National Institute of Environmental Health Sciences
Division of Translational Toxicology

Curating Chemical Use Categories and Exposure Predictions to Inform Chemical Assessment

V. Hull[1], T. LaPratt[1], A.L. Karmaus[1]*, K.T. To[1], A. Unnikrishnan[1], D.G. Allen[1]*, N.C. Kleinstreuer[2]

[1]*Inotiv, Research Triangle Park, NC;* [2]*NIH/NIEHS/DTT/NICEATM, Research Triangle Park, NC*

Abstract 4322
Poster P106

Introduction

- Understanding how human populations interact with, and are exposed to, chemicals is key to prioritizing chemicals for risk assessment, as risk is the intersection of chemical hazard and exposure (Wambaugh et al. 2019).
  - Many chemicals lack measured exposure estimates and chemical use information.
- High-throughput exposure simulations and structure-based chemical-use models can inform exposure scenarios for data-poor chemicals but can be difficult to interpret for users with limited computational experience.
- NICEATM seeks to curate complex datasets and present approachable and interactive use and exposure data in the Integrated Chemical Environment (ICE; https://ice.ntp.niehs.nih.gov/), an open-access resource containing toxicologically relevant data and computational tools.
  - Functional use (role a chemical plays in a product) and product use (products a chemical has been found in) categories were curated from the Environmental Protection Agency's (EPA's) Chemicals and Products Database (CPDat, Williams 2017).
  - Chemical exposure predictions from EPA's Systematic Empirical Evaluation of Models (SEEM3, Ring et al. 2019) SEEM3 were retrieved and curated.

SEEM3 Exposure Predictions

SEEM3 (**Fig. 1A**) was developed by the ExpoCast group at EPA's Center for Computational Toxicology and Exposure. It predicts both population-level exposure in mg/kg/day for over 600,000 chemicals and the potential pathway of exposure (**Fig. 1B**).

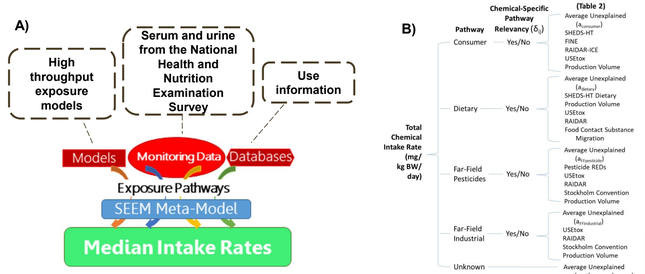

Figure 1. Overview of SEEM3 Model (Adapted from Ring et al. 2019). A) An overview of the SEEM3 model inputs and outputs. B) Exposure pathways predicted by the SEEM3 model and their respective high-throughput exposure models.

Exposure Predictions for Chemicals in ICE

SEEM3 predictions were pulled for over 600,000 chemicals in November 2022 from EPA's GitHub page (https://github.com/HumanExposure/SEEM3RPackage). Outputs of the model include the 5th, 50th, and 95th percentile of exposure in mg/kg/day. The data set was further refined to ~480,000 chemicals that were within the model's domain of applicability (**Fig. 2**).

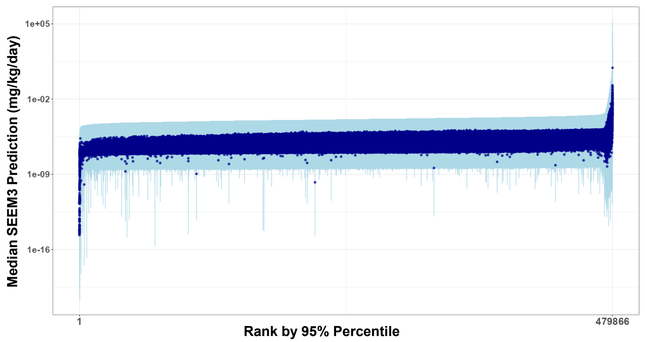

Figure 2. SEEM3 Predictions. Median SEEM3 predictions in mg/kg/day for 479,866 chemicals within the model's domain of applicability. Error bars represent the 5th and 95th percentile of exposure. Chemicals are ranked by 95% percentile.

Pathways of Exposure for Chemicals in ICE

For the ~480,000 chemicals within the domain of applicability, SEEM3 exposure pathway predictions were used to create near-field and far-field annotations in the ICE data set (**Fig. 3**). >80% of chemicals do not have a known pathway, as chemical use information is often limited.

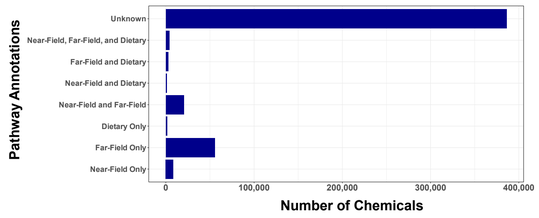

Figure 3. Chemical Exposure Pathway Annotations.

