Variability of Reference Data

Agnes L. Karmaus
Principal Toxicologist, Inotiv in support of NICEATM

SACATM Meeting
September 22, 2022

Disclaimer: Inotiv staff provide technical support for NICEATM, but do not represent NIEHS, NTP, or the official positions of any federal agency.
Introduction

• Data from guideline studies are used by regulatory agencies to make decisions about chemical classification and labeling

• In vivo guideline studies have been the reference upon which alternative method performance is often assessed
  – Do we reproduce the same outcome (sufficiently sensitive alternatives)?
  – Affects our confidence and context for interpreting results

• Better characterizing the in vivo guideline study reproducibility could provide additional insight to set an appropriate expectation for alternatives

• Reproducibility evaluation has been conducted for 3 regulatory endpoints:
  – Eye Irritation, Skin Sensitization, and Acute Oral Toxicity
Evaluating Reproducibility

Assessing Impact on Categorical Endpoints

- Many guideline studies are interpreted by hazard category classification
- Variability cannot be assessed quantitatively (e.g., by standard deviation)
- Instead, reproducibility is evaluated to determine how often the same category is identified across replicate studies

### Chemical X

<table>
<thead>
<tr>
<th>Prior type</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1: category 3</td>
<td>1</td>
<td>25%</td>
<td>50%</td>
<td>25%</td>
<td>-</td>
</tr>
<tr>
<td>Study 2: category 2</td>
<td>2</td>
<td>25%</td>
<td>50%</td>
<td>25%</td>
<td>-</td>
</tr>
<tr>
<td>Study 3: category 2</td>
<td>3</td>
<td>25%</td>
<td>50%</td>
<td>25%</td>
<td>-</td>
</tr>
<tr>
<td>Study 4: category 1</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Reproducibility of Categorical Outcomes

Rabbit Draize Eye Test

GHS Classification

- **Category 1**: Effects on the cornea, iris or conjunctiva that are not expected to reverse or that have not fully reversed within 21 days.
- **Category 2A**: Effects on the cornea, iris or conjunctiva that fully reverse within 21 days.
- **Category 2B**: Effects on the cornea, iris or conjunctiva that fully reverse within 7 days.

<table>
<thead>
<tr>
<th>Prior type</th>
<th>1</th>
<th>2A</th>
<th>2B</th>
<th>NC</th>
<th>Total Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>73%</td>
<td>16.1%</td>
<td>0.4%</td>
<td>10.4%</td>
<td>46</td>
</tr>
<tr>
<td>2A</td>
<td>4.2%</td>
<td>32.9%</td>
<td>3.5%</td>
<td>59.4%</td>
<td>138</td>
</tr>
<tr>
<td>2B</td>
<td>0.2%</td>
<td>4%</td>
<td>15.5%</td>
<td>80.2%</td>
<td>86</td>
</tr>
<tr>
<td>NC</td>
<td>1.1%</td>
<td>3.5%</td>
<td>1.5%</td>
<td>93.9%</td>
<td>400</td>
</tr>
</tbody>
</table>

- ECHA database evaluation
- GHS hazard categories
- 491 substances with at least 2 Draize eye studies

Luechtefeld et al., 2016. ALTEX 33(2)
# Reproducibility of Categorical Outcomes

## Acute Dermal Skin Irritation/Corrosion

<table>
<thead>
<tr>
<th>EPA Category</th>
<th>Category I</th>
<th>Category II</th>
<th>Category III</th>
<th>Category IV</th>
<th>Total Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDII</td>
<td>Corrosive</td>
<td>&gt;5.0</td>
<td>2.1-5.0</td>
<td>0-2.0</td>
<td></td>
</tr>
<tr>
<td>Signal Word</td>
<td>DANGER</td>
<td>WARNING</td>
<td>CAUTION</td>
<td>CAUTION</td>
<td></td>
</tr>
<tr>
<td>PPE Required</td>
<td>Coveralls worn over long-sleeved shirt and long pants</td>
<td>Coveralls worn over short-sleeved shirt and short pants</td>
<td>Long-sleeved shirt and long pants</td>
<td>Long-sleeved shirt and long pants</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Socks; chemical-resistant footwear</td>
<td>Socks; chemical-resistant footwear</td>
<td>Socks; shoes</td>
<td>Socks; shoes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Waterproof or chemical-resistant gloves</td>
<td>Waterproof or chemical-resistant gloves</td>
<td>Waterproof or chemical-resistant gloves</td>
<td>No minimum</td>
<td></td>
</tr>
</tbody>
</table>

- ECHA database evaluation
- EPA hazard categories
- 425 substances with at least two studies

Rooney et al., 2021. Reg Tox Pharm 122:104920
### Reproducibility of Categorical Outcomes

#### Rat Acute Oral Toxicity

**EPA Categories**

- I (≤ 50 mg/kg)
- II (>50 ≤ 500 mg/kg)
- III (>500 ≤ 5000 mg/kg)
- IV (>5000 mg/kg)

<table>
<thead>
<tr>
<th>Prior type</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>Total Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>57.9%</td>
<td>34.5%</td>
<td>6.2%</td>
<td>1.3%</td>
<td>446</td>
</tr>
<tr>
<td>II</td>
<td>5.7%</td>
<td>66.5%</td>
<td>27.5%</td>
<td>0.4%</td>
<td>1694</td>
</tr>
<tr>
<td>III</td>
<td>0.5%</td>
<td>11%</td>
<td>79.8%</td>
<td>8.7%</td>
<td>4646</td>
</tr>
<tr>
<td>IV</td>
<td>0.1%</td>
<td>0.6%</td>
<td>44.7%</td>
<td>54.6%</td>
<td>788</td>
</tr>
</tbody>
</table>

- Comprehensive compilation of data from multiple global resources
- Data heavily curated manually
- Includes limit tests and point estimate data

*Karmaus et al., 2021. Tox Sci 188(1)*
Reproducibility of Categorical Outcomes

Rat Acute Oral Toxicity

- Comprehensive compilation of data from multiple global resources
- Data heavily curated manually
- Includes limit tests and point estimate data

Karmaus et al., 2021. Tox Sci 188(1)
Defining a Margin of Uncertainty

- Curated point estimate LD50 values were used to compute a margin of uncertainty.
- Bootstrapping across MADs derived from replicate LD50 values per chemical.
- Blue shading shows defined range $0.24 \log_{10}(\text{mg/kg})$ encompasses most experimental LD50 values.

Karmaus et al., 2021. Tox Sci 188(1)
Summary

• Replicate study data are available for many guideline in vivo studies

• Evaluating variability and reproducibility across existing in vivo studies can:
  – Provide context on existing guideline studies to better characterize reference data
  – Help set expectations for evaluating new alternative methods
  – Define a margin of uncertainty which can be applied to in silico predictions and alternative methods