Deriving Performance Baselines for Predictive Modeling of Systemic Toxicity using ToxRefDB

Matt Martin[1], Keith McLaurin[1], Lisa Truong[1], Gladys Ouedraogo[2], Sophie Loisel-Joubert[2]

[1] NCCT/ORD, US EPA, RTP NC, USA | [2] L'Oréal, Paris, France

Matt Martin | martin.matt@epa.gov | 919-541-0502

Abstract

A primary goal of computational toxicology is to generate predictive models of toxicity. An elusive target of alternative test methods and models has been the accurate prediction of systemic toxicity points of departure (PoD). We aim to scope the problem by generating floor and ceiling baseline uncertainty bounds for which to judge future models. EPA's ToxRefDB, originally populated with pesticide registration data, has grown to incorporate guideline-like studies from the pharmaceutical industry, National Toxicology Program, and publicly available research literature. Over 6000 high quality animal studies on 1071 chemicals were captured using standardized study design, treatment and effect vocabulary. For model development, a subset of 500 chemicals was identified by the EPA. Systemic lowest effect levels (LEL) were obtained for each study across a diverse set of study types including systemic sub-acute (SAC), subchronic (SUB), chronic (CHR) studies as well as systemic adult effects observed in developmental (DEV) and reproductive (MGR) studies. Species and study type adjusted chemical-level LEL were derived demonstrating a floor baseline of roughly 4.25 orders of magnitude uncertainty (OMU; 95% CI-Range) based on the default distribution of LEL. Using SUB to predict CHR rat and mouse to predict rat CHR NEL, ceiling baselines were established of 3.47 and 3.35 OMU, respectively. Further classification of study types based on exposure duration (short = SAC, DEV; medium = SUB, MGR; long = CHR), established a ceiling baseline for short vs medium, and long vs medium to be 2.84 and 3.55 OMU, respectively. Thusly, the goal of any predictive model of systemic toxicity is to improve upon the 4.25 OMU and approach 2.84 OMU, but cannot be expected to exceed the inherent uncertainty in toxicological testing and evaluation. *This abstract does not necessarily reflect US EPA policy.*

Objectives

- Examine variability and uncertainty found in *in vivo* animal toxicity studies
- Generate floor and ceiling baseline uncertainty bounds to evaluate modeling efforts

ToxRefDB Overview

- 40+ years of OPP guideline and high quality traditional guideline-like animal toxicity studies
- Original public release of 1978 studies covering 474 chemicals
- Now contains 6799 total studies with inclusion of NTP, public literature, and pharmaceutical studies
- Filter incomplete, deficient, unacceptable studies, unidentified chemicals (~800)
- Focus on seven main data rich study type species combinations
- Approximately 1071 chemicals with high quality studies

Systemic Toxicity Normalization

Table 1. Normalization parameters used to adjust **(A)** species and **(B)** study type to be a subchronic rat study. **(A)** Five species body surface area conversion values to normalize to a rat. **(B)** Adjustment factors to normalize exposure duration to a 90 day (subchronic) study.

(A)

Study Type (Adjusted to sub-chronic)	Operation	Adjustment Factor
Multigeneration Reproductive & Subchronic	Multiple	1
Chronic	Multiple	3
Subacute, Developmental	Divide	3

(B)

Species (Adjusted to rat)	Body Surface Area Conversion
Mouse	3/6
Primate	12/6
Dog	20/6
Rabbit	12/6
Hamster	5/6

Profile of ToxRefDB LELs

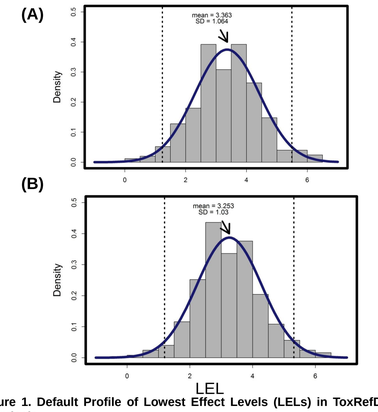

Figure 1. Default Profile of Lowest Effect Levels (LELs) in ToxRefDB for 500 chemicals.

A histogram illustrating the distribution of $-\log_{10}$ **(A)** raw LELs and the **(B)** normalized LELs for 500 chemicals in ToxRefDB. LELs displayed are negative log 10 values. Black arrow depicts the mean LEL and standard deviation for the distribution to be 3.363 ± 1.064. Dotted lines show the 95% confidence interval (OMU). Dark blue line illustrates a normal distribution for the data.