Visualize Exposure Data

In the ICE In Vitro to In Vivo Extrapolation (IVIVE) tool (**Fig. 4**), exposure predictions can be overlayed onto boxplots of equivalent administered dose (EAD). Visualizing predicted exposure beside converted hazard ranges (as characterized by EAD derived from in vitro assays) provides important context to interpreting in vitro points of departure.

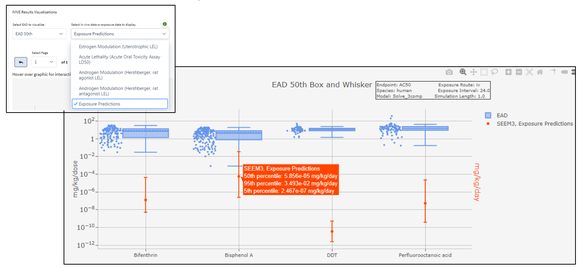

Figure 4. ICE IVIVE Tool Output with Boxplots Showing EAD and SEEM3 Predictions. 4 chemicals with EAD, calculated using curated High Throughput Screening (cHTS) in vitro data, and human exposure prediction overlays.

Visualize Product Use Data

- The ICE Chemical Characterization tool (**Fig. 5**) contains curated chemical use categories that are derived from EPA's CPDat v3, a document-based database that contains exposure-relevant data on chemicals in consumer products (Dionisio et al. 2018).
- Within ICE there are over 300 different product use categories for 4,896 unique chemicals.
  - Knowing product use can help determine factors like chemical composition, exposure frequency, and route of exposure.
- Users can visualize and extract categories through summary tables and circle plots that help represent hierarchy of terminology and relative abundance of terms in a query (**Fig. 5**).

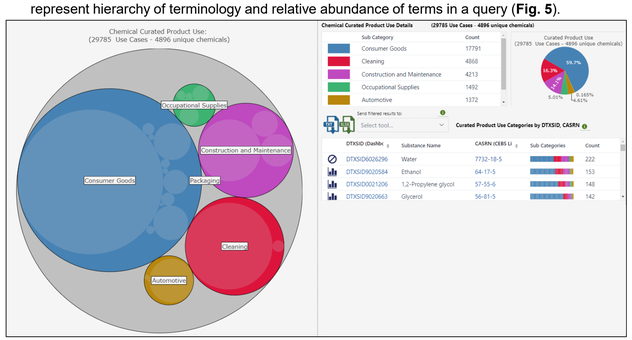

Figure 5. ICE Chemical Characterization Tool Visualizations of Product Use.

Curation of Functional Use Data

- Functional use can inform what products a chemical might be used in if product use category is not known and help estimate how much of a chemical is in a product and potential alternatives.
- Two types of functional use information were pulled from CPDat, including reported functional use harmonized to Organization for Economic Co-operation and Development (OECD) categories and predicted functional use as predicted by quantitative structure–use relationship models (QSUR, Phillips et al. 2017; **Fig. 6**).
  - As part of the curation process, reported use was harmonized to OECD use by the NICEATM team for ~2,000 chemicals.
  - For predicted functional use, the prediction probability was limited to ≥80% to ensure high-confidence results.
- The functional use dataset in ICE has 77 OECD uses for 9,395 chemicals and 37 QSUR-predicted uses for 192,438 chemicals.

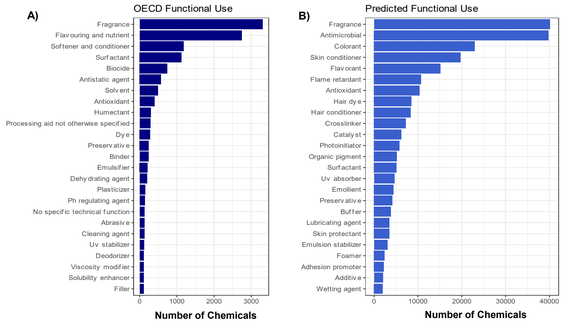

Figure 6. Functional Use Categories Most Highly Represented in ICE. The 25 most highly represented functional use categories in ICE for A) OECD and B) QSUR-predicted categories.

Visualize Functional Use Data

The ICE Chemical Characterization tool allows users to visualize and extract OECD functional use and predicted functional use via heatmaps (**Fig. 7**). Plot titles indicate the number of substances; color of the cell corresponds with the chemical input list (red or blue for individual chemical input lists; purple for both lists), and a figure legend is generated to help identify submitted query input chemical lists.

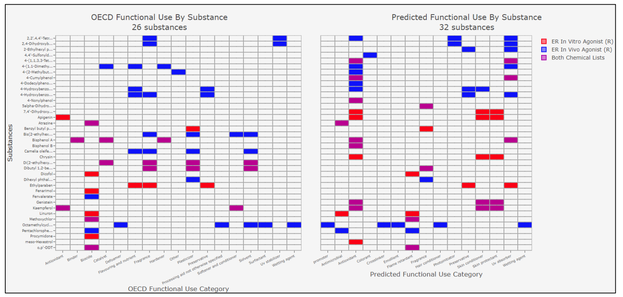

Figure 7. Visual Summaries of Functional Use in ICE.

