Screening for Contaminants of Concern in Surface Water Using an Automated Exposure-Focused Workflow

Jonathan T. Wall[1], Christopher Greene[2], Helen Goeden[2] , Antony Williams[1], Jill Franzosa[1], John Wambaugh[1], Amar Singh[1], Monica Linnenbrink[1], Jason Lambert[1], and Kristin Isaacs[1]

[1]Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency

[2]Minnesota Department of Health

Jonathan Wall | wall.jonathan@epa.gov | 0000-0002-4610-8645

United States Environmental Protection Agency

EPA at 50 — Progress for a Stronger Future

Abstract

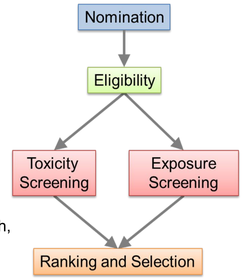
Characterization of exposure risk for chemicals found in the environment is a task that often exceeds the capacity of government agencies due to the sheer number of chemicals found in the environment. Often, a manual screening process is performed using scientifically sound approaches to prioritize chemicals for further assessment. The Minnesota Department of Health (MDH), under its Contaminants of Emerging Concern (CEC) initiative, uses a standardized process to screen potential contaminants based on exposure potential. Recently, MDH partnered with the U.S. EPA Office of Research and Development (ORD) to accelerate this process via development of an automated workflow to collect and report relevant exposure data, including New Approach Methodologies (NAMs) for exposure from ORD's ExpoCast project and relevant chemical data found on ORD's CompTox Chemicals Dashboard. The workflow pulls from 23 broad data sources, covering five domains: chemical identity and use, properties, emission and disposal, environmental occurrence, and human exposure. The collected data were used to score chemicals on exposure potential and data availability using quantitative algorithms previously developed by MDH (score range of 0-10 for exposure potential; 0-5 for data availability). A validation test, comparing the output to 88 manually screened chemicals, confirmed agreement between the processes in most data domains. Here, the automated workflow was applied to a case study of 1,762 chemicals to identify potential candidates for the CEC program. The case study data was pulled from the U.S. EPA's Multimedia Monitoring Database (MMDB) for all unique chemicals that were detected in surface water samples. The average time for a chemical to complete the workflow (pull relevant data, score, generate report) was approximately five minutes, with the quickest finishing in under two minutes and the longest taking two hours. This is an exponential increase in chemical screening rate compared to the manual process (4-10 hours per chemical). Some higher-ranking chemicals (high adjusted exposure score; data availability) included Phenol (6.30; 3.81) and Acetophenone (5.21; 3.79). This case study provides an example of how promising automated workflows can be for accelerating chemical screening and prioritization processes by leveraging existing data structures and NAMs.

Background

- Through its Contaminants of Emerging Concern (CEC) initiative, the Minnesota Department of Health (MDH) collaborates with partners and the public to identify potential drinking water contaminants of interest.
- MDH identifies contaminants of emerging concern as substances that have been released to, found in, or have the potential to enter Minnesota waters, and:
  - Pose a real or perceived health threat,
  - Do not already have Minnesota human health-based guidance, or
  - Have new or changing health or exposure information that increases the level of concern.
- Substances are chosen for the CEC initiative via a stakeholder nomination process, followed by a screening level evaluation of toxicity and exposure potential, and final ranking and selection for nominated contaminants.
  - Based on the exposure and toxicity screening results, MDH assigns a preliminary ranking of high, medium, or low to each contaminant.
  - MDH uses the preliminary ranking to inform selection of contaminants for an in-depth toxicological review and guidance development.
- Over time, MDH has modified this process to be more focused on how the contaminant will be released to water, how easily it moves through the environment, how long it might persist, how likely it is to occur in Minnesota sources of drinking water, and detection frequency and measured concentration.
- Recently, MDH partnered with the U.S. EPA Office of Research and Development (ORD) to accelerate its exposure screening process and incorporate New Approach Methodologies (NAMs) for exposure from ORD's Exposure Forecasting (ExpoCast) project, along with other relevant chemical data included in ORD's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard).