Ceiling Baseline Comparisons

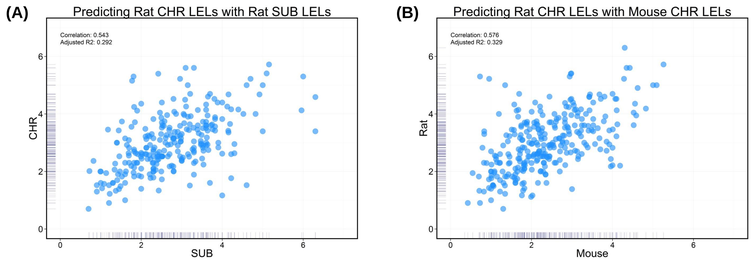

Figure 2. Predicting (A) Rat CHR LELs with SUB LELs and (B) CHR Rat LELS with Mouse LELs.

(A) Each dot presents a chemical assessed in both study types (296) with SUB LELs displayed on the x-axis, and CHR LELs on the y-axis. A correlation analysis revealed a coefficient of 0.542, with an adjusted R^2 of 0.292. **(B)** A correlation analysis of Mouse CHR LELs (x-axis) vs Rat CHR LELs (y-axis) for 335 chemicals (represented as each point) computed a coefficient of 0.576 with an adjusted R^2 of 0.329. On each axis, a density rug is present to illustrate density of chemicals at a LEL. Each vertical line represents a chemical and its corresponding LEL.

Baseline Comparisons of Exposure Duration

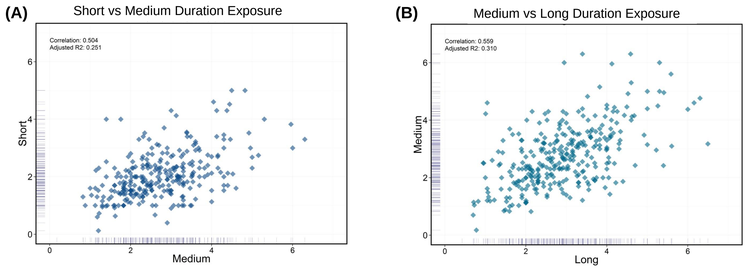

Figure 3. Comparison of (A) Short vs Medium and (B) Long vs Medium exposure duration.

(A) A total of 368 chemicals were evaluated in both short and medium exposure duration. The correlation coefficient is 0.504 with an adjusted R^2 of 0.251. **(B)** Comparison of long vs medium exposure duration for 373 chemicals revealed a correlation coefficient of 0.559 and an adjusted R^2 of 0.311. Each point represents one of these chemicals. A density rug where each vertical line represents a chemical and its corresponding LEL is displayed on the axis of the figure.

Performance Comparison of Each Analysis

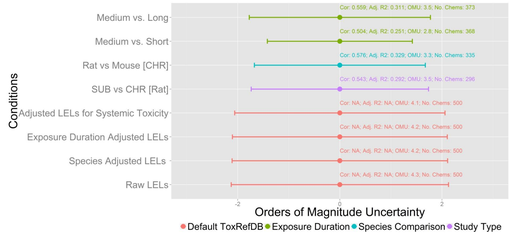

Figure 4. Performance plot comparing each analysis based on their orders of magnitude uncertainty (95% confidence interval; OMUs).

The default ToxRefDB was analyzed based on four conditions (red), resulting in OMUs ranging from 4.1 to 4.3 (raw, unaltered). Comparisons based on predictability of study types (purple) using animal data resulted in an OMU of 3.5. Species comparison of rat and mouse (teal) was approximately 3.3 OMU. Exposure duration comparison had an OMU of 2.8 and 3.5 (green). For each comparison, the correlation coefficient, adjusted R^2, OMU and number of chemicals used for the analysis is annotated.

Conclusions

- These data illustrate the large variability and uncertainty found in most *in vivo* animal toxicity studies
- Using generalized and default measure of accuracy and predictability to judge systemic toxicity models is too stringent
- It is necessary to generate appropriate floor and ceiling baseline of uncertainty bounds to measure future models
- Models predicting systemic toxicity using *in vivo* studies should have a OMU higher than 4.3, and the expectation of a good model has a OMU around 2.8

Future Directions

- Develop a framework to predict systemic toxicity and utilize the floor and ceiling baseline to evaluate the accuracy of the model
- Derive the adjustment factors for exposure duration
- Improve the normalization parameters to reduce the OMU surrounding the studies in ToxRefDB