans and/or animals. Chemical substances that have not been evaluated for their ability to cause cancer in humans, or where there are no available human data, are evaluated for the ability to cause cancer in animals. This is a binary result, not based on the dose required to produce carcinogenicity, but based on the presence or absence of carcinogenicity in a study (Figure 5/Table 2). Chemical substances with evidence of known human carcinogenicity as determined by an authoritative source are given a value of 4; chemical substances that have been determined to have evidence as possible or probable human carcinogens are given a value of 3; chemical substances that have been shown to cause cancer in animals but have not otherwise been assessed for their ability to cause cancer in humans are given a value of 2; and chemical substances with evidence indicating a low likelihood of carcinogenicity in either humans or rodents based on negative data (e.g., a negative rodent cancer bioassay) are given a value of 1. This category may also be termed as having inadequate or insufficient evidence of carcinogenicity in an authoritative carcinogenesis assessment. A value of 0 is given in the absence of data and an information gathering flag is included.

Table 2. Criteria used to calculate the carcinogenicity domain metric	
Metric	**Carcinogenicity Determination**
0	No available data for carcinogenicity
1	Evidence of low likelihood of carcinogenicity; inadequate or insufficient data
2	Evidence for animal carcinogenicity but not assessed for human carcinogenicity
3	Evidence of possible or probable human carcinogenicity based on either human epidemiology or animal toxicology data
4	Known human carcinogen
Information Gathering (IG) Flags: predicted data; secondary source data; determination by authoritative source	

Limitations and Longer-term Options

There are limited data available for carcinogenicity of chemical substances. One limitation of this approach is the lack of a published peer-reviewed automated predictive model for the determination of carcinogenicity. OncoLogic™ [61], a computer system that evaluates the carcinogenic potential of chemical substances, has not yet been modified to analyze large numbers of chemical substances in a manner that can be readily incorporated into this approach. In the future, the OncoLogic system could be adapted to meet this need and could be incorporated into this workflow in a tiered manner. Indeed, there is an activity within the OECD Toolbox Management Group which is investigating the feasibility of implementing the decision logic of selected OncoLogic chemical classes into the OECD Toolbox. Currently, OncoLogic is used as part of the expert review of compounds and incorporated into the weight-of-evidence assessment of specific compounds.

Genotoxicity Domain

Genotoxicity is an important component of understanding chemical substance hazards. Genotoxicity refers to the ability of agents to induce DNA damage, such as DNA strand breaks or DNA adducts, as well as the ability to induce mutations, i.e., heritable changes in DNA sequence. In the absence of carcinogenicity data, genotoxicity is often used as a surrogate. This document evaluates chemical substances for genotoxicity by considering data from assays that collectively detect mutations in bacteria or mammalian cells, as well as DNA damage in mammalian cells or rodents. For the PICS approach, some consideration was given to including genotoxicity within the same domain as carcinogenicity. However, it was determined that these should be considered

[61] https://www.epa.gov/tsca-screening-tools/oncologictm-computer-system-evaluate-carcinogenic-potential-chemicals

separately to incorporate not only the impact of nongenotoxic carcinogens, but also capture genotoxic chemical exposures that may not have been assessed for cancer.

Since the initial EPA implementation of TSCA in 1976, many studies have assessed which combinations of genotoxicity tests are the most predictive[62], resulting in testing schemes recommended by the OECD Genetic Toxicology Test Guidelines[63], the International Conference on Harmonization[64], and the NTP[65]. Additional consideration has been given to entirely new testing approaches, which do not rely on traditional assays[66,67].

Most regulatory bodies in the U.S., such as the EPA and FDA, recommend the OECD genetic toxicology guidelines. This testing includes a set of bacterial assays for gene mutation using strains of *Salmonella* (the Ames strains) and strains of *Escherichia coli* WP2 and assays for chromosomal mutation (*in vitro* chromosome aberration assay, mouse bone-marrow micronucleus assay, and the mouse lymphoma $Tk^{+/-}$ assay). The combination of assays identifies genotoxic agents that produce primarily gene mutations, chromosomal mutations or both gene and chromosomal mutations. A smaller number of chemical substances produce aneuploidy (chromosome gain or loss), which is also detected by the chromosome aberration (CA) or mouse bone-marrow micronucleus assays (MNT).

In the PICS approach, we considered that the genotoxicity of an agent can be sufficiently assessed by evaluating data in the standard bacterial mutation assays (the Ames *Salmonella* and *E. coli* WP2 strains) and the three principal assays for chromosomal mutation (*in vitro* chromosome aberration assay, mouse bone-marrow micronucleus assay, and the mouse lymphoma $Tk^{+/-}$ assay). The selection of a subset of genotoxicity assays used in this approach was based on recent work

---

[62] Eastmond DA, Hartwig A, Anderson D, Anwar WA, Cimino MC, Dobrev I, Douglas GR, Nohmi T, Phillips DH, Vickers C. (2009). Mutagenicity testing for chemical risk assessment: update of the WHO/IPCS Harmonized Scheme. Mutagenesis 24:341-349.

[63] Guidance Document on Revisions to OECD Genetic Toxicology Test Guidelines (2015) http://www.oecd.org/chemicalsafety/testing/Genetic%20Toxicology%20Guidance%20Document%20Aug%2031%202015.pdf

[64] International Conference on Harmonisation. (2012). Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended for Human Use. S2(R1). http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S2_R1/Step4/S2R1_Step4.pdf

[65] NTP Genetic Toxicology. https://ntp.niehs.nih.gov/testing/types/genetic/index.html

[66] Thomas RS, Philbert MA, Auerbach SS, Wetmore BA, Devito MJ, Cote I, Rowlands JC, Whelan MP, Hays SM, Andersen ME, Meek ME, Reiter LW, Lambert JC, Clewell HJ 3rd, Stephens ML, Zhao QJ, Wesselkamper SC, Flowers L, Carney EW, Pastoor TP, Petersen DD, Yauk CL, Nong A. (2013). Incorporating new technologies into toxicity testing and risk assessment: Moving from 21st century vision to a data-driven framework. Toxicological Sciences. 136:4–18.

[67] Dearfield KL, Gollapudi BB, Bemis JC, Benz RD, Douglas GR, Elespuru RK, Johnson GE, Kirkland DJ, LeBaron MJ, Li AP, Marchetti F, Pottenger LH, Rorije E, Tanir JY, Thybaud V, van Benthem J, Yauk CL, Zeiger E, Luijten M. (2017). Next generation testing strategy for assessment of genomic damage: A conceptual framework and considerations. Environmental and Molecular Mutagenesis 58:264-283.

by Williams et al. 2019[68] which demonstrated that this subset of assays is sufficient to identify 99% of mutagens tested.

In the absence of experimental data, genotoxicity may be predicted using *in silico* Quantitative Structure-Activity Relationship (QSAR) models for Ames mutagenicity or *in silico* structural alerts for clastogenicity. This evaluation is similar for measured data but is tagged with an IG flag for predicted data. The EPA Toxicity Estimation Software Tool (TEST) was used to predict Ames mutagenicity, along with the OECD Toolbox, which includes DNA alerts for Ames, CA and MNT; protein binding alerts for CA; *in vitro* mutagenicity (Ames test) alerts by Instituto Superiore di Sanita (ISS), and *in vivo* mutagenicity (micronucleus) alerts by ISS. The Ames mutagenicity module within the TEST software is based on a dataset of 6,512 chemical substances that was compiled from several different sources as described in Hansen *et al*.[69]. After removal of salts, mixtures, ambiguous compounds, and compounds without CAS numbers, the final dataset consisted of 5,743 chemical substances. Several different approaches were used to derive TEST predictions, including a hierarchical-clustering approach[70], a nearest-neighbor approach, a Food and Drug Administration approach, and a single-model approach. The profilers within the OECD Toolbox are collections of structural alerts that have been compiled and developed by various researchers and organizations. Most of the profilers incorporate the alerts devised by Ashby and Tennant (1991)[71], but additional alerts are included depending on the experimental data available. DNA_OASIS profilers include alerts derived from the training sets used for the TIMES expert system, whereas the ISS alerts rely on the ISSCAN[72] database. A consensus outcome from the individual models culminates in the overall prediction conclusion generated for a given chemical substance.

[68] Williams RV, DM DeMarini, LF Stankowski Jr, PA Escobar, E Zeiger, J Howe, R Elespuru, KP Cross. (2019). Are all bacterial strains required by OECD mutagenicity test guideline TG471 needed? Mutation Research 848:503081.
[69] Hansen K, Mika S, Schroeter T, Sutter A, ter Laak A, Steger-Hartmann T, Heinrich N, Müller KR. (2009). Benchmark data set for *in silico* prediction of Ames mutagenicity. Journal of Chemical Information and Modelling 49(9):2077-81.
[70] More details can be found at https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test
[71] Ashby J, Tennant RW. (1991). Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. Mutation Research 257(3):229-306.
[72] Carcinogens database developed by ISS. http://old.iss.it/publ/anna/2008/1/44148.pdf

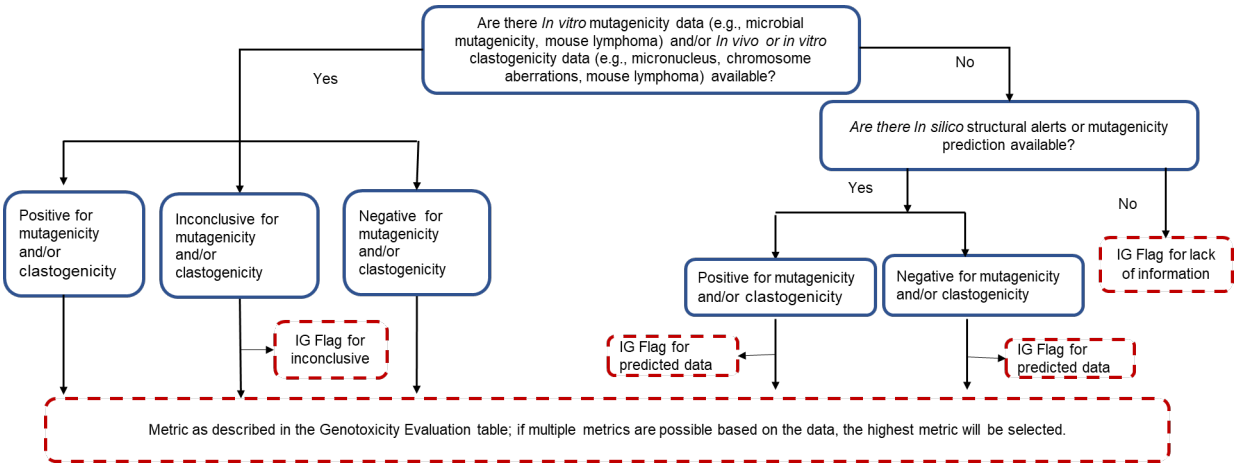


Figure 6. Tiered evaluation process associated with the genotoxicity domain. The process evaluates the potential mutagenicity and DNA damaging potential of a chemical substance as well as the potential clastogenicity. IG, information gathering.

Genotoxicity Evaluation

The process for screening chemical substances for genotoxicity is shown in Figure 6/Table 3. Chemical substances with evidence of genotoxicity are evaluated based on results for gene mutation in bacterial mutagenicity assays and any of the three assays for chromosomal mutations (clastogenicity as described above). Chemical substances that have been determined to be genotoxic experimentally (either as a gene or chromosomal mutagen) are given a value of 4; chemical substances that are predicted to be genotoxic are given a value of 3; chemical substances with inconclusive data are given a value of 2; chemical substances with data showing that the chemical substance is not likely to be genotoxic are given a value of 1; and chemical substances with no data are given a value of 0. Chemical substances with inconclusive results are also tagged with an IG flag. If there are multiple data sources for a chemical substance, the highest metric is selected, e.g., if a chemical substance is a predicted clastogen (metric of 3) and a mutagen based on measured data (metric of 4), an overall metric of 4 was used. No attempts were made to evaluate the quality and design of the studies, and the IG flag gives the end-user an opportunity to evaluate the weight-of-evidence in situations that are deemed inconclusive. Combinations of the Ames or clastogenicity information was used to determine the genotoxicity, which was converted to a numeric scale as described in Table 3.

Table 3. Criteria used to calculate the genotoxicity domain metric	
Metric	**Genotoxicity Determination**
0	No available data for genotoxicity
1	Evidence of low likelihood of genotoxicity – predicted or measured data
2	Inconclusive evidence of genotoxicity (predicted or measured)
3	Evidence of genotoxicity - predicted data
4	Evidence of genotoxicity - measured data
Information Gathering (IG) Flags: predicted data; conflicting results in situations of varied results in same assay type; secondary source data	

Limitations and Longer-term Options

A limitation of the genotoxicity analysis was the reliance on secondary source data for specific genotoxicity endpoints. The sources included authoritative assessments, data compilation summaries, and publicly available review papers. For the purposes of the automated PICS approach, secondary data sources were deemed acceptable but given an IG flag so that the expert reviewer would have an awareness of the potential limitations of the data sources. Although weight-of-evidence results from authoritative sources may be considered, this approach does not perform a weight-of-evidence analysis for genotoxicity. For many of the chemical substances in the database, there were genotoxicity data available from a variety of data sources. This screening approach does not explicitly address the absolute number of positive or negative results or contradictory results. These issues should be considered during the downstream expert review of any candidate chemical substances (Figure 2). Similarly, a recent effort to develop a genotoxicity hazard assessment framework using *in silico* tools was published[73]. This method was developed to rapidly assess chemical hazard but is not designed as an automated approach for analysis of a large number of chemicals at one time. While the published approach goes further to incorporate expert review, many of the underlying assumptions on incorporating specific genotoxicity assays support decisions made in the PICS approach. Longer-term options should also include recent advances in the development of a consensus model using combinations of QSAR models and structural alert predictions to strengthen the use of predictive results in the genotoxicity assessment of compounds[74].

[73] Hasselgren C, Ahlberg E, et al. (2019). Genetic toxicology in silico protocol. Regulatory Toxicology and Pharmacology 107:104403.

[74] Pradeep P, Judson R, DeMarini, DM, Keshava N, Martin TM, Dean J, Gibbons CF, Sima A, Warren SH, Gwinn MR, Patlewicz, G. (2021). Evaluation of existing QSAR models and structural alerts and development of new ensemble models for genotoxicity using a newly compiled experimental dataset. Computational Toxicology 18:100167.

Ecological Hazard Domain

The ecological hazard domain is intended to account for potential toxicity to a broad diversity of wildlife and plants. Under typical approaches for ecological hazard classification of chemical substances, for example under the Globally Harmonized System (GHS)[75], only aquatic toxicity is considered. However, it is common to consider data from at least three trophic levels of organisms, generally a fish, an invertebrate (crustacean), and at least one plant or algae species[76]. In cases where data for all three trophic levels are unavailable, additional uncertainty factors are often applied to account for the fact that potentially sensitive classes of organisms with highly distinct life histories and physiology have not been considered. Consistent with GHS, experimentally-derived test data are preferred for derivation of the ecological hazard metric. However, in cases where no experimentally derived PODs are available, QSAR models are used. This well-established method for evaluating ecological hazards is the basis of the workflow developed for the ecological hazard domain of the PICS approach.

The ecological hazard domain is a hazard-only approach based on measured or estimated chronic and acute aquatic toxicity. *In vivo* aquatic toxicity data are collected from US EPA's ECOTOX Knowledgebase[77], eChemPortal[78], and EFSA[79]. In cases where *in vivo* data are absent, QSAR-based predictions of aquatic acute toxicity are derived from EcoSAR[80] or TEST[81]. All data are compiled into ToxValDB.

In the PICS approach (Figure 7), additional uncertainty factors are not applied when the three trophic levels typically considered in GHS classification[82] are not represented in the dataset. However, there is a computational evaluation of the presence/absence of data for each trophic level. In cases where one or more trophic levels of organisms are not represented, an alert is provided in the form of an information gathering flag. This alert provides an indication that at least one major group of potentially sensitive taxa are not currently considered as part of the ecological hazard domain metric.

---

[75] United Nations. (2017). Globally harmonized system of classification and labeling of chemical substances. Seventh revised edition. United Nations, New York, NY, USA. ST/SG/AC.10/30/Rev.7

[76] United Nations. (2017). Globally harmonized system of classification and labeling of chemical substances. Seventh revised edition. United Nations, New York, NY, USA. ST/SG/AC.10/30/Rev.7

[77] https://cfpub.epa.gov/ecotox/

[78] https://www.echemportal.org/

[79] http://www.efsa.europa.eu/

[80] https://www.epa.gov/tsca-screening-tools/ecological-structure-activity-relationships-ecosar-predictive-model

[81] https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test

[82] United Nations. (2017). Globally harmonized system of classification and labeling of chemical substances. Seventh revised edition. United Nations, New York, NY, USA. ST/SG/AC.10/30/Rev.7

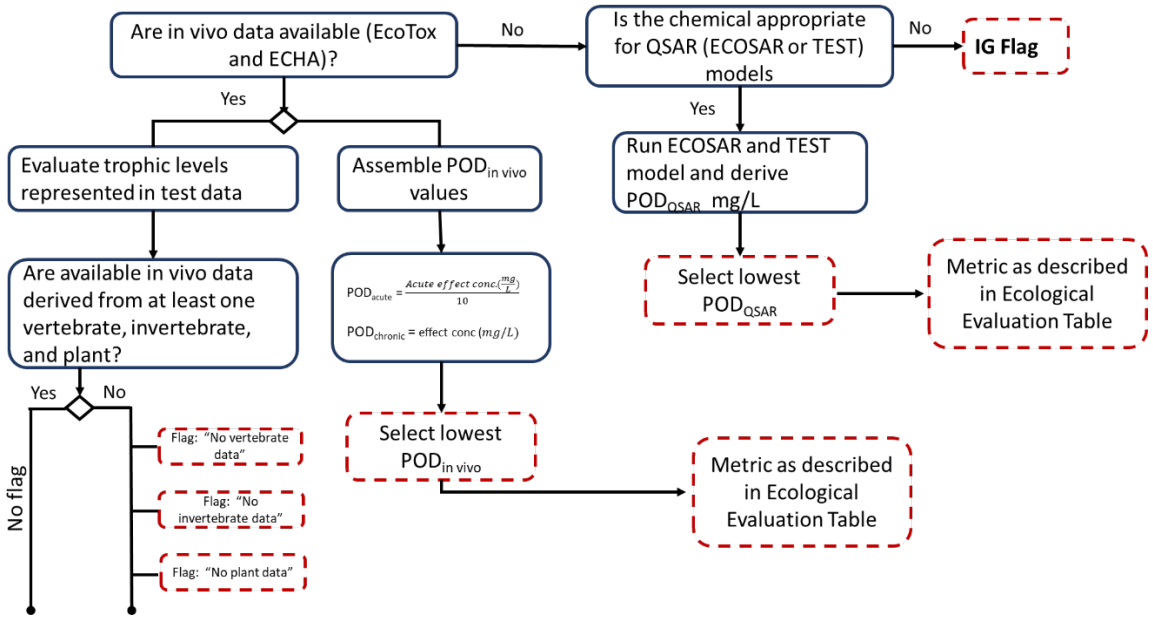


Figure 7. Workflow for derivation of the ecological hazard domain metric. Boxes with dashed red borders indicate decision points and information gathering alerts that are appended to the metric. Acute POD values are divided by 10 prior to comparison to chronic values. This aligns with a 10-fold difference in the category cut-off concentrations for acute versus chronic values applied under the Globally Harmonized System[83]. QSAR predicted values are considered only when no *in vivo* data are available and are used, unadjusted.