DEPARTMENT OF HEALTH

CEC Initiative



Approach

MDH and EPA-ORD worked in collaboration to plan and implement an automated workflow to collect relevant exposure data from a wide variety of existing data sources, for lists of chemical nominees. Based on these data, nominated chemicals were scored in an automated fashion according to MDH criteria and summary reports and detailed supplemental data tables were generated for each chemical. To compare workflow results against manual scoring, the workflow was applied to 88 chemicals previously evaluated by MDH. The automated workflow was then applied to a case study of 1,762 chemicals previously detected in surface water to identify potential candidates for the CEC program. The case study chemicals were identified by querying the U.S. EPA's Multimedia Monitoring Database (MMDB) for all unique chemicals that were detected in surface water samples.

Methods

Data Sources

Table 1. Data from a variety of EPA-ORD databases, other public information streams, and Minnesota-specific sources, are incorporated into 5 domains (chemical identity and property, emission and release, environmental media occurrence, and exposure) for use in the workflow

Chemical Identity and Use	
Chemical Identifiers and Synonyms	EPA-ORD's CompTox Chemicals Dashboard/Underlying Databases
Uses	EPA-ORD's Chemicals and Products Database[1] (CPDat)
Uses	EPA's Chemical Data Reporting[2] (CDR) Consumer, Commercial, Industrial uses
National Production Volume	EPA-ORD's CompTox Chemicals Dashboard (Underlying data)
Uses	EPA Safer Chemical Ingredients List[3]
Chemical Properties	
Measured Properties	EPA-ORD's CompTox Chemicals Dashboard/Underlying Databases
Predicted Properties	EPA-ORD's CompTox Chemicals Dashboard (OPERA QSAR Models[4])
Predicted Wastewater Treatment Removal	EPA's Estimation Program Interface Suite[5] (EPI-Suite)
Transformation Products	EPA-ORD's CompTox Chemicals Dashboard/Underlying Databases
Chemical Emissions and Disposal	
Pesticide Releases	National Agricultural Statistics Service[6]
Chemical Releases	EPA's Toxics Release Inventory[7]
Down-the-Drain Releases	EPA's SHEDS-HT model[8]
Chemical Occurrence in Environment, Drinking Water, and Food	
Occurrence in Environmental Media, Including Drinking and Surface Water	EPA-ORD Multimedia Monitoring Database (MMDB)
Occurrence in US Water	US Geological Survey (USGS) Water Quality Portal[9] (WQP) data, via its application programming interface (API)
Occurrence in MN Water	Custom Database developed by USGS for MDH
Occurrence in MN Water	MN-specific reports, curated into EPA's chemical databases
Occurrence in Food	US Department of Agriculture (USDA) Pesticide Data Program[10]
Occurrence in Food	US Food and Drug Administration (FDA) Substances Added to Food Database[11]
Occurrence in Food	US Food and Drug Administration (FDA) Indirect Food Additives Database[12]
Human Exposure	
Intake Exposures Inferred from Biomonitoring Data	EPA-ORD's CompTox Chemicals Dashboard/Underlying Databases
Biomonitoring Data	EPA-ORD Multimedia Modeling Database (MMDB)
Consumer Exposure Predictions	EPA-ORD's SHEDS-HT Model[2]
General Population Exposures	EPA-ORD's Systematic Empirical Evaluation of Models (SEEM) Consensus Predictions[13]
Presence on Biomonitoring Lists	Biomonitoring California[14]

