Rat Acute Systemic Toxicity Testing: Evaluating Reproducibility and Inherent Variability

A Karmaus[1], K Mansouri[2], J Fitzpatrick[3], J Strickland[1], G Patlewicz[5], D Allen[1], W Casey[4] , N Kleinstreuer[2]

[1]ILS/NICEATM, RTP, NC, USA; [2]NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA; [3]ScitoVation, RTP, NC, USA; [4]NIH/NIEHS/DNTP, RTP, NC, USA; [5]USEPA/ORD/CCTE, RTP, NC, USA

Introduction

- Regulatory agencies rely on rodent in vivo acute oral lethality data to determine hazard categorization, assign appropriate precautionary labeling, and perform quantitative risk assessments. As toxicology testing moves towards animal-free new approach methodologies (NAMs), there is need to develop reliable and robust reference data sets to:
  - Contextualize results.
  - Set expectations regarding NAM performance.
  - Train and evaluate computational models.
- Rat acute oral LD50 (dose corresponding to 50% lethality) data from multiple international databases were compiled and curated yielding an inventory of 1885 chemicals with at least two point-estimate LD50 experimental values, or an expanded inventory of 2441 chemicals with at least two independently derived hazard categories.
- Data were analyzed to characterize variability and reproducibility of results across a set of more than 2400 chemicals with multiple independent study records to help better define and characterize variability of the in vivo rat acute systemic toxicity test.

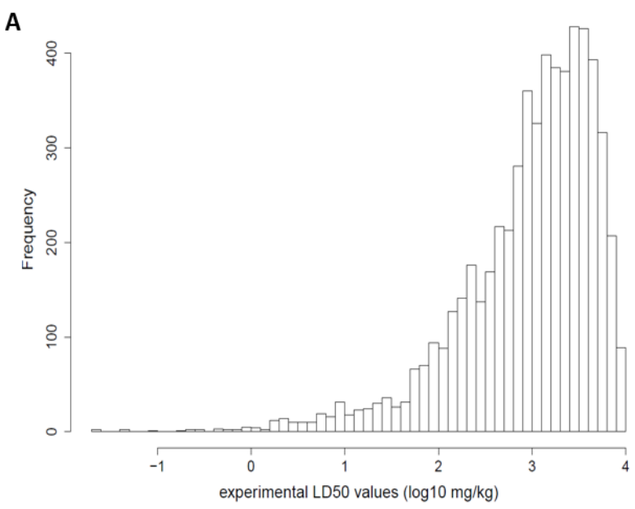
Conclusions

- We could not attribute the observed variability to any chemical-specific physiochemical characteristics, structural features (defined by ToxPrints), or the number of times a study was repeated.
- Chemical potency defined by GHS categories was correlated to variability: the more potently toxic chemicals had higher variability among repeated studies.
- Inherent biological or protocol variability is most likely underlying the variance in rat oral acute systemic toxicity LD50 replicate studies.
- Bootstrapping across computed chemical-specific standard deviations was used to define a 95% confidence interval of ±0.25 log10(mg/kg).
- The computed 95% confidence interval was used to define the uncertainty associated with discrete in vivo rat acute oral LD50 values and may serve as a benchmark to apply to future NAM performance assessments.

Characterizing the Dataset and Establishing Representative LD50 Values for Each Chemical

Figure 1. The curated rat acute oral systemic toxicity LD50 data set comprised chemicals with at least two point-estimate LD50 values, resulting in an inventory of 5826 LD50 values representing 1885 chemicals. (A) Histogram of the distribution of LD50 values in the dataset. The LD50 values ranged from 0.02 to 10000 mg/kg, with most values between 1000-10000 mg/kg. (B) Summary of replicate LD50 values in the dataset per chemical. Most chemicals (1742/1885 chemicals) had five or fewer LD50 values. (C) For each of the 1885 chemicals with at least two discrete point estimate LD50 values, a median value was used as the chemical-specific representative LD50. This plot correlates each experimentally derived in vivo rat acute oral LD50 replicate value with the representative LD50 value for that chemical. The relatively high correlation coefficient (r2 = 0.927) suggests that the computed representative values are a suitable approximation of the experimental data.