Ecological Hazard Evaluation

An ecological domain metric is based on the lowest POD value derived from available *in vivo* data (acute and chronic; any endpoint or life stage) or QSAR predictions when *in vivo* POD estimates are unavailable (Figure 7). Using the same rationale as developed for GHS classification, *in vivo* acute values (expressed as LC50 or LD50) are divided by a factor of 10 to derive a final *in vivo* POD estimate ($POD_{in\ vivo}$). Chronic toxicity data (e.g., reported as NOEC, LOEC, NOEL, LOEL, NOAEL, LOAEL) and QSAR-based estimates are used unadjusted. The minimum POD_{QSAR} value is the minimum across all EcoSAR and TEST predictions (including both acute and chronic models). In general, the chronic values are lower than the acute values, so chronic QSAR values are most often used. The resulting minimum $POD_{in\ vivo}$ across all three categories are then compared with the ecological hazard domain metric table (Table 4) to assign the final value on a scale of 1-4 (low hazard to very high hazard). A value of 1 is also given when the POD

[83] United Nations. (2017). Globally harmonized system of classification and labeling of chemical substances. Seventh revised edition. United Nations, New York, NY, USA. ST/SG/AC.10/30/Rev.7

is greater than the water solubility of the chemical[84]. In the absence of *in vivo* data, the minimum POD_{QSAR} is selected. The domain metric was calculated in the same way as for human HER, i.e. by scaling the continuous chemical level POD values onto the 1-4 scale, using Formula 2:

$$Domain\ metric = 4 - 3 \times \frac{log_{10}POD - log_{10}POD_{min}}{log_{10}POD_{max} - log_{10}POD_{min}} \quad (2)$$

Here the POD_{min} and POD_{max} are the minimum and maximum PODs across all chemicals. As with HER, this assigns a metric value of 4 to the chemical with lowest POD and a metric value of 1 to the chemical with the largest POD. The minimum and maximum POD values will be somewhat sensitive to the set of chemicals included, but these values are taken from the 5031 out of the TSCA active inventory with *in vivo* PODs for either acute or chronic aquatic studies. The largest POD is 163,709 mg/L (phosphonic acid, [1,6-hexanediylbis [nitrilobis(methylene)]]tetrakis) and the smallest value is $1.6x10^{-19}$ (propanoic acid, 3-(dodecylthio)-, 2,2-bis[[3-(dodecylthio)-1-oxopropoxy]methyl]-1,3-propanediyl ester).

Table 4. Criteria used to calculate the ecological hazard domain metric

Metric	Minimum Aquatic POD (mg/L)
0	No available data
1	Result is on a continuum based on Formula 2, i.e., 1 = highest POD (lowest concern); 4 = lowest POD (highest concern)
2	
3	
4	
Information Gathering (IG) Flags: predicted data; secondary source data	

Limitations and Longer-term Options

One of the limitations of the ecological hazard domain metric is that, unlike the human hazard-to-exposure ratio metric, the ecological hazard domain metric is solely based on hazard, without consideration of exposure. This is primarily due to limited peer-reviewed, automated ecological exposure estimation tools. While procedures to derive ecological exposure estimates are routinely implemented, those approaches cannot currently be automated for thousands of chemical substances. Consequently, the development of an appropriate automated framework for

[84] Water solubility is predicted by OPERA (Mansouri et al., 2018; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5843579/)

ecological exposure estimation is viewed as a high priority for long-term improvement of the ecological domain metric. If appropriate ecological exposure estimates could be generated or obtained in an automated fashion, the ecological hazard domain metric could be adapted to use a hazard-to-exposure ratio approach that parallels that used in the human health domain described earlier.

A second limitation of the current ecological hazard domain metric is that it only considers aquatic ecotoxicity data. Terrestrial ecotoxicity data for mammals is considered in the human health hazard domain. However, ecotoxicity data for other terrestrial organisms (e.g., amphibians, birds, reptiles, earthworms, insects, terrestrial plants) are not considered which is consistent with the GHS approach[85]. The 2012 TSCA Work Plan Chemical substances methods[86] only considered aquatic ecotoxicity data in deriving a hazard metric. The greater reliance on aquatic ecotoxicity data over terrestrial ecotoxicity data is based in part on the assumption that the aquatic compartment is maximally vulnerable as a final receiving environment for many chemical substances. Aquatic organisms are continuously immersed in the aqueous exposure media and tend to have high exposure levels[87]. The focus on aquatic species is also based in part on the greater availability of aquatic ecotoxicity data and the availability of well-established QSARs for predicting aquatic toxicity.

Finally, consistent with the GHS approach, experimentally-derived PODs are preferred over QSAR predictions. This is expected to be robust when a modest number of experimentally-derived POD values are available. However, in cases where the experimental data are sparse, one or a few poorly designed studies could lead to underestimation of hazard. Given that the PICS approach is intended to assist in efficiently selecting chemical substances for subsequent expert review, it is assumed that study quality and data sufficiency would be evaluated in the expert-driven part of the process.

Susceptible Human Population Domain

The modernization of TSCA included a requirement for increased attention to susceptible subpopulations, such as infants, children, pregnant women, workers, or the elderly. In the context of TSCA, a "potentially exposed or susceptible subpopulation" is defined as a group of individuals within the general population who, due to either greater susceptibility or greater exposure, may be at greater risk than the general population of adverse health effects from exposure to a chemical substance or mixture. Currently, children are the only susceptible subpopulation given separate

[85] United Nations. (2017). Globally harmonized system of classification and labeling of chemical substances. Seventh revised edition. United Nations, New York, NY, USA. ST/SG/AC.10/30/Rev.7
[86] https://www.epa.gov/sites/production/files/2014-03/documents/work_plan_methods_document_web_final.pdf
[87] United Nations. (2017). Globally harmonized system of classification and labeling of chemical substances. Seventh revised edition. United Nations, New York, NY, USA. ST/SG/AC.10/30/Rev.7

consideration in the PICS approach, although additional subpopulations (e.g., pregnant women, elderly) could be incorporated in the future if appropriate to the decision context.

Children may have higher exposure levels to environmental chemical substances than adults, and life-stage dependent exposure sources and pathways can contribute to this differential[88]. Infants and young toddlers have unique exposure sources such as breast milk and formula. Children play close to the ground, and thus increased contact with the floor and a lower height of the breathing zone results in increased exposure to chemical substances in dust and to chemical substances emitted from flooring or from products applied to floors. Children display increased hand and object mouthing behaviors, and thus can be more highly exposed to chemical substances in consumer products applied to the body, residential surfaces, or in articles such as toys. Children can also be more heavily exposed to environmental pollutants than adults due to physiological factors; they consume more food and water and have higher inhalation rates per pound of body weight than adults[89].

Susceptible Population Evaluation

A susceptible population domain metric was developed that characterizes the potential for differential exposure between children and the general population (Figure 8/Table 5). The proposed susceptible subpopulation domain metric incorporates exposure from multiple sources that contribute to an increased exposure of children relative to the general population (Figure 8). Each exposure source is given an exposure differential score to semi-quantitatively represent the magnitude of potential exposure differential between children and adults. Each chemical is assessed using available data to determine whether it occurs in each exposure source and is evaluated accordingly. The exposure source definitions, the data sources used to identify associated chemicals, and the mechanism(s) by which the exposure source contributes to increased exposure for children are summarized in Appendix D. These definitions are consistent with the pathways used by ORD in high-throughput models of exposure[90]. Information in EPA's Chemical and Product (CPDat)[91] and Chemical Product Category (CPCat)[92] databases or in EPA Chemical Data Reporting (CDR) is used to determine which exposure sources are relevant for each chemical

[88] Environmental Protection Agency (2006). A Framework for Assessing Health Risks of Environmental Exposures to Children, EPA/600/R-05/093F.
[89] Environmental Protection Agency (2002). Child Specific Exposure Factors Handbook, EPA-600-P-00-002B.
[90] Ring CL, Arnot JA, Bennett DH, Egeghy PP, Fantke P, Huang L, Isaacs KK, Jolliet O, Phillips KA, Price PS, Shin HM, Westgate JN, Setzer RW, Wambaugh JF. (2019). Consensus modeling of median chemical intake for the U.S. population based on predictions of exposure pathways. Environmental Science and Technology 53(2):719-732.
[91] Dionisio KL, Phillips K, Price PS, Grulke CM, Williams A, Biryol D, Hong T, Isaacs KK. (2018). The Chemical and Products Database, a resource for exposure-relevant data on chemicals in consumer products, Scientific Data, 5: 1-9.
[92] Dionisio KL, Frame AM, Goldsmith MR, Wambaugh JF, Liddell A, Cathey T, Smith D, Vail J, Ernstoff AS, Fantke P, Jolliet O, Judson RS. (2015). Exploring consumer exposure pathways and patterns of use for chemicals in the environment. Toxicology Reports 2:228-37.

substance. CPDat and CPCat contain use information from hundreds of data sources; CPDat contains reported chemical substance data on thousands of consumer products (obtained from product Safety Data Sheets or ingredient lists), while CPCat contains general chemical substance use information for over 75,000 chemical substances from public manufacturer, government, or industry chemical substance lists. Two exposure sources, breast milk and residential dust, are characterized here via a recently published meta-analyses[93,94].

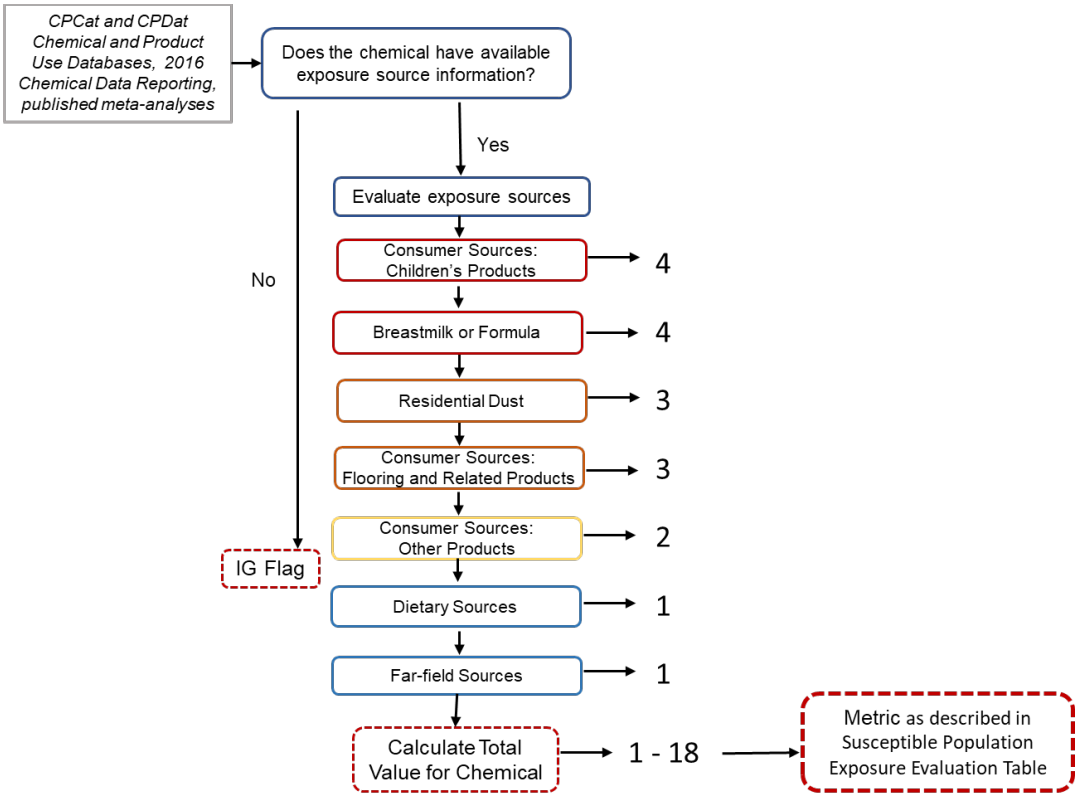


Figure 8. Workflow of susceptible population domain metric based on the relevance of multiple exposure pathways. The sources associated with individual chemical substances were identified using information in EPA's Chemical and Product (CPDat) and Chemical Product Category (CPCat) databases, in EPA Chemical Data Reporting (CDR) results, or in a published meta-analysis of chemical substances in residential dust. Exposure sources are shown here in the middle of the figure, with the assigned value to the right of the source representing the exposure differential score, which semi-quantitatively represents the magnitude of contribution to the exposure differential between children and adults. IG: information gathering flag.

[93] Lehmann GM, LaKind JS, Davis MH, Hines EP, Marchitti SA, Alcala C, Lorber C. (2018). Environmental chemicals in breast milk and formula: Exposure and risk assessment implications. Environmental Health Perspectives 126:96001.

[94] Mitro SD, Dodson RE, Singla V, Adamkiewicz G, Elmi AF, Tilly MK, Zota AR. (2016). Consumer product chemicals in indoor dust: A quantitative meta-analysis of U.S. studies. Environmental Science and Technology 50: 13611-11.

Once it is determined that a chemical substance has available exposure source information, exposure sources are evaluated for the chemical substance. If the exposure source is determined for a substance, the value assigned to the source (represented in Figure 8 to the right of each potential source) is added to the total for that chemical substance. The value associated with each exposure source represents the relative magnitude of potential exposure differential between children and adults. The total for each chemical substance can range from 1 – 18 and is used to assign the susceptible population domain metric based on the ranges specified in Table 5.

Table 5. Criteria used to evaluate the susceptible population exposure domain metric.

Metric	Total Exposure Source Value
0	Chemical substance had no information in the exposure source data sources
1	Chemical substance had information in at least 1 data source but the reported sources were not associated with evidence of potential for higher exposure for children (i.e., not associated with the sources in Figure 8).
2	Chemical had information in at least 1 data source with a combined exposure differential value corresponding to value ≤ to the 50th percentile score for all chemicals (using the workflow in Figure 8)
3	Chemical had information in at least 1 data source with a combined exposure differential value corresponding to value between the 50th and 90th percentiles scores (using the workflow in Figure 8)
4	Chemical had information in at least 1 data source with a combined exposure differential value corresponding to value ≥ the 90th percentile score (using the workflow in Figure 8)
Information Gathering (IG) Flags: predicted data; secondary source data	

Limitations and Longer-Term Options

A limitation of the susceptible population exposure domain is that only children's exposure was included. Future updates may expand this component to include other susceptible populations (e.g., workers, elderly). Further, the metric developed herein was designed to capture information about exposure sources relevant to children, using publicly available exposure-relevant data that has been compiled and curated by ORD. Collection and curation of product and monitoring data by ORD is ongoing, and new data or data streams can be incorporated into this workflow when they reach an acceptable level of quality review. It is likely that the available product and monitoring data used to develop the susceptible population metric may not be representative of the total chemical substance landscape to which children are exposed. For example, the children's

products for which data are available in CPDat may not be representative of all products used by children, and the chemical substances reported on safety data sheet for these products are not necessarily representative of all the chemical substances in those products. Therefore, the lack of a positive for a chemical substance data source does not necessarily imply a global negative (only a negative for these data sources). Additionally, this domain could be further expanded through incorporation of any future ExpoCast models as they are developed, including those that generated consensus exposure predictions for individual cohorts (e.g., children, the elderly).

Persistence and Bioaccumulation Domain

Persistence refers to the tendency of a chemical substance to remain in the environment in its original form, potentially resulting in exposures that last for a long period of time. Bioaccumulation refers to the tendency of a chemical substance to accumulate in biota. Although generally considered separately, these properties overlap to a substantial degree because molecular features that tend to increase chemical persistence, such halogenation and/or steric features that limit microbial degradation, often promote increased bioaccumulation. Chemical bioaccumulation may occur in both terrestrial and aquatic environments; however, bioaccumulation assessments are often focused on the potential for chemical accumulation in fish. This focus reflects the well-known tendency of hydrophobic substances to partition out of the water column and into aquatic biota. Fish may accumulate chemicals directly from water and by consuming contaminated food items.

Persistence Evaluation

The workflow for evaluating persistence is based on ultimate biodegradation, which is defined as complete mineralization resulting in the formation of carbon dioxide, water, mineral salts and biomass and is measured in weeks (Figure 9). If measured biodegradation half-lives are available, the persistence domain metric is derived based on this data. In the absence of measured data, calculated ratings for ultimate biodegradability are obtained from BIOWIN3 (Ultimate Survey Model) in EPI Suite[95].

[95] US EPA. (2019). Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. United States Environmental Protection Agency, Washington, DC, USA.

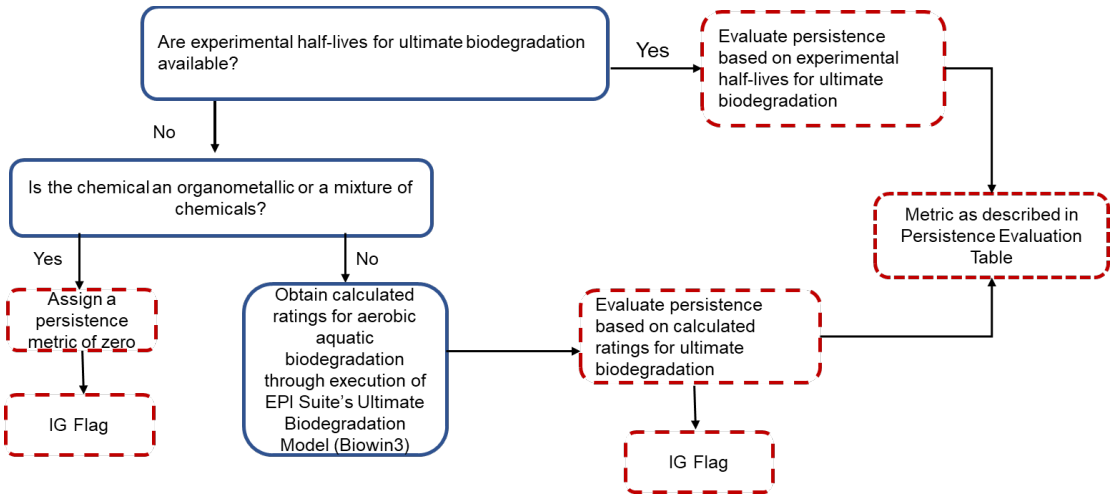


Figure 9. Tiered evaluation process associated with the persistence domain metric. Persistence evaluation is performed by comparing a half-life value for ultimate biodegradation. IG = information gathering.

Persistence evaluation is performed by comparing a half-life value from the workflow for overall persistence to the criteria recommended in the 2012 TSCA Work Plan Chemicals: Methods Document[96] (Table 6).

Table 6. Criteria used to calculate the persistence domain metric

Metric	Experimental Half-Lives or Calculated Rating for Ultimate Biodegradation	Persistence Criteria
0	No data available	
1	<1.75 - 2.25 weeks	$t_{½}$ < 60 days
2	>2.25 - 2.75 weeks	60 days < $t_{½}$ < 180 days
3	>2.75 – 5 weeks	$t_{½}$ > 180 days
Information Gathering (IG) Flags: predicted data; secondary source data		

Bioaccumulation Evaluation

Bioaccumulation is typically evaluated using a steady-state fish bioaccumulation factor (BAF) or bioconcentration factor (BCF). The BAF (chemical substance concentration in fish/chemical substance concentration in water; L/kg) quantifies chemical substance accumulation

[96] U.S EPA. (2012). TSCA Work Plan Chemicals: Methods Document. U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics. https://www.epa.gov/sites/production/files/2014-03/documents/work_plan_methods_document_web_final.pdf

occurring in fish by all possible routes of exposure and is generally determined in field-collected animals. The BCF (chemical substance concentration in fish/chemical substance concentration in water; L/kg) quantifies accumulation occurring in a water-only exposure and is usually measured in controlled laboratory studies. Because BCFs can be determined experimentally, they are often used as a surrogate measure of bioaccumulation potential when measured BAFs are unavailable. However, a measured BAF better represents the potential for chemical substance bioaccumulation in a real-world setting.

A workflow for evaluating chemical substance bioaccumulation potential in fish is shown in Figure 10. In most cases, this workflow leads to a BAF or BCF that can be evaluated against previously defined criteria from the 2012 TSCA Work Plan Chemical Methods Document[97] (Table 7). Possible exceptions include chemical substances that are outside the applicability domain of predictive models (see below). Relative confidence in these BAF and BCF values is represented by flags indicating "high," "intermediate," or "low" confidence. The structure of the workflow represents several general considerations. First, preference is given to chemical substances for which measured BAFs are available. If a measured BAF is not available, chemical substances for which measured BCFs are available receive preference. A database of measured fish BAFs and BCFs was published by Arnot and Gobas[98]. The BCFs assembled by these authors were evaluated for data quality based on aspects of study design and the collection of supporting analytical information. This database did not provide criteria for evaluating fish BAFs; however, methods used to measure reported BAFs were compared to existing guidance for analysis of environmental samples. For the purposes of the bioaccumulation domain metric, a "high quality BCF" refers to a BCF that was scored 1 or 2 in the Arnot-Gobas database across the entire set of evaluation criteria (labeled "acceptable confidence" by the authors), while a "high quality BAF" refers to a BAF that was judged by the same authors to be of "acceptable quality." Either of these bioaccumulation metrics can be used as the basis for evaluating bioaccumulation potential, and the resulting values are flagged as "high confidence."

