Variability within Systemic *In Vivo* Toxicity Studies

Pham, Ly L.[1]; Woodrow, Setzer[2]; Martin, Matt[2]

[1]ORISE Fellow, RTP/EPA

[2]National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, RTP, NC

Ly Ly Pham | pham.lyly@epa.gov | 919-541-5569

Introduction

In vivo studies have long been considered the gold standard for toxicology safety assessment. Often time models developed in silico and/or using *in vitro* data to estimate points of departures (POD) are compared to the *in vivo* data to benchmark and evaluate quality and goodness-of-fit. However, recent work has illustrated that currently available in vivo data are not without flaws and inherent variance presents a challenge in predictive modeling[1,2,3]. **The goal of the current work was to quantify the amount of variance that exists within systemic *in vivo* PODs (explained and unexplained).**

We hypothesize that the variance between observed POD from study to study can be characterized by the equation:

Var(Observed POD) = Var("True" POD)
+ Var(Study Conditions)
+ Unexplained Variance

POD is defined as the Log10 of the lowest dose in which a treatment related effect was observed per study.

Methods

Data Preparation

Data taken from: US EPA's Toxicity Reference Database (ToxRefDB)

- Contains over 5,000 in vivo toxicity studies covering over 1,000 chemicals.
- Guideline or guideline comparable studies from various sources.

Data was filtered to only include:

- Adult animals in the F0 generation
- Systemic toxicity studies (CHR, SUB, DEV, MGR, and SAC)
- Administration Route: Oral
- Species: mouse, rat, dog, and rabbit
- Non-control group data

Three datasets were created:

- Two or More Studies Per Chemical.
- Two or More Studies & Study Type Per Chemical.
- Two or More Studies, Study Type, & Species Per Chemical

Analysis

Variance Calculations

- Multilinear Regression and ANOVA is used to estimate overall variance in the observed POD.
- Residual mean square (RMS) error were used to estimate the variance that could not be explained by study conditions.
- Percent of variability that can be explained in a given data set was calculated by

$$\left(\frac{\text{Var(Observed POD)} - \text{Unexplained Variance}}{\text{Var(Observed POD)}}\right) \text{X100}$$

Importance of Each Study Condition

- Leave one out (LOO) method is used to test each study condition's contribution to the explainable variance.
- K-means clustering of toxprint chemotypes were used to assess chemical groupings contribution to the unexplained variance.

Results

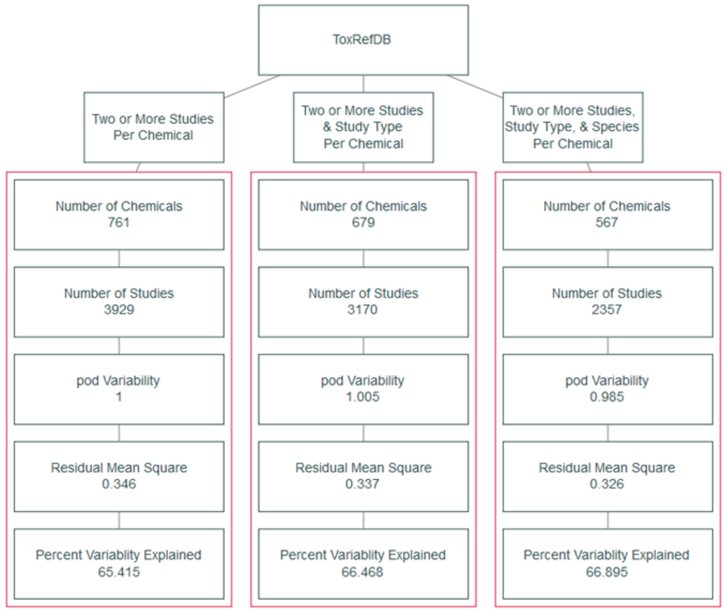

Figure 1: Flow chart of the three dataset used along with results of the variance analysis. For each dataset, the number of unique chemical and studies are shown along with the calculated variance of the POD.

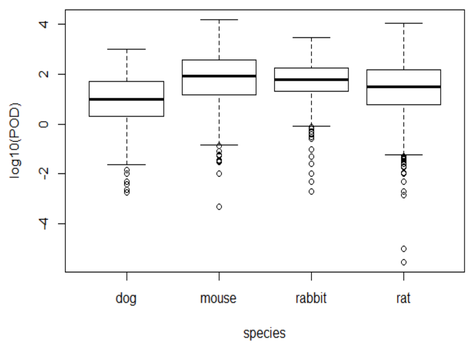

Figure 2: Box plot of the $\log_{10}$(POD) by species on dataset containing chemicals with at least two studies.

