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

Maureen R. Gwinn, PhD
US Environmental Protection Agency
CoP Webinar
July 22, 2021

The views presented are those of the author and do not necessarily represent the views of the US EPA.
Toxic Substances Control Act (TSCA)

- The Toxic Substances Control Act (TSCA) regulates the introduction of new and existing chemicals.
- TSCA was amended by the Frank R. Lautenberg Chemical Safety for the 21st Century Act (June 22, 2016)
- EPA required to make a determination if a chemical substance presents an unreasonable risk of injury to human health or the environment. Determinations are risk-based.

Near-Term Strategy

- High-priority candidates selected from TSCA workplan based on priorities, quality and quantity of information, and workload
- Low-priority candidates selected from EPA SCIL, ChAMP, and OECD SIDS based on quality and quantity of information for hazard and exposure for each condition of use

Long-Term Strategy

- Identify chemicals based on a combination of risk-related scoring and information availability
- Committed to subsequent release of proof-of-concept with a small number of substances that provides operational details on the data integration, scoring, and identification of information gaps
Defining Intended Application of Public Information Curation and Synthesis (PICS) Approach

The PICS approach was intended to:

- Understand the landscape of publicly-available information on the over 33,000 substances on the active inventory
- Provide a transparent and reproducible process for integrating available information and identifying potential information gaps
- Increase efficiency and manage workload by focusing expert review on 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 safety-related information
- Organize the process into modular workflows that can be readily updated or adapted to address scientific advances and prioritization needs under other mandates

The PICS approach was not intended to:

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

https://www.epa.gov/chemical-research/translation-and-knowledge-delivery
Schematic of PICS Approach Within the Candidate Selection Process

TSCA Active Inventory (~33,000 chemicals)

Public Information Curation and Synthesis (PICS) Approach
- Scientific Domain Metric (SDM)
- Information Availability Metric (IAM)

Subset of the TSCA Active Inventory

Expert review and analysis
- Other tools
- Systematic Approach to Data Analysis
- Weight-of-Evidence evaluation

Identification of Candidate Chemical Substances
A total of 238 substances selected from the curated, non-confidential active TSCA inventory

Selection based on the following:
- Proposed set of 20 high- and 20 low-priority candidate substances
- Substances from the 2014 update to the TSCA Work Plan
- Substances with known relevance to each of the scientific domains
- Subset of chemical substances listed in the FDA’s Substances Added to Food inventory and EPA’s Safer Chemical Ingredients List (SCIL)
Proof-of-Concept

Proof-of-Concept (238 Chemicals)

Data QA/QC

- Specific data domain and data source error rates
- Data QA plan for TSCA active inventory
- FTE estimates for data QC
- QC Tool (beta)

Scientific Domain Metric

Information Availability Metric

![Proof of Concept Chart]

8
Data extracted from “Type 1” data sources
  - Type 1 data sources are publicly available and readily searchable, enabling data extraction in structured form
  - Consistent with approach outlined in the Near-term Strategy

Quality control (QC) was performed on the data for the proof-of-concept chemicals in order to:
  - Estimate the accuracy of the data used in this case study
  - Inform the development of formal quality assurance (QA) plan
  - Obtain information on the scope and resources needed to perform QC for the entire active TSCA inventory or for other sets of chemicals
Source Traceability and Error Rates:
- Transcription error rates were typically <1%.
- Lack of primary and secondary sources was ~6%.
- Lack of primary source was higher (5 – 60%)

Time Investment:
- QC review time ranged between 1 – 10 min/data point.
- For human health data, there are >2,200,000 data points for all TSCA actives requiring ~100 person years to review.
- For eco data, there are >2,700,000 data points for all TSCA actives requiring 25 person years.
- If applied to all TSCA actives, development of customized data QC tool would decrease these time frames
Proof-of-Concept

Proof-of-Concept (238 Chemicals)

Data QA/QC

- Specific data domain and data source error rates
- Data QA plan for TSCA active inventory
- FTE estimates for data QC
- QC Tool (beta)

Scientific Domain Metric

Information Availability Metric

![Proof of Concept Diagram](image-url)
Public Information Curation and Synthesis (PICS) Approach

Chemicals to be Evaluated

Human Health-to-Exposure Ratio
Carcinogenicity
Genotoxicity
Ecological Hazard
Susceptible Populations
Persistence and Bioaccumulation
Skin Sensitization Skin/Eye Irritation

Scientific Domain Metric (SDM)

Relevant Studies and Information

Modifying criteria

Information Availability Metric (IAM)