If a measured BAF or BCF is not available, the chemical substance is passed on to predictive BCF modeling. The workflow assumes that all structures are neutral at environmental pH values. In most cases, this represents a conservative assumption; that is, BCFs predicted under this assumption are likely to be higher than the actual BCF values associated with these compounds. However, exceptions to this general rule are known to exist.

---

[97] U.S EPA. (2012). TSCA Work Plan Chemical Methods Document. U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics. https://www.epa.gov/sites/production/files/2014-03/documents/work_plan_methods_document_web_final.pdf

[98] Arnot JA, Gobas, FA. (2006). A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms. Environmental Reviews 14:257-297.

The two models used to predict bioconcentration of neutral chemical substances are the Arnot-Gobas QSAR model[99] (henceforth "Arnot-Gobas model" without citation) included in the BCFBAF module of EPI Suite ver. 4.11, and the OPERA BCF QSAR model[3] (henceforth "OPERA BCF model" without citation) developed by the EPA ORD. The Arnot-Gobas model is a one-compartment mass-balance description that predicts bioconcentration from competing rates of uptake and elimination, while the OPERA BCF model employs a *k*-nearest neighbor approach to calculate a BCF from measured values for chemical substances that exhibit molecular similarity to the chemical substance. Before using either model, the chemical substance is evaluated using the KOWWIN model in EPI Suite to determine whether it possesses a predicted log K_{OW} value > 9. If the predicted log K_{OW} is > 9, the chemical substance is flagged as "low confidence." This designation reflects uncertainty in both the log K_{OW} estimate and the modeled bioconcentration prediction.

The workflow is structured so that the Arnot-Gobas model is implemented first. This model provides several BCF metrics. For this workflow, we focused on BCF prediction for lower trophic level fish, assuming biotransformation. This focus reflects the fact that most standardized in vivo BCF tests are performed using small fish species or juveniles of larger species. If the Arnot-Gobas returns a value, the chemical substance is passed on for evaluation of bioaccumulation potential. If not, the chemical substance is passed to the OPERA BCF model.

If the OPERA model does not return a value, the process terminates with an IG flag, which indicates that there are no data (measured or predicted) available. This outcome is anticipated for some inorganic chemical substances, metals, organometallic chemical substances and mixtures. If OPERA returns a value, a determination is made whether the chemical substance falls within the model's applicability domain (AD). A chemical substance that falls within the AD is passed on for evaluation. A chemical substance that falls outside the AD is flagged as "low confidence" and then passed on for evaluation. Any chemical substance for which the evaluation is based on a predicted BCF is flagged "medium confidence" unless it has been flagged "low confidence" at some earlier step (e.g., because it possesses a log K_{OW} value > 9). This designation reflects the fact that BCF prediction models have been developed and calibrated using data for a relatively small number of industrial chemical substances (< 1000). In addition, predicted BCFs do not account for potential food web effects (i.e., biomagnification).

[99] Arnot JA, Gobas F. (2003). A generic QSAR for assessing the bioaccumulation potential of organic chemical substances in aquatic food webs. QSAR and Combinatorial Science. 22:337-345.

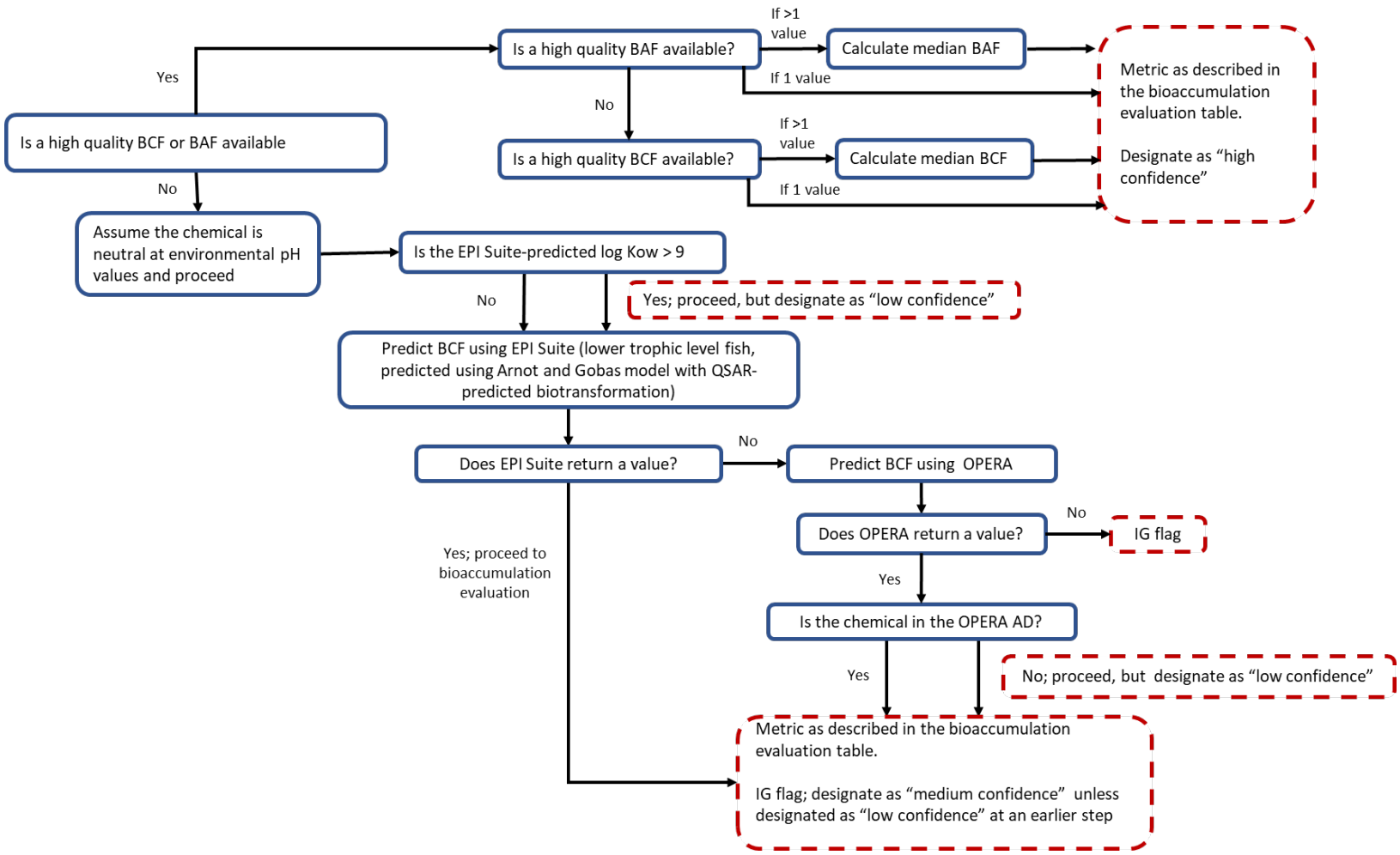


Figure 10. Tiered evaluation process associated with the bioaccumulation domain; high quality BAF or BCF refer to scoring in the Arnot and Gobas database (see text for details). BCF = bioconcentration factor; BAF = bioaccumulation factor; K_{OW} = octanol/water partition coefficient; QSAR = quantitative structure activity relationship; AD = applicability domain; and IG = information gathering flag.

Table 7. Criteria for bioaccumulation domain metric	
Metric	**BCF or BAF[1]**
0	No data available
1	< 1000
2	1000-5000
3	≥ 5000
Information Gathering (IG) Flags: High quality measured BCF or BAF; Predicted BCF, with exceptions noted under "Low confidence"; Predicted log K_{ow} > 9 and/or BCF is predicted by OPERA, but the chemical substance is outside OPERA's applicability domain [1]The BCF and BAF are expressed as the concentration of the test substance in fish (mg/kg)/concentration of the test substance in water (L/kg), resulting in units of L/kg.	

Limitations and Longer-Term Options

EPA is currently in the process of adopting a new approach for evaluating persistence based on the potential half-life in air, water, soil, and sediment. The approach factors the partitioning characteristics of the chemical substances and potential removal pathways based on standard physical-chemical substance properties and environmental fate parameters. Once adopted, this approach may be included in the persistence component of this workflow.

Measured BAFs for some poorly metabolized compounds may exceed measured BCFs by an order of magnitude or more due to biomagnification of chemical residues (i.e., higher concentrations at successively higher trophic levels)[100]. These differences are also apparent in modeled BAF and BCF values. Given this fact, as well as the preference of measured BAFs over measured BCFs noted above, it is reasonable to ask whether preference should be given to predicted BAFs over predicted BCFs. Because predicted BAFs may exceed modeled BCFs, particularly for chemicals of special concern from a bioaccumulation perspective, they represent a more conservative metric of bioaccumulation. Predicted BAFs are well suited, therefore, to performing risk assessments for individual chemicals. The current focus on predicted BCFs was motivated by the fact that the number of published BCFs for fish greatly exceeds the number of published BAFs. The use of a BCF prediction model therefore results in comparable modeled and measured values, which is important if both data sources are used to perform a relative evaluation of bioaccumulation potential for many chemicals. To evaluate this question further, Costanza et al.[101] predicted BCFs and BAFs for 6,034 chemicals using the Arnot-Gobas model, and then

[100] Arnot JA, Gobas, FA. (2006). A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms. Environmental Reviews 14:257-297.
[101] Costanza J, Lynch DG, Boethling RS, Arnot JA. (2012). Use of the bioaccumulation factor to screen chemicals for bioaccumulation potential. Environmental Toxicology and Chemistry 31:2261-2268.

compared these values to criteria used by EPA to screen chemicals for bioaccumulation potential ("not bioaccumulative" = BCF or BAF < 1000; "bioaccumulative" = 1000 < log BCF or log BAF < 5000; "highly bioaccumulative" = BCF or BAF > 5000). The results showed that for 86% of chemicals there was no change in bioaccumulation rating when using the BAF rather than the BCF. This finding suggests that for screening-level assessments, modeled BCFs and BAFs generally lead to the same conclusion.

The inability of current modeling approaches to adequately predict bioaccumulation of ionizable compounds represents a well-recognized research need. Process-based models that describe uptake and accumulation of ionizable compounds in fish have been described by several authors[102,103]. Additional models have been developed specifically for per- and polyfluorinated alkylated substances (PFAS), many of which are > 99% ionized at environmental pH values[104]. Ionizable chemicals represent a particularly difficult challenge in the field of predictive bioaccumulation modeling. Guidance provided in EPI Suite (BCFBAF Help menu, 7.1 Bioconcentration Factor (BCF), 7.1.1 Estimation Methodology) indicates that the Arnot-Gobas model is not intended for use with ionizable compounds. Instead, the guidance recommends using an alternative model which bins chemicals based on their estimated log K_{OW} values (neutral form). BCFs predicted by this model for such compounds tend to be relatively low (maximum log BCF of 1.75). Additional guidance indicates, however, that this model should not be used for compounds possessing specific molecular features. These features include the presence of an aromatic azo group (a structural component of many pigments and dyes), a charged metal species (esp., mercury or tin), or a long chain alkyl group (e.g., many cationic and anionic surfactants). Not given on this list of molecular features is the presence of a fluorine group. Nevertheless, this model appears to be poorly suited for ionizable PFAS compounds, several of which have been shown to accumulate in fish (log BCFs > 3)[105].

The BCF dataset used to train the OPERA BCF model contains a small (< 70) number of ionizable compounds. In principal, these measured BCFs reflect the net result of all processes responsible for bioconcentration including chemical speciation in the water and fish, and uptake and accumulation of all relevant chemical species. Given the small number of chemicals in the

[102] Erickson RJ, McKim JM, Lien GJ, Hoffman AD, Batterman, SL. (2006) Uptake and elimination of ionizable organic chemicals at fish gills. II. Observed and predicted effects of pH, alkalinity, and chemical properties. Environmental Toxicology and Chemistry 25:1522-1532.
[103] Armitage JM, Arnot JA, Wania F, Mackay D. (2013). Development and evaluation of a mechanistic bioconcentration model for ionogenic organic chemicals in fish. Environmental Toxicology and Chemistry 32:115-128.
[104] Ng CA, Hungerbuhler K. (2013). Bioconcentration of perfluorinated alkyl acids: How important is specific binding? Environmental Science and Technology 47:7214-7223.
[105] Martin JW, Mabury SA, Solomon KR, Muir DCG. (2003). Bioconcentration and tissue distribution of perfluorinated acids in rainbow trout (Oncorhynchus mykiss). Environmental Toxicology and Chemistry 22:196-204.

training set, however, there is a relatively high likelihood that a given ionizable chemical will fall outside the applicability domain of the QSAR. Moreover, as indicated by the exclusionary criteria given in EPI Suite, some ionizable compounds may require special attention.

Anticipating future acceptance of a process-based model for ionizable compounds, we may consider possible modifications to the decision tree used to evaluate chemical bioaccumulation potential. The first step in a revised tree would be to determine whether a chemical is substantially ionized at environmental (5 − 9) pH values. In a recent review, it was suggested that pH effects on uptake and accumulation of ionizable chemicals by fish are likely to be minor unless the extent of ionization in bulk water exceeds 90%[106]. Of special concern are weak acids and bases for which an accurate estimate of pKa would be required. Ideally, this estimate would be obtained using an open-source software tool that has been evaluated against measured data as well as existing proprietary software. An OPERA pKa model may provide such a tool[107]. If exclusionary criteria are still required, they would be applied at an early stage in the decision process. For some chemical classes (e.g., PFAS), the decision tree may direct the user to employ a model specifically designed for such compounds. BCFs predicted by a general model for ionizable compounds could be handled in a manner analogous to that for BCFs presently generated by the Arnot-Gobas model. Assuming further that the OPERA BCF model will be updated and trained using newly available data for ionized chemicals, we may anticipate a situation wherein BCFs can be predicted by a process-based model and by the OPERA BCF model. This would require some type of process for averaging these predictions or choosing one in preference to the other. Presently, due to a lack of data for model calibration and evaluation, these models cannot be applied with confidence to the broad range of ionizable structures contained on the TSCA inventory[108]. It is anticipated that these or similar models can be incorporated into future bioaccumulation evaluation efforts.

Combined Persistence and Bioaccumulation Domain Metric

The combined persistence and bioaccumulation domain metric (Table 8) is obtained by adding the separate metrics from the persistence and bioaccumulation workflows (Figures 9 and 10) and

[106] Armitage JM, Erickson RJ, Luckenbach T, Ng CA, Prosser RS, Arnot JA, Schirmer K, Nichols JW. (2017). Assessing the bioaccumulation potential of ionizable organic compounds: current knowledge and research priorities. Environmental Toxicology and Chemistry 36(4):882-897.
[107] Mansouri K, Cariello, NF, Korotcov A, Tkachenko V, Grulke CM, Sprankle CS, Allen D, Casey WM, Kleinstreuer NC, Williams AJ. (2019) Open-source QSAR models for pKa prediction using multiple machine learning approaches. Journal of Cheminformatics 11:60.
[108] Franco A, Ferranti A, Davidsen C, Trapp S. (2010) An unexpected challenge: Ionizable compounds in the REACH chemical space. International Journal of Life Cycle Assessment. 15:321-325.

described in Tables 6 and 7. This process is consistent with the method recommended in the 2012 TSCA Work Plan Chemical Methods document[109].

Table 8. Combined persistence/bioaccumulation domain metric

Metric	Combined Score	
0	0	No data available
1	1 - 2	Low
2	3 – 4	Moderate
3	5	High
4	6	Very High

Skin Sensitization and Skin/Eye Irritation Domain

Skin sensitization and skin/eye irritation are important potential hazards of chemical substances that are of concern for human health. These are addressed here in a separate domain due to different routes of exposure and different data sources for these endpoints. Ocular and dermal exposures can occur through a variety of sources, particularly occupational exposures as well as consumer exposures

Local effects are changes at the site of contact (skin, eye, mucous membrane/gastrointestinal tract, or mucous membrane/respiratory tract) as a result of exposure to a chemical substance. Such changes after a single exposure may be categorized as irritant or corrosive, depending on the severity and reversibility of the outcomes. Corrosive substances are those which may destroy living tissues with which they come into contact. Irritant substances are non-corrosive substances which, through immediate contact with the tissue may cause inflammation.

Skin sensitization denotes the immune-mediated hazards associated with human allergic contact dermatitis and/or rodent contact hypersensitivity. Allergic contact dermatitis is the clinical term that indicates the presence of skin erythema and edema that result from delayed type IV cell-mediated skin hypersensitivity.

[109] U.S EPA. (2012) TSCA Work Plan Chemicals: Methods Document. U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics. https://www.epa.gov/sites/production/files/2014-03/documents/work_plan_methods_document_web_final.pdf

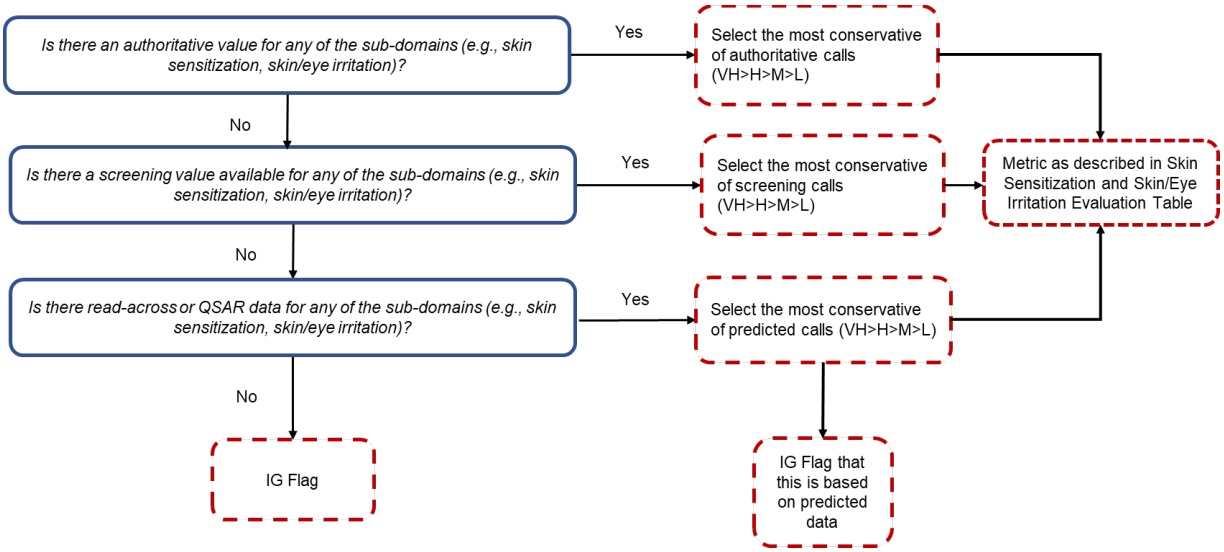


Figure 11. Tiered evaluation process associated with the skin sensitization, skin/eye irritation domain metric. IG = information gathering flag; L = low; M = medium; H = high; VH = very high.

Skin Sensitization and Skin/Eye Irritation Evaluation

The proposed skin sensitization and skin/eye irritation domain metric incorporates GHS hazard codes or hazard categories from the ECHA Classification, Labelling and Packaging (CLP) Regulation and agencies of several countries (e.g., Canada, Japan, Denmark), as well as study-level REACH registration data from ECHA via OECD's eChemPortal. GHS classification and labeling information was extracted from websites of the environmental or occupational health agencies of individual countries. The study-level data in eChemPortal were obtained by searching for the endpoints of eye irritation (*in vivo* and *in vitro*), skin corrosion (*in vitro*), skin irritation (*in vivo*), and skin sensitization (*in vivo* local lymph node assay (LLNA), *in vivo* non-LLNA and *in vitro*).

The GHS classifications were converted to a 4-level ranking of Low (L), Moderate (M), High (H), and Very High (VH), which was then converted to a numerical scale of 1-4 (L=1, VH=4). The ECHA experimental data were converted to the same scale using a mapping from the result summary sentences provided by each study. An expert-derived dictionary was created that mapped each unique result sentence to a scoring level. The criteria for determining metrics are shown in Table 9 below. These criteria were based on the EPA's Design for the Environment Program (DfE) Alternatives Assessment Criteria for Hazard Evaluation[110,111].

