Identifying, Estimating, and Communicating Uncertainty in *In Silico* Approaches to Chemical Safety Assessment

Jim Rathman, PhD

Chemical & Biomolecular Engineering
The Ohio State University, Columbus, OH
rathman.1@osu.edu

MN-AM, Nürnberg, Germany, and Columbus, OH
james.rathman@mn-am.com
Conflict of Interest Statement

- No conflicts of interest to declare
Uncertainty

- Types of uncertainty
  - Aleatory (statistical, “random error”)
  - Epistemic (knowledge limits, “ignorance”)

- Epistemic uncertainty may be reduced by improving models through incorporation of relevant mechanistic knowledge from chemistry and biology.

- Although we routinely deal with uncertainty in everyday life, many people do not expect or accept uncertainty when it come to scientific outcomes.
In Silico Approaches to Chemical Toxicity Prediction

- Quantitative structure-activity (QSAR) models
  - Global and local mode-of-action models
  - Descriptors
    - ToxPrint chemotypes (expert-defined fragments)
    - Physicochemical properties: logP, logS, TPSA, shape descriptors, etc.
    - Quantum mechanical properties: HOMO, LUMO, heat of formation

- Structural rules

- Read-across

- Weight-of-evidence combination of information
Sources of Uncertainty in QSAR and Rule-based Approaches (Cronin et al., 2019)

● Chemical structures and descriptors
  - Inaccurate structures
  - Chemical reactivity, metabolism (computational form differs from active form)

● Biological data
  - Secondary or tertiary data sources (e.g., databases, safety assessment reports) may not be precise or exhaustive, or may introduce mistakes
  - Lack of information on guideline, study design parameters, or critical effects
  - Inconsistent calls (e.g., same study with different calls depending on the regulatory body/organization responsible for the call)
  - Limited knowledge of mechanism of action
  - Relevance of data to endpoint being modeled
Descriptors
- Structural features used to fingerprint molecules
- Coverage
- Experimental/calculated properties
- Relevance of descriptors

Undefined (or poorly-defined) domain of applicability

- Some of these can be addressed by careful curation of the chemical and biological data.
- For others, we want to quantify uncertainty to estimate how these factors affect reliability of the prediction results.
Reliability Measures for QSAR Classification Models

- Performance statistics from model validation
  - Accuracy (concordance, Matthews correlation coeff)
  - Sensitivity and specificity
  - Positive and negative predictive values

- Training set and validation
  - Size and balance
  - Coverage of relevant chemical space

- Domain of applicability
Example of a chemotype alert for skin sensitization

$\alpha,\beta$-unsaturated ketone (Michael acceptor)

odds ratio = 5.28
Reliability Measures for Toxicity Studies

- ECVAM-validated methods with reliability estimates (e.g., DPRA, Keratinosens™, and h-CLAT assays for skin sensitization).

- Klimisch scoring based on assessment of how well a toxicity study conforms to internationally accepted testing guidelines.
  
  1 = reliable without restriction
  2 = reliable with restriction
  3 = not reliable
  4 = not assignable

(Klimisch et al., 1997)

- Klimisch scores can only be determined and verified if the original study data are available.
Strategies for Combining Multiple Pieces of Evidence

Bacterial reverse mutation assay (Ames test)

Mesotrione

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority wins (voting)</td>
<td>NEGATIVE</td>
</tr>
<tr>
<td>One or more ($N$ or more)</td>
<td>POSITIVE</td>
</tr>
<tr>
<td>Probability bounds analysis</td>
<td>?</td>
</tr>
</tbody>
</table>

Ames assay images: www.mun.ca/biology/scarr/4241_Ames_test_reversion.html
Probability Bounds Methodology (Rathman et al., 2018) (Based on Dempster-Shafer Theory, DST)

- Provides a rigorous approach for:
  - Estimating uncertainty
  - Combining multiple sources of evidence to make a decision