SDM

High Information, High concern

Low Information, Low concern

IAM
Seven scientific domains were selected based on:
- Previous use in TSCA prioritization activities (i.e., TSCA workplan)
- Statutory language in the amended TSCA
- Consultation with OCSPP management and staff
- Tiered workflows for each scientific domain designed based on the current state of the science
- The overall scientific domain metric is determined by summing the results from the individual scientific domain workflows
Example SDM Workflow: Human Hazard-to-Exposure Evaluation

1. Inhalation data used only if converted to mg/kg-day.
2. Tiering of sources described in the text.
**Example SDM Workflow:**
Human Hazard-to-Exposure Evaluation

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

<table>
<thead>
<tr>
<th>Metric</th>
<th>HER, BER, or TER value&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No available data (hazard or exposure)</td>
</tr>
<tr>
<td>1</td>
<td>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)</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

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

<sup>1</sup> 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 estimate; TER, TTC-to-exposure ratio calculated based on the TTC divided by the median ExpoCast exposure estimate.
Example SDM Workflow: Susceptible Populations

1. Does the chemical have available exposure source information?
   - Yes: Evaluate exposure sources
     - Consumer Sources: Children’s Products
     - Breastmilk or Formula
     - Residential Dust
     - Consumer Sources: Flooring and Related Products
     - Consumer Sources: Other Products
     - Dietary Sources
     - Far-field Sources
   - IG Flag: Metric as described in Susceptible Population Exposure Evaluation Table

   Metric: 1 - 18

2. No: Next steps based on available information.
Table 5. Criteria used to evaluate the susceptible population exposure domain metric.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Total Exposure Source Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Chemical substance had no information in the exposure source data sources</td>
</tr>
<tr>
<td>1</td>
<td>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).</td>
</tr>
<tr>
<td>2</td>
<td>Chemical had information in at least 1 data source with a combined exposure differential value corresponding to value = 1-2 (using the workflow in Figure 8)</td>
</tr>
<tr>
<td>3</td>
<td>Chemical had information in at least 1 data source with a combined exposure differential value corresponding to value = 3-7 (using the workflow in Figure 8)</td>
</tr>
<tr>
<td>4</td>
<td>Chemical had information in at least 1 data source with a combined exposure differential value corresponding to value = 8 - 18 (using the workflow in Figure 8)</td>
</tr>
</tbody>
</table>

IG Flags: predicted data; secondary source data
Public Information Curation and Synthesis (PICS) Approach

Chemicals to be Evaluated

Human Health-to-Exposure Ratio
- Carcinogenicity
- Genotoxicity
- Ecological Hazard
- Susceptible Populations
- Persistence and Bioaccumulation
- Skin Sensitization Skin/Eye Irritation

Scientific Domain Metric (SDM)

Modifying criteria

Relevant Studies and Information

Information Availability Metric (IAM)

<table>
<thead>
<tr>
<th>SDM</th>
<th>IAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Information, High concern</td>
<td>Low Information, Low concern</td>
</tr>
</tbody>
</table>
Information Availability Metric

• Included in PICS approach to evaluate the amount of information available for use in any future chemical substance risk evaluation
• Based on the potentially relevant information for exposure, human health and ecological hazard
• Modifying criteria (based on OPPT new chemicals program and consultation with OPPT technical staff) applied to make the metric context-specific
• Incorporates information gathering flags to highlight data types used in specific scientific domain metrics as well as possible data gaps
## Information Availability Metric Calculation

