National Institute of Environmental Health Sciences
Division of Translational Toxicology

Compilation and Standardization of Rat Acute Inhalation Study Data to Support Predictive Modeling

Victoria Hull[1], Emily N. Reinke[1], Amber B. Daniel[1], Kimberly T. To[1], Agnes L. Karmaus[1]*, David G. Allen[1], Kamel Mansouri[2], Nicole C. Kleinstreuer[2]

[1]Inotiv, Research Triangle Park, NC; [2]NIH/NIEHS/DTT/NICEATM, Research Triangle Park, NC

Introduction

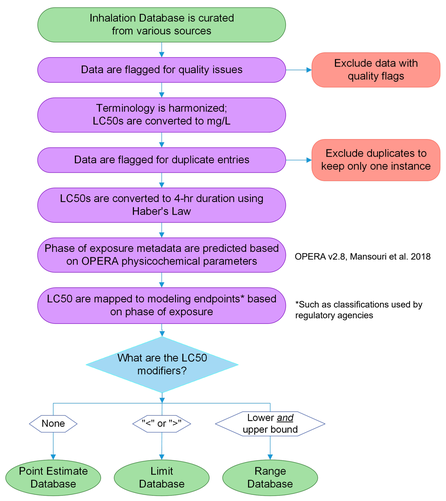
- Multiple U.S. federal and international agencies require acute inhalation toxicity data to determine occupational exposure safety limits, personal protective equipment, consumer safety levels, and packaging and transportation requirements.
- Computational models to predict acute inhalation toxicity have been proposed as alternatives to animal tests to support regulatory decision-making.
  - Developing such models requires robust, well-curated, and chemically diverse training data that is easily accessible.
- A working group established at an "Alternative Approaches for Acute Inhalation Toxicology Testing Workshop" (Clippinger et al. 2018) was tasked with establishing an acute inhalation toxicity database both to address agency information needs and support modeling efforts.
- The working group asked the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) to compile a rat acute inhalation toxicity database.
  - This poster describes the database, which contains data from about 1200 chemicals.
  - This data are available in the Integrated Chemical Environment (ICE; https://ice.ntp.niehs.nih.gov). ICE is an open-access resource developed by NICEATM to provide toxicologically relevant data and computational tools.
- To assess reproducibility of LC50 point estimates and U.S. Environmental Protection Agency (EPA) Office of Pesticide Programs (OPP) hazard categorizations, we analyzed variability in this database for chemicals having at least two reported LC50 values.

Methods: Data Sources

Data Source	Data Records	Unique Substances
Legacy data from ChemIDplus (now integrated into PubChem)	2036	1249
National Institute for Occupational Safety and Health (NIOSH) Pocket Guide	136	649
European Chemicals Agency (ECHA) Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) Database	3016	611
Environmental Protection Agency (EPA) Acute Exposure Guideline Levels (AEGL)	1682	271
Department of Defense	3016	13

- Where available, data were collected on:
  - Chemical names and identifiers (CASRNs, DTXSIDs, SMILES, InChiKeys)
  - Study information, including duration of exposure, LC50, and units of LC50 (mg/L, ppm, or mg/m^3)
  - Study metadata such as species, sex, strain of species, route/phase of exposure (aerosol, gas, vapor), the exposure type (nose-only or whole-body), and vehicle
  - Any additional clarifying data, such as additional details on study design or interpretation of results

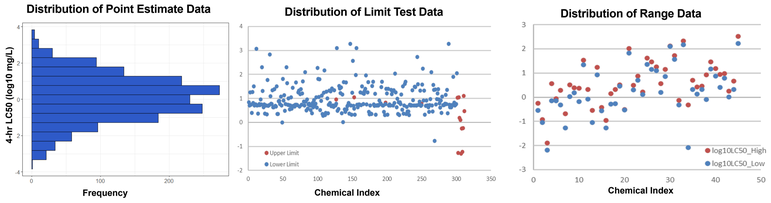
Methods: Data Curation Process



- **Data quality flags** included missing or incorrect units for LC50 values, missing study duration, species measured other than rat, and study type indicated as a read-across study.
- Data were **flagged as duplicates** if two data points met all the following criteria: LC50 values differed by 0.1 mg/L or less; duration of both studies was equal or unreported; sex was the same or unreported; and the route of administration matched.

