Estimation of the upper bound of predictive performance for alternative models that use *in vivo* reference data

Ly Ly Pham[1,2], Christopher Grulke[1], R. Woodrow Setzer[1], Matthew Martin[1], Katie Paul Friedman[1]
[1] National Center for Computational Toxicology, Office of Research and Development, US EPA; RTP, NC
[2] [1]Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee; Post-Doctoral Fellow

www.epa.gov

This poster may not reflect U.S. EPA policy.

Ly Ly Pham | pham.lyly@epa.gov | ORCID:0000-0001-8467-2645

INTRODUCTION

The large number of chemicals with limited toxicological information for chemical risk decision-making has motivated accelerated development of alternative models. Predictivity of these models is often evaluated via referencing animal toxicology studies, which are generally considered the standard for hazard assessment and point-of-departure (POD) determinations. However, variability in these *in vivo* reference data may limit the upper bound of predictivity for alternative models. To bound the expected predictive performance of models that reference *in vivo* studies, this work quantified variance within *in vivo* toxicity studies. Systemic toxicity POD values were extracted from the US EPA Toxicity Reference Database (ToxRefDB)[1] along with associated study parameters. **The goal of the current work was to quantify the amount of variance that exists within systemic *in vivo* PODs (both explained and unexplained).**

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

$$\text{Var(Observed POD)} = \text{Var(“True” POD)} + \text{Var(Study Conditions)} + \text{Unexplained Variance}$$

POD is defined as the Log_{10} mg/kg/day of the lowest dose in which a treatment related effect was observed per study, and includes lowest effect level (LEL) and lowest observable adverse effect (LOAEL) values. This work was further refined by calculating the variance per the equation above for LOAEL values only in order to understand the variance more specifically for adverse effects. LOAEL is defined as the Log_{10} mg/kg/day of the lowest dose in which critical effect was observed per study.

METHODS

Data Source and Preparation

Source: US EPA’s Toxicity Reference Database (ToxRefDBv1.3)
- Contains 5,890 in vivo toxicity studies for 1,144 chemicals.
- Guideline or guideline studies from various sources.

Data were filtered to only include:
- Adult animals in the F0 generation
- Systemic toxicity studies
  - Chronic (CHR), subchronic (SUB), developmental (DEV), multigeneration reproductive (MGR), and subacute (SAC)
- Administration Route: Oral
- Species: mouse, rat, dog, and rabbit
- Non-control group data
- The resulting 3,929 studies were used to create three datasets: Dataset A, B, and C (Figure 1 & 2)

Dataset A
Two or More Studies Per Chemical
Example:

Chemical	Study	Study Type	Species
1	1	CHR	Rat
1	2	CHR	Rat
1	3	CHR	Mouse
1	4	SUB	Mouse
2	1	CHR	Rat

Dataset B
Two or More Studies Per Study Type and Chemical
Example:

Chemical	Study	Study Type	Species
1	1	CHR	Rat
1	2	CHR	Rat
1	3	CHR	Mouse
1	4	SUB	Mouse
2	1	CHR	Rat

Dataset C
Two or More Studies, Per Study Type, Species, and Chemical
Example:

Chemical	Study	Study Type	Species
1	1	CHR	Rat
1	2	CHR	Rat
1	3	CHR	Mouse
1	4	SUB	Mouse
2	1	CHR	Rat

Shaded areas indicate studies that were removed from dataset

Figure 1: Filtering example for dataset A, B, and C. ***Dataset C is a subset of dataset B, and dataset B is a subset of Dataset A.***

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 parameters
- ANOVA was used to compare the significance of individual study parameters
- The same variance estimation was performed on LOAEL values only (a subset of the POD values). For chemicals where no adverse effect was observed, the LEL values were used.

Importance of Each Study Parameter
- Nested models using a leave one out (LOO) approach were used to test each study parameter’s contribution to the explainable variance.
- Toxicokinetics
  - Estimation of plasma steady state concentrations(Css) for 281 chemicals obtained from Wetmore in the HTTK package. Css values are estimated from an oral infusion in a 3 compartment model[2].
- MSE was then calculated for the subset of chemicals.

Evaluating Chemical Structures and Descriptors (Dataset A)
- Stratification of Data by Chemical Class
  - The top 3 classes with the most number of chemicals were stratified (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.
- Assessing naïve chemical groupings
  - Toxprint chemotypes (https://chemotyper.org/) were substituted for chemical treatment and then clustered using K-means and hierarchical methods
    - MSE was calculated at each clustering interval
  - Maximum Common Substructures (MCS)
    - Tanimoto coefficient was used to calculate MCS distance pairwise. The similarity matrix was then clustered using hierarchical methods.
    - MSE was calculated at each clustering interval.