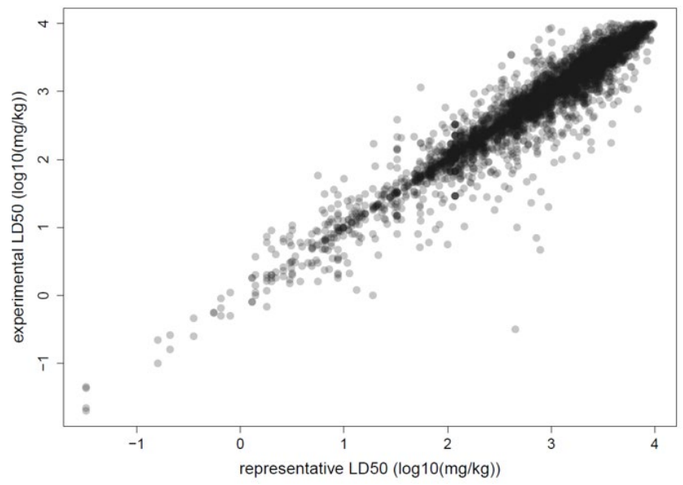
A



B

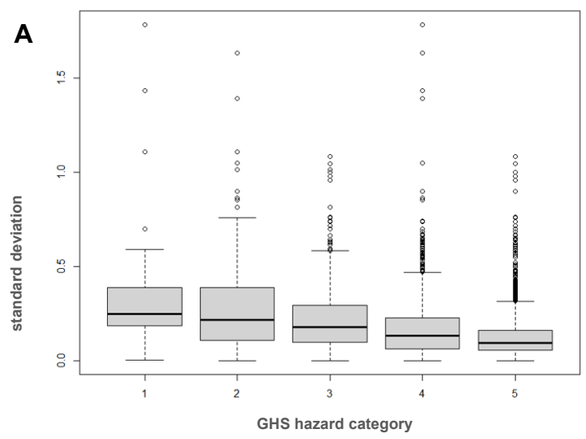
Number of LD50s	Number of chemicals
2	915
3	503
4	222
5	102
6	51
7	42
8	25
9	7
10	5
11	4
12	2
13	1
14	1
15	2
16	1
17	1
18	1

C

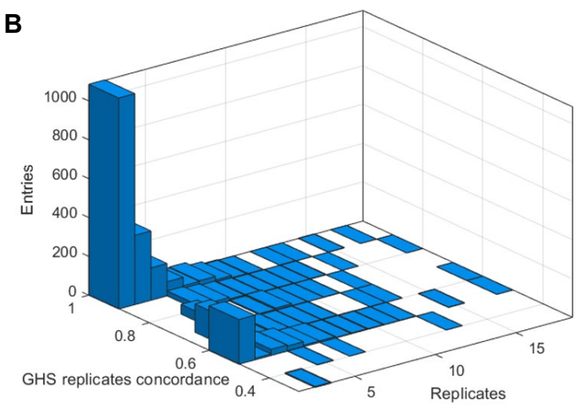


Evaluating the Impact of Variability on GHS Hazard Categorization

A



B



C

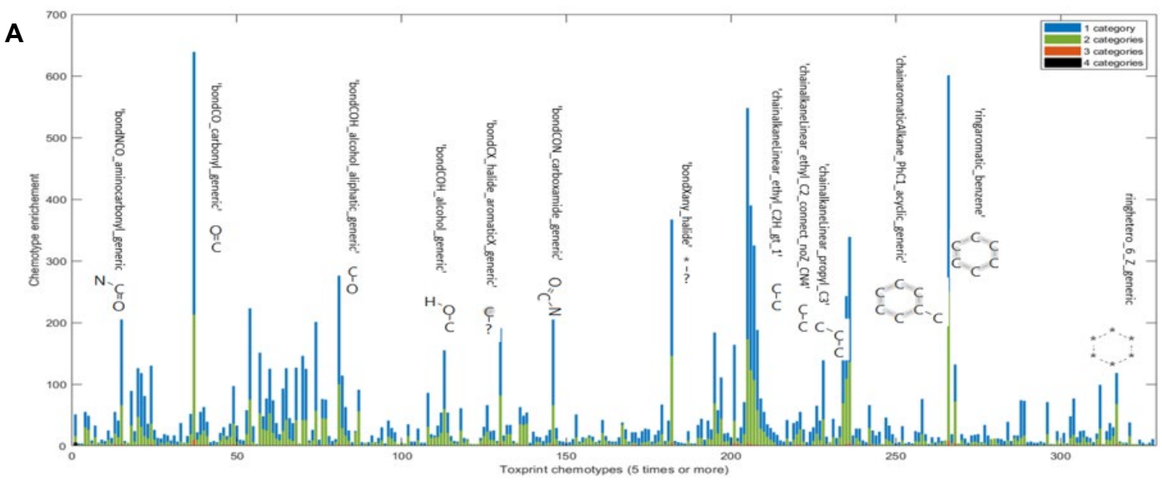
		Conditional Probability of Subsequent Study Categorization				
		1	2	3	4	5
First Study Hazard Category	1	0.533	0.349	0.015	0.051	0.051
	2	0.077	0.489	0.332	0.089	0.013
	3	0.002	0.071	0.619	0.289	0.019
	4	0.001	0.010	0.110	0.661	0.218
	5	0.000	0.002	0.010	0.238	0.750

