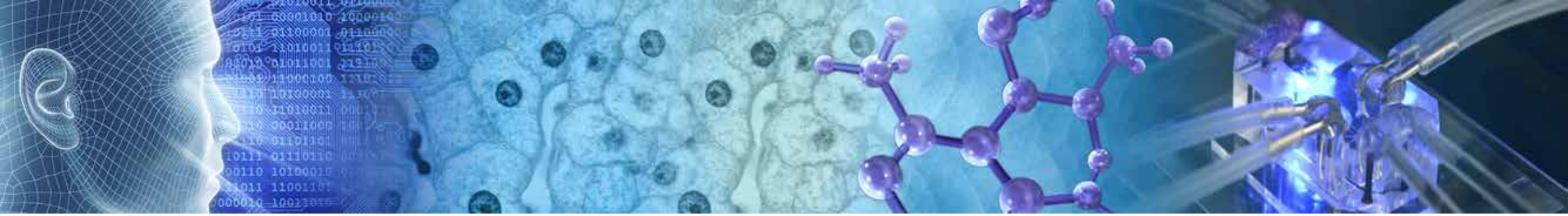


Variability of Reference Data

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

Study 1: category 3

Study 2: category 2

Study 3: category 2

Study 4: category 1

Prior type	1	2	3	4	Total Studies
1	25%	50%	25%	-	1
2	25%	50%	25%	-	2
3	25%	50%	25%	-	1
4	-	-	-	-	0

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.

Prior type	1	2A	2B	NC	Total Studies
1	73%	16.1%	0.4%	10.4%	46
2A	4.2%	32.9%	3.5%	59.4%	138
2B	0.2%	4%	15.5%	80.2%	86
NC	1.1%	3.5%	1.5%	93.9%	400

- 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

	Irritant		Non-irritant	
EPA	**Category I**	**Category II**	**Category III**	**Category IV**
PDII	Corrosive	>5.0	2.1-5.0	0-2.0
Signal Word	DANGER	WARNING	CAUTION	CAUTION
PPE Required	Coveralls worn over long-sleeved shirt and long pants	Coveralls worn over short-sleeved shirt and short pants	Long-sleeved shirt and long pants	Long-sleeved shirt and long pants
	Socks; chemical-resistant footwear	Socks; chemical-resistant footwear	Socks; shoes	Socks; shoes
	Waterproof or chemical-resistant gloves	Waterproof or chemical-resistant gloves	Waterproof or chemical-resistant gloves	No minimum

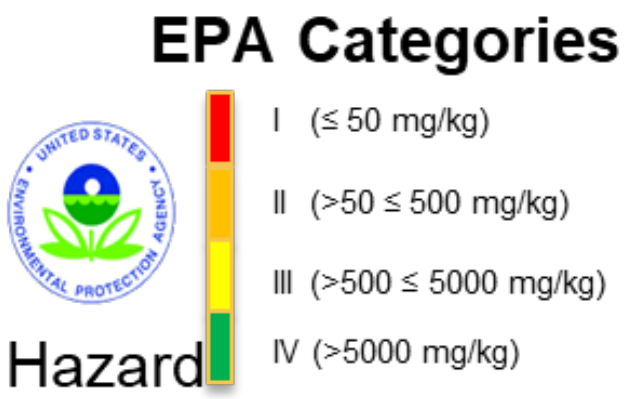
- ECHA database evaluation
- EPA hazard categories
- 425 substances with at least two studies

Prior type	I (Corrosive)	II	III	IV	Total Studies
I (Corrosive)	86.3%	4.2%	7.1%	2.5%	207
II	14.1%	44.9%	20.5%	20.5%	35
III	6.9%	5.2%	53.6%	34.3%	133
IV	0.9%	2.0%	9.1%	88.0%	690

Rooney *et al.*, 2021. Reg Tox Pharm 122:104920

Reproducibility of Categorical Outcomes

Rat Acute Oral Toxicity



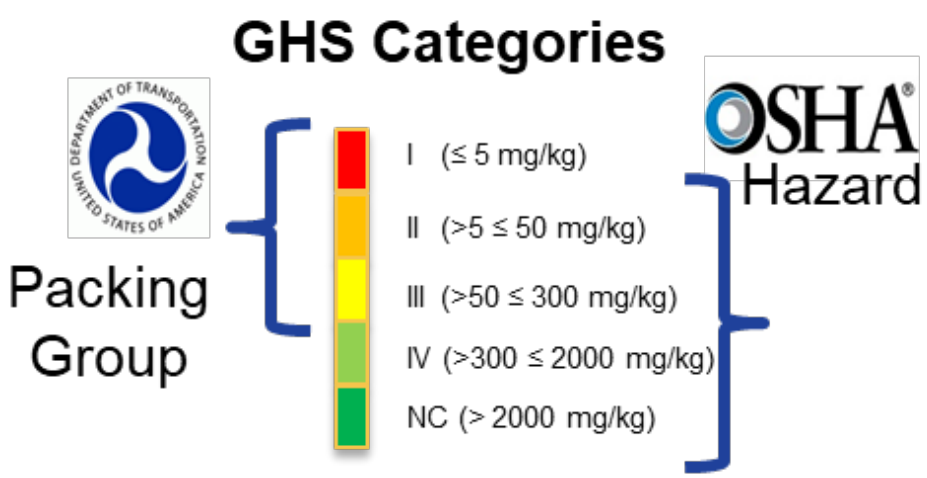
Prior type	I	II	III	IV	Total Studies
I	57.9%	34.5%	6.2%	1.3%	446
II	5.7%	66.5%	27.5%	0.4%	1694
III	0.5%	11%	79.8%	8.7%	4646
IV	0.1%	0.6%	44.7%	54.6%	788