RESULTS

Variance Calculations

Dataset A	Dataset B	Dataset C
Number of Chemicals: 761	Number of Chemicals: 679	Number of Chemicals: 567
Number of Studies: 3,929	Number of Studies: 3,170	Number of Studies: 2,357
POD	POD	POD
Variance: 1	Variance: 1.005	Variance: 0.985
MSE: 0.346	MSE: 0.337	MSE: 0.326
RMSE: 0.588	RMSE: 0.581	RMSE: 0.571
Variance Explained : 65.40 %	Variance Explained : 66.47 %	Variance Explained : 66.90 %
LOAEL	LOAEL	LOAEL
Variance: 0.887	Variance: 0.879	Variance: 0.847
MSE: 0.280	MSE: 0.274	MSE: 0.263
RMSE: 0.529	RMSE: 0.523	RMSE: 0.513
Variance Explained: 68.43 %	Variance Explained: 68.83 %	Variance Explained: 68.94 %

Figure 2: 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 2), 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. For the $\log_{10}$(LOAEL) analysis, the total variance was ~0.88, the residual MSE after adjusting for study parameters was ~0.26, and the root mean squared error (RMSE) was ~0.52. Using the LOAEL values reduced the total variance, as expected since the adverse effects observed for a chemical would be anticipated to be more consistent across studies, but the percent explained variance remained approximately the same as with the broader POD analysis.

Contribution of Study Parameters to Variance

Importance of Each Study Parameters Used in Full Model

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 & 2). Chemical had the largest impact on the amount of explained variability, accounting for upwards of 50% or an MSE range of 0.74-0.84. The results were consistent across datasets A, B, and C for both the POD and the LOAEL analysis. The removal of other study conditions (using LOO methods) did not have as large an impact, but some of the covariates were statistically significant. 7/10 of study conditions in the POD and 3/10 of study conditions in LOAEL were consistently significant covariates across datasets A, B, and C.

POD	*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*	
Models	MSE	p-value	MSE	p-value	MSE	p-value
Full Model	0.346		0.337		0.326	
Chemical Removed	0.871	< 0.00E-4	0.844	< 9.88E-4	0.790	< 6.43E-4
Strain group Removed	0.391	< 0.00E-4	0.389	< 3.23E-4	0.356	< 9.81E-4
Study Type Removed	0.368	< 2.81E-4	0.354	< 1.54E-4	0.350	< 3.34E-4
Admin Method Removed	0.347	2.42E-2	0.338	2.92E-2	0.327	9.16E-2
Dose Spacing Removed	0.348	< 1.11E-4	0.339	< 8.17E-4	0.330	< 9.64E-4
Number of Dose Removed	0.349	< 4.24E-4	0.341	< 1.08E-4	0.331	< 2.67E-4
Study Year Removed	0.346	2.14E-1	0.337	4.37E-1	0.326	1.45E-1
Substance Purity Removed	0.346	1.58E-1	0.337	1.90E-1	0.326	2.76E-1
Study Source Removed	0.347	5.99E-3	0.338	1.33E-2	0.327	4.17E-2
Gender Removed	0.347	1.68E-3	0.339	1.02E-4	0.330	< 4.29E-4

Table 1: MSE estimates of PODs for full model and full model with one study condition taken out for all three datasets of ToxRefDB. ANOVA was used to compare each leave one out model back to the full model.

LOAEL	*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*	
Models	MSE	p-value	MSE	p-value	MSE	p-value
Full Model	0.280		0.274		0.263	
Chemical Removed	0.801	<1.00E-04	0.790	<1.00E-04	0.742	<4.39E-4
Strain group Removed	0.319	<6.64E-04	0.316	<8.81E-4	0.290	<1.36E-4
Study Type Removed	0.290	<3.09E-04	0.283	<1.29E-4	0.279	<1.12E-21
Admin Method Removed	0.280	2.49E-01	0.274	1.27E-01	0.263	2.44E-01
Dose Spacing Removed	0.280	7.58E-01	0.274	5.02E-01	0.263	1.37E-01
Number of Dose Removed	0.281	1.12E-03	0.275	1.58E-03	0.265	1.89E-04
Study Year Removed	0.280	4.78E-01	0.273	8.35E-01	0.263	5.33E-01
Substance Purity Removed	0.280	5.26E-01	0.274	2.93E-01	0.263	3.02E-01
Study Source Removed	0.280	1.71E-01	0.274	1.52E-01	0.264	1.41E-01
Gender Removed	0.280	1.88E-01	0.274	2.61E-02	0.265	9.18E-04

Table 2: MSE estimates of LOAELs for full model and full model with one study condition taken out for all three datasets of ToxRefDB. ANOVA was used to compare each leave one out model back to the full model.

Toxicokinetics as a Study Parameter

- Css values were obtained for 281 chemicals to understand if toxicokinetic parameters associated with each chemical might contribute to the observed variance.
- MSE for the 281 chemical subset was ~0.37.

Alternative Chemical Descriptors to Predicting POD

Stratification of Data by Chemical Class

Three chemical classes with the most representation within dataset A, (phenols, conazoles, and carbamates) were used to stratify the dataset and MSE calculated for each group (Figure 3). 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.

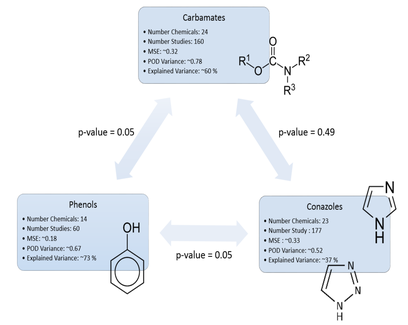

Figure 3: Variance estimation of three chemical class. A comparison of their variance were performed and results shown by the p-value.

Assessing naïve chemical groupings

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 of Toxprint chemotypes and Hierarchal clustering of MCS were not comparable to the 0.33 MSE found when using chemical treatment (Figure 4). 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. Even with the smaller clusters, the MSE did not equal that of individual chemicals.

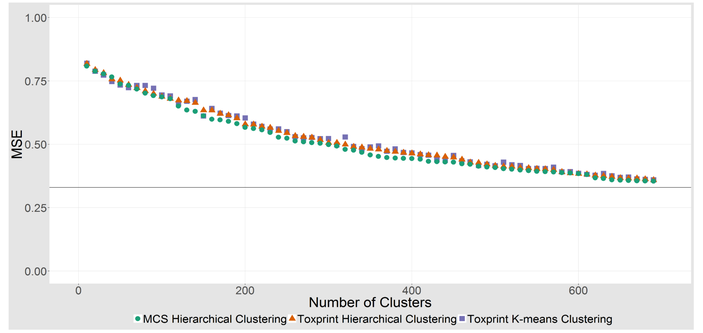

Figure 4: Plot showing the within MSE of the ANOVA analysis for K-Means and Hierarchal clustering of ToxPrint chemotypes and the Hierarchial clustering for MCS. The coefficient, chemical treatment, were replaced by their cluster group number for each analysis. Each analysis is defined as one run of “Number of Clusters”, as shown on the x-axis.

SUMMARY & CONCLUSIONS

In a linear regression analysis of data from approximately 3,500 in vivo studies,

Variance Calculation – Proportion of Explained and Unexplained
- Estimated unexplained variance for PODs across all datasets is ~0.33. Even when datasets were made more homogeneous, the unexplained variance in the POD values was still approximately one-third of the total variance.
- **The RMSE was ~0.52 to 0.58 in log10(mg/kg/day) units, indicating the minimum predictive interval for the POD data used in this analysis.**
  - For a POD of 10 mg/kg/day, the minimum confidence interval would be 3.02-33.11 mg/kg/day.

Contribution of Study Parameters
- Chemical explained ~50% of the total variance, and so chemical features were explored further to understand if it would be feasible to predict POD values in this set using chemical groupings or descriptors.
  - 7/10 of the study parameters evaluated were a contributor to POD variance – but proportionately much less than chemical
  - Adding predicted Css as a study parameter did not explain additional variance

Chemical Descriptors to Predict POD
- Stratifying chemical treatment across common classes failed to explain additional percentage of the total variance, outside of phenols (which demonstrated a marginal improvement, ~20% unexplained variance)
- Replacing individual chemical treatments with chemical groups (clustered toxprint, chemotypes, and MCS) did not show that clusters can serve as chemical surrogates

Reference
1. Martin, M.T., Judson, R.S., *et al.* (2009) Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef database. *Environ. Health Perspect.*, **117**, 392–399.
2. Pearce, R., Strope, C., Setzer, R.W., Sipes, N., Wambaugh, J.F. httk: R package for high-throughput toxicokinetics. *J. Stat. Softw. (in press)*