# ToxRefDB 2.0: Improvements in Capturing Qualitative and Quantitative Data from *in vivo* Toxicity Studies

Sean Watford[1,2], Angelica Adrian[3], Jessica Wignall[4], Josephine Brown[3] and Matt Martin[5]

1 University of North Carolina Chapel Hill
2 Student Services Contractor, USEPA National Center for Computational Toxicology
3 National Student Services Contractor ORAU
4 ICF
5 USEPA, National Center for Computational Toxicology

Sean Watford
watford.sean@epa.gov
919-541-7655
ORCID: 0000-0003-0888-5029

## ToxRefDB Overview

Toxicity Reference Database (ToxRefDB) serves as a resource for retrospective and predictive toxicology

- ToxRefDB stores large sets of guideline and guideline-like *in vivo* chemical toxicological data
  - o Aids in validation of *in vitro* high throughput screening of chemicals
  - o Used in predictive model development

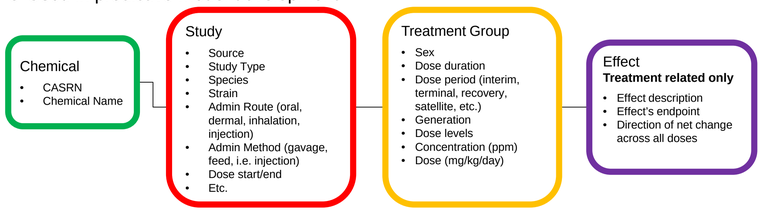


**Figure 1. ToxRefDB 1.0 general schema.** ToxRefDB 1.0 captures basic study design, treatments, and treatment-related effects (Only captures "positive" results)

### Building ToxRefDB 2.0:

- Chemical library expansion to >1,100 chemicals
- Study type expansion (Fig. 2)
- Standardized terminology for endpoints and effects
- **Quantitative Data Entry**
- **Endpoint observation status**
- Study reliability evaluation
- Guideline level information

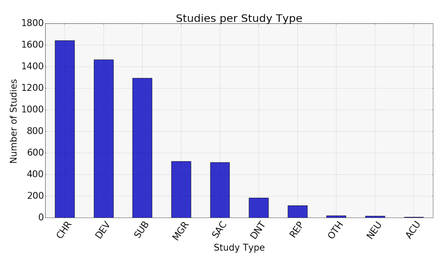


**Figure 2. Study types.** CHR: Chronic, DEV: Developmental, SUB: Subchronic, MGR: Multigenerational, SAC: Subacute, DNT: Developmental Neurotoxicity, REP: Reproductive, OTH: Other, NEU: Neurological, ACU: Acute

## Endpoint Observation Status

- Endpoint observation status (not in ToxRefDB 1.0) distinguishes between negative vs. missing (not tested) effects
- Two binary fields were created to represent the testing status: "reported" and "tested"
- If an endpoint was required to be tested according to the study's specific guideline, then the database reflects this status as a 1, unless stated otherwise in the report (Table 1)

**Table 1. Observation status.**

| Reported | Tested | Rationale |
|---|---|---|
| 1/true | 1/true | The study reported that the endpoint was tested |
| 1/true | 0/false | The study reported that the endpoint was not tested<br>• Triggered endpoints will start with this notation, and should be updated to R1,T1 if there is evidence that the endpoint was measured. |
| 0/false | 1/true | The study did not report that an endpoint was tested.<br>• For example, if gross pathology was tested, but specific organs were not listed – this notation is for those organs indicated as "required" by guideline |
| 0/false | 0/false | The study did not provide any information on this endpoint whatsoever and it is not required by guidelines (no assumptions to be made; not listed in a table or text at all) |

## Data Entry Method

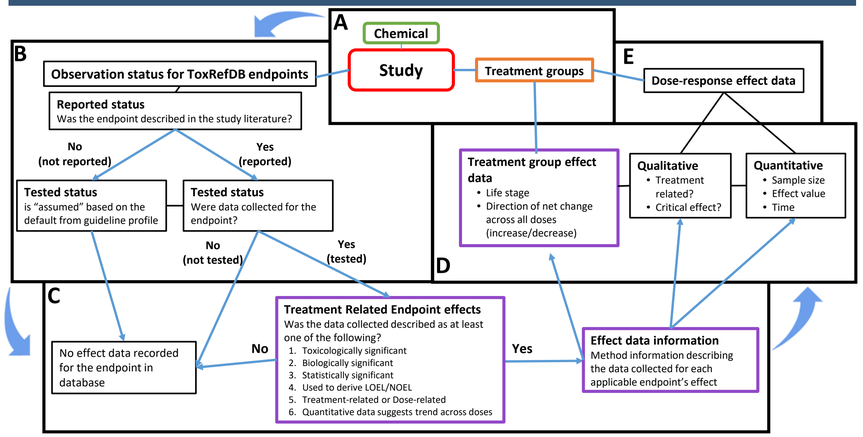


**Figure 3A-D. ToxRefDB 2.0 general schema**
A. Portion of ToxRefDB 1.0 that carried over to version 2.0 unchanged.
B. New profiling portion of database. Uses decision tree to classify 400 standardized database endpoints as described in study reports.
C. Observed Endpoints classified as "tested" are evaluated for treatment-related effects. Treatment-related effects are indexed by endpoint and method information pertaining to the data collected.
D. Treatment group effect data: qualitative data from ToxRefDB 1.0 and does not contain dose-response data.
E. Dose-response effect data: qualitative and quantitative data for each dose.

