Variability in organ-level effects in repeat dose animal studies

Katie Paul Friedman[1], Richard S. Judson[1]

[1]Center for Computational Toxicology and Exposure, US EPA, Research Triangle Park, NC USA

www.epa.gov

Katie Paul Friedman | paul-friedman.katie@epa.gov

Abstract 2546/Poster P284

March 22, 2021

Society of Toxicology Virtual Annual Meeting

This poster does not necessarily reflect EPA policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Building scientific confidence in the use of new approach methodologies (NAMs) in safety assessment may include performance comparison to *in vivo* study outcomes. This work defines the variability in organ-level effects and suggests qualitative and quantitative benchmarks for maximum NAM performance for prediction of organ-level effects in repeat dose studies of adult animals. Previous work suggests that the