<table>
<thead>
<tr>
<th>Available data categories</th>
<th>Modifying Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mammalian Acute</td>
<td>None</td>
</tr>
<tr>
<td>2. One of (mammalian</td>
<td>Is there a high-</td>
</tr>
<tr>
<td>subchronic, mammalian</td>
<td>quality public</td>
</tr>
<tr>
<td>repeat dose, mammalian</td>
<td>risk assessment</td>
</tr>
<tr>
<td>chronic)</td>
<td>(cancer or non-</td>
</tr>
<tr>
<td>3. Mammalian reproductive</td>
<td>cancer)?</td>
</tr>
<tr>
<td>4. Mammalian developmental</td>
<td>Is this a chemical</td>
</tr>
<tr>
<td>5. Mammalian neurotoxicity</td>
<td>intermediate AND</td>
</tr>
<tr>
<td>6. Mammalian cancer</td>
<td>a short environmen-</td>
</tr>
<tr>
<td>7. Mammalian genotoxicity</td>
<td>tal half-life</td>
</tr>
<tr>
<td>8. Skin Sensitization or</td>
<td>(hours)?</td>
</tr>
<tr>
<td>eye corrosivity</td>
<td>Is this a chemical</td>
</tr>
<tr>
<td>9. Exposure</td>
<td>with low water</td>
</tr>
<tr>
<td>10. Eco aquatic plant</td>
<td>solubility (&lt; 0.1</td>
</tr>
<tr>
<td>acute</td>
<td>mg/L)?</td>
</tr>
<tr>
<td>11. Eco aquatic invertebrate acute</td>
<td>Add 1 for categories 1-9 with available data</td>
</tr>
<tr>
<td>12. Eco aquatic vertebrate acute</td>
<td>Add 1 for categories 1-8 with available data</td>
</tr>
<tr>
<td>13. Eco aquatic plant repeat dose</td>
<td>Add 1 for categories 8 and 9 with available data</td>
</tr>
<tr>
<td>14. Eco aquatic invertebrate repeat dose</td>
<td>Divide by the denominator (15)</td>
</tr>
<tr>
<td>15. Eco aquatic vertebrate repeat dose</td>
<td>Divide by the denominator (15)</td>
</tr>
<tr>
<td></td>
<td>Divide by the</td>
</tr>
<tr>
<td></td>
<td>denominator (9)</td>
</tr>
<tr>
<td></td>
<td>Divide by the</td>
</tr>
<tr>
<td></td>
<td>denominator (8)</td>
</tr>
<tr>
<td></td>
<td>Divide by the</td>
</tr>
<tr>
<td></td>
<td>denominator (2)</td>
</tr>
</tbody>
</table>

Scale to percent.

IAM
Public Information Curation and Synthesis (PICS) Approach

Chemicals to be Evaluated

Human Health-to-Exposure Ratio
Carcinogenicity
Genotoxicity
Ecological Hazard
Susceptible Populations
Persistence and Bioaccumulation
Skin Sensitization Skin/Eye Irritation

Scientific Domain Metric (SDM)

Relevant Studies and Information

Modifying criteria

Information Availability Metric (IAM)

SDM

High Information, High concern

Low Information, Low concern

IAM
Proof-of-Concept Results

Plot showing distributions of metric scores for selected chemical substance lists. For each list, the point shows the median scientific domain and information availability metrics. The whiskers span 90% of the distributions. Data here is taken from the lists across the TSCA Active Inventory.

Plot of the information availability vs. scientific domain metrics for the POC238 set of chemical substances. Positions of points are staggered for ease of visualization.
Proof-of-Concept Results

Di-isobutyl phthalate (DIBP)

- Scientific domain metric related to lack of human hazard assessment in the public domain.
- Near-final IRIS assessment is available (but not considered an authoritative assessment for calculating the metric).
- This result seems comparable to what was found in the TSCA evaluation scoping document.
Calcium D-Gluconate

- Scientific domain metric related to lack of human hazard assessment and ecological hazard data in the public domain.
- Similar to other chemicals on the low priority list.
- Read across may have been used for this determination.
Proof-of-Concept Results Compared to Non-Confidential TSCA Active Inventory
The proof-of-concept study demonstrated that the PICS approach:

- Generally resulted in higher metrics for the high-priority candidates compared to the low-priority candidates
  - Different data sources
  - Conflicting results from multiple studies
- Identified areas for potential information gathering
  - Informs potential areas of uncertainty
  - Focusing resources and data gathering on key gaps
- Can be scaled and applied to the TSCA inventory
Key Takeaways!

• The PICS approach:
  • Increases understanding of the landscape of publicly available information
  • Efficiently identifies high and low priority candidates among large chemical inventories for expert review
  • Provides a transparent and reproducible process for integrating available information and identifying potential information gaps
  • Incorporates results from domain-specific workflows that can be readily updated or adapted to address scientific advances and prioritization needs under other mandates
Data Curation and QC Tiger Team

- **General** – John Cowden, Richard Judson, Amar Singh
- **QC Data Integration and QA Automation Workgroup** - Richard Judson, Jeremy Dunne, Amar Singh, Chris Grulke
- **Human Health Hazard/Risk Assessment Workgroup** - Johanna Congleton, Urmila Kodavanti, Chris Lau, Mary Gilbert, Yu-Sheng Lin, Dan Vallero, Kelly Garcia, Carolyn Gigot, Andrew Greenhalgh, Allison Eames
- **Ecological Toxicity Data Workgroup** - Dale Hoff, Colleen Elonen, Leslie Hughes, Anita Pascocello
- **Exposure Data Workgroup** - Katherine Phillips, Janet Burke, Abhisheek Komandur, Ashley Jackson, Lauren Koval
- **Genotoxicity Data Workgroup** - David DeMarini, Maureen Gwinn, Catherine Gibbons, Sarah Warren, Jeff Dean, Anita Simha, Nagu Keshava
- **Chemistry Data Workgroup** - Kent Thomas, Michael Gonzalez, Doug Young, Chris Grulke
Proof-of-Concept Tiger Team