## Summary Statistics

- Processed over 2100 Chronic and Subchronic studies
  - Includes OPP DERs and NTP studies
- 200K Effect-treatment group-dose quantitative data points

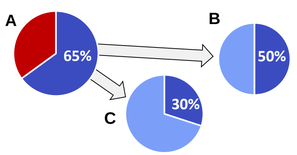


**Figure 4A-C: Quantitative data summary.** (A) Of all reported effects, 65% have quantitative data. (B) Of the quantitative data entered, over 50% is dichotomous or incidence-type. (C) 30% of the quantitative data has variance reported.

## Study Reliability with ToxRTool

- ToxRTool: **Tox**icological data **R**eliability Assessment **Tool**
  - Assigns Klimisch score of 1-4 to assess the reliability of studies
  - Provides comprehensive criteria and guidance for evaluating the inherent quality of toxicological data
  - Only applied to non-guideline (non-NTP and non-DER) studies

**Table 2: Sample questions from ToxRTool covering 5 criteria for assessing study reliability**

| Criteria Group | | Question |
|---|---|---|
| I | 1 | Was the test substance identified? |
| II | 5 | Is the species given? |
| | 10 | Is the administration route given? |
| III | 11 | Are doses administered or concentrations in application media given? |
| | 12 | Are frequency and duration of exposure as well as time-points of observations explained? |
| | 14 | Is the number of animals (in case of experimental human studies: number of test persons) per group given? |
| IV | 17 | Are the study endpoint(s) and their method(s) of determination clearly described? |
| V | 20 | Is the study design chosen appropriate for obtaining the substance-specific data aimed at? |

## Guideline Profiles

- Endpoint language was updated to adhere to series 870 Health Effects Testing Guidelines created by the Office of Chemical Safety and Pollution Prevention (OCSPP).
- Profiles were created for cholinesterase, systemic, reproductive, and developmental endpoints. Clinical chemistry, hematology, urinalysis, pathology (gross and microscopic), organ weight and in life observation are all types of systemic endpoints (Fig. 5).

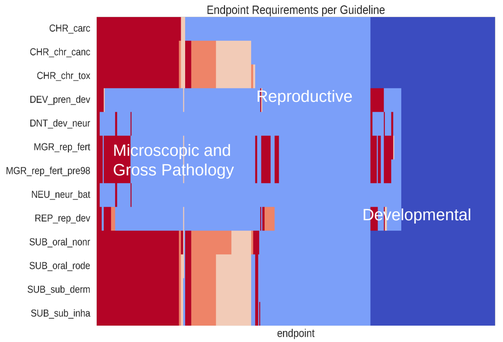


**Figure 5. Guideline Profiles.**

Required: guideline stated data should be collected for a particular endpoint.

Recommended: guideline stated that it recommends data to be collected for a particular endpoint.

Trigger: an endpoint is considered required or recommended under specified circumstances in the guideline.

Not required: the endpoint was not included in the guideline.

Na: the endpoint could not possibly be tested in the guideline's study type

NULL: Endpoint was too specific or general to be applied to guideline specifications (i.e. Blood cell subtype, diagnosis).

## Benchmark Dose Modeling

- Calculate BMD and BMDL for all effects, not just critical effects
- Batch BMDS with python package bmds (https://github.com/shapiromatron/bmds)
- User selectable BMR

**A**

| Dose | Sample Size | Mean | Standard Deviation |
|---|---|---|---|
| 0 | 25 | 2.61 | 0.81 |
| 10 | 25 | 2.81 | 1.19 |
| 50 | 24 | 2.96 | 1.37 |
| 150 | 24 | 4.66 | 1.72 |
| 400 | 24 | 11.23 | 2.84 |

**B**

| Model Name | BMD | BMDL | AIC | Recommended |
|---|---|---|---|---|
| Linear | 51.12 | 42.90 | 228.24 | FALSE |
| Polynomial | 88.35 | 61.52 | 221.68 | FALSE |
| Power | 94.74 | 66.75 | 221.89 | FALSE |
| Hill | 94.78 | 67.27 | 223.89 | FALSE |
| Exponential | 93.78 | 82.45 | 220.02 | TRUE |

**C**

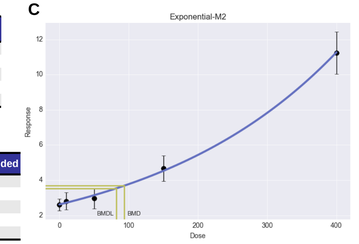


**Figure 6: BMD example.** (A) Input data for kidney weight percentage relative to body weight. (B) Five of the 10 modeled outputs from BMDS. (C) Plot of the recommended model output.

## Conclusions and Future Directions

- **Quantitative data entry completed for Subchronic and Chronic studies with MGR and DEV next**
- Guideline profiling improves modeling sets for predictive toxicology with a clearer delineation between endpoints with no observed treatment related effects and endpoints that were not tested
- Evaluate study reliability using ToxRTool for non-DER and non-NTP studies
- Default testing statuses generated from guideline profiles will allow for systematic evaluation of guideline adherence
- **Batch BMD pipeline for systematic BMD modeling of quantitative data**
- **Mapping of endpoints and effects to CDISC/SEND ontology**
  - CDISC SEND: **S**tandard for **E**xchange of **N**onclinical **D**ata
  - Linking ToxRefDB to CDISC SEND terminology will provide a standardized language for collaboration between many organizations and regulatory agencies