National Institute of Environmental Health Sciences
Division of Translational Toxicology

Variability of In Vivo Toxicology Studies: Impact on NAMs

O.B. Oyetade[1], A.L. Karmaus[1*], E. Reinke[1], D.G. Allen[1], N. Kleinstreuer[2]

[1]Inotiv, RTP, NC, USA; [2]NIH/NIEHS/DTT/NICEATM, RTP, NC, USA

Introduction

- Guideline in vivo toxicology studies are the default approach for chemical safety assessments for regulatory decision-making and the reference against which new approach methodologies (NAMs) are evaluated.
- Retrospective analyses have revealed substantial variability in data from many in vivo study types, which can confound the use of in vivo studies as reference data for developing or establishing confidence in NAMs.
  - Many in vivo studies are used based on historical practice, without any scientific or statistical support through validation.
- In this study, we investigated the variability and reproducibility of in vivo study data to provide a more realistic context to existing data streams and help set appropriate expectations for the overall performance of alternatives.

Summary

- Unlike more recently adopted guidelines for in vitro tests, in vivo toxicity tests described in guidelines issued by the Organisation for Economic Co-operation and Development (OECD) generally lacked comprehensive validation studies.
  - Inadequate validation of a method can result in data generated using that method lacking robustness and reproducibility.
  - Data generated using assays described in internationally harmonized test guidelines for ocular and dermal irritation and acute oral toxicity show substantial variability.
  - Conditional probabilities for replicating a regulatory classification outcome were low in all test methods evaluated, particularly for classifications of mild and moderate toxicity.
  - In many cases the probability that two tests of the same chemical would result in the same classification was less than 50%.
- Inherent variability across rat acute oral toxicity LD50 values was quantitatively characterized by computing a margin of uncertainty using statistical bootstrapping.
- Future efforts will include assessing how variability and validation can impact confidence in NAMs.

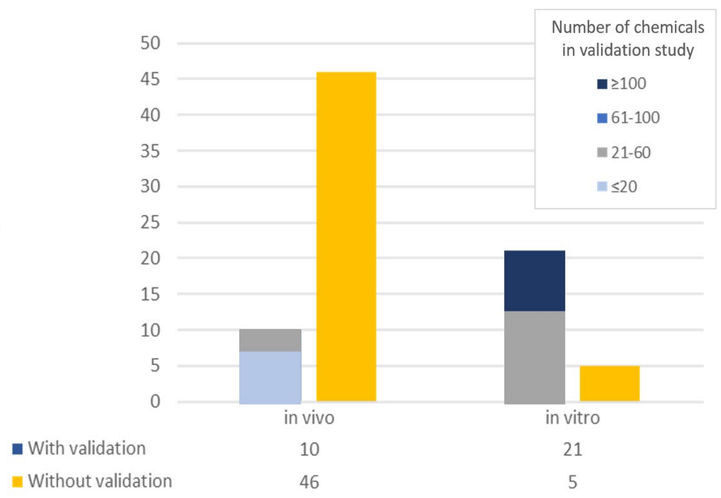
Literature Review for Existing Variability Assessments and Validation Status of Guideline Studies

Literature Search Methods

- PubMed (including MEDLINE) and Causaly databases were used to search for MeSH terms and text words including “variability”, “reproducibility”, and their variants, combined with “in vivo”, “animal studies”, “experimental studies” and other relevant synonyms.
- Only variability analyses, systematic reviews, and meta-analyses of in vivo toxicological studies were retrieved.
- Applicable test guidelines were pulled from the OECD Guidelines for the Testing of Chemicals (Section 4, Health Effects) for assessment of applicable validation studies that were conducted prior to development of the test guideline.

Literature Search Summary

- A total of 8530 manuscripts were initially identified.
- Duplicate studies and studies prior to 1990 were removed.
- Manual screening of title, abstract, and full text was used to exclude non-relevant articles.
- This curation resulted in approximately 100 papers being brought forward for further review and inclusion, in addition to the applicable OECD Test Guidelines for both in vivo and in vitro test methods.
- Variability analyses were evaluated, and validation studies were summarized to characterize established methods.
- In all, 56 in vivo and 26 toxicologically relevant in vitro test guidelines were analyzed for dates/existence of validation studies along with total numbers of chemicals utilized in the validation study (summarized in bargraph at right).



Summary of validation status for 56 in vivo and 26 in vitro assays. Literature sources and the OECD Test Guideline Library were mined to evaluate which relevant assays had had validation studies conducted. Results are sorted both by whether validation studies were conducted and by the number of reference chemicals included in the validation study.

Categorical Reproducibility

Three study types were selected to assess the **conditional probabilities** of a test chemical being in the same hazard category (GHS or EPA) if tested multiple times.

Acute Ocular Irritation/Corrosion: In Vivo Rabbit Eye Test

- 491 substances with at least 2 in vivo eye studies retrieved from the ECHA database.

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.

Likelihood of GHS category classification in independently conducted replicate assays

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

Luechtefeld et al., 2016. ALTEX 33(2)

Acute Dermal Irritation/Corrosion: In Vivo Rabbit Skin Test

- 425 substances with at least 2 in vivo skin studies retrieved from the ECHA database.