[110] US EPA (2011) Design for the Environment Program Alternatives Assessment Criteria for Hazard Evaluation Version 2.0. https://www.epa.gov/saferchoice/alternatives-assessment-criteria-hazard-evaluation. Accessed 09/24/18
[111] Vegosen L, Martin TM. (2020). An Automated Framework for Compiling and Integrating Chemical Hazard Data. Clean Technologies and Environmental Policy 22(2):441–58.

For each associated sub-domain (skin sensitization, skin irritation, eye irritation) a chemical substance could have one or more hazard determinations from different sources. The evaluation method used for combining these values into an overall metric for each sub-domain is shown in Figure 11. Briefly, the evaluation was based on the source of the information. Similar to the approach used by the Clean Production Action's GreenScreen List Translator[112], this evaluation method selects the highest value (highest hazard level) from the most authoritative source as the final output. In the method that was presently implemented (Figure 11), authoritative sources take precedence over screening sources, which take precedence over QSAR models. Within each of those three levels, the source that produces the highest value takes precedence. Using this method, an overall value across all sources was determined for each sub-domain. Then, the most conservative of the three sub-domain values was used as the final value for the skin/eye domain.

Table 9a. Criteria for Skin Sensitization Sub-Domain Metric

Metric	Description	Classification	GHS Code	ECHA (eChemPortal)
0	No data	No Data Available	--	--
1	Low	Not Likely to be Sensitizing	Not Classified	Not sensitizing, Not Classified
2	Moderate	Low to Moderate Frequency of Sensitization		Category 1B, Moderate Sensitizer, Mild Sensitizer, Weak Sensitizer
3	High	High Frequency of Sensitization	H317, SkinSens1, Sah/Sh	Category 1A, Sensitizing

Table 9b. Criteria for Skin Irritation Sub-Domain Metric

Metric	Description	Classification	GHS Code	ECHA (eChemPortal)
0	No data	No Data Available	--	--
1	Low	Studies Indicate No Significant Irritation	Not Classified	Not Irritating, Not Classified
2	Moderate	Moderate or Mild Irritation	H316, 6.3B	Category 3, Moderately Irritating, Mildly Irritating, Slightly Irritating
3	High	Severe Irritation	H315, SkinIrr2, 6.3A	Category 2, Highly Irritating, Irritating

[112] Clean Production Action (2018). Greenscreen for Safer Chemicals Hazard Assessment Guidance version 1.4. https://greenscreenchemicals.org/method/full-greenscreen-method

4	Very High	Corrosive	H314, 8.2A, 8.2B, 8.2C	Category 1, Corrosive

Table 9c. Criteria for Eye Irritation Sub-Domain Metric

Metric	Description	Classification	GHS Code	ECHA (eChemPortal)
0	No data	No Data Available	--	--
1	Low	Studies Indicate No Significant Irritation	Not Classified	Not Irritating, Not Classified
2	Moderate	Moderate or Mild Irritation	H320	Category 2B, Moderately Irritating, Mildly Irritating, Slightly Irritating
3	High	Severe Irritation	H319, Category 6.4A	Category 2A, Severely Irritating, Highly Irritating, Irritating
4	Very High	Corrosive or Irritation Persists for > 21 days	H314, H318, Category 8.3A	Category 1, Corrosive, Serious Eye Damage
IG Flag: No information for sub-domain				

Limitations and Longer-term Options

A limitation of the skin sensitization and skin/eye irritation evaluation was the availability of data sources, particularly a lack of large databases of validated non-animal data. Thus, the data sources that were used have some limitations. While there are non-animal test methods to assess both skin sensitization and skin/eye irritation and corrosion, the data for large numbers of chemical substances remains limited. Moreover, the ability of alternative eye irritation assessment methods to discriminate between GHS categories remains a limitation. The source of the ECHA data in eChemPortal is industry-submitted REACH registration dossiers. REACH requires that at least 5% of the registration dossiers of each tonnage band of chemical substances must be checked for compliance with legal requirements for chemical substance identity descriptions and safety information[113]. Because up to 95% of REACH registration dossiers may not be checked for compliance, the quality assurance of the data from these dossiers is limited. The majority of chemical substances (60-80%) had information from only the ECHA REACH dossiers or a GHS source but not both. For chemical substances that had data available from the REACH dossiers and another source, the results from the REACH dossiers often were consistent with the results from the other source. However, there were instances in which a chemical substance had a value of VH (4) based on the GHS classification data, but the ECHA data indicated a value of L (1). The

[113] ECHA, REACH compliance checks: https://echa.europa.eu/regulations/reach/evaluation/compliance-checks

GHS classifications were generally more severe on average than those from ECHA, which might be partially attributable to GHS classification being more likely to be conducted when there is prior cause for concern. The GHS categorization of "Not Classified", which indicates that a chemical substance does not meet the requirements to be classified as hazardous under the GHS, was reported by Japan's National Institute of Technology and Evaluation (NITE) but was not reported by the other sources of GHS data. An additional limitation relates to the lack of reporting of GHS classifications for chemicals which do not meet the classification criteria. For this reason, lack of a classification is ambiguous, meaning either no data or not meeting criteria. For the purpose of the PICS approach, "not classifiable" or "classification not possible" were interpreted as no data (or insufficient data or ambiguous). In contrast, we interpreted "not classified" as not meeting GHS criteria for being classified as hazardous. Further in-depth review would be considered part of the expert review for compounds of interest that may follow the application of the PICS approach, which would involve evaluating the mapping of summary sentences from REACH dossiers. The method of determining an overall value for each endpoint by selecting the most hazardous value from the most authoritative source reduces the influence of differences in data quality between different data sources, but the inability to check all primary sources for data quality remains a limitation. Potential longer-term improvements include additional quality assurance checks as well as the inclusion of additional data sources if such sources become available. Furthermore, new QSAR models are being developed to fill in missing data for these endpoints.

5.3 Scientific Domain Metric Calculation

The overall scientific domain output is calculated by summing the metrics from each of the 7 domains. Any individual domain metric of zero is given a value of 1, which helps to normalize the values across chemical substances and domains. The lack of data is captured in the IAM and is visualized in Figure 14 based on the size of the point representing the chemical. The maximum possible output is 27 and the minimum output is 7. The summed results are scaled to values from 0 to 100 to match the IAM. Therefore, the minimum value of 7 is converted to zero, with the maximum value of 27 equal to 100.

5.4. Information Availability Metric

The second dimension of the PICS approach is a metric that represents the information available for use in any future chemical substance risk evaluation. Under TSCA, there is no minimum data requirement necessary to perform a chemical substance risk evaluation, as decisions about what would be considered a sufficient amount of hazard or exposure data are typically context specific and would require expert judgment to determine. While this would be possible during the formal prioritization and risk evaluation processes, expert judgment is not part of the

automated approach described here. The IAM is designed to automatically evaluate chemical substances based on the number and type of studies available to inform this analysis. To partially address the context-specific aspect of the data, this metric includes a relatively simple set of four modifying criteria of potentially relevant exposure, human health and ecological toxicity information. The criteria include a combination of primary use as a chemical substance intermediate, environmental half-life, water solubility, molecular weight, whether the chemical substance is an exempt polymer and whether the chemical substance has been assessed for human risk by an authoritative source (Figure 12). Following application of the criteria, the IAM is calculated as a function of information in the associated lists. Missing information is flagged for potential future information gathering, but these IG flags do not directly impact the IAM and only identify specific information gaps. The IAM is calculated by giving a chemical substance one point for having experimental data in each of the domains corresponding to the appropriate box in Figure 12. This metric does not take into account the quality or quantity of studies for each chemical. Predicted data is also not incorporated into this metric with the exception of the SEEM3 exposure model.

If the chemical substance has an authoritative human risk assessment from one of these specific sources (IRIS, EFSA, ATSDR, SIDs, OPPT, OPP), it is given a point for each of the 8 human information availability study types (mammalian values plus carcinogenicity, genotoxicity and skin and eye). The output is then calculated by dividing by the total number of domains for the appropriate box and multiplying by 100. A detailed scheme for the calculation is given in Appendix G, Figure G-1.

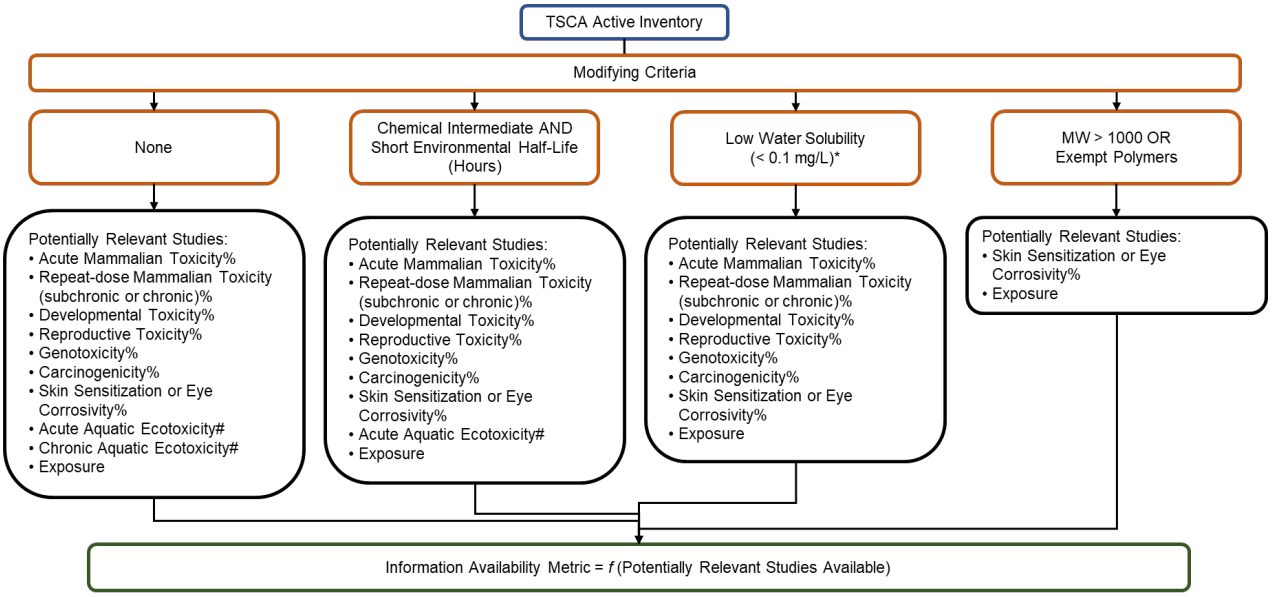


Figure 12. Flow chart for determining the IAM for each chemical substance based on a small number of physicochemical and use criteria for identifying potentially relevant human health and ecological toxicity information. Modifying criteria as shown here are used to inform the set of potentially relevant exposure, human health and ecological toxicity information for specific types of chemical substances.

5.5 Results of the Proof-of-Concept Analysis

Overall Evaluation

The number of chemical substances in the current non-confidential TSCA active inventory is 33,364; however, only 14,017 of these are unique organic chemical substances with defined structures (Table 10). The majority of chemical substances in the inventory are mixtures of varying complexity. Chemical substances that have some *in vivo* mammalian and ecotoxicological data constitute 11-13% of the overall inventory and 3% have experimental cancer data. The data included in the PICS approach is public and excludes industry submitted CBI studies. The PICS approach also does not include data extracted from the literature beyond what is included in the Type 1 data sources currently being utilized.

Table 10. Number of chemical substances in the non-confidential active TSCA inventory with specific types of experimental data.		
Experimental Information	**Number of Chemical substances**	**Percentage**
Human Exposure	14,477	43
Mammalian Repeat Dose Toxicity	4,109	12
Ecological Toxicity (Acute)	3,963	12
Ecological Toxicity (Repeat Dose)	3,466	10
Carcinogenicity	765	2.3
Genotoxicity	3,027	9.1
Skin Sensitization and Skin/Eye Irritation	8,689	26
Total TSCA Active Inventory[1]	33,364	---
[1]The Non-confidential TSCA Active inventory contains 33,364 chemical substances but the study only included the subset that can be mapped to the DSSTox database.		

As noted earlier, the POC238 was also selected to test the PICS approach using a subset of the TSCA active inventory and spanned a range of potential concern and information availability (Fig. 13).

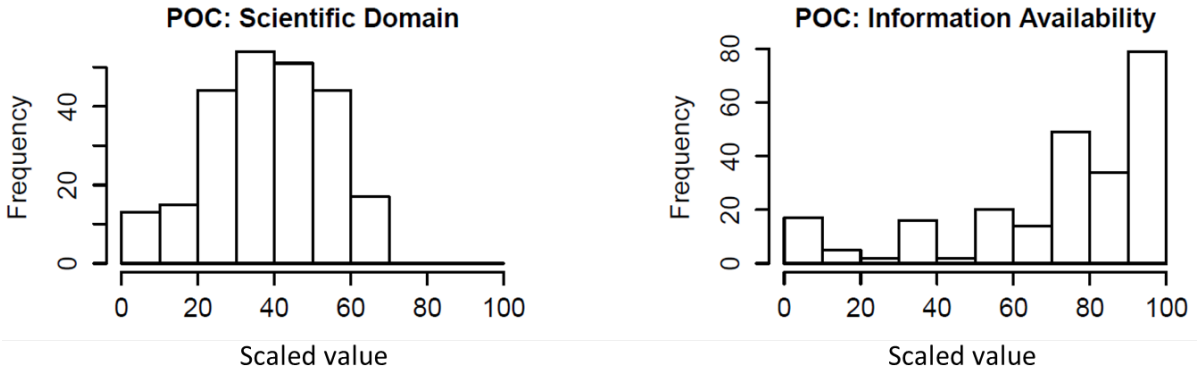


Figure 13. Distributions of the scaled SDM and IAM for the POC238 subset. The x-axis displays the scaled value and the y-axis shows the frequency of that value in the subset. Similar histograms for the TSCA active inventory can be found in Appendix F.

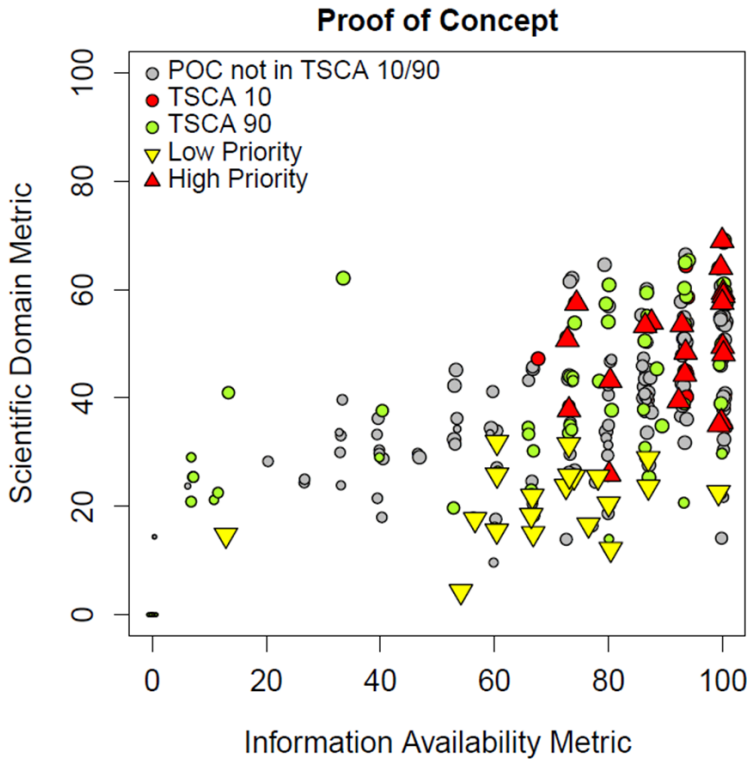


Figure 14. Plot of the Information Availability vs. Scientific Domain Metrics for the POC238 set of chemical substances. Each dot represents one chemical substance, with the size of the dot representing the number of domains with data for the specific chemical. The red dots represent the first ten TSCA Work Plan chemical substances selected for risk evaluation in 2016 (TSCA 10). The green dots represent the TSCA Work Plan chemical substances from the 2014 update (TSCA 90). The red triangles represent the high priority chemical substances and the yellow triangles represent the low priority chemical substances released in March 2019. Positions of points are staggered for ease of visualization. A similar graph for the entire non-confidential TSCA active inventory is found in Appendix F.

The two-dimensional representation of the SDM and IAMs can be summarized for the POC238 (Figure 14; results used to inform this figure can be found in Appendix E). There is an association between IAM and SDM (i.e., more information tends to produce a higher value). This may be a result of potential testing or publication bias. Chemical substances that are expected to show or have previously shown indications of potential hazard will lead to more data being generated, while those that are not expected to show high hazard are less likely to be tested. Additionally, there is a publication bias towards positive results as most peer-reviewed publications do not describe negative results. Further, a lack of available data does not indicate a lack of toxicity.

Using the recently released TSCA high and low priority chemical substance candidates selected by EPA[114], the high priority candidates generally having higher metrics than the low priority candidates when analyzed by the PICS approach (Fig 14). In part, these candidates were selected by expert reviewers examining both publicly available and CBI data for each chemical substance using a systematic review process[115], which takes into consideration study quality and consistency in the database. Further, this review would take into consideration various policy aspects and scientific judgment that are not part of the automated PICS approach. Discrepancies between the conclusions of expert reviewers and the results of the PICS approach may be related to the different data sources used, but also may be related to the more in-depth review of the studies used as a basis for the candidate selection. For example, in some cases, the conservative decisions used in the genotoxicity domain of the PICS approach (i.e., assigning a positive genotoxicity score in the presence of one positive study regardless of the results of other studies) may give a chemical substance a higher domain metric than a weight of evidence analysis as the latter would take into account the full dataset.

[114] https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/list-chemical-undergoing-prioritization
[115] https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/application-systematic-review-tsca-risk-evaluations

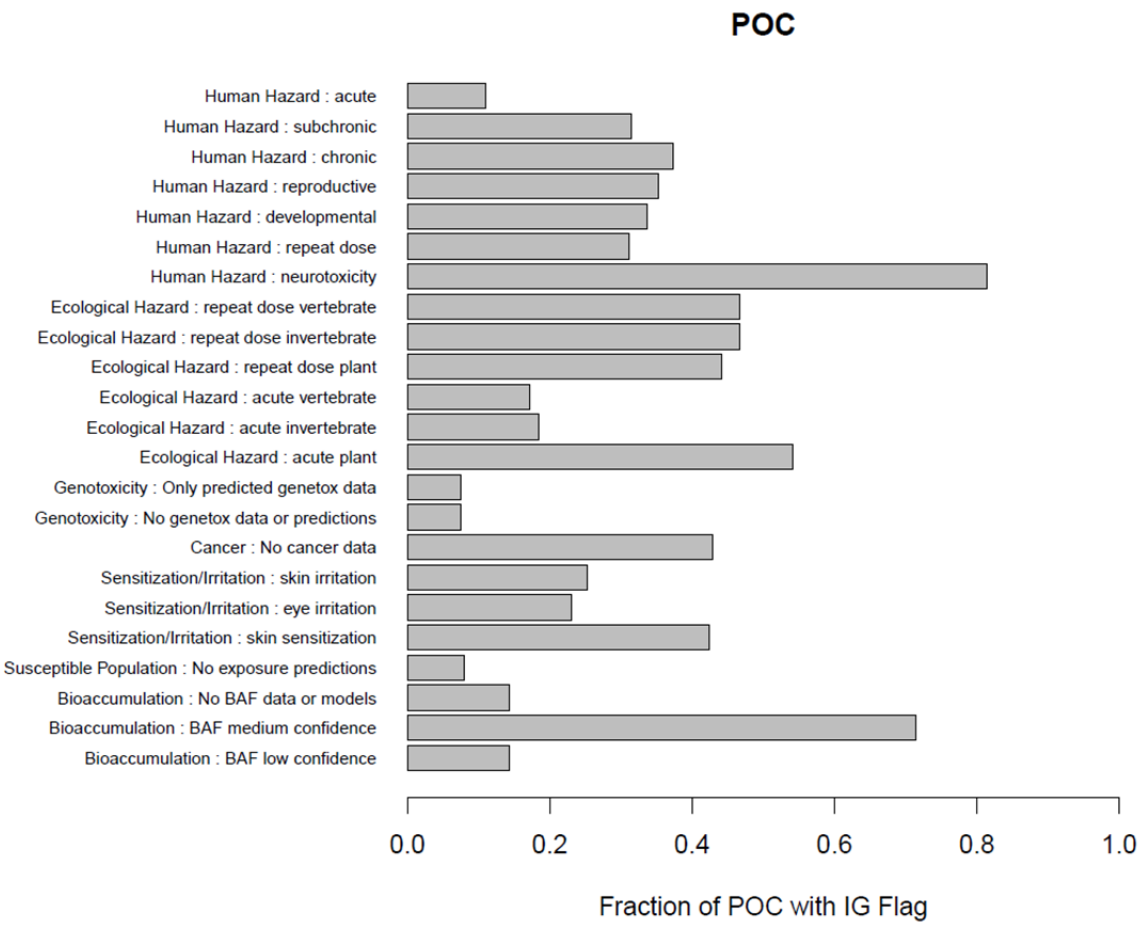


Figure 15. Plot of the frequency distribution of IG flags for each SDMs for the POC238 set of chemical substances. IG flags are designed to highlight data types used in specific SDMs as well as possible data gaps. IG = information gathering.