Data Distribution

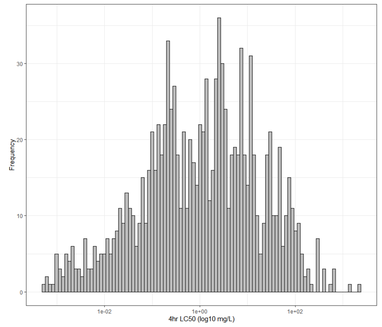
- After being processed through the curation workflow, the dataset had 2109 entries for 1025 unique chemicals (see graphs below).
  - Point estimate database had 1635 entries and 780 chemicals (histogram, far left).
  - Limit test database had 420 entries and 312 chemicals, with 301 lower limits and 15 upper limits (center plot).
  - Range database had 54 entries and 45 chemicals (far right plot).
- Reported duration of exposure ranged from 10 seconds to 24 hours.
- For point estimate data, the mean 4-hr LC50 was 35.75 mg/L with a standard deviation of 253 mg/L. The 4-hr LC50 values ranged from 0.0005 mg/L to 6600 mg/L.



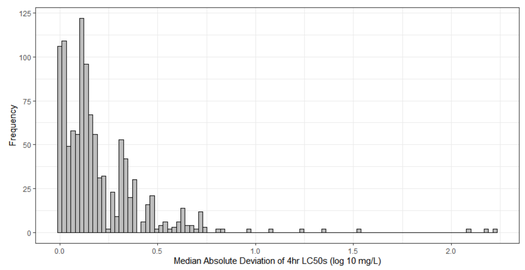
Variability Analysis of Point Estimate LC50s

- We had at least two LC50 point estimates for 231 chemicals (table below left). The remaining 549 chemicals in the database with LC50 point estimates had only a single LC50 value available.
- The distribution of the chemicals with two or more point estimate 4-hr LC50s (log10 mg/L) was approximately normal (histogram below right).

Number of LC50s	Number of Chemicals	Number of LC50s	Number of Chemicals
2	75	12	2
3	49	13	1
4	31	14	1
5	15	15	2
6	19	16	2
7	7	17	1
8	10	19	2
9	5	20	1
10	3	22	1
11	3	29	1

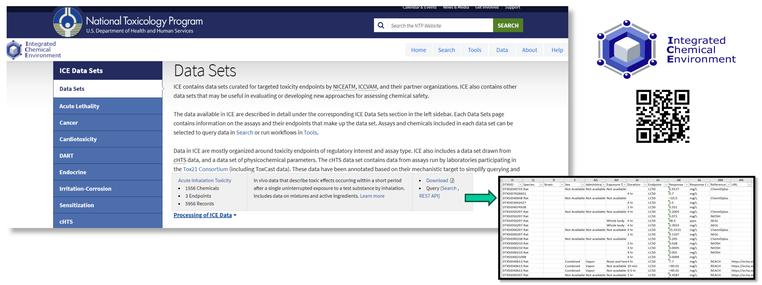


- The mean absolute deviation of repeat 4-hr LC50s ranged from 0.0 to 2.2 (histogram below).
- There was more variation between chemicals than within chemicals (ANOVA, $F_{230,855} = 61.52$, $p < 0.001$).



Access the Database of LC50s

- LC50 point estimate, limit, and range data are can be downloaded directly through the ICE Data Sets page (https://ice.ntp.niehs.nih.gov/DATASETDESCRIPTION) or explored through the ICE Search tool.



Integrated Chemical Environment

Assigning Hazard Categories

- Hazard categorizations are essential for regulatory decision-making. Validation of alternative models can often involve evaluating how these models align with classifications based on reference data. However, hazard categorizations for acute inhalation toxicity are complex, as illustrated in the tables below.
  - The limits of a schema can differ by exposure phase, and there is little consistency across schema on how phases are considered.
  - Despite the importance of exposure phase for categorizations, less than 28% of extracted data in our acute inhalation database indicated the phase of exposure.
  - Instances where chemicals have more than one phase of exposure add an additional level of complexity, as there is increased potential for conflicting categorizations within a schema.
- NICEATM is actively working on these issues and categorizations will be added to the public version of the database once high-confidence solutions are in place.