- **General** - Maureen Gwinn, Richard Judson
- **Information availability** - Tony Williams, Jeremy Dunne, Jason Lambert, Amar Singh
- **Human Hazard-to-Exposure Ratio** - Katie Paul-Friedman, John Wambaugh, Elaina Kenyon, Kristin Isaacs, Jason Lambert
- **Susceptible Population Exposure** - Kathie Dionisio, Kristin Isaacs, John Wambaugh
- **Carcinogenicity/Genotoxicity** - Grace Patlewicz, David DeMarini, Catherine Gibbons, Jeffry Dean, Anita Simha, Nagu Keshava, Todd Martin, Sarah Warren
- **Eco Hazard** - Dan Villeneuve, Carlie LaLone, Todd Martin
- **Persistence/bioaccumulation** - John Nichols, Lawrence Burkhard, Eric Weber
- **Skin sensitization/irritation and Eye irritation** - Todd Martin, Leora Vegosen
Example SDM Workflow: Carcinogenicity

1. **Determination of human carcinogenicity potential from expert review (e.g., IRIS cancer descriptor)?**
   - Yes: Evidence of carcinogenicity or low likelihood of carcinogenicity [see Metric Criteria Table (Table 2)]
   - No: Evidence of animal carcinogenicity (e.g., two-year cancer bioassay)?
2. **Evidence of animal carcinogenicity (e.g., two-year cancer bioassay)?**
   - Yes: Evidence of carcinogenicity or low likelihood of carcinogenicity [see Metric Criteria Table (Table 2)]
   - No or inconclusive: IG Flag; see Metric Criteria Table (Table 2)
# Example SDM Workflow: Carcinogenicity

Table 2. Criteria used to calculate the carcinogenicity domain metric

<table>
<thead>
<tr>
<th>Metric</th>
<th>Carcinogenicity Determination</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No available data for carcinogenicity</td>
</tr>
<tr>
<td>1</td>
<td>Evidence of low likelihood of carcinogenicity; inadequate or insufficient data</td>
</tr>
<tr>
<td>2</td>
<td>Evidence for animal carcinogenicity but not assessed for human carcinogenicity</td>
</tr>
<tr>
<td>3</td>
<td>Evidence of possible or probable human carcinogenicity based on either human epidemiology or animal toxicology data</td>
</tr>
<tr>
<td>4</td>
<td>Known human carcinogen</td>
</tr>
</tbody>
</table>

**Information Gathering (IG) Flags:** predicted data; secondary source data; determination by authoritative source
Example SDM Workflow: Genotoxicity

Are there \textit{in vitro} mutagenicity data (e.g., microbial mutagenicity, mouse lymphoma) and/or \textit{in vivo} or \textit{in vitro} clastogenicity data (e.g., micronucleus, chromosome aberrations, mouse lymphoma) available?

- Yes
  - Positive for mutagenicity and/or clastogenicity
  - Inconclusive for mutagenicity and/or clastogenicity
  - Negative for mutagenicity and/or clastogenicity

- No
  - Are there \textit{in silico} structural alerts or mutagenicity prediction available?
    - Yes
      - Positive for mutagenicity and/or clastogenicity
      - Negative for mutagenicity and/or clastogenicity
    - No
      - IG Flag for lack of information

- IG Flag for inconclusive
- IG Flag for predicted data

Metric as described in the Genotoxicity Evaluation table; if multiple metrics are possible based on the data, the highest metric will be selected.
Example SDM Workflow: Genotoxicity

<table>
<thead>
<tr>
<th>Value</th>
<th>Genotoxicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No available data for genotoxicity</td>
</tr>
<tr>
<td>1</td>
<td>Evidence of nongenotoxicity – predicted or measured data</td>
</tr>
<tr>
<td>2</td>
<td>Inconclusive evidence of genotoxicity</td>
</tr>
<tr>
<td>3</td>
<td>Evidence of genotoxicity - predicted data</td>
</tr>
<tr>
<td>4</td>
<td>Evidence of genotoxicity - measured data</td>
</tr>
</tbody>
</table>

 IG Flags: predicted data; inconclusive or lack of information; secondary source data data