In addition to the TSCA high and low priority candidates, the POC238 also includes a selection of chemical substances from other lists that had a prior expectation of higher or lower potential concern. For instance, the chemical substances from the 2014 TSCA Work Plan[116] are generally expected to be of higher than average concern with an existing authoritative hazard assessment. The chemical substances on the SCIL list or chemical substances that are intentional food ingredients are expected to be of lower than average concern. Figure 16 summarizes the metric distributions for selected chemical substance lists from across the full TSCA active inventory. From this plot we can see that the SDM values are largely consistent with expectations. The TSCA high priority candidates and the 2014 TSCA Work Plan chemical substances are relatively high while the TSCA low priority candidates, SCIL, and food ingredients are relatively low. The median scientific domain metr6ic for the full TSCA active inventory is very low, but this generally reflects the overall low information availability. Comparatively low information

[116] https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/tsca-work-plan-chemicals

availability is also seen in the SCIL and food ingredient lists. The POC238 list is enriched in the high priority regulatory chemical substances, and the remaining chemical substances were largely selected because of knowledge of some toxicological concern. As a result, the POC238 has a distribution similar to the high concern lists and is not reflective of the overall TSCA active inventory.

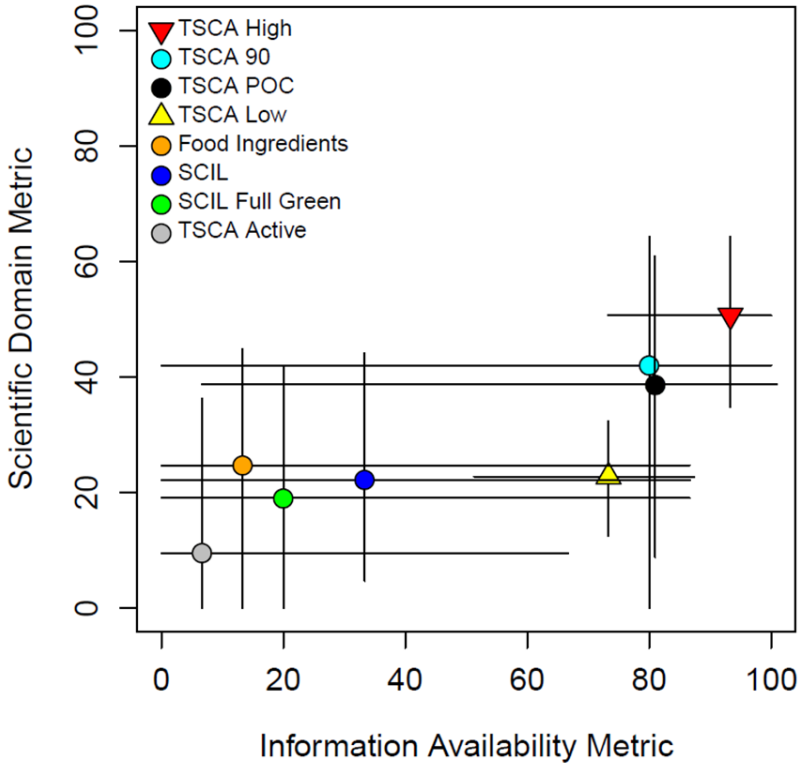


Figure 16. Plot showing distributions of metric scores for selected chemical substance lists. For each list, the point shows the median scientific domain and IAMs. The whiskers span 90% of the distributions. Data here are taken from the lists across the non-confidential TSCA active inventory. TSCA High = high priority candidates; TSCA 90 = chemical substances from 2014 TSCA Work Plan; TSCA POC = 238 chemical substances from TSCA POC; TSCA Low = low priority candidates; Food Ingredients = chemical substances from the FDA food ingredients list; SCIL = Safer Choice Ingredients List; SCIL Full Green = SCIL labeled low concern based on experimental and modeled data; TSCA Active = nonconfidential TSCA active inventory.

To illustrate how the process works for individual chemical substances, we show information for two chemical substances, one with a relatively high value for the SDM (benzene, 68) and one with a relatively low value (3-methoxybutyl acetate, 14.7). Both chemical substances have relatively high values for the IAM (benzene = 93% and 3-methoxybutyl acetate = 67%). Table 11 shows values for individual domains. Following the overall metric scores are the

information gathering flags that indicate the types of *in vivo* data that are lacking and information about the BAF values used. There were no IG flags for the other domains. Next are the seven individual domain metrics, whose sum could range from 7 to 27. As described above, the overall score is equal to 100 x (sum-7)/(27-7). Benzene has a 2.5 (out of 4) for the human hazard-to-exposure ratio metric, based on the HER value. 3-Methoxybutyl acetate has a 2.3 of out 4 for the human hazard-to-exposure ratio metric.

The human health repeat dose POD for benzene is 0.015 mg/kg-bw/day, which is the chronic NOAEL from an authoritative human hazard assessment (ATSDR / CDC). The corresponding value for 3-methoxybutyl acetate is 100 mg/kg-bw/day, which is a NOAEL from an ECHA repeat dose study in guinea pigs. For the ecological hazard metric, the minimum PODs for the two chemical substances are very similar (0.71 and 0.49 mg/L), which leads to the same ecological hazard metric value. The minimum ecological POD for benzene is 0.49 mg/L, derived from an acute study (96 hours) with POD of a 4.9 mg/L (divided by 10 in our process) in sockeye salmon [ECHA / eChemPortal[117]. For 3-methoxybutyl acetate, the minimum ecological POD is 7.1 mg/L in an acute zebrafish study from ECHA / eChemPortal. This POD was then divided by 10 using the acute-to-repeat dose factor. In the OCSPP evaluation of this chemical, the zebrafish study was disregarded because it did not meet the minimum quality criteria. OCSPP identified a repeat-dose study with a NOAEL of 74 mg/L from a source not included in the current Type 1 data sources.

Benzene has a maximum value for the cancer metric because it is classified as a Group 1 human carcinogen by IARC, while 3-methoxybutyl acetate had no cancer-related information. For genotoxicity, benzene is classified as genotoxic and 3-methoxybutyl acetate is classified as non-genotoxic, based on a single negative Ames assay. Benzene has 27 total assays in the genotoxicity database, including 2 positive Ames tests, leading to the positive overall classification. However, there are also 3 negative and 2 ambiguous Ames results. Both chemical substances had elevated values for the susceptible population metric (4 for benzene, 3 for 3-methoxybutyl acetate), indicating that there is a high probability that children may be exposed to these chemical substances. Benzene has a moderate metric for persistence / bioaccumulation based on a persistence score from EpiSuite of 2.4, while 3-methoxylbutyl acetate has a low metric for persistence/bioaccumulation based on low persistence and bioaccumulation values. Benzene has a sensitization / irritation score of 3 based on High scores for both skin and eye irritation. These are based on GHS classifications, which were consistent between ECHA, New Zealand, Canada, Malaysia and Japan. 3-Methoxybutyl acetate has low scores for skin and eye irritation.

[117] Black JA, Birge WJ, McDonnell WE, Westerman AG, Ramey BA, Bruser DM. (1982). The Aquatic Toxicity of Organic Compounds to Embryo-Larval Stages of Fish and Amphibians. Research Report No.133, Water Resources Research Institute, University of Kentucky, Lexington, KY.

Table 11. **Results for benzene and 3-methoxybutyl acetate.**		
CASRN	4435-53-4	71-43-2
Name	3-Methoxybutyl acetate	Benzene
Scientific Domain Metric	14.6	67.9
Information Availability Metric	67	93
IG flag human hazard	Mammalian *in vivo* hazard data missing: subchronic, chronic	Mammalian *in vivo* hazard data missing: developmental
IG flag ecological hazard	Eco *in vivo* hazard data missing: acute plant, repeat dose invertebrate, repeat dose vertebrate	Eco *in vivo* hazard data missing: acute plant
IG flag BAF	BAF medium confidence (modeled value)	BAF medium confidence (modeled value)
Human hazard-to-exposure ratio metric	2.3	2.5
Ecological hazard metric	1.8	1.8
Carcinogenicity metric	0 (no data)	4
Genotoxicity metric	1	4
Susceptible population metric	2	4
Persistence bioaccumulation metric	1	2
Sensitization / irritation metric	1	3
HER repeat dose	1,325,3000	909,925
POD *in vivo* oral repeat dose	100 mg/kg-day	0.015 mg/kg-day
Human exposure (SEEM3)	0.0000075 mg/kg-day	0.0000013 mg/kg-day
Ecological min POD	0.71 mg/L[1]	0.49 mg/L
Bioaccumulation EpiSuite	1.3	8.9
Bioaccumulation OPERA	2.4	7.1
Persistence EpiSuite	3.0	2.4
Persistence OPERA	4.6	10.3
Genotoxicity call	non-genotoxic	genotoxic
Carcinogenicity call		Group I: carcinogenic to humans
Skin sensitization metric		L
Eye irritation metric	L	H
Skin irritation metric	L	H
Volatile	No	Yes
Water soluble	Yes	Yes

[1] The minimum ecological POD is 7.1 mg/L from an acute toxicity study, which was then converted to a value of 0.71 using the acute-to-repeat dose factor. OCSPP did not use this value in their evaluation because the study did not meet their minimum quality criteria. IG = information gathering flag.

Each of the workflows also includes the use of IG flags to identify missing experimental information, even in the case where predicted values could be used (Figure 15). For example, there are a large number of study types that may be used in calculating the human hazard-to-exposure domain metric. However, even if one or more such values are available, the schema depicted in Figure 12 includes flags for study types that are missing. For instance, Figure 15 shows that a large fraction (0.85) of POC238 chemical substances are missing neurotoxicity data, although most had at least one other acceptable mammalian study to be used in calculating an HER value. Similarly, for the ecological hazard domain, this figure shows that that most of the POC were lacking an acute plant study, although in most cases there was still at least one *in vivo* study that could be used to provide an ecotoxicological POD.

5.6 Overall Limitations and Long-term Options

The PICS approach described here was designed under consideration for use in support of TSCA. However, this approach was designed to be adaptable to other decision contexts. The main limitations in adapting the PICS approach is the availability of state of the science methods and access to curated datasets. Addressing these areas would allow for the incorporation of other specific endpoints (e.g., reproductive or developmental toxicity), pathways (e.g., estrogenic), and additional data sources. As the science progresses, changes to this approach could also address applicability to data poor compounds by increasing focus on the use of NAMs to help fill specific data gaps. Future advances across the scientific domains, including the development and incorporation of additional NAMs (e.g., *in chemico* and alternative species models) which could also aid in incorporating specific chemical classes that are not easily addressed with the methods in the current PICS approach (e.g., volatile chemicals). The adaptability of the approach also applies to how the impact of specific domains may be adjusted. Depending on the decision context, the user may want to weight some scientific domains or data sources differently than we have done in this proof-of-concept case study in order to focus on scientific endpoints of specific concern for that decision context. Alternatively, the decisionmaker may want to incorporate additional IG flags to highlight aspects important to that decision context (e.g., IG flag for GHS classification for carcinogenicity). Longer term efforts could also help to address how this work could be applied to mixtures, although more research is needed to determine how best to address this issue. As noted earlier in the document, a limitation for this case study is the focus of the susceptible exposure domain only on children's exposure. However, if data sources are available, they could be incorporated to include additional populations (e.g., workers, elderly) as appropriate for future applications. As with the hazard domains, as the research and data evolve, additional populations can be incorporated as appropriate for the decision context.

6. Summary

Historical approaches that search, compile, and manually evaluate relevant information would be very time and resource intensive to implement for all ~33,000 chemical substances in the non-confidential, active TSCA inventory. The EPA developed the PICS approach to integrate information from a variety of sources to better understand the landscape of publicly available information for these chemical substances. The PICS approach uses a large information management and technology infrastructure to synthesize traditional and NAM information in key scientific domains including human health hazard-to-exposure ratio, ecological hazard, carcinogenicity, genotoxicity, exposure to susceptible populations, persistence/bioaccumulation, and skin sensitization and skin/eye irritation. The output is a display of the chemical substances from the TSCA active inventory that reflects the overall degree of potential concern related to human health and the environment and the relative coverage of potentially relevant human health and ecological toxicity and exposure information. Behind this visual display is a quantitative summary of the individual domain metrics. This information could aid in determining the level of effort and resources that may be needed to evaluate specific chemical substances together with flags to identify potential information needs.

A proof-of-concept case study was performed by applying the PICS approach to a subset of the TSCA active inventory. The design of the scientific domain workflows was an iterative process using the results for chemical substances of known/expected hazard or exposure. For example, the results of the analyses for chemical substances with previous genotoxicity assessments helped to refine the genotoxicity domain workflow and determine where and why the workflow may vary from past assessments. The results of this case study showed that the overall SDM was generally correlated with the IAM, suggesting potential testing or reporting bias. However, the PICS approach was able to segregate the recently released TSCA high- and low priority candidate chemical substances, with some differences related to important aspects of expert review. Expert review would include data and study quality analysis, which may lead to removal of some studies or endpoints included in the PICS approach. Further, expert review would include a weight-of-evidence analysis and take into account the breadth of the available data unlike the PICS approach that focuses on selecting the more conservative result in order to limit the number of false negatives.

Apart from the TSCA high and low priority candidate chemical substances, most of the remaining chemical substances from the 2014 TSCA Work Plan were juxtaposed with TSCA high priority candidates. The chemical substances from the 2014 TSCA Work Plan were expected to have both a high SDM and IAM due to the rigorous selection process lead up to the Work Plan. However, a small subset had limited information availability suggesting that the Type 1 data sources may not capture all of the information sources utilized in the selection process. The

POC238 also included chemical substances from the SCIL and intentional food ingredients lists. The PICS approach generally resulted in these chemical substances having a lower SDM, with some exceptions related to the conservative approach that may be addressed during a more systematic review (e.g., study quality). However, the chemical substances from the SCIL and intentional food ingredients lists also had lower than expected IAMs suggesting either missing information sources or the information collected on these chemicals may be targeted towards the specific uses and exposures.

As described above, the PICS approach has caveats and limitations. To accelerate the process of integrating publicly available data for a large number of chemical substances, the evaluation of the Scientific Domain and IAMs are performed using an automated process that may not account for all potential exceptions or contexts that may occur for a specific chemical substance or chemical substance group. The PICS approach relies on a large database of chemical substance properties, hazard, exposure, persistence, and bioaccumulation information that have been integrated from multiple publicly available sources and models. As the databases and methodologies are updated, the PICS approach can be applied again to update the results based on the latest available information. Although efforts have been taken to ensure the accuracy of the information, the database may contain errors propagated from the source databases. The cleaning and curation of the information will be an ongoing process and require significant resource investment to iteratively improve and develop new systems that avoid regeneration of legacy data. In many cases, data used in this analysis were not able to be verified back to primary source information. Data points that were verified from authoritative secondary sources were flagged with an information gathering flag and individual study quality was not considered. The quality control effort relied on the acceptance of data and information from authoritative sources. Finally, the domain workflows were designed to select the more conservative options unless otherwise stated; this likely results in a higher incidence of false positives, activities reported at lower doses, and exposures reported at higher doses. This was done to create the most comprehensive group of potential candidates for prioritization with the potential false positives identified in the subsequent expert review phase.

7. Conclusion

The EPA has developed the PICS approach to integrate information from a variety of sources to better understand the landscape of publicly available information for these chemical substances. The automated approach provides a systematic and reproducible process for integrating available information and identifying potential information gaps. Over time, the PICS approach will increase efficiency and workload management by focusing expert review on substances that may have a greater potential for selection as high- or low priority potential

candidates. The domain-specific workflows embedded in the approach can be adapted to scientific advances or the availability of new information to create a flexible and sustainable process. The proof-of-concept study suggests that the PICS approach can help inform chemical prioritization, identify possible data gaps which can inform data needs, and provide other information related to the EPA's TSCA program. The PICS approach is designed in discrete domains and utilizes well-documented data sources which allows flexibility for future adaptation and customization as needed to meet program requirements or needs that might be different to those of the TSCA. In addition, the approach may also be useful in identifying common data gaps across large groups of chemicals, which could facilitate research efficiencies.

Appendix A. Proof-of-Concept (POC) Subset of the Non-confidential TSCA Active Inventory

CASRN	DTXSID	PREFERRED_NAME
156-60-5	DTXSID7024031	(E)-1,2-Dichloroethylene
79-33-4	DTXSID6034689	(S)-2-Hydroxypropionic acid
54464-59-4	DTXSID5052200	1-(1,2,3,4,5,6,7,8-Octahydro-2,3,5,5-tetramethyl-2-naphthyl)ethan-1-one
68155-66-8	DTXSID9052397	1-(1,2,3,5,6,7,8,8a-Octahydro-2,3,8,8-tetramethyl-2-naphthyl)ethan-1-one
68155-67-9	DTXSID6041923	1-(2,3,8,8-Tetramethyl-1,2,3,4,6,7,8,8a-octahydronaphthalen-2-yl)ethanone
79-34-5	DTXSID7021318	1,1,2,2-Tetrachloroethane
79-00-5	DTXSID5021380	1,1,2-Trichloroethane
75-34-3	DTXSID1020437	1,1-Dichloroethane
1163-19-5	DTXSID9020376	1,1'-Oxybis[2,3,4,5,6-pentabromobenzene]
110-98-5	DTXSID7026863	1,1'-Oxybis-2-propanol
96-18-4	DTXSID9021390	1,2,3-Trichloropropane
120-82-1	DTXSID0021965	1,2,4-Trichlorobenzene
3194-55-6	DTXSID4027527	1,2,5,6,9,10-Hexabromocyclododecane
106-93-4	DTXSID3020415	1,2-Dibromoethane
95-50-1	DTXSID6020430	1,2-Dichlorobenzene
107-06-2	DTXSID6020438	1,2-Dichloroethane
78-87-5	DTXSID0020448	1,2-Dichloropropane
6920-22-5	DTXSID40863959	1,2-Hexanediol
57-55-6	DTXSID0021206	1,2-Propylene glycol
106-99-0	DTXSID3020203	1,3-Butadiene
99-65-0	DTXSID9024065	1,3-Dinitrobenzene
102-06-7	DTXSID3025178	1,3-Diphenylguanidine

106-46-7	DTXSID1020431	1,4-Dichlorobenzene
123-91-1	DTXSID4020533	1,4-Dioxane
106-94-5	DTXSID6021874	1-Bromopropane
71-36-3	DTXSID1021740	1-Butanol
88-73-3	DTXSID0020280	1-Chloro-2-nitrobenzene
100-00-5	DTXSID5020281	1-Chloro-4-nitrobenzene
661-19-8	DTXSID4027286	1-Docosanol
629-96-9	DTXSID0027272	1-Eicosanol
36653-82-4	DTXSID4027991	1-Hexadecanol
112-92-5	DTXSID8026935	1-Octadecanol
111-87-5	DTXSID7021940	1-Octanol
118-96-7	DTXSID7024372	2,4,6-Trinitrotoluene
732-26-3	DTXSID2021311	2,4,6-Tris(tert-butyl)phenol
51-28-5	DTXSID0020523	2,4-Dinitrophenol
121-14-2	DTXSID0020529	2,4-Dinitrotoluene
108-31-6	DTXSID7024166	2,5-Furandione
96-29-7	DTXSID1021821	2-Butanone oxime
110-44-1	DTXSID3021277	2E,4E-Hexadienoic acid
183658-27-7	DTXSID9052686	2-Ethylhexyl 2,3,4,5-tetrabromobenzoate
149-30-4	DTXSID1020807	2-Mercaptobenzothiazole
109-86-4	DTXSID5024182	2-Methoxyethanol
78-83-1	DTXSID0021759	2-Methyl-1-propanol
534-52-1	DTXSID1022053	2-Methyl-4,6-dinitrophenol
55583-69-2	DTXSID70873187	2-Methylallyl alcohol ethoxylate
88-74-4	DTXSID1025726	2-Nitroaniline
79-94-7	DTXSID1026081	3,3',5,5'-Tetrabromobisphenol A
91-94-1	DTXSID6020432	3,3'-Dichlorobenzidine