- 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

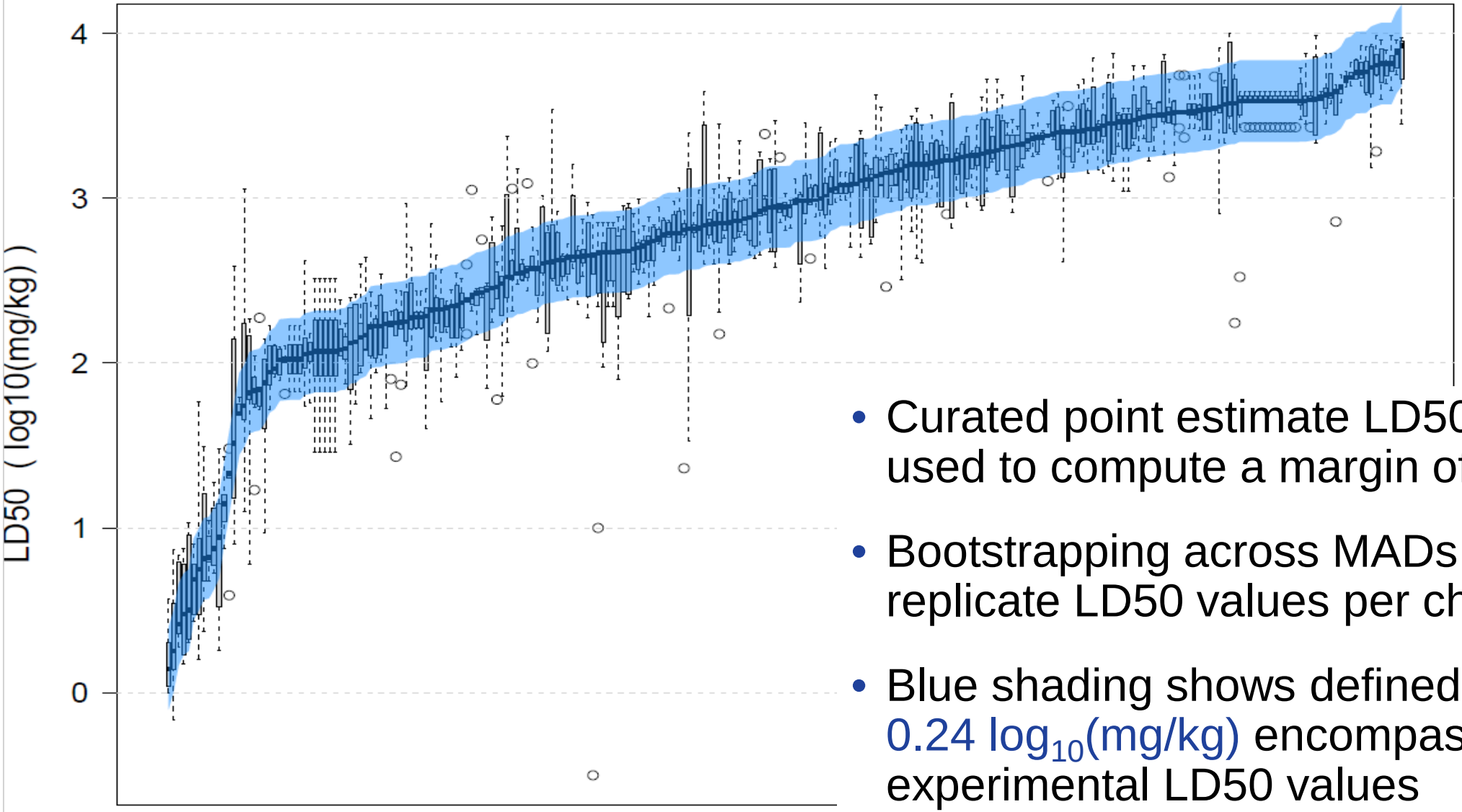


Prior type	1	2	3	4	5	Total Studies
1	53.3%	34.9%	1.5%	5.1%	5.1%	104
2	7.7%	48.9%	33.2%	8.9%	1.3%	342
3	0.2%	7.1%	61.9%	28.9%	1.9%	1166
4	0.1%	1%	11%	66.1%	21.8%	3095
5	0%	0.2%	1%	23.8%	75%	2867

- 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 $\log_{10}$(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

NTP
National Toxicology Program




National Toxicology Program
U.S. Department of Health and Human Services

Integrated Chemical Environment

NICEATM
ntp.niehs.nih.gov/go/niceatm