Accessing Curated ICE Data Sets

Exposure predictions and functional use data can be downloaded directly through the ICE data sets page (https://ice.ntp.niehs.nih.gov/DATASETDESCRIPTION).

Integrated Chemical Environment

Download Data from ICE.
https://ice.ntp.niehs.nih.gov/DATASETDESCRIPTION

Future Additions: ClassyFire Chemical Taxonomies

- ClassyFire (Djoumbou et al. 2016, Wishart Research Group) is an automated, structure-based, hierarchical chemical taxonomy with 4,285 classifications organized into as many as 11 levels of classification.
- We are linking chemical use categories within ICE to ClassyFire taxonomies, allowing ICE users to identify the most abundant chemical classifications.
  - Identifying chemical classifications of interest can help focus follow-up investigations or aid in the selection of alternative chemicals.
- As an example, we applied the ClassyFire taxonomy to 100 chemicals within the OECD functional use of "biocides", which mapped to 88 chemicals within the ClassyFire Kingdom "Organic Compounds" (shown below).

Table 1. ClassyFire Taxonomies Applied to OECD Biocides in ICE. Includes SuperClass and Class levels of the ClassyFire hierarchy.

Superclass	Class	Number of Chemicals
Benzenoids	Benzene and substituted derivatives	27
	Naphthalenes	1
	Phenol ethers	4
	Phenols	4
	Triphenyl compounds	1
Lipids and lipid-like molecules	Fatty Acyls	2
	Glycerolipids	1
	Prenol lipids	3
Organic acids and derivatives	Carboxylic acids and derivatives	2
	Organic carbonic acids and derivatives	2
	Organic sulfuric acids and derivatives	1
Organic nitrogen compounds	Organonitrogen compounds	10
Organic oxygen compounds	Organooxygen compounds	9
Organoheterocyclic compounds	Azoles	1
	Azolidines	3
	Benzothiazoles	1
	Diazinanes	1
	Isoindoles and derivatives	1
	Metalloheterocyclic compounds	1
	Oxazinanes	1
	Pyridines and derivatives	3
	Quinolines and derivatives	3
	Triazinanes	1
Organosulfur compounds	Thioethers	1
	Thioureas	1
Phenylpropanoids and polyketides	Cinnamic acids and derivatives	1
	Depsides and depsidones	1

Summary and Discussion

- In our continuing efforts to provide high-confidence, high-quality toxicologically relevant data, we have added curated exposure predictions from EPA's SEEM3 model and chemical use data from EPA's CPDat to ICE.
  - This includes exposure predictions for ~480,000 chemicals, OECD functional use categories for nearly 10,000 chemicals, and predicted functional use data for approximately 192,000 chemicals.
- Data can be directly downloaded from the ICE data sets page or visualized in the IVIVE and Chemical Characterization tools.
- With the inclusion of these data, users will be able to better explore how human populations may interact with chemicals and their potential levels of exposure.
- ClassyFire chemical taxonomies will be incorporated into future releases of ICE to connect chemical structures with exposure scenarios and chemical use cases.
- The addition of exposure and use data into ICE provides additional context to assay data and facilitates the potential addition of new data sources, exposure models, and use types.

References

Dionisio et al. 2018 Sci Data 10;5:180125. https://doi.org/10.1038/sdata.2018.125.

Djoumbou et al. 2016. J Cheminfo 8:16. https://doi.org/10.1186/s13321-016-0174-y.

Phillips et al. 2017. Green Chem 19(4): 1063–1074. https://doi.org/10.1021/acs.est.8b04056.

Ring et al. 2019. Environ Sci Technol 15;53(2):719-732. https://doi.org/10.1021/acs.est.8b04056.

Wambaugh et al. 2019. Curr Opin Toxicol 15:76-92. https://doi.org/10.1016/j.cotox.2019.07.001.

Williams. 2017. The Chemical and Products Database (CPDat) MySQL Data File. EPA Center for Computational Toxicology and Exposure. Dataset. https://doi.org/10.23645/epacomptox.5352997.

Acknowledgments

The Intramural Research Program of the National Institute of Environmental Health Sciences (NIEHS) supported this poster. Technical support was provided by Inotiv under NIEHS contract HHSN273201500010C. Since the poster was written as part of the official duties of the authors, it can be freely copied. We thank Catherine Sprankle, Barbara Stevenson, and Christine Burillo-Kirch, Inotiv, for editorial input.

*A.L. Karmaus is currently affiliated with Syngenta, Greensboro, NC. D.G. Allen's current affiliation is with the International Collaboration on Cosmetics Safety, New York, NY.

Subscribe to NICEATM News email list
https://ntp.niehs.nih.gov/go/niceatm