612-83-9	DTXSID1020433	3,3'-Dichlorobenzidine dihydrochloride
591-35-5	DTXSID2025006	3,5-Dichlorophenol
4435-53-4	DTXSID2052106	3-Methoxybutyl acetate
99-08-1	DTXSID5021831	3-Nitrotoluene
140-66-9	DTXSID9022360	4-(1,1,3,3-Tetramethylbutyl)phenol
17540-75-9	DTXSID8029315	4-(Butan-2-yl)-2,6-di-tert-butylphenol
16090-02-1	DTXSID0027777	4,4'-Bis(2-morpholino-4-anilino-s-triazinyl-6-amino)stilbene-2,2'-disulfonic acid disodium salt
101-14-4	DTXSID5020865	4,4'-Methylenebis(2-chloroaniline)
80-51-3	DTXSID7026499	4,4'-Oxybis(benzenesulfohydrazide)
101-80-4	DTXSID0021094	4,4'-Oxydianiline
136-85-6	DTXSID1038743	5-Methyl-1H-benzotriazole
51-52-5	DTXSID5021209	6-Propyl-2-thiouracil
75-07-0	DTXSID5039224	Acetaldehyde
103-90-2	DTXSID2020006	Acetaminophen
79-06-1	DTXSID5020027	Acrylamide
79-10-7	DTXSID0039229	Acrylic acid
107-13-1	DTXSID5020029	Acrylonitrile
3825-26-1	DTXSID8037708	Ammonium perfluorooctanoate
62-53-3	DTXSID8020090	Aniline
NOCAS_872414	DTXSID30872414	Antimony & Antimony Compounds
NOCAS_872415	DTXSID90872415	Arsenic & Arsenic Compounds
137-66-6	DTXSID3041611	Ascorbyl palmitate
50-78-2	DTXSID5020108	Aspirin
1912-24-9	DTXSID9020112	Atrazine
25057-89-0	DTXSID0023901	Bentazone

1302-78-9	DTXSID6030782	Bentonite
71-43-2	DTXSID3039242	Benzene
65-85-0	DTXSID6020143	Benzoic acid
119-61-9	DTXSID0021961	Benzophenone
85-68-7	DTXSID3020205	Benzyl butyl phthalate
26040-51-7	DTXSID7027887	Bis(2-ethylhexyl) tetrabromophthalate
103-23-1	DTXSID0020606	Bis(2-ethylhexyl)hexanedioate
80-05-7	DTXSID7020182	Bisphenol A
75-25-2	DTXSID1021374	Bromoform
128-37-0	DTXSID2020216	Butylated hydroxytoluene
17852-99-2	DTXSID2066270	C.I. Pigment Red 52, calcium salt (1:1)
5567-15-7	DTXSID1021453	C.I. Pigment Yellow 83
7440-43-9	DTXSID1023940	Cadmium
NOCAS_872417	DTXSID10872417	Cadmium & Cadmium Compounds
58-08-2	DTXSID0020232	Caffeine
62-54-4	DTXSID0020234	Calcium acetate
299-28-5	DTXSID2029618	Calcium D-gluconate
105-60-2	DTXSID4020240	Caprolactam
10605-21-7	DTXSID4024729	Carbendazim
56-23-5	DTXSID8020250	Carbon tetrachloride
513-77-9	DTXSID1029623	Carbonic acid, barium salt (1:1)
1698-60-8	DTXSID3034872	Chloridazon
108-90-7	DTXSID4020298	Chlorobenzene
143-28-2	DTXSID0022010	cis-Oleyl alcohol
77-92-9	DTXSID3020332	Citric acid
1702-17-6	DTXSID9029221	Clopyralid

NOCAS_872419	DTXSID30872419	Cobalt & Cobalt Compounds
7646-79-9	DTXSID9040180	Cobalt chloride
64-86-8	DTXSID5024845	Colchicine
8001-58-9	DTXSID2023987	Creosote
420-04-2	DTXSID9034490	Cyanamide
NOCAS_872420	DTXSID40872420	Cyanide salts
1222-05-5	DTXSID8027373	Cyclopenta[g]-2-benzopyran, 1,3,4,6,7,8-hexahydro-4,6,6,7,8,8-hexamethyl-
134-62-3	DTXSID2021995	DEET
50-02-2	DTXSID3020384	Dexamethasone
81-13-0	DTXSID3022906	Dexpanthenol
50-70-4	DTXSID5023588	D-Glucitol
526-95-4	DTXSID8027169	D-Gluconic acid
50-99-7	DTXSID7022910	D-Glucose
117-81-7	DTXSID5020607	Di(2-ethylhexyl) phthalate
131-17-9	DTXSID7020392	Diallyl phthalate
109-43-3	DTXSID1041847	Dibutyl decanedioate
84-74-2	DTXSID2021781	Dibutyl phthalate
75-09-2	DTXSID0020868	Dichloromethane
62-73-7	DTXSID5020449	Dichlorvos
99-30-9	DTXSID2020426	Dicloran
84-61-7	DTXSID5025021	Dicyclohexyl phthalate
105-53-3	DTXSID7021863	Diethyl propanedioate
111-77-3	DTXSID3025049	Diethylene glycol monomethyl ether
35367-38-5	DTXSID1024049	Diflubenzuron
84-69-5	DTXSID9022522	Diisobutyl phthalate

26761-40-0	DTXSID4025082	Diisodecyl phthalate
28553-12-0	DTXSID4022521	Diisononyl phthalate
60-51-5	DTXSID7020479	Dimethoate
108-59-8	DTXSID4029145	Dimethyl malonate
108-01-0	DTXSID2020505	Dimethylaminoethanol
117-84-0	DTXSID1021956	Di-n-octyl phthalate
25265-71-8	DTXSID0027856	Dipropylene glycol
88917-22-0	DTXSID4029062	Dipropyleneglycol methyl ether acetate
330-54-1	DTXSID0020446	Diuron
69-65-8	DTXSID1023235	D-Mannitol
112-85-6	DTXSID3026930	Docosanoic acid
577-11-7	DTXSID8022959	Docusate sodium
9004-32-4	DTXSID2020555	Edifas B
759-94-4	DTXSID1024091	EPTC
64-17-5	DTXSID9020584	Ethanol
91-53-2	DTXSID9020582	Ethoxyquin
141-78-6	DTXSID1022001	Ethyl acetate
100-41-4	DTXSID3020596	Ethylbenzene
107-21-1	DTXSID8020597	Ethylene glycol
110-71-4	DTXSID0025286	Ethylene glycol dimethyl ether
75-21-8	DTXSID0020600	Ethylene oxide
60168-88-9	DTXSID2032390	Fenarimol
50-00-0	DTXSID7020637	Formaldehyde
446-72-0	DTXSID5022308	Genistein
106-24-1	DTXSID8026727	Geraniol
90-80-2	DTXSID0026549	Gluconolactone
111-30-8	DTXSID6025355	Glutaraldehyde

25637-99-4	DTXSID8025383	Hexabromocyclododecane
87-68-3	DTXSID7020683	Hexachloro-1,3-butadiene
118-74-1	DTXSID2020682	Hexachlorobenzene
51235-04-2	DTXSID4024145	Hexazinone
123-31-9	DTXSID7020716	Hydroquinone
54464-57-2	DTXSID7031290	Isocyclemone E
97-54-1	DTXSID7022413	Isoeugenol
78-79-5	DTXSID2020761	Isoprene
1332-58-7	DTXSID6049640	Kaolin
50-21-5	DTXSID7023192	Lactic acid
63-42-3	DTXSID2023193	Lactose
52-90-4	DTXSID8022876	L-Cysteine
NOCAS_872421	DTXSID00872421	Lead & Lead Compounds
63-68-3	DTXSID5040548	L-Methionine
NOCAS_872422	DTXSID60872422	Long-chain chlorinated paraffins (C18-20)
6915-15-7	DTXSID0027640	Malic acid
NOCAS_872423	DTXSID20872423	Medium-chain chlorinated paraffins (C14-17)
7487-94-7	DTXSID5020811	Mercuric chloride
67-56-1	DTXSID2021731	Methanol
625-45-6	DTXSID1031591	Methoxyacetic acid
74-83-9	DTXSID8020832	Methyl bromide
9004-67-5	DTXSID1036919	Methyl cellulose
99-76-3	DTXSID4022529	Methylparaben
NOCAS_872424	DTXSID80872424	Molybdenum & Molybdenum Compounds
31138-65-5	DTXSID7027966	Monosodium D-glucoheptonate

108-38-3	DTXSID6026298	m-Xylene
99-97-8	DTXSID0021832	N,N,4-Trimethylaniline
91-20-3	DTXSID8020913	Naphthalene
NOCAS_872425	DTXSID40872425	Nickel & Nickel Compounds
872-50-4	DTXSID6020856	N-Methyl-2-pyrrolidone
86-30-6	DTXSID6021030	N-Nitrosodiphenylamine
25154-52-3	DTXSID3021857	n-Nonylphenol
95-48-7	DTXSID8021808	o-Cresol
1843-05-6	DTXSID9027441	Octabenzone
556-67-2	DTXSID7027205	Octamethylcyclotetrasiloxane
124-07-2	DTXSID3021645	Octanoic acid
112-80-1	DTXSID1025809	Oleic acid
95-47-6	DTXSID3021807	o-Xylene
133-49-3	DTXSID3044540	Pentachlorobenzenethiol
87-86-5	DTXSID7021106	Pentachlorophenol
3296-90-0	DTXSID9020164	Pentaerythritol dibromide
375-73-5	DTXSID5030030	Perfluorobutanesulfonic acid
335-76-2	DTXSID3031860	Perfluorodecanoic acid
335-67-1	DTXSID8031865	Perfluorooctanoic acid
108-95-2	DTXSID5021124	Phenol
85-44-9	DTXSID2021159	Phthalic anhydride
1918-02-1	DTXSID1021160	Picloram
81-33-4	DTXSID9052555	Pigment Violet 29
6528-34-3	DTXSID0052336	Pigment Yellow 65
298-14-6	DTXSID0021177	Potassium bicarbonate
299-27-4	DTXSID7029617	Potassium D-gluconate
29420-49-3	DTXSID3037707	Potassium perfluorobutanesulfonate

106-42-3	DTXSID2021868	p-Xylene
108-46-3	DTXSID2021238	Resorcinol
68-26-8	DTXSID3023556	Retinol
127-47-9	DTXSID6021240	Retinol acetate
90-02-8	DTXSID1021792	Salicylaldehyde
122-34-9	DTXSID4021268	Simazine
497-19-8	DTXSID1029621	Sodium carbonate
4418-26-2	DTXSID7026029	Sodium dehydroacetate
527-07-1	DTXSID7027170	Sodium D-gluconate
7632-00-0	DTXSID0020941	Sodium nitrite
1344-09-8	DTXSID9029647	Sodium silicate
10102-17-7	DTXSID6044197	Sodium thiosulfate, pentahydrate
111-01-3	DTXSID0046513	Squalane
100-42-5	DTXSID2021284	Styrene
57-50-1	DTXSID2021288	Sucrose
994-05-8	DTXSID8024521	tert-Amyl methyl ether
127-18-4	DTXSID2021319	Tetrachloroethylene
58-55-9	DTXSID5021336	Theophylline
62-55-5	DTXSID9021340	Thioacetamide
137-26-8	DTXSID5021332	Thiram
7772-99-8	DTXSID8021351	Tin(II) chloride
126-73-8	DTXSID3021986	Tributyl phosphate
1461-22-9	DTXSID3027403	Tributyltin chloride
79-01-6	DTXSID0021383	Trichloroethylene
101-20-2	DTXSID4026214	Triclocarban
55335-06-3	DTXSID0032497	Triclopyr
77-93-0	DTXSID0040701	Triethyl citrate

2451-62-9	DTXSID4026262	Triglycidyl isocyanurate
115-86-6	DTXSID1021952	Triphenyl phosphate
68937-41-7	DTXSID4028880	Triphenyl phosphates isopropylated
24800-44-0	DTXSID7027837	Tripropylene glycol
55934-93-5	DTXSID8042503	Tripropylene glycol butyl ether
115-96-8	DTXSID5021411	Tris(2-chloroethyl) phosphate
75-01-4	DTXSID8021434	Vinyl chloride

Appendix B. Detailed Information on Data Sources used in the PICS Approach

Sources	Human Hazard	Carcinogenicity	Genotoxicity	Ecological Hazard	Persistence and Bioaccumulation	Skin Sensitization and Skin/Eye	Exposure	Susceptible Populations	Description	Reference	URL
Compiled PODs, Toxicity Values and Cancer Classifications											
Alaska Department of Environmental Conservation		*							Cancer slope factors and unit risk compiled by State of Alaska	NA	https://dec.alaska.gov/spar/csp/guidance/cleanuplevels.pdf
Agency for Toxic Substances and Diseases Registry (ATSDR)	*								NOAEL values derived from CDC / ATSDR risk assessments	NA	https://www.atsdr.cdc.gov/mrls/mrllist.asp

California Environmental Protection Agency, Office of Environmental Health Hazard Assessment (OEHHA)		*							Cancer slope factors and unit risk compiled by State of California	NA	https://oehha.ca.gov/chemicals
California Environmental Protection Agency		*							Cancer classifications from California EPA	NA	https://oehha.ca.gov/proposition-65/proposition-65-list
COSMOS	*		*	*					Data compiled by the COSMOS project, a collaboration between the US FDA and Cosmetics Europe using integrated *in silico* models for the prediction of human repeated dose toxicity of COSMetics to Optimize Safety.	NA	http://www.cosmostox.eu/what/COSMOSdb/

Department of Energy (DOE) Wildlife Benchmarks	*			*					PODs from ecological risk assessments performed by DOE on both mammalian and aquatic species	Sample, B.E., Opresko, D.M., Suter, G.W. (1996) Toxicological Benchmarks for Wildlife: 1996 Revision. Springfield, VA: National Technical Information Service, U.S. Department of Commerce	https://rais.ornl.gov/documents/tm86r3.pdf
European Chemicals Agency (ECHA) eChemPortal	*		*	*		*			Data compiled by ECHA and made available through eChemPortal	NA	https://www.echemportal.org/echemportal/index.action
ECHA (IUCLID)	*			*					Data compiled by ECHA (European Chemicals Agency) and made available via an IUCLID data file	NA	https://echa.europa.eu/information-on-chemicals/registered-substances
European Union Reference Laboratory for Alternatives to Animal Testing (ECVAM EURL) Genotoxicity and Carcinoge			*						Genotoxicity data compiled by EURL ECCVAM	NA	https://data.europa.eu/euodp/data/dataset/jrc-eurl-ecvam-genotoxicity-carcinogenicity-ames

nicity Consolidated Database											
European Food Safety Authority (EFSA)	*			*					POD values compiled by EFSA (European Food Safety Agency)	NA	https://zenodo.org/record/1252752#.W-WNgDNReHs
EPA Health Effects Assessment Summary Tables (HEAST)	*			*					POD values compiled by EPA HEAST	NA	https://epa-heast.ornl.gov/heast.php
EPA High Production Volume Information System (HPVIS)	*			*					POD values compiled by EPA OPPT High Production Volume Information System	NA	Data was initially fully public, but access is now restricted. Data used here is a download from HPVIS from 2015
EPA Office of Pesticide Programs (OPP) Assessments		*							Cancer slope factors and classifications compiled by EPA OPP	NA	https://iaspub.epa.gov/apex/pesticides/f?p=HHBP:home

EPA Office of Pollution Prevention and Toxics (OPPT)	*								POD values from EPA OPPT Risk Assessment documents	NA	Data extracted from pdf files provided by OPPT, which should reflect data from ChemView https://chemview.epa.gov/chemview
Health Assessment Workspace Collaborative (HAWC)	*			*					POD values compiled from public HAWC projects	NA	https://hawcproject.org/
Health Canada		*							Cancer slope factors, unit risk and classifications compiled by Health Canada	NA	http://publications.gc.ca/collections/collection_2012/sc-hc/H128-1-11-638-eng.pdf
Hazard Evaluation Support System (HESS)	*								POD values compiled by HESS Japan (Hazard Evaluation Support System Integrated Platform, National Institute of Technology and Evaluation)	NA	http://www.nite.go.jp/en/chem/qsar/hess-e.html
International Agency for Research on Cancer (IARC)		*							Cancer classifications derived by IARC	NA	https://monographs.iarc.fr/list-of-classifications-volumes/

Integrated Risk Information System (IRIS)	*	*							POD values, cancer slope factors, unit risk and cancer classifications from EPA IRIS	NA	https://cfpub.epa.gov/ncea/iris_drafts/simple_list.cfm
National Institute for Occupational Safety and Health (NIOSH)		*							Cancer classifications compiled by NIOSH (CDC, National Institute for Occupational Safety and Health)	NA	https://www.cdc.gov/niosh/topics/cancer/npotocca.html
National Toxicology Program (NTP) Report on Carcinogens (ROC)		*							Cancer classifications compiled by NTP (National Toxicology Program)	NA	https://ntp.niehs.nih.gov/pubhealth/roc/index-1.html#toc1
EPA Provisional Peer-Review Toxicity Values (PPRTV) [NCEA database]	*								POD Values from EPA PPRTV documents, provided by EPA NCEA	NA	data provided by NCEA as an MS Access database
EPA Provisional Peer-Review	*	*							POD Values from EPA PPRTV documents, cancer classifications	NA	https://hhpprtv.ornl.gov/

Toxicity Values (PPRTV) [ORNL database]									extracted from ORNL PPRTV web site		
Data compilations											
EPA Chemical Data Reporting (CDR)							*	*	TSCA Chemical data reporting rule (CDR) production volume information	NA	https://www.epa.gov/chemical-data-reporting/2016-chemical-data-reporting-results
EPA Chemical And Product Categories (CPCat)								*	Chemical and product categories	Dioniso et al. Exploring consumer exposure pathways and patterns of use for chemicals in the environment, Tox. Reports vol 2, pp 28-237 (2015)	https://actor.epa.gov/cpcat/faces/home.xhtml
EPA Chemical and Products Database (CPDat)								*	Chemical and product database	Dionisio KL, Phillips K, Price PS, Grulke CM, Williams A, Biryol D, Hong T, Isaacs KK. The Chemical and Products Database, a resource for exposure-relevant data on chemicals in consumer products. Sci Data. 2018 Jul 10;5:180125. doi: 10.1038/sdata.2018.125. PubMed PMID: 29989593; PubMed Central PMCID: PMC6038847	https://www.epa.gov/chemical-research/chemical-and-products-database-cpdat