EPA Classification

- **Category I**: Irritant, Corrosive (DANGER)
- **Category II**: Irritant, PDII>5.0 (WARNING)
- **Category III**: Non-Irritant, PDII 2.1-5.0 (CAUTION)
- **Category IV**: Non-Irritant, PDII 0-2.0 (CAUTION)

Likelihood of EPA category classification in independently conducted replicate assays

Prior Type	I	II	III	IV	Total Studies
I	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

Acute Oral Toxicity: Rat Acute Lethality Test

- 2441 substances with at least 2 rat acute toxicity studies retrieved from numerous international databases.

GHS Classification

- **Category I**: LD50 ≤ 5 mg/kg
- **Category II**: 5 mg/kg < LD50 ≤ 50 mg/kg
- **Category III**: 50 mg/kg < LD50 ≤ 300 mg/kg
- **Category IV**: 300 mg/kg < LD50 ≤ 2000 mg/kg
- **Category V** (NC): LD50 > 2000 mg/kg

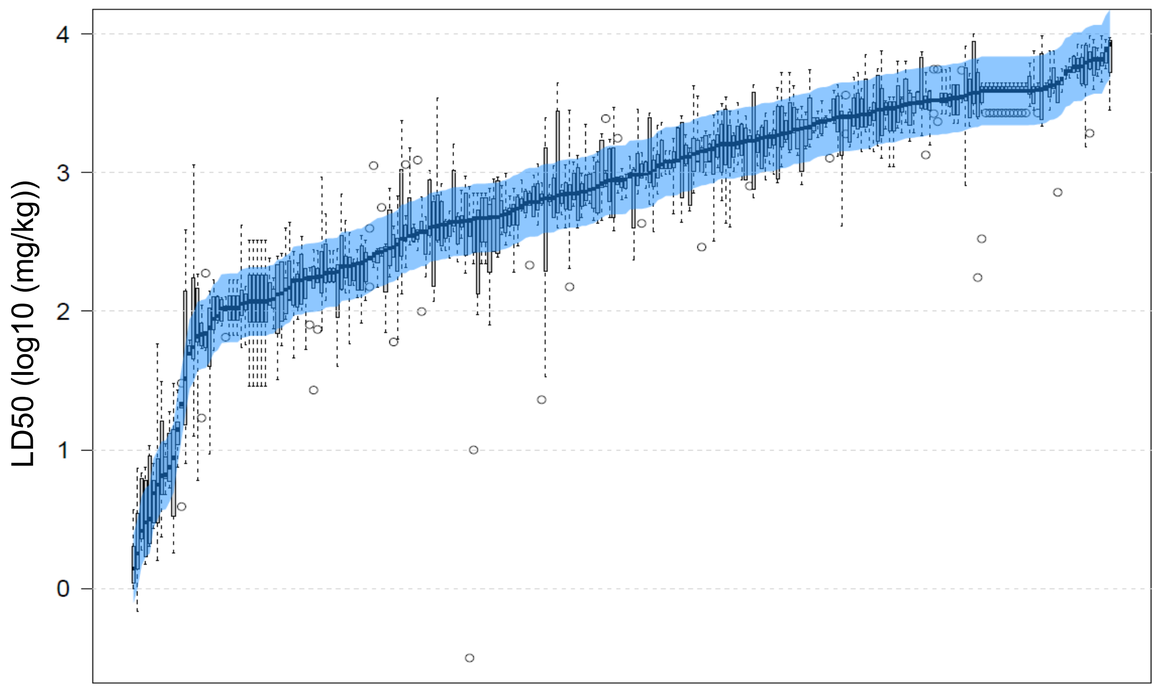
Likelihood of GHS category classification in independently conducted replicate assays

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

Karmaus et al., 2022. Toxicol Sci 188(1)

ECHA, European Chemicals Agency; EPA, U.S. Environmental Protection Agency; GHS, United Nations Globally Harmonized System for Classification and Labelling of Chemicals; NC, Not Categorized; PDII, primary dermal irritation index.

Defining a Margin of Uncertainty for Acute Oral LD50s



The boxplots in the graph above summarizes the distribution of experimental LD50 values for 467 chemicals with at least four independently derived rat acute oral LD50 values (Karmaus et al., 2022). The median from bootstrapping the mean absolute deviation, per chemical, was used to derive a margin of uncertainty equating to ±0.24 (in log10 mg/kg units), visualized by blue shading around the median LD50 value of each chemical. The defined margin of uncertainty generally encompasses the distribution of experimental LD50 values and serves as a reasonable range for determining a high-confidence prediction of where the most likely estimate of LD50 for each chemical would lie.

Acknowledgments

This project was funded with federal funds from the National Institute of Environmental Health Sciences, National Institutes of Health, under Contract No. HHSN273201500010C.

The views expressed above do not necessarily represent the official positions of any federal agency. Since the poster was written as part of the official duties of the authors, it can be freely copied.

*A.L. Karmaus currently works for Syngenta, Greensboro, NC.

A summary of NICEATM and ICCVAM activities at the 12th World Congress is available online at https://ntp.niehs.nih.gov/go/niceatm-wc12.