Variance Calculations

With the creation of the dataset, as the criteria became more strict, the number of chemicals dropped by about 100 and the studies by around 1,000 (Figure 1). The variance in the POD did not vary much. Using multilinear regression to calculate the residual sum of squares, we accounted for known variability in study conditions to quantify the unexplained variance of the POD to be about 0.35 for all three dataset. With the calculated variance in the dataset (observed(POD)) and the RMS, we can calculated the explainable variance in the dataset. The amount of variance that can be explained remain consistent showing no significant improvement in the model's ability to account for more variance even when the data becomes more homogenous. Boxplot of the POD stratified by species does not show much difference in the spread of the data (Figure 2).

Importance of Each Study Condition

The leave-one-out method was used to assess the amount of variance explained by various study conditions (e.g., species, purity of test material) and chemicals were found to be the biggest contributor, explaining ~50% of the variance (Table 1). This was observed across all three data subsets. The importance of chemical as a variable in the model was further assess by identifying chemicals by their structure similarities using toxprint chemotypes. Chemicals were defined as similar using k-means clustering methods. Similar chemicals did not explain a significant amount of the variance until the number of clusters got close to the number of chemicals (Figure 3).

Table 1: RMS results for the LOO analysis with ANOVA between each full model and a LOO model.

	Two or More Studies Per Chemical		Two or More Studies & Study Type Per Chemical		Two or More Studies, Study Type, & Species Per Chemical	
Models	RMS	p-value	RMS	p-value	RMS	p-value
Full Model	0.326		0.337		0.326	
Chemical Removed	0.790	6.43E-213	0.844	9.88e-323	0.790	6.43E-213
Strain group Removed	0.356	9.81E-29	0.389	3.23E-69	0.356	9.81E-29
Study Type Removed	0.350	3.34E-26	0.354	1.54E-25	0.350	3.34E-26
Admin Method Removed	0.327	9.16E-02	0.338	2.92E-02	0.327	9.16E-02
Dose Spacing Removed	0.330	9.64E-07	0.339	8.17E-05	0.330	9.64E-07
Number of Dose Removed	0.331	2.67E-07	0.341	1.08E-07	0.331	2.67E-07
Study Year Removed	0.326	1.45E-01	0.337	4.37E-01	0.326	1.45E-01
Substance Purity Removed	0.326	2.76E-01	0.337	1.90E-01	0.326	2.76E-01
Study Source Removed	0.327	4.17E-02	0.338	1.33E-02	0.327	4.17E-02
Gender Removed	0.330	4.29E-05	0.339	1.02E-04	0.330	4.29E-05

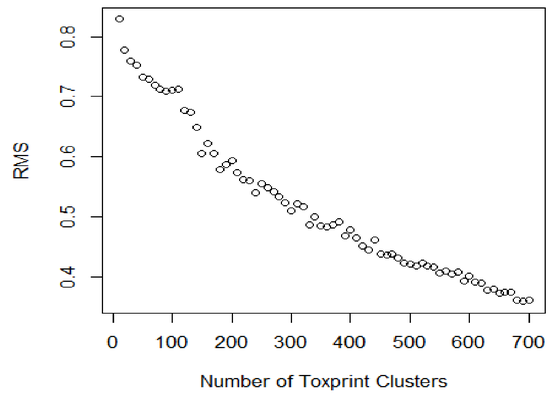

Figure 3: Plot of the RMS by the number of k-means clusters on datasets containing chemicals with at least two studies. K-Means was used to cluster the toxprint chemotypes of the chemical and each chemical is assigned a chemical cluster number. ANOVA was performed with the chemical name replaced by cluster number.

Conclusion & Future Work

Conclusion

- We have quantified the amount of variance in systemic POD within the ToxrefDB as ~65%.
- We estimate the unexplained variance for systemic POD toxicity study is ~0.35.
- The unexplained variance provides a benchmark and lower bounds on the mean-square-error for development of predictive toxicity models.
- This work also provides an upper bound on the level of precision predictive models can attain when trained on conventional PODs.

Future Steps

- Analysis a subset of well studied class of chemicals to see if the unexplained variance as calculated by the mean-square-error stays ~0.35.
- Assess the variability of the qualitative data (biological effects) across studies per chemical.

Reference

1. Hoffmann et al. 2010. *Regulatory Toxicology and Pharmacology* 58(3): 395–407.
2. Jensen and Ritskes-Hoitinga. 2007. *Laboratory Animals* 41(1).
3. Kleinstreuer et al. 2015. *Environmental Health Perspectives* 124(5))