Multimedia Monitoring Database **MMDB**

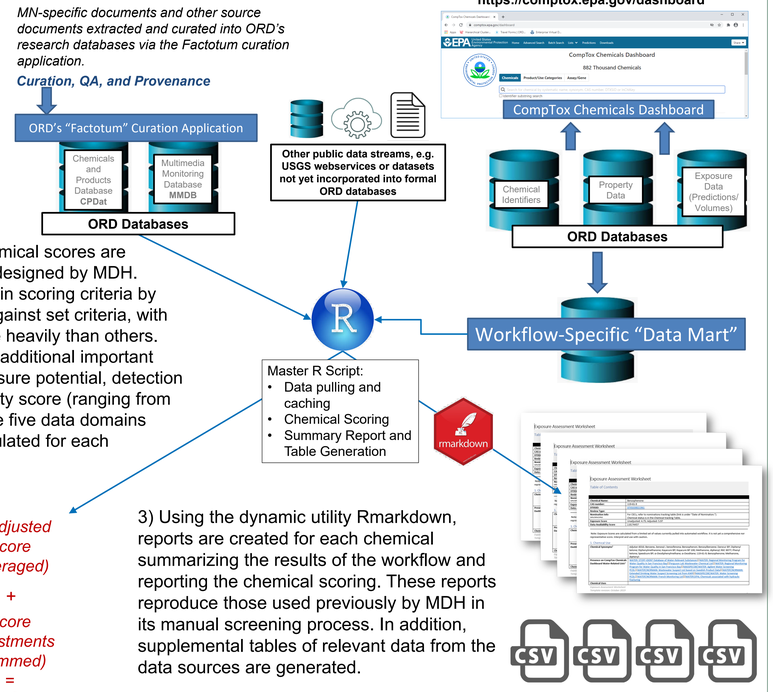
Database of measurements from over 20 public data sources, harmonized to chemical identifier and media (e.g., drinking water, surface water, human blood or urine, soil, food, and ecological species). Includes data from several EPA programs, California state monitoring programs, the FDA, the Comparative Toxicogenomics Database, the EU's Information Platform for Chemical Monitoring Data (IPCHEM), the National Health and Nutrition Examination Survey (NHANES), the USDA, The International Council for the Exploration of the Sea (ICES), International Council of Chemical Associations' Long-Range Research Initiative (ICCA-LRI).

Workflow Design and Implementation

1) For any chemical list, relevant information is pulled from data sources in an automated fashion, based on standard chemical substance identifiers (DTXSID for EPA databases, or CASRN for external sources)

MN-specific documents and other source documents extracted and curated into ORD's research databases via the Factotum curation application.

Curation, QA, and Provenance



2) Based on the available data, chemical scores are calculated according to algorithms designed by MDH. Scores are determined for three main scoring criteria by comparing data from the sources against set criteria, with some data elements weighted more heavily than others. Scores are then adjusted based on additional important factors (e.g. chemical identity, exposure potential, detection frequency). An information availability score (ranging from 0-5), quantifying the coverage of the five data domains described in Table 1, was also calculated for each chemical.

Main Scoring Criteria

Persistence and Fate
Release Potential
Occurrence
→ *Unadjusted Score (averaged)*

\+

***Scoring Adjustments* (+/-)**

*Chemical Identity**
Exposure Potential
Detection Frequency
→ *Score Adjustments (summed)*

=

Final Score

*Refers to specific questions: 1) Are there major new/expanded uses that could increase exposure? 2) Are anticipated decreases in exposure due to decreasing use, regulations, or bans?

3) Using the dynamic utility Rmarkdown, reports are created for each chemical summarizing the results of the workflow and reporting the chemical scoring. These reports reproduce those used previously by MDH in its manual screening process. In addition, supplemental tables of relevant data from the data sources are generated.

Results

Exposure Screening Case Study Surface Water Detected Chemicals

- The automated workflow was applied to the 1,762 case study chemicals (60 evaluated by MDH previously) identified in surface water from MMDB; all data collection, scoring, and report/table writing were completed in approximately 86 hours.
  - On average, chemicals completed the workflow in under five minutes, with data-rich chemicals (especially those having large amounts of USGS data to be obtained via the WQP API) taking longer (up to 2 hours).
- The automated workflow results compared favorably with manual scores for Persistent/Fate and Release Potential domains. To improve Occurrence domain comparability, additional data sources related to chemical environmental occurrence were identified for future implementation.
- There were 92 chemicals that did not have enough data for main unadjusted scores to be calculated, these chemicals could be candidates for collection or identification of additional data.
- Scoring results (final score versus information availability score) for the surface water case study chemicals are shown in Figure 1; score breakdowns are provided in Figure 2 as stacked bars to convey score domain proportionality.
- Only ~16% (12/76) of the chemicals with the highest scores (>5) had already been screened by MDH, with only 1 previously screened chemical making the top 20 (Nicotine).
  - 26% of high scoring chemicals (score >5) lacked enough data to calculate the "Occurrence" based score domain, meaning their true final score could be greater if more occurrence data was curated.
- Case study chemicals increased the breadth and spread of the chemical space represented by the original 92 MDH screened chemicals using the scoring algorithms.
- Examination of highly-scored chemicals could also be used to refine current screening algorithms (e.g., elucidation of additional criteria for refining chemical scores)