EPA ECOTOXicology knowledgebase (ECOTOX)	*			*					EPA ECOTOX database. Data imported to ToxValDB	NA	https://cfpub.epa.gov/ecotox/
EPA Toxicity Forecaster (ToxCast) v3.0	*								High-throughput data for a variety of high level cellular responses.	Judson et al. *In vitro* Screening of Environmental Chemicals for Targeted Testing Prioritization: The ToxCast Project, Environmental Health Perspectives, volume 118, p 485	https://figshare.com/articles/ToxCast_Database_invitroDB_/6062623/2
EPA Toxicity Reference Database (ToxRefDB)	*								POD values from EPA ToxRefDB	Watford, S., A. Adrian, J. Wignall, J. Brown, AND M. Martin. ToxRefDB 2.0: Improvements in Capturing Qualitative and Quantitative Data from *in vivo* Toxicity Studies (SOT). Presented at SOT Annual Meeting, Baltimore, MD, March 12 - 16, 2017. https://doi.org/10.23645/epacomptox.5178622	https://epa.figshare.com/articles/Animal_Toxicity_Studies_Effects_and_Endpoints_Toxicity_Reference_Database_-_ToxRefDB_files_/6062545
Istituto Superiore di Sanita Chemical Toxicity database			*						Database of genotoxicity data compiled by Istituto Superiore di Sanita (ISS), Italy	Begnini et al. "a novel approach: chemical relational databases,and the role of the ISScaN database on assessing chemical carcinogenicity", Ann Ist	http://old.iss.it/publ/anna/2008/1/44148.pdf

on long-term carcinogenicity bioassay on rodents (rat and mouse) (ISSCAN)										super sAnItà 2008 | Vol. 44, no. 1: 48-56	
Toxicology Data Network (TOXNET)			*						Genetoxicity data downloaded from National Library of Medicine (NLM) TOXNET	NA	https://toxnet.nlm.nih.gov/newtoxnet/genetox.htm
Prediction Models											
EPA Tool for High-Throughput Toxicokinetics (HTTK)	*								High-throughput toxicokinetic data and models used to predict *in vivo* administered equivalent doses from *in vitro* bioactive concentrations	Pearce et al. "httk: R Package for High-Throughput Toxicokinetics", J Stat Softw. 2017 Jul 17; 79(4): 1–26. doi: 10.18637/jss.v079.i04	https://cran.r-project.org/web/packages/httk/index.html (data are in the RData within the package)
EPA Exposure Forecaster (ExpoCast) Systemic Empirical Evaluation of Models 3 (SEEM3)							*		Systematic Empirical Evaluation of Models (SEEM) framework includes calibration and evaluation of the models using chemical concentrations found in blood and urine samples from the National Health and Nutrition Examination Study.	Ring et al., "Consensus Modeling of Median Chemical Intake for the U.S. Population Based on Predictions of Exposure Pathways", Environ. Sci. Technol., 2019, 53 (2), pp 719–732 DOI: 10.1021/acs.est.8b04056	https://pubs.acs.org/doi/10.1021/acs.est.8b04056

Ecological Structure Activity Relationship Prediction Model (ECOSAR)				*					SAR model to predict aquatic PODs	NA	https://www.epa.gov/tsca-screening-tools/ecological-structure-activity-relationships-ecosar-predictive-model
EPI Suite					*				QSAR software to estimate physicochemical and fate and transport properties		https://www.epa.gov/tsca-screening-tools/download-epi-suitetm-estimation-program-interface-v411
EPA Toxicity Estimation Software Tool (TEST)			*			*			Skin and Eye irritation and sensitization data derived from GHS documents, and genotoxicity QSAR model	Vegosen and Martin, "An automated framework for compiling and integrating chemical hazard data", Clean Tech. Environ. Policy, 2020, 22, pp. 441-458 DOI 10.1007/s10098-019-01795-w	https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test
OPEn structure–activity/property Relationship App (OPERA)					*				OPEn structure–activity/property Relationship App for predicting physicochemical and environmental fate properties	Mansouri et al."OPERA models for predicting physicochemical properties and environmental fate endpoints", Journal of Cheminformatics201810:10	https://jcheminf.biomedcentral.com/articles/10.1186/s13321-018-0263-1
ToxTree	*								Toxic Hazard Estimation by decision tree approach used to predict threshold of	Patlewicz, G., et al. (2008). "An evaluation of the implementation of the Cramer classification scheme	http://toxtree.sourceforge.net/

									toxicological concern (TTC)	in the Toxtree software." SAR QSAR Environ Res 19(5-6): 495-524.	
World Health Organization (WHO) International Programme on Chemical Safety (IPCS)	*								Acute toxicity values from WHO Pesticides Classification	See URL	http://www.who.int/ipcs/publications/pesticides_hazard_2009.pdf
Publications											
Arnot and Gobas (2006)					*				BAF and BCF values compiled from experimental studies	Arnot and Gobas, "A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms", Environmental Reviews, 2006, 14(4): 257-297, https://doi.org/10.1139/a06-005	https://www.nrcresearchpress.com/doi/10.1139/a06-005#.XL9OkehKg2w
Chiu et al. (2018)	*								POD values compiled by Chiu *et al.*	Chiu W, et al. "Beyond the RfD: Broad Application of a Probabilistic Approach to Improve Chemical Dose–Response Assessments for Noncancer Effects", Env.	https://doi.org/10.1289/EHP3368

										Health. Persp. Vol 126 (2018)	
Wignall et al. (2014)	*								BMD values compiled or derived by Wignall *et al.*	Wignall J. et al. "Standardizing Benchmark Dose Calculations to Improve Science-Based Decisions in Human Health Assessments", Env. Health Persp. Vol. 22 p 499 (2014)	https://ehp.niehs.nih.gov/doi/10.1289/ehp.1307539

Appendix C. Quality Assurance Recommendations to Efficiently Review Datasets to Support Candidate Chemical Identification for TSCA

Overview

This appendix describes the approaches that ORD implemented to ensure a high standard of review of compiled publicly available information from Type 1 data sources on the proof-of-concept 238 chemical substances (POC238). The data curation effort reviewed chemical, toxicological, and exposure data to confirm accuracy. Specific details about the review processes for each data domain are described below.

The QC review approaches were developed in a "learn-by-doing" pilot study using the POC238. The pilot study developed methods for data aggregation, curation, and evaluation, as well as recommendations to efficiently review large Type 1 datasets. The TSCA data curation team consisted of scientists from ORD organized into workgroups based on expertise and given data for review.

Procedures for QC review

Chemical data were sorted into six data domains: human hazard, exposure, genotoxicity, ecological hazard, skin sensitization and skin/eye irritation, and bioaccumulation. Data on the chemicals were collected from Type 1 data sources [as defined in *A Working Approach for Identifying Potential Candidate Chemicals for Prioritization*[118]*;* see Appendix B for Data Source list]. Type 1 data sources are publicly available and readily searchable, enabling data extraction in a structured form. To review these data, workgroups were organized according to scientific expertise. Each workgroup established a process for reviewing their data domain and these processes are summarized below.

No study quality considerations were evaluated during QC review. Reviewers did not perform a critical analysis of experimental design, statistical analyses, or data interpretation. Rather, reviewers compared the collected Type 1 data to primary and secondary sources. A primary source was defined as the study, report, or manufacturer report with health safety data. A secondary source was defined as a database or source that provided data aggregated from multiple primary sources.

Reviewers flagged data that could not be confirmed to the primary source, even if the aggregated data matched the secondary source. However, certain secondary sources, such as the

[118] https://www.epa.gov/sites/production/files/2018-09/documents/preprioritization_white_paper_9272018.pdf

ECOTOX Knowledgebase[119], the Integrated Risk Information System (IRIS)[120], or exposure data[121] have existing QC processes or peer review processes. For these select databases, confirmation to secondary source was sufficient to pass QC review. Reviewers recorded reasons for QC flags and developed QC metrics for data errors. "QC flag" measured the percentage of data without confirmation to the primary source. "Error rate" determined how often a secondary source incorrectly reported a value from a primary source.

The workgroups collected metrics on the data sources. The QC flag metric reported the percentage of data points that could not be confirmed to a primary source, for example, in instances where the primary source was not available. QC flag metric was calculated by dividing the number of data points that were flagged by the total number of data points reviewed. Error rate measured how often a secondary database did not match the primary source, that is when both the primary source and secondary source are available, but the values do not match. Error rate was calculated by dividing the number of data points that did not match the primary source by the number of data points that have both a primary and secondary source.

Human Hazard Domain Workgroup Review Approach

Human hazard data consisted of *in vivo* data aggregated from publicly available databases. Data were provided in a spreadsheet for the workgroup to review. A single scientist was assigned to review an individual data source. The human hazard workgroup QC review focused on a subset of the aggregated Type 1 data: chemical identifier, route of exposure, study duration, and point-of-departure (POD) data.

Each reviewer compared the chemical identifier, route of exposure, study duration, and POD data to the secondary source - i.e., the Type 1 source from which data were extracted. In addition, each reviewer attempted to link data points to a primary source – i.e., the original reference. If data could not be confirmed to a primary reference, the data was flagged. However, the workgroup recognized that certain secondary sources, such as IRIS[3], had existing QC processes. For these databases, the secondary source was sufficient to pass QC review.

Study quality considerations were not evaluated during QC review. Reviewers did not review experimental design, statistical analyses, or data interpretation. Rather QC review was limited to comparing the data with primary and secondary sources. The QC review scope was limited for three reasons. First, the study quality would be evaluated during the expert review for candidate selection. However, if a reviewer noted a potential study quality issue, the issue could

[119] https://cfpub.epa.gov/ecotox/
[120] https://www.epa.gov/iris
[121] https://www.epa.gov/chemical-research/rapid-chemical-exposure-and-dose-research

be recorded for consideration during the subsequent expert review. Secondly, systematic data quality review requires a minimum of two reviewers for each data point. The human hazard workgroup did not have sufficient personnel to have two reviewers per data point within the accelerated timeline of the pilot study. Lastly, determining a POD is often a fit-for-purpose process that may differ between academic, industrial, and regulatory groups. Harmonizing POD selection across studies was beyond the scope of the human hazard workgroup.

Exposure Domain Workgroup Review Approach

Two datasets were reviewed by the exposure workgroup: exposure model data (including model parameters and outputs) and selected parameters related to susceptible population exposure. Exposure model data estimated potential human exposures via chemical use parameters from publicly available data sources (Wambaugh et al., 2014[122], Ring et al., 2018[123]). Susceptible population data was limited to potential exposure in children. Susceptible population data consisted of chemical occurrence in the following media: consumer products with which children either directly (children's products) or indirectly (other household or personal care products) come into contact, flooring product, house dust, breast milk, foods or food-contact materials, or far-field sources. Chemical occurrence data of compounds in house dust[124] or breast milk[125] were collected from two primary references. Occurrence data for the remainder of the media were collected from three different secondary data sources: Chemical Data Reporting (CDR)[126], Chemical and Products Database (CPDat)[127], and Chemical Product Categories Database (CPCat)[128]. These secondary sources aggregate government and/or manufacturer reported information on chemical presence in various types of consumer products or industrial processes. The exposure data and susceptible life stage data were provided in separate spreadsheets for the workgroup to review. All data were reviewed by at least two workgroup members.

---

[122] Wambaugh JF, Wang A, Dionisio KL, Frame A, Egeghy P, Judson R, Setzer RW. (2014). High throughput heuristics for prioritizing human exposure to environmental chemicals. Environmental Science and Technology 48(21):12760-7.

[123] Ring CL, Arnot J, Bennett DH, Egeghy P, Fantke P, Huang L, Isaacs KK, Jolliet O, Phillips K, Price PS, Shin HM, Westgate JN, Setzer RW, Wambaugh JF. (2018). Consensus Modeling of Median Chemical Intake for the U.S. Population Based on Predictions of Exposure Pathways. Environmental Science and Technology. 53(2):719-732.

[124] Mitro SD, Dodson RE, Singla V, Adamkiewicz G, Elmi AF, Tilly MK, and Zota AR. (2016). Consumer product chemicals in indoor dust: A quantitative meta-analysis of U.S. studies. Environmental Science and Technology 50: 13611-11.

[125] Lehmann GM, LaKind JS, Davis MH, Hines EP, Marchitti SA, Alcala C, and Lorber C. (2018). Environmental chemicals in breast milk and formula: Exposure and risk assessment implications. Environmental Health Perspectives, 126: 96001.

[126] https://www.epa.gov/chemical-data-reporting/2016-chemical-data-reporting-results

[127] https://comptox.epa.gov/dashboard/downloads, CPDATdownload

[128] https://comptox.epa.gov/dashboard/downloads, CPCATARCHIVE

For the exposure model data, the workgroup focused on a subset of chemical use parameters (pesticide active, pesticide inert, production volume) and estimated exposure outputs from the exposure model median. These values were checked for accuracy against the Wambaugh et al. (2014)[6] and Ring et al. (2018)[7] references. The workgroup did not do a QC review of all model inputs. Rather, the workgroup focused their review on the 3 of the 5 most predictive heuristics in Wambaugh et al. 2014[6]. The heuristics that were not reviewed were the industrial and/or consumer use of a chemical. The workgroup did not re-run the exposure models to confirm outputs, but instead focused on accurate transposition from the source material.

Susceptible population occurrence data were confirmed by reviewing either the primary sources for house dust[8] or breast milk[9] occurrence or the primary sources cited in the secondary sources (i.e., reported information of chemicals contained in consumer products or used in industrial processes in primary references found in secondary sources). When primary sources were derived from secondary sources, the workgroup reviewed chemicals reported in consumer products related to child use, flooring, food or food-contact material, or far-field exposure sources until a primary source was found that met QC requirements. OCSPP uses presence/absence in children's products is used as an indicator for potential susceptible population exposure during candidate identification. As this metric is binary (i.e., yes/no) rather than weighted (i.e., occurrence in 10 products versus present only in 1 product), the workgroup focused QC review on (a) ease of record access and (b) confirming occurrence in each medium related to susceptible population exposure. House dust and breast milk were confirmed by checking to respective primary sources. However, for all other media, CDR[10] records were most readily accessible and reviewable in an automated fashion. In addition, CDR[10] records are manufacturer reported under TSCA, so these records will likely be relevant for candidate chemical identification. For these reasons, CDR[10] records were reviewed first, and if CDR[10] records passed QC review, then available CPDat[11] or CPCat[12] records were not reviewed. CPDat[11] records were reviewed if no CDR[10] data was available or if CDR[9] records did not pass QC, as CPDat[11] records link a chemical directly to manufacturer-reported information of compounds in a consumer product. Finally, if neither CDR[10] nor CPDat[11] records were available or did not pass QC review, CPCat[12] data records were reviewed.

As most data for this domain consisted of manufacturer-reported data, no study quality considerations were evaluated during the QC review. The reviewers did not analyze model design, statistical analyses, or data interpretation. Rather the workgroup focused spreadsheet data accuracy relative to primary and secondary sources.

Genotoxicity Domain Workgroup Review Approach

Genotoxicity data was aggregated from publicly available databases and provided in a spreadsheet for the workgroup to review. A single reviewer was assigned to group of chemicals and asked to review the genotoxicity data for those chemicals. The genotoxicity workgroup focused their QC review on a subset of the aggregated Type 1 data: mutagenicity data and clastogenicity data. The workgroup evaluated data in the standard bacterial mutation assays (the Salmonella and E. coli WP2 strains), as well as three main assays for chromosomal mutation (*in vitro* chromosome aberration assay, mouse bone-marrow micronucleus assay, and the mouse lymphoma Tk+/- assay). These data were selected based on a comparison of the predictivity of combination of genotoxicity assays to the Salmonella (Ames) mutagenicity assay alone[129]. Using a database of >10,000 compounds, genotoxicity data from two bacterial strains (TA98 and TA100 of Salmonella) identified 93% of the mutagens. When chromosomal mutation assay data were included, 99% of the mutagens were identified. These findings suggest that bacterial and chromosomal mutation data are sufficient for evaluating genotoxicity potential. Each reviewer confirmed selected data back to the secondary source - i.e., the database from which spreadsheet data were extracted. In addition, each reviewer tried to confirm data back to the primary source – i.e., the original reference cited in the secondary source. If data could not be confirmed to a primary reference, the data was flagged.

The genotoxicity workgroup categorized chemicals as "genotoxic," "non-genotoxic," or "inconclusive" based on preliminary review of available genotoxicity data. This categorization was not intended to represent a final determination on the genotoxicity of these chemicals. For example, chemical substances with at least one positive genotoxic assay were categorized as genotoxic, with the understanding that further evaluation of study quality and design may lead to a different determination.

Bioaccumulation Subdomain Workgroup Review Approach

Bioaccumulation data were aggregated from public databases and provided in a spreadsheet for review. All data were reviewed by the workgroup members. The QC review focused on a subset of the bioaccumulation chemical data: bioconcentration factor (BCF) and bioaccumulation factor (BAF).

Chemical data consisted of two categories: experimental data and modeled data. Although more limited, experimental data were given priority. Reviewers attempted to trace experimental data back to a primary source – i.e., the original reference with measured values. If the reviewers

[129] Williams RV, DM DeMarini, LF Stankowski Jr, PA Escobar, E Zeiger, J Howe, R Elespuru, KP Cross. (2019). Are all bacterial strains required by OECD mutagenicity test guideline TG471 needed? Mutation Research 848:503081.

could not confirm the data to a primary source, the data was flagged. The workgroup was only able to confirm a small percentage of the data to a secondary source (i.e., the database from which spreadsheet data were extracted) because secondary source no longer existed or contained proprietary data. The QC review pilot study was limited to publicly available data.

For the predicted BCF and BAF values, the workgroup only reviewed publicly available models. Models were re-run and compared with the spreadsheet for accuracy. Where the models produced different values, the values were actively corrected. The workgroup did not QC review the model inputs. For some data, underlying model inputs and predicted values had published QC review processes. Other model outputs did not provide information on model inputs or QC review. The workgroup flagged models that did not have a publicly accessible QC process as a decision point for further consideration.

Ecological Hazard Domain Workgroup Review Approach

Ecological hazard data consisted of data extracted from two Type 1 sources – the US EPA ECOTOX Knowledgebase[2] and the European Chemicals Agency (ECHA) database[130]. These data were provided in a spreadsheet for the workgroup to review. A single reviewer was assigned to a group of chemicals and asked to review the ecological hazard data for those chemicals. The ecotoxicology workgroup reviewed a large amount of data across a variety of species.

Each reviewer was instructed to confirm the selected data back to the secondary source - i.e., the database from which data on the spreadsheet were extracted. Each reviewer also tried to trace the data point back to a primary source – i.e., the original reference cited in the secondary source. If data could not be confirmed to a primary reference, the data was flagged. However, the ECOTOX Knowledgebase[2] has a robust QC process ensuring data are verified using reliable source and reflect what was reported in the publication. Once quality assurance steps have been completed, the data are released to the ECOTOX Knowledgebase[2]. Therefore, confirming data to the ECOTOX Knowledgebase[2] (i.e., secondary source) was equivalent to a primary source.

No study quality considerations were evaluated during the QC review. The workgroup did not review experimental design, statistical analyses, or data interpretation. Rather, the ecological hazard workgroup focused only on the data accuracy relative to the primary and secondary sources.

Skin Sensitization and Skin/Eye Irritation Workgroup Review Approach

Skin sensitization and skin/eye irritation data were collected from publicly available sources and aggregated for review. Information standardized with the Global Harmonized System

[130] https://echa.europa.eu/

of Classification and Labeling of Chemicals (GHS)[131] was extracted. The GHS classification codes -- including hazard codes or H-codes; numerical hazard categories; signal word codes; and classification, labeling, and packaging (CLP) hazard class -- were included and used for designating the chemicals as sensitizing or irritating. A single reviewer was assigned to review the skin and eye irritation data.

A reviewer was asked to verify data in the resulting spreadsheet for quality control of the programmatic data collection process. The reviewer performed an automated check of the data to ensure that transposition of secondary source data was correct by comparing numbers in the spreadsheet to the secondary source material. The reviewer then manually reviewed 10% of the data back to a primary source. This limited manual review was necessary owing to time and resource issues.

Summary

In summary, the POC238 chemical substances were reviewed for transcription from primary and secondary sources into the database used for the PICS approach. This review did not take into account study quality or data validity, as that was determined to be part of an expert review process separate from this effort. For the PICs approach, data were deemed 'acceptable' if confirmed to a secondary source; primary source confirmation was not required. The case study developed methods for data aggregation, curation, and evaluation, as well as QA recommendations to efficiently review Type 1 datasets. The case highlighted various challenges in data quality and availability of primary sources in addition to the changing landscape of online resources worldwide and urgent need for specialized curation/quality resources, data quality tool(s), common data dictionary, process to store the documents for data provenance and quality flags for various data usage. These lessons learned have informed the QC process moving forward, storage and linking of secondary or primary sources to data records wherever available for data provenance, inclusion of data audit capability, development of quality flags to enable fit-for-purpose data aggregation, and the development of a QC tool for future use with large datasets as part of continuous curation effort.

[131] https://www.unece.org/trans/danger/publi/ghs/ghs_welcome_e.html

Appendix D. Definition of Exposure Pathways for Calculating the Susceptible Population Domain Metric.