Globally Harmonized System of Classification and Labelling of Chemicals (GHS) Hazard Categories

GHS Category	Gases (ppm)	Vapors (mg/L)	Dust and Mists (mg/L)
1	LC50 ≤ 100	LC50 ≤ 0.5	LC50 ≤ 0.05
2	100 < LC50 ≤ 500	0.5 < LC50 ≤ 2.0	0.05 < LC50 ≤ 0.5
3	500 < LC50 ≤ 2500	2.0 < LC50 ≤ 10.0	0.5 < LC50 ≤ 1.0
4	2500 < LC50 ≤ 20,000	10.0 < LC50 ≤ 20.0	1.0 < LC50 ≤ 5.0
Not classified	LC50 > 20,000	LC50 > 20.0	LC50 > 5.0

Environmental Protection Agency Office of Pesticide Programs (EPA OPP) Hazard Categories

EPA OPP Category	Criteria for Gases, Dusts, Mists, or Vapors (mg/L)
I	LC50 ≤ 0.05
II	0.05 < LC50 ≤ 0.5
III	0.5 < LC50 ≤ 2.0
IV	LC50 > 2.0

Environmental Protection Agency Office of Pollution Prevention and Toxics (EPA OPPT) Hazard Categories

EPA OPPT Category	Gas or Vapor (mg/L)	Dusts/mists (mg/L)
3 (high)	<2.0 LC50 ≤ 10.0	<0.5 < LC50 ≤ 1.0
2 (moderate)	10.0 < LC50 ≤ 20.0	1.0 < LC50 ≤ 5.0
1 (low)	LC50 > 20.	LC50 > 5.0

Consumer Product Safety Commission (CPSC) Hazard Categories

CPSC Category	Gas or Vapor (ppm)	Dusts/mists (mg/L)
Highly toxic	LC50 ≤ 200	LC50 ≤ 2
Toxic	200 < LC50 ≤ 20,000	2 < LC50 ≤ 200
Nontoxic	LC50 > 20,000	LC50 > 200

Binary Hazard Categories

Binary	Gases (ppmV)	Vapors (mg/L)	Dust and Mists (mg/L)
Toxic	LC50 ≤ 200	Toxic ≤ 2	LC50 ≤ 2
Else	LC50 > 200	LC50 > 2	LC50 > 2

EPA OPP Hazard Category

- EPA OPP does not consider different phases and thus is a suitable case study for how point estimate, limit test, and range data in the acute inhalation database can be categorized and assessed for variability.
- We used the following rules to assign a hazard category for limit test data and range data:
  - If modifier is "<" , LC50 must be in the "most toxic" category.
  - If modifier is ">" or "≥", LC50 must be in the "least toxic" category.

LC50 ≤ 0.05	0.05 < LC50 ≤ 0.5	0.5< LC50 ≤ 2.0	LC50 > 2.0
EPA OPP Cat I	EPA OPP Cat II	EPA OPP Cat III	EPA OPP Cat IV

Example : 4-hr LC50 < 0.04 mg/L — Example: 4-hr LC50 > 3.5 mg/L

  - If modifier is "<" but LC50 is not within the "most toxic" category, it is not assigned.
  - If modifier is ">" or "≥", but LC50 is not within the "least toxic" category, it is not assigned.

LC50 ≤ 0.05	0.05 < LC50 ≤ 0.5	0.5< LC50 ≤ 2.0	LC50 > 2.0
EPA OPP Cat I	EPA OPP Cat II	EPA OPP Cat III	EPA OPP Cat IV

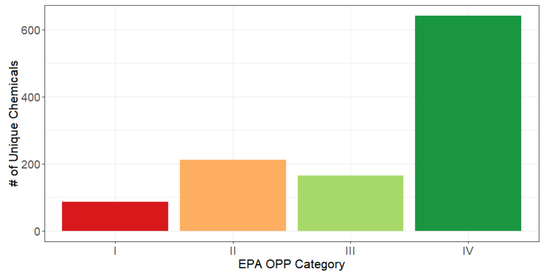
Example 4-hr LC50 > 0.35 mg/L — Example 4-hr LC50 < 1.5 mg/L

- Lower and upper bound of range must be entirely within a category to be assigned to that category.

LC50 ≤ 0.05	0.05 < LC50 ≤ 0.5	0.5< LC50 ≤ 2.0	LC50 > 2.0
EPA OPP Cat I	EPA OPP Cat II	EPA OPP Cat III	EPA OPP Cat IV

Example: 0.06 mg/L < 4-hr LC50 < 0.2 mg/L

- Across point estimate, limit, and range data, 1019 chemicals could be assigned to EPA OPP categories.
- Most chemicals fell within Category IV (least toxic) but there were at least 80 chemicals in each category.



Variability of EPA OPP Categories

- Of the 1019 chemicals with an EPA OPP category, 339 chemicals had at least two reported LC50s.
- Conditional probabilities (the probability of a chemical being within a category given that it was previously categorized in that same category) were calculated for those chemicals with n ≥ 2.
  - Recategorization to the same category was always the most likely outcome.
  - A Category IV categorization is the most consistent, with an 85.9% probability of being recategorized as Category IV.
  - A Category III categorization is the least consistent, with a 46.9% probability of being recategorized as Category III.

EPA OPP Category	I	II	III	IV
I	70.3%	24.6%	--	5.2%
II	10.7%	68.0%	13.8%	7.5%
III	--	25.8%	46.9%	27.2%
IV	0.6%	3.9%	9.7%	85.9%

Summary and Discussion

- NICEATM compiled and curated a database of rat acute inhalation study data to support predictive modeling and regulatory decision-making.
  - LC50s, duration of exposure, and associated metadata within the database are publicly accessible through ICE.
- High-level variability analyses show that 4-hr LC50 point estimate data are fairly consistent, and there is more variability between chemicals than within chemicals.
- Efforts are still ongoing to assign 4-hr LC50s to different hazard schema and assess variability within those categories.
  - There are many complexities that need to be considered before these data are made publicly available and full variability analyses can be conducted.
- Conditional probabilities for EPA OPP categories show that Category IV categorization is most consistent and Category III categorization is least consistent.
- Assessment of effects of animal sex or exposure (head-only or whole-body) could not be conducted due to a paucity of adequate repeat study data.

Future Development and Applications

- A future version of the database will include both reported and predicted phase of exposure so that hazard categories can be assigned.
  - Exposure phase will be modelled based on melting point, boiling point, and vapor pressure utilizing EPA Pollution Prevention Framework (2012) rules.
- We will continue to evaluate reproducibility and variability of the acute inhalation toxicity database.
- We are scoping the feasibility of using the database to support a multi-stakeholder modeling project, similar to models developed to predict estrogen activity (CERAPP, Mansouri et al. 2016), androgen activity (CoMPARA, Mansouri et al. 2020), and acute oral toxicity (CATMoS, Mansouri et al. 2021).

References

Clippinger et al. 2018. Toxicol In Vitro 48:53-70. https://10.1016/j.tiv.2017.12.011.

EPA. 2012. Sustainable Futures / P2 Framework P2 Framework Manual. https://www.epa.gov/sustainable-futures/sustainable-futures-p2-framework-manual.

Mansouri K. et al. 2016. Environ Health Perspect. 24(7):1023-33. https://doi.org/10.1289/ehp.1510267.

Mansouri K. et al. 2018. J Cheminform. 10: 10. https://doi.org/10.1186/s13321-018-0263-1

Mansouri K. et al. 2020. Environ Health Pespect. Feb;128(2):27002. https://doi.org/10.1289/EHP5580.

Mansouri K. et al. 2021. Environ Health Pespect. Apr;129(4):47013. https://doi.org/10.1289/EHP8495.

Acknowledgments

The Intramural Research Program of the National Institute of Environmental Health Sciences (NIEHS) supported this poster. Technical support was provided by Inotiv under NIEHS contract HHSN273201500010C.

Since the poster was written as part of the official duties of the authors, it can be freely copied. We thank Catherine Sprankle and Barbara Stevenson, Inotiv, for editorial input.

*A.L. Karmaus is currently affiliated with Syngenta, Greensboro, NC.

Visit ICE
https://ice.ntp.niehs.nih.gov/