- Allows us to explicitly take into account:
  - Reliability of quantitative structure-activity (QSAR) models
  - Reliability of structural rules ("alerts")
  - Reliability of experimental results from in vitro assays and toxicity studies
Example: Skin Sensitization Hazard Prediction

QSAR Model A
PPV = 0.60
NPV = 0.50

QSAR Model B
PPV = 0.80
NPV = 0.80

\[ p(POS) = 0.70 \]

\[ p(POS) = 0.60 \]
Prediction and Combination

Source 1
(QSAR Model A)

Reliabilities: 60% 50%

Prediction: equivocal
Probability Bounds Analysis of Single Sources

Evidence

Source 1

0.70  0.30

Reliabilities: 60%  50%

0.42  0.43  0.15

Source 2

0.60  0.40

80%  80%

0.48  0.20  0.32
Dempster Combination Rule

Source 1: 0.42 \times 0.48

Source 2: 0.48 \times 0.32
Dempster Combination Rule

Source 1

0.42 × 0.43 = 0.15

Source 2

0.48 × 0.20 = 0.32

0.42 × 0.20 = 0.08

0.42 + 0.32 + 0.08 = 0.82
Dempster Combination Rule

Source 1

0.42 0.43 0.15

Source 2

0.48 0.20 0.32

0.42 × 0.48 + 0.42 × 0.20 + 0.43 × 0.48 = 0.49
Dempster Combination Rule

Source 1

\[
\begin{align*}
0.42 & \times 0.48 \\
0.15 & \times 0.43
\end{align*}
\]

Source 2

\[
\begin{align*}
0.48 & \times 0.20 \\
0.32 & \times 0.42
\end{align*}
\]

\[
\begin{align*}
0.42 \times 0.48 & + 0.43 \times 0.48 = 0.49 \\
0.15 \times 0.32 & + 0.15 \times 0.20 + 0.43 \times 0.32 = 0.21
\end{align*}
\]
Dempster Combination Rule

Source 1

\[
\begin{align*}
0.42 & \times 0.48 + 0.43 & \times 0.20 + 0.15 & \times 0.32 \\
+ 0.43 & \times 0.48 & = 0.49 \\
+ 0.15 & \times 0.20 & + 0.43 & \times 0.32 & = 0.21 \\
& + 0.43 & \times 0.20 & = 0.09
\end{align*}
\]
Dempster Combination Rule

\[
\begin{align*}
\text{Source 1} & : 0.42 \times 0.48 + 0.42 \times 0.20 + 0.43 \times 0.48 = 0.49 \\
\text{Source 2} & : 0.48 \times 0.32 + 0.48 \times 0.20 + 0.20 \times 0.32 = 0.21
\end{align*}
\]
Dempster Combination Rule

Source 1

0.42 0.43 0.15

Source 2

0.48 0.20 0.32

0.42 \times 0.48 + 0.42 \times 0.20

+ 0.43 \times 0.48 = 0.49

0.15 \times 0.32 + 0.15 \times 0.20 + 0.43 \times 0.32 = 0.21

0.43 \times 0.20 = 0.09

Normalize

0.49 0.09 0.21

0.62 0.11 0.27

Prediction: POSITIVE
Yager Combination Rule

Conflicts increase uncertainty
Yager Combination Rule

\[
\begin{align*}
\text{Source 1} & : 0.42 \times 0.48 + 0.43 \times 0.48 + 0.15 \times 0.32 \\
\text{Source 2} & : 0.48 \times 0.20 + 0.42 \times 0.20 + 0.32 \times 0.48 \\
\end{align*}
\]

Prediction: EQUIVOCAL
Inagaki Combination Rule

**Source 1**
- 0.42
- 0.43
- 0.15

**Source 2**
- 0.48
- 0.20
- 0.32

**Dempster Rule**
- 0.62
- 0.11
- 0.27
Prediction: **POSITIVE**

**Yager Rule**
- 0.49
- 0.30
- 0.21
Prediction: **EQUIVOCAL**

**Inagaki Rule**
- 0.56
- 0.19
- 0.25
Prediction: **POSITIVE**
When Evidence Sources Disagree

Source 1

Evidence

Source 2

Reliabilities: 90% 90% 90% 90%

Inagaki Rule

Prediction: EQUIVOCAL
Ordinal Classification

Consider a four-level classification model for skin sensitization:

