Variability within Systemic *In Vivo* Toxicity Points-of-Departure

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 | ORCID:0000-0001-8467-2645

Background

Animal toxicology studies have long been considered the gold standard for hazard identification and characterization, including point-of-departure (POD) determinations. Due to regulatory mandates and the sparsity of animal toxicology data for thousands of environmentally-relevant chemicals, alternative approaches have been increasingly relied upon for chemical safety decision-making. However, comparing alternative method performance to traditional approaches without better understanding the underlying variability of the traditional methods is difficult and often misleading because reproducibility within animal studies have long been a topic of concern[1-4]. Using the USEPA Toxicity Reference Database (ToxRefDB) systemic toxicity POD values and associated study parameters, multilinear regression and analysis of variance was performed to quantify the explained variability due to various study parameters, (e.g., chemical treatment, study type, species, strain, dose spacing) and to estimate the remaining unexplained variability.

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 Log_{10} mg/kg/day 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 v1.3)
- 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:
- Dataset A: Two or More Studies Per Chemical.
- Dataset B: Two or More Studies & Study Type Per Chemical.
- Dataset C: Two or More Studies, Study Type, & Species Per Chemical

Analysis

Variance Calculations
- Multilinear Regression was used to partition the total variance in the observed POD into an unexplained component and a component attributable to different study design factors, and ANOVA was used to compare the significance of individual components.
- Percent of variability that can be explained in a given data set was calculated by:

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

Importance of Each Study Condition
- Nested models using a Leave one out (LOO) approach were used to test each study condition's contribution to the explainable variance

Assessing naïve chemical groupings (Dataset A only)
- Toxprint chemotypes were substituted for chemical treatment and then clustered using K-means and Hierarchical methods

Stratification of Data by Chemical Class (Dataset A only)
- Chemicals were stratified into 3 classes (Conazoles, Phenols, and Carbamates), and MSE was estimated for each.
- Significance of the difference between variances was calculated by computing the F-distribution between the classes, pairwise. This is calculated as the ratio of the greater variance over the smaller. The upper confidence limit was then calculated for each pair.

Results

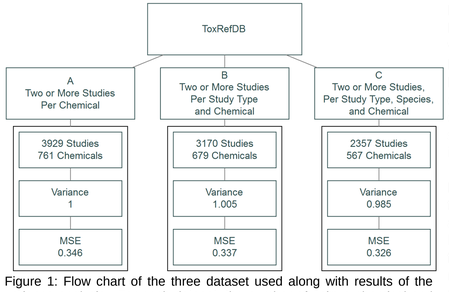

Figure 1: Flow chart of the three dataset used along with results of the variance analysis. For each dataset, the number of unique chemical and studies are shown along with the calculated variance of the POD.

In all three datasets, the POD variance was approximately 1, the MSE was approximately 0.33 (Figure 1), and the percent of variability that can be explained is ~66% (not shown). Using the MSE, we can calculate the RMSE ($\sqrt{MSE}$) to be about 0.58. MSE remained constant across all three datasets even as the datasets became more homogeneous, indicating that the amount of variance that can be accounted for is constant. This provides some level of confidence that the underlining unknown error is inherit across all systemic toxicology studies.

By comparing the nested model with the full model, we quantified the contribution of each study variable to the total variance across all three datasets (Table 1). Chemical treatment had the largest impact on the amount of explained variability, accounting for upwards of 50% (0.790 – 0.84) or an MSE of ~0.8. The results were consistent across datasets A, B, and C. The removal of other study conditions (using LOO methods) did not have as large an impact, but the covariates were statistically significant >70% of the time. Study type, strain group, and dose spacing were all consistently significant covariates across datasets A, B, and C.

Three chemical classes (>15 chemical per class) represented within dataset A, phenols, conazoles, and carbamates, were used to stratify the dataset and MSE calculated for each group (Figure 2). Carbamate and conazole datasets produced MSE comparable to the MSE of the complete dataset A, despite having a smaller variance. However, the phenol dataset had an MSE of 0.18 potentially due to fewer chemical and study numbers. The significance of the variances between the chemical class and all comparisons produced a p-value > or = 0.05.

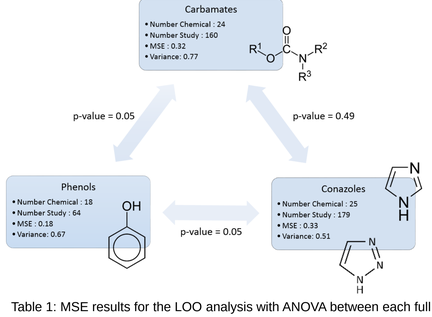

Figure 2: Variance estimation of three chemical class. A comparison of their variance were performed and results shown by the p-value.