Exposure Source	Definition	Data Source	Primary Contributors to Increased Children's Exposure versus the General Population	Exposure Differential Metric
Consumer Sources -Children's Products	Known occurrence (via reporting or measurement) in consumer products used more commonly (or exclusively) by children (e.g., arts and crafts formulations, baby preparations, car seats and other gear, toys, and marketed to children such as children's sunscreens). Known occurrence in products to which infants or children have closer contact than adults due to behavior (i.e., carpet, flooring)	CPDat: reported presence of chemical in a product used primarily by children; CPCat: chemicals directly reported or measured in children's products; chemicals reported or detected in flooring (e.g., carpet, carpet padding); 2016 CDR Consumer/Commercial Use Information: flag for use in children's products	Increased prevalence of use by children versus adults; closer proximity of source to children compared to adults	4
Breast Milk or Formula	Chemical detected in breast milk or formula	Lehmann GM, LaKind JS, Davis MH, Hines EP, Marchitti SA, Alcala C, and Lorber C. (2018) Environmental chemicals in breast milk and formula: Exposure and risk assessment implications. Environ Health Perspect, 126: 96001	Source unique to children	4
Dust	Measured in residential house dust in at least two studies in published meta-analysis.	Mitro SD, Dodson RE, Singla V, Adamkiewicz G, Elmi AF, Tilly MK, and Zota AR. (2016). Consumer product chemicals in indoor dust: A quantitative meta-analysis of U.S. studies. Environ Sci Technol, 50: 13611-11.	Increased contact by children; increased hand-to-mouth behaviors (and thus chemical ingestion); closer proximity of source to children (e.g., within children's breathing zone)	3
Consumer Sources – Flooring and Related Products	Known occurrence (via reporting or measurement) in flooring or floor coverings, or in products used on these items (such as cleaners)	CPDat: reported presence of chemical in a flooring related product, CPCat: presence in consumer product categories associated with flooring or related products; 2016 CDR Consumer/Commercial Use Information: reported in product categories associated with flooring	Increased contact by children; Increased hand-to-mouth behaviors (and thus chemical ingestion); closer proximity of source to children (e.g.,	3

			within children's breathing zone)	
Consumer Sources – Other Products (General Population)	Known occurrence (via reporting or measurement) in consumer products not captured elsewhere that may be either used by children or transferred to children by adults	CPDat: reported presence in consumer products; CPCat: presence in general consumer product categories; 2016 CDR Consumer/Commercial Use Information: reported as having "Consumer" use or "Both" (consumer and commercial)	Increased hand-to-mouth behaviors (and thus chemical ingestion); increased inhalation rates of children versus adults	2
Dietary Sources	Chemicals in food packaging (indirect food additives), direct food additives, agricultural chemicals, chemicals measured in drinking water	CPCat: presence in related categories; 2016 CDR Consumer/Commercial Use Information: presence in food-related product categories	Increased food consumption per unit body weight by children	1
Far-field Sources	Industrial pollutants (which may be released the environment and result in ultimate exposures via contact with contaminated media)	CPCat: presence in industrial use categories; 2016 CDR Industrial Process and Use Information: chemical was reported	Increased hand-to-mouth behaviors (and thus chemical ingestion); increased inhalation rates of children versus adults	1

Appendix E. Public Information Curation and Synthesis (PICS) Output for Proof-of-Concept (POC) Subset of the Non-confidential TSCA Active Inventory

Attached is the output results for the proof-of-concept subset used to inform the Public Information Curation and Synthesis (PICS) Approach. This data can also be viewed at https://ccte-tscapoc.epa.gov. Results were determined as described in the text and displayed visually in Figure 14 of the report.

Order	Column	Description
1	DTXSID	DSSTox generic substance ID
2	CASRN	Chemical Abstracts Registry Number
3	Name	Chemical Name
4	TSCA Active	Is the chemical in the TSCA Active Inventory?
5	TSCA 90	Is the chemical in the 2014 update to the TSCA Workplan?
6	TSCA POC	Is the chemical in the TSCA POC (this list)?
7	TSCA 10	Is the chemical in the 2016 Work Plan Chemicals?
8	SCIL	Is the chemical in the EPA Safer Chemicals Ingredients List (SCIL)?
9	SCIL Green Circle	Is the chemical in the SCIL Green Circle List?
10	SCIL Half Green Circle	Is the chemical in the SCIL Half Green Circle List?
11	SCIL Yellow Triangle	Is the chemical in the SCIL Yellow Triangle List?
12	SCOGS GRAS	Is the chemical in the FDA Generally Regarded as Safe (GRAS) List?
13	Intentional Food Ingredient	Is the chemical in the FDA Substances Added to Food Inventory?
14	IRIS	Is the chemical in the EPA IRIS Inventory?
15	PPRTV	Is the chemical in the EPA PPRTV Inventory?
16	SIDS	Is the chemical in the OECD Screening Information Data Set List?
17	TSCA Low Priority	Is the chemical in the High Priority TSCA list?
18	TSCA High Priority	Is the chemical in the Low Priority TSCA list?
19	Scientific Domain Metric	The value of the scientific domain metric (range from 0-100)
20	Information Availability Metric	The value of the information availability metric (range from 0-100)

21	Public Risk Assessment Noncancer	If there is a public non-cancer risk assessment, lists whether there is an RfD and/or RfC available
22	Public Risk Assessment Cancer	If there is a public cancer risk assessment, lists whether there is an slope factor and/or unit risk available
23	IG Flag (human hazard)	Information gathering flags for human hazard
24	IG Flag (ecological hazard)	Information gathering flags for ecological hazard
25	IG Flag (genotoxicity)	Information gathering flags for genotoxicity
26	IG Flag (cancer)	Information gathering flags for cancer
27	IG Flag (childrens exposure)	Information gathering flags for children's exposure
28	IG Flag (persistence / bioaccumulation)	Information gathering flags for persistence / bioaccumulation
29	IG Flag (sensitization / irritation)	Information gathering flags for sensitization / irritation
30	Score human hazard to exposure ratio	Score for human hazard to exposure ratio (range of 0-4)
31	Score ecological hazard	Score for ecological hazard (range of 0-4)
32	Score cancer	Score for cancer (range of 0-4)
33	Score genotoxicity	Score for genotoxicity (range of 0-4)
34	Score children	Score for children's exposure (range of 0-4)
35	Score persistence / bioaccumulation	Score for persistence / bioaccumulation (range of 0-4)
36	Score sensitization irritation	Score for sensitization / irritation (range of 0-4)
37	Human hazrd to exposure ratio (repeat dose)	Minimum repeat-dose mammalian POD / exposure estimate
38	POD mammalian in vivo oral repeat dose (mg/kg/day)	Minimum repeat-dose mammalian POD
39	Estimated human exposure (mg/kg/day)	Human exposure estimate from the SEEM3 model
40	Minimum aquatic ecological POD (mg/L)	Minimum aquatic ecological POD (experimental repeat dose or acute or predicted)
41	Ecological aquatic POD in vivo acute (mg/L)	Minimum aquatic ecological POD in vivo, acute
42	Ecological aquatic POD in vivo repeat dose (mg/L)	Minimum aquatic ecological POD in vivo, repeat dose
43	Data available (mammalian acute)	Is there an experimental mammalian POD for an acute toxicity study?
44	Data available (mammalian subchronic)	Is there an experimental mammalian POD for a subchronic toxicity study?
45	Data available (mammalian chronic)	Is there an experimental mammalian POD for a chronic toxicity study?
46	Data available (mammalian repeat dose)	Is there an experimental mammalian POD for a repeat dose toxicity study?
47	Data available (mammalian developmental)	Is there an experimental mammalian POD for a developmental toxicity study?

48	Data available (mammalian reproductive)	Is there an experimental mammalian POD for a reproductive toxicity study?
49	Data available (mammalian neurotoxicity)	Is there an experimental mammalian POD for a neurotoxicity study?
50	Data available (ecological acute plant)	Is there an experimental ecological POD for an acute plant study?
51	Data available (ecological acute invertebrate)	Is there an experimental ecological POD for an acute invertebrate study?
52	Data available (ecological acute vertebrate)	Is there an experimental ecological POD for an acute vertebrate study?
53	Data available (ecological repeat dose plant)	Is there an experimental ecological POD for a repeat dose plant study?
54	Data available (ecological repeat dose invertebrate)	Is there an experimental ecological POD for a repeat dose invertebrate study?
55	Data available (ecological repeat dose vertebrate)	Is there an experimental ecological POD for a repeat dose vertebrate study?

Appendix F. Comparison of Individual Scientific Domain Metrics for the POC238 and Non-confidential TSCA Active Inventory

Below are the results for the full PICS approach and the individual metrics for the POC238 subset as compared to the non-confidential TSCA active inventory. As expected, these comparisons show that the POC238 subset is more data-rich than the full non-confidential TSCA active inventory. This level of detail could allow the decision-maker to examine if any specific endpoint is potentially of more concern than any other and to explore potential data gaps for the chemical group of interest.

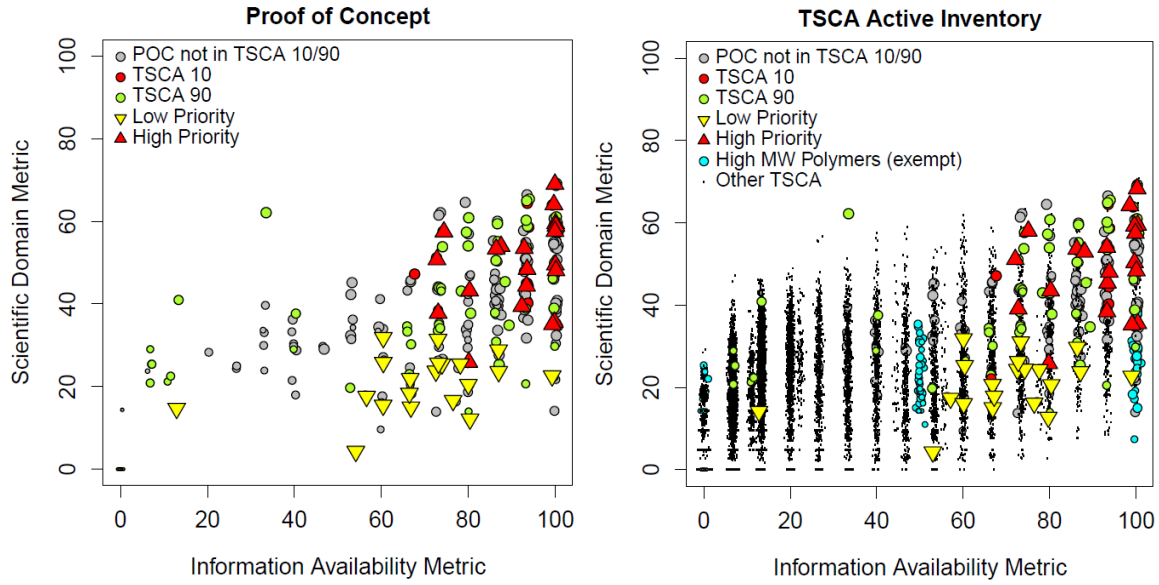


Figure F-1. Plot of the Information Availability vs. Scientific Domain Metrics for the POC238 set of chemical substances (left) and non-confidential TSCA active inventory (right). Each dot represents one chemical substance, with the size of the dot representing the number of domains with data for the specific chemical. The red dots represent the first ten TSCA Work Plan chemical substances selected for risk evaluation in 2016 (TSCA 10). The green dots represent the TSCA Work Plan chemical substances from the 2014 update (TSCA 90). The red triangles represent the high priority chemical substances and the yellow triangles represent the low priority chemical substances released in March 2019. The blue dots represent the high molecular weight compounds and exempt polymers and the small black dots represent the remaining compounds in the inventory. Positions of points are staggered for ease of visualization.

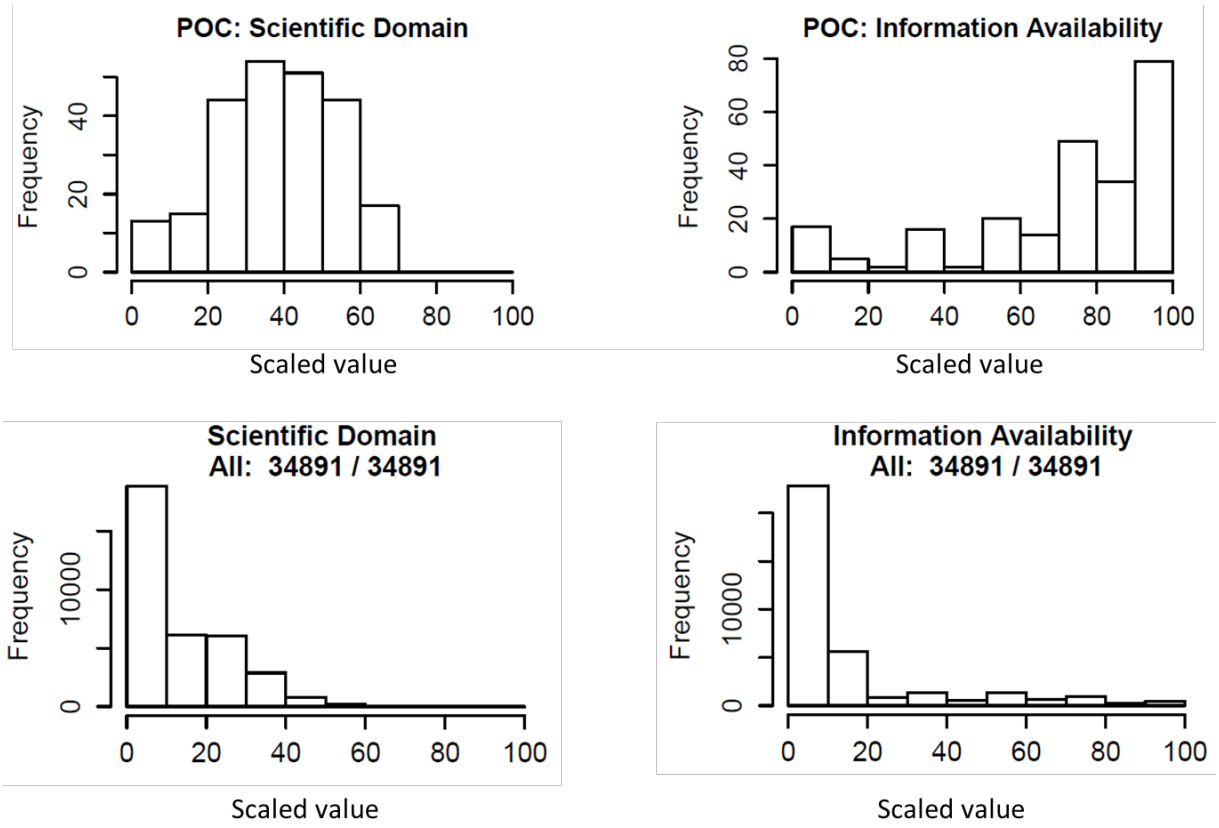
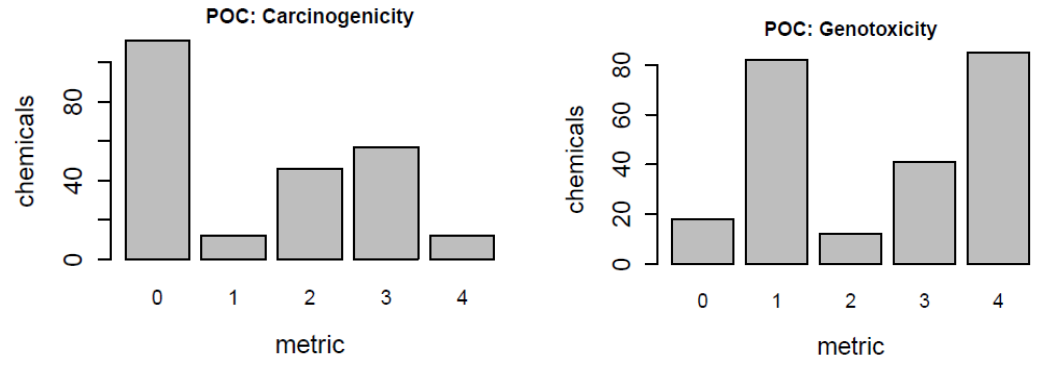


Figure F-2 – Distributions of the scaled SDM and IAM for the POC238 subset and non-confidential TSCA active inventory. The x-axis displays the scaled value and the y-axis shows the frequency of that value in the subset.



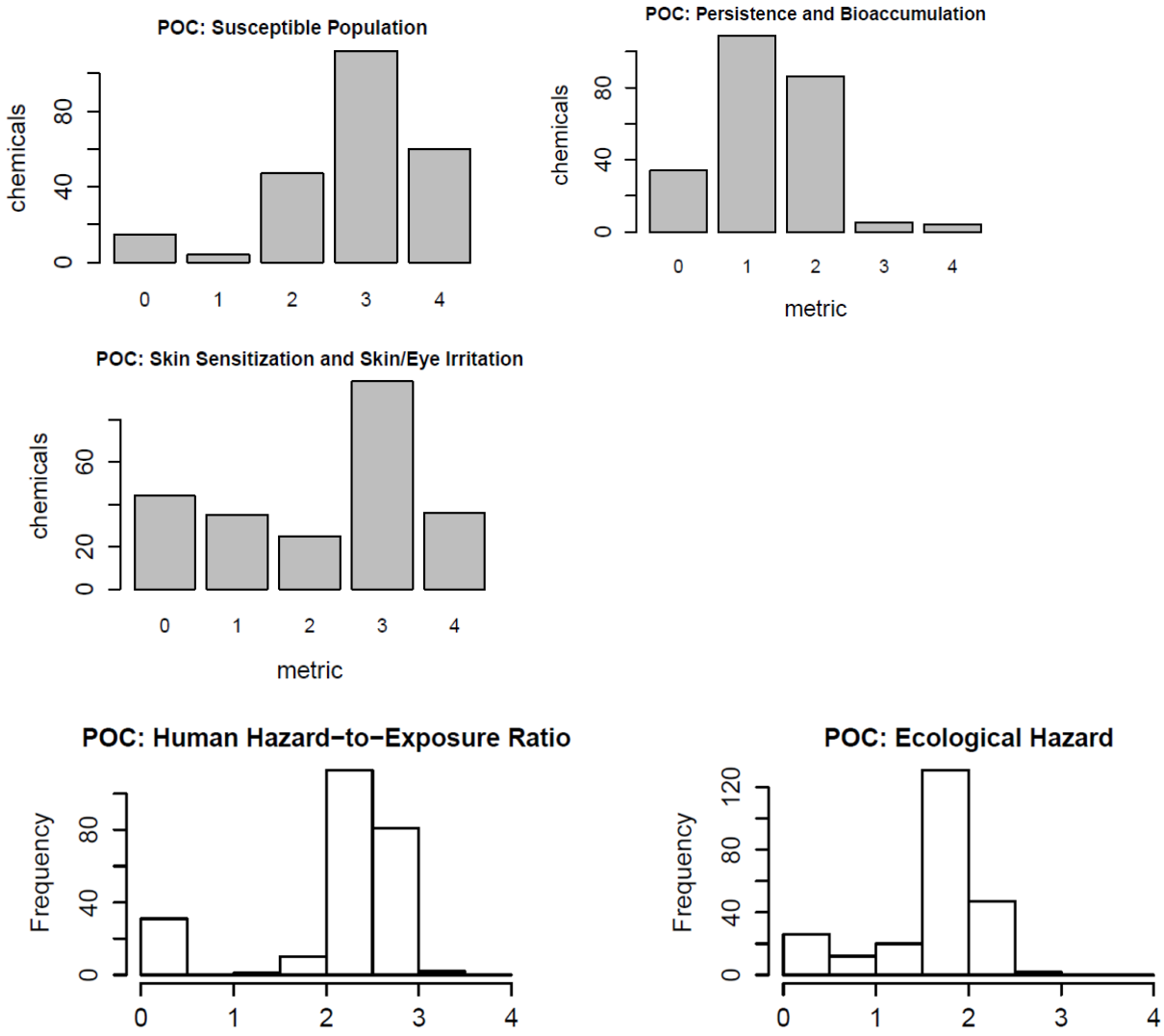


Figure F-3 – Distributions of the scaled SDM for the individual domains for the POC238 subset. The x-axis displays the scaled value and the y-axis shows the frequency of that value in the subset.

Appendix G. Information Availability Metric Calculation

Figure G-1. Flow chart explaining the Information Availability Metric (IAM) calculation used in the PICS approach.