- non-sensitizer (N)
- weak (W)
- moderate (M)
- strong (S)

The focal elements can be defined such that the model has 8 possible prediction outcomes:

- N
- N or W
- W
- W or M
- M
- M or S
- S
- N or W or M or S

DST allows us to capture different degrees of uncertainty.
## Quantifying Uncertainty of *In Vitro* Assays

### Example: skin sensitization (LLNA) for benzaldehyde

![Benzaldehyde](image)

### Performance Metrics

<table>
<thead>
<tr>
<th>In vitro assay</th>
<th>Assay result</th>
<th>Probability Bounds</th>
<th>LLNA Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DPRA</strong></td>
<td>negative</td>
<td>0.57 - 0.43</td>
<td>negative</td>
</tr>
<tr>
<td><strong>KeratinoSens</strong></td>
<td>positive</td>
<td>0.15 - 0.85</td>
<td>positive</td>
</tr>
<tr>
<td><strong>h-CLAT</strong></td>
<td>positive</td>
<td>0.15 - 0.85</td>
<td>positive</td>
</tr>
</tbody>
</table>

(Urbisch et al., 2015)
## Combining Evidence from Multiple Ames Assays

### Mesotrione

<table>
<thead>
<tr>
<th>Ames Assay Result</th>
<th>Study Reliability</th>
<th>Probability Bounds</th>
<th>DST Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative Ames test</td>
<td>0.50</td>
<td>0.50</td>
<td>0.74</td>
</tr>
<tr>
<td>Negative Ames test</td>
<td>0.95</td>
<td>0.95</td>
<td>0.74</td>
</tr>
<tr>
<td>Negative Ames test</td>
<td>0.80</td>
<td>0.80</td>
<td>0.74</td>
</tr>
</tbody>
</table>

WoE Outcome NEGATIVE
Reliability Measures for Read-Across

- Read-across: using data available for suitable analogs to infer toxicity of a target compound
- Reliability depends on (Schultz et al., 2019)
  - Similarity of analog(s) to the target
    - structure similarity
    - property similarity
    - metabolic similarity
    - toxicodynamic similarity
    - toxicokinetic similarity
  - reliability of tox data available for analogs
Combining Evidence from an Analog for Read-Across

<table>
<thead>
<tr>
<th>Ames Assay Result</th>
<th>Study Reliability</th>
<th>Analog Quality</th>
<th>Probability Bounds</th>
<th>DST Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative Ames test</td>
<td>0.50</td>
<td>0.62</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>Negative Ames test</td>
<td>0.95</td>
<td></td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>Negative Ames test</td>
<td>0.80</td>
<td></td>
<td>0.50</td>
<td></td>
</tr>
</tbody>
</table>

WoE Outcome: NEGATIVE

**Target:** metabolite of Mesotrione

**Analog:** Mesotrione
Read-Across Example with Multiple Analogs

Read-across for repeated-dose toxicity of dihydro-\textit{\alpha}-terpineol from menthol and menthanediol.

<table>
<thead>
<tr>
<th>dihydro-\textit{\alpha}-terpineol</th>
<th>menthol</th>
<th>menthanediol</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Dihydro-\textit{\alpha}-terpineol" /></td>
<td><img src="image2" alt="Menthol" /></td>
<td><img src="image3" alt="Menthanediol" /></td>
</tr>
</tbody>
</table>
## Compound Summary

<table>
<thead>
<tr>
<th>CMS ID</th>
<th>Target</th>
<th>Analog 1</th>
<th>Analog 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Chemical Structures" /></td>
<td><img src="image2.png" alt="Chemical Structures" /></td>
<td><img src="image3.png" alt="Chemical Structures" /></td>
<td></td>
</tr>
</tbody>
</table>