Table 1: MSE 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	MSE	p-value	MSE	p-value	MSE	p-value
Full Model	0.326		0.337		0.326	
Chemical Removed	0.790	$< 6.43x10^{-4}$	0.844	$< 9.88x10^{-4}$	0.790	$< 6.43x10^{-4}$
Strain group Removed	0.356	$< 9.81x10^{-4}$	0.389	$< 3.23x10^{-4}$	0.356	$< 9.81x10^{-4}$
Study Type Removed	0.350	$< 3.34x10^{-4}$	0.354	$< 1.54x10^{-4}$	0.350	$< 3.34x10^{-4}$
Admin Method Removed	0.327	$9.16x10^{-2}$	0.338	$2.92x10^{-2}$	0.327	$9.16x10^{-2}$
Dose Spacing Removed	0.330	$< 9.64x10^{-4}$	0.339	$< 8.17x10^{-4}$	0.330	$< 9.64x10^{-4}$
Number of Dose Removed	0.331	$< 2.67x10^{-4}$	0.341	$< 1.08x10^{-4}$	0.331	$< 2.67x10^{-4}$
Study Year Removed	0.326	$1.45x10^{-1}$	0.337	$4.37x10^{-1}$	0.326	$1.45x10^{-1}$
Substance Purity Removed	0.326	$2.76x10^{-1}$	0.337	$1.90x10^{-1}$	0.326	$2.76x10^{-1}$
Study Source Removed	0.327	$4.17x10^{-2}$	0.338	$1.33x10^{-2}$	0.327	$4.17x10^{-2}$
Gender Removed	0.330	$< 4.29x10^{-4}$	0.339	$1.02x10^{-4}$	0.330	$< 4.29x10^{-4}$

Results

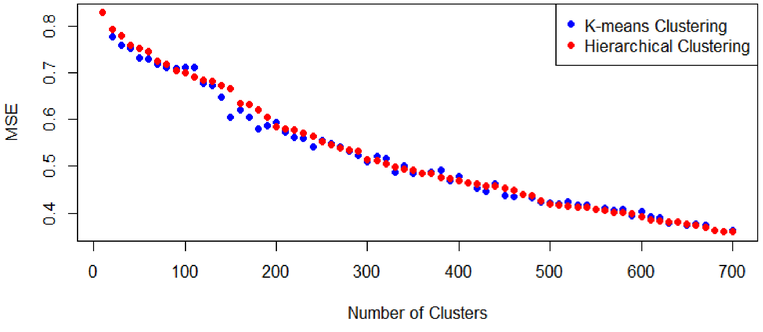

Figure 3: Plot showing the within MSE of the ANOVA analysis for both K-Means and Hierarchal clustering of toxprint chemotypes clustering methods. Data set for the ANOVA contained systemic toxicity studies from ToxRefDB of chemicals that were studied at least twice. The coefficient, chemical treatment, were replaced by their cluster group number for each analysis. Each analysis is defined as one run of "Number of Clusters", as shown on the x-axis.

Replacing chemical treatment with groupings based on structural similarities did not account for as much variance as using chemical treatment. The MSE for both K-means and Hierarchal clustering was not comparable to the 0.33 found when using chemical treatment until over 600 clusters were created (Figure 3). The MSE is equal to the residual sum of squares (RSS) divided by the degrees of freedom. The relationship between the MSE and RSS indicates that as the number of clusters go up, both the MSE and RSS go down. At 600 clusters, most clusters contained around one chemical, thereby mirroring the original analysis using chemical treatment.

Conclusions

In a linear regression analysis of data from more than 3,500 in vivo studies,
- **The spread around a predicted POD value is ~0.58**
  - Since the standard deviation of the log10 transformed PODs is about 0.5, the 95% prediction interval for a POD covers more than 2 orders of magnitude (10^-2*0.58, 10^2*0.58)
- Estimated unexplained variance across all datasets: ~0.33
- Chemical treatment explained: ~50% of the total variance, and so chemical features were explored further in an effort to account for additional variance.
- The estimated unexplained variance was consistent even when source data was stratified to be more homogeneous
- Stratifying chemical treatment across common classes with ≥15 members failed to explain additional variance, outside of phenols (which demonstrated ~20% unexplained variance)
- Replacing individual chemical treatments with chemical groups (clustered toxprint chemotypes) explained little additional variance

Reference
1. Hoffmann, S. *et al.* Acute oral toxicity: Variability, reliability, relevance and interspecies comparison of rodent LD50 data from literature surveyed for the ACuteTox project. *Regul. Toxicol. Pharmacol.* **58,** 395–407 (2010).
2. Kleinstreuer, N. C. *et al.* A Curated Database of Rodent Uterotrophic Bioactivity. *Environ. Health Perspect.* **124,** (2015).
3. Grisham, J. W. Interspecies comparison of liver carcinogenesis: implications for cancer risk assessment. *Carcinogenesis* **18,** 59–81 (1997).
4. Gold, L. S., Manley, N. B., Slone, T. H. & Ward, J. M. Compendium of chemical carcinogens by target organ: results of chronic bioassays in rats, mice, hamsters, dogs, and monkeys. *Toxicol Pathol* **29,** 639–652 (2001).