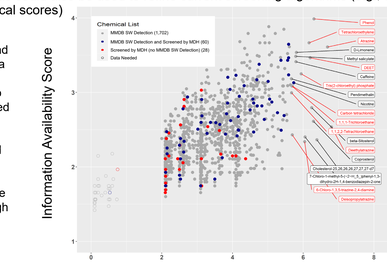


Figure 1. Final adjusted scores and information availability scores for a case study list of chemicals detected in surface water. The top 20 final adjusted scores are labeled with chemical names. Red labels indicate MDH has already developed health-based rules and guidance for the chemical. The "Data Needed" category reflects chemicals with a positive exposure scoring adjustment, but not enough data to calculate an unadjusted score.

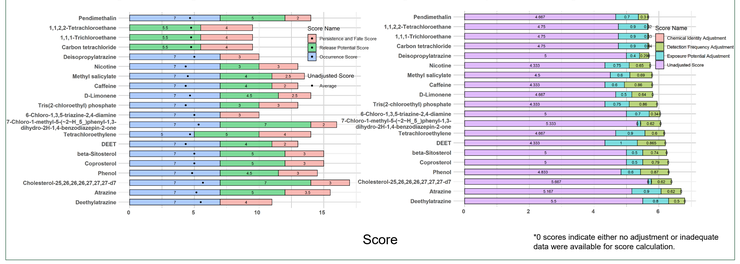


Figure 2. Scoring details for the 20 chemicals with the highest overall exposure screening scores.

Impact and Next Steps

- This workflow will allow MDH scientists to speed up screening evaluations and expand the number of chemicals assessed, freeing resources to complete the more complex aspects of exposure assessment.
- The workflow shows promise for rapid *a priori* screening of large libraries of chemicals for potential initiative nominees which would allow MDH to be more proactive in contaminant identification.

References

1. Isaacs KK, Dionisio K, Phillips K, Bevington C, Egeghy P, Price PS. Establishing a system of consumer product use categories to support rapid modeling of human exposure. J Expo Sci Environ Epidemiol. 2020;30(1):171-183. doi:10.1038/s41370-019-0187-5.
2. https://www.epa.gov/chemical-data-reporting
3. https://www.epa.gov/saferchoice/safer-ingredients
4. Mansouri K, Grulke CM, Judson RS, Williams AJ. OPERA models for predicting physicochemical properties and environmental fate endpoints. J Cheminform. 2018;10(1):10. Published 2018 Mar 8. doi:10.1186/s13321-018-0263-1
5. https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface
6. https://www.nass.usda.gov/
7. https://www.epa.gov/toxics-release-inventory-tri-program
8. Isaacs KK, Glen WG, Egeghy P, et al. SHEDS-HT: an integrated probabilistic exposure model for prioritizing exposures to chemicals with near-field and dietary sources. Environ Sci Technol. 2014;48(21):12750-12759. doi:10.1021/es502513w.
9. https://www.waterqualitydata.us/
10. https://www.ams.usda.gov/datasets/pdp
11. https://www.fda.gov/food/food-additives-petitions/substances-added-food-formerly-eafus
12. https://www.fda.gov/food/packaging-food-contact-substances-fcs/list-indirect-additives-used-food-contact-substances
13. Ring CL, Arnot JA, Bennett DH, 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):719-732. doi:10.1021/acs.est.8b04056
14. https://biomonitoring.ca.gov/