Figure 2. Summary of LD50 variability across GHS hazard categorization. **(A)** The standard deviation across all LD50 values, per chemical in log10(mg/kg) units, was computed for each of the 1885 chemicals having at least two independently reported LD50s. The boxplots reflect the distribution of standard deviations for each chemical classified into each GHS hazard category. These distributions demonstrate that standard deviation, as a measure of variability, may be correlated with hazard category (i.e., more potently toxic chemicals have higher variability; Jonckheere-Terpstra trend test *p*-value 0.0002). **(B)** Distribution of hazard categorization replicate concordance compared to number of replicates per chemical and to total hazard category counts using the expanded inventory of 2441 chemicals with categorical acute toxicity data revealed no significant trends. **(C)** Conditional probabilities were computed for GHS hazard category reproducibility to assess likelihood of subsequent in vivo study resulting in the same hazard category outcome. Category 5 was most likely to be reproduced in subsequent studies.

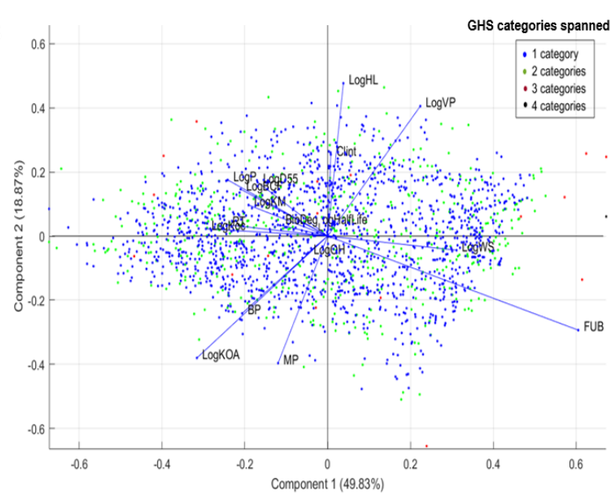
Evaluation of LD50 Variability and Chemical Structure or Physiochemical Properties

Figure 3. The expanded categorical inventory of 2,441 chemicals was evaluated for properties associated with variability. **(A)** Chemicals were mapped to 408 ToxPrint chemotypes, of which 224 had at least five chemicals representing the feature and are represented in this figure. Enrichment of ToxPrint chemotypes was proportional to the number of chemicals per variability class, rather than the class itself (1 category low variability class, 2 moderate, and 3-4 categories high variability class). **(B)** Physiochemical properties were retrieved from OPERA and used to conduct principal component analysis revealing no discernable physiochemical property related to variability.

A

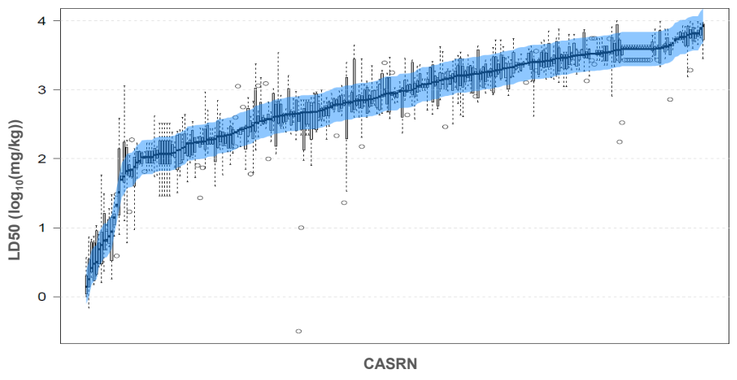


B



Defining a Rat Acute Oral Toxicity LD50 95% Confidence Interval

Figure 4. The standard deviations across point-estimates for each chemical were used as input for bootstrapping (sampling 1 million times). From this, the median was used to compute a 95% confidence interval. This interval equates to ±0.25 (in log10 mg/kg units) and is shown centered around the median of LD50 values per chemical (blue). For illustration purposes, only chemicals with at least four LD50 values (467 chemicals) are shown in this plot. The defined range generally encompasses the distribution of experimental LD50 values and serves as a reasonable range for evaluating acceptable LD50 estimates per chemical.



Acknowledgements

The Intramural Research Program of the National Institute of Environmental Health Sciences (NIEHS) supported this poster. Technical support was provided by ILS under NIEHS contracts N01-ES 35504 and HHSN27320140003C.
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 summary of NICEATM and ICCVAM activities at the Ninth World Congress is available on the National Toxicology Program website at **https://ntp.niehs.nih.gov/go/wc11**