## Data Summary

<table>
<thead>
<tr>
<th></th>
<th>Target</th>
<th>Analog 1</th>
<th>Analog 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studies</td>
<td>0</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

## Fingerprints

<table>
<thead>
<tr>
<th>Fingerprint</th>
<th>Tanimoto</th>
<th>Tanimoto</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDKit MolFingerprint</td>
<td>0.8</td>
<td>0.76</td>
</tr>
<tr>
<td>ToxPrint Fingerprint</td>
<td>0.5</td>
<td>0.71</td>
</tr>
</tbody>
</table>

## Skyline Profiles

<table>
<thead>
<tr>
<th>Skyline Terpineol</th>
<th>Skyline</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image4.png" alt="Skyline Bar Chart" /></td>
<td><img src="image5.png" alt="Skyline Bar Chart" /></td>
</tr>
<tr>
<td>Pearson correlation coefficient</td>
<td>1</td>
</tr>
</tbody>
</table>

## Analogue Quality

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.74</td>
<td>0.81</td>
</tr>
<tr>
<td>Study Type</td>
<td>Description</td>
<td>Outcome</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>Short-term RDT</td>
<td>Rat, oral-gavage, 28 days</td>
<td>LOEL = 200 mg/kg BW/day, Liver</td>
</tr>
<tr>
<td>Subchronic RDT</td>
<td>Rat, oral, 91 days</td>
<td>NOEL = 50 mg/kg BW/day, Organ Weight</td>
</tr>
<tr>
<td>Chronic RDT</td>
<td>Rat, oral, 730 days</td>
<td>NOAEL = 750 mg/kg BW/day, Body Weight</td>
</tr>
<tr>
<td>DART Study</td>
<td>Rat, DART</td>
<td>NOAEL = 400 mg/kg BW/day, Pub Weight</td>
</tr>
<tr>
<td>Predicted Toxicity</td>
<td>target</td>
<td>analog 1</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------</td>
<td>----------</td>
</tr>
<tr>
<td>Cleft Palate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability Bar</td>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>Oral hDILI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Call</td>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>Probability Bar</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analogue Quality</th>
<th></th>
<th>0.74</th>
<th>0.81</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIER 1 (Analogue+Exp)</td>
<td>negative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIER 2 (Analogue+Exp+In silico)</td>
<td>negative</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dealing with Uncertainty in the Real World

We are all experienced safety and risk assessors!
Uncertainty About Uncertainty

- We want to make good decisions using methods that are
  - Transparent
  - Interpretable and mechanistic
  - As simple as possible

- Although we often reason probabilistically and understand that certain evidence may be more reliable than other, we may not feel confident in our ability to quantify these.

- We are often uncomfortable reporting decisions with any appreciable uncertainty, or if there are conflicting pieces of evidence.
Subjectivity

- When dealing with complex problems, we should expect that experts evaluating the same evidence will not always agree.
- Differences will in part be due to how each person evaluates the reliability and relevance of different sources of evidence.
- Probability bounds analysis provides a systematic and rigorous approach to help focus the conversation and identify why we reach different conclusions.
Communicating Uncertainty: Language Matters!

- Explore different ways to communicate
  - pictograms
  - verbal qualifiers: “likely,” “perhaps”
  - numeric ranges: “50 to 70%,” “less than 70%,” “at most 70%”…

- Studies suggest communicating uncertainty using natural frequencies instead of probabilities can be effective: “3 out of 5” or “36 out of 60” instead of 60%.

Communicating uncertainty is an interesting and active area of study in Psychology, Public Policy, and Engineering.

International Conference on Uncertainty in Risk Analysis: Challenges and Advances in Assessing, Managing and Communicating Uncertainty, Sponsored by Bundesinstitute für Risikobewertung (BfR) and European Food Safety Authority (EFSA), Berlin, Feb 20-22, 2019.


Acknowledgements

MN-AM (Molecular Networks GmbH and Altamira LLC)

The Ohio State University