Data Evaluation Criteria for Animal Toxicity and *In Vitro* Studies to Support Human Health Hazard Under TSCA

Amy Benson, MS, DABT*
Office of Pollution Prevention and Toxics, Risk Assessment Division, US Environmental Protection Agency, Washington, DC, United States

benson.amy@epa.gov, 202-564-7624

www.epa.gov

Systematic Review Process

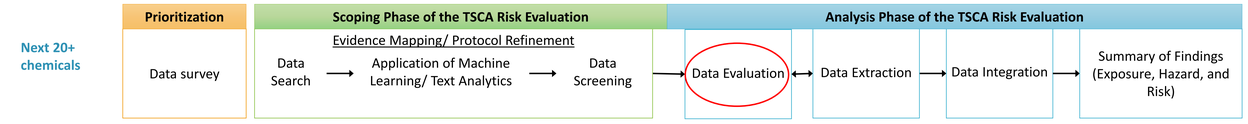
Systematic Review is a comprehensive, unbiased, transparent and reproducible way to identify relevant literature on a topic.

On June 22, 2016, the Frank R. Lautenberg Chemical Safety for the 21st Century Act was signed into law amending the Toxic Substance Control Act (TSCA), the Nation's primary chemicals management law. The U.S. EPA's Office of Pollution Prevention and Toxics (EPA/OPPT) intends to apply systematic review in developing risk evaluations under TSCA.

This involves implementing a structured process to identify, evaluate, and integrate evidence for the hazard and exposure assessments developed for risk evaluation. This poster describes the data evaluation process assessing the quality of the animal toxicity and *in vitro* data types supporting the human health hazard assessment.

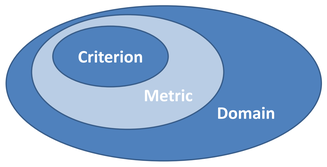
Data Evaluation Process within Systematic Review under TSCA

Key Stages of the Systematic Review Process in TSCA Risk Evaluations

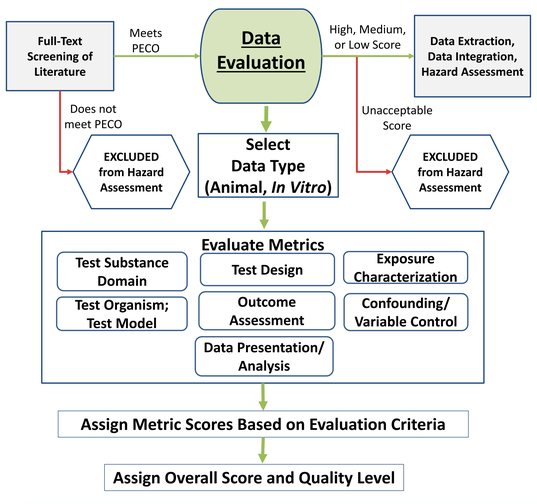


Key Terms in Data Evaluation

- **Domain** – the general categories of data/information attributes intended to assess methodological conduct and risk of bias
- **Metric** – the sub-categories of domain attributes
- **Criteria** – Specific criteria are developed for each metric, which express conditions of the quality level assigned to the metric (high, medium, low, or unacceptable)
- **Data Quality Score** – Quantitative score calculated following evaluation of discipline-specific and data type-specific data evaluation domains and metrics according to predefined scoring criteria and accounting for metric weighting factors.



Data Evaluation Workflow



Data Evaluation for Animal Toxicity and *In Vitro* Studies

Development of criteria

- Reviewed existing study quality and risk of bias evaluation tools for toxicity studies (EC, 2018; Cooper et al., 2016; Lynch et al., 2016; Moermond et al., 2016b; Samuel et al., 2016; NTP, 2015a; Hooijmans et al., 2014; Koustas et al., 2014; Kushman et al., 2013; Hartling et al., 2012; Hooijmans et al., 2010)**
- Consulted with other EPA offices (ORD) and applied professional judgment based on experience reviewing animal toxicity and *in vitro* studies.

Method overview

- **Metrics:** Scores based on strengths and limitations - high, medium, low, or unacceptable (1, 2, 3 or 4)
- Some metrics are not applicable, depending on the study
- Weighting factors: critical metrics = 2, other metrics = 1
- Critical metrics based on:
  - ❑ Review of other risk of bias tools for animal toxicity studies (Lynch et al., 2016; Samuel et al., 2016)**
  - ❑ Greatest ability to inform hazard identification, dose-response and professional judgment
- **Overall study scores:** Combination of metric scores and weights (see "Weighted Scoring System" diagram to right)
- Numerical scores translate to high, medium, low, or unacceptable quality rankings
  - ❑ Qualitative descriptors used in risk evaluations
- Reviewers can adjust overall score to capture professional judgment, if appropriately justified.

Selected statistics from the first 10 risk evaluations

❖ Animal toxicity studies	❖ *In vitro* studies
• Overall study scores: **H:** 50%; **M:** 27%; **L:** 7%; **U:** 16%	• Overall study scores: **H:** 60%; **M:** 12%; **L:** 2%; **U:** 27%
• Overall scores changed (professional judgement) **Downgraded:** 12%; **Upgraded:** 4%	• Overall scores changed (professional judgement) **Downgraded**: 3%; **Upgraded:** 1%

Example justifications for downgrades: study duration too short to fully evaluate carcinogenicity; unusually high incidence of studied health effect in negative control group; doses differed within a treatment group

Evaluation Domains and Critical Metrics

Evaluation Domain	Definition	Critical Metrics [a]
Test Substance	Metrics evaluate whether the test substance used in a study can reliably be determined to have the same (or sufficiently similar) identity, purity and properties as the substance of interest.	**Test substance identity**
Test Design	Metrics evaluate whether the experimental design can distinguish the effect of exposure from other factors, and address use of control groups, randomization.	**Negative/vehicle controls** Positive controls
Exposure Characterization	Metrics assess the validity and reliability of methods used to characterize exposure, including, for example, whether exposure remained consistent over the duration of the experiment and whether exposure levels were appropriate for the outcome.	**Reporting of doses/ concentrations** Exposure duration
Test Organism (Animal Toxicity); Test Model (In Vitro)	These metrics assess the appropriateness of the organism(s), number of organisms and/or replicates per exposure group, and the organism/model conditions.	**Test animal characteristics/model**
Outcome Assessment	Metrics assess the validity and reliability of methods, including sensitivity of methods, that are used to measure or otherwise characterize the outcome(s) of interest.	**Outcome assessment methodology** Sampling adequacy
Confounding/Variable Control	Metrics assess the potential impact of various factors that may affect the outcome and whether the studies identify, account for, and/or control for such factors.	**Confounding variables in test design and procedures**
Data Presentation/ Analysis	Metrics assess whether appropriate statistical methods were used and if data for all outcomes are presented.	**Reporting of data** Data interpretation

[a] For critical metrics, **bold text** = animal and *in vitro*; no bold = *in vitro* only

Risk of Bias by Domain (Metric)

- **Selection bias**
  Test design (randomized allocation)
- **Performance/detection bias**
  Test design (negative/vehicle controls)
  Outcome assessment (Blinding, methodology)
  Test substance (purity)
- **Attrition/exclusion bias**
  Confounding/variable control (health outcomes unrelated to treatment)
- **Confounding bias**
  Confounding/variable control (during test design, procedures)
- **Selective reporting/reporting bias**
  Data presentation and analysis (reporting of data)

Similarities with Established Tools

- ***Risk of Bias/Methodological Quality/Sensitivity:*** Metrics address biases covered by OHAT, IRIS, Navigation Guide, sensitivity measures within IRIS and several method quality measures from SciRAP.
- ***Reporting:*** Many metrics address the level of data reporting within sources.

Scoring System: Metrics for Study Quality

High	No notable deficiencies or concerns are identified in the domain metric that are likely to influence results [score of 1]
Medium	Minor uncertainties or limitations are noted in the domain metric that are unlikely to have a substantial impact on results [score of 2]
Low	Deficiencies or concerns are noted in the domain metric that are likely to have a substantial impact on results [score of 3]
Unacceptable	Serious flaws are noted in the domain metric that consequently make the data/ information source unusable [score of 4]
Not rated/ applicable	This rating means the metric is not applicable to the data/information source being evaluated [no score]

Scoring System: Overall Data Quality Levels/Scores

A numerical scoring method is used to convert the quality level for each metric into the overall quality level for the data/information source. A study is disqualified from further consideration if the confidence level of one or more metrics is rated as Unacceptable [score of 4].

Weighted Scoring System by Metrics

Overall Score=

$$\sum(\text{Metric Score} \times \text{Weighting Factor}) \div \sum(\text{Metric Weighting Factors})$$

High	Medium	Low	Range of Overall Score:
≥1 and < 1.7	≥1.7 and <2.3	≥2.3 and ≤3	**1 to 3**

Updates/Improvements

- Clarify individual metrics (e.g., add more specific language, consistently rate metrics where data is not reported).
- Use fewer quality bins (e.g., eliminate high bins) for metrics that have a limited number of criteria/response options.
- Make criteria for the high and unacceptable bins more stringent.
- Specify when not applicable can be used for a metric.
- Delete scoring for any new dichotomous metrics.

U.S. Environmental Protection Agency

NASEM Review of the Environmental Protection Agency's Toxic Substance Control Act Systematic Review Guidance Document

Acknowledgements: EPA/OPPT/RAD - Fran Branch, Iris Camacho, William Irwin, Keith Jacobs, Susanna Wegner; EPA/OPP – Jocelyn Hospital; EPA/ORD – Kris Thayer and others; SRC – Teresa Manyin, Jennifer Rayner, Marc Odin

**Full citations are within U.S. EPA, 2018. *Application of Systematic Review in TSCA Risk Evaluations.* Appendix G. EPA Document# 740-P1-8001. Office of Chemical Safety and Pollution Prevention. May. Available at: https://www.epa.gov/sites/production/files/2018-06/documents/ final_application_of_sr_in_tsca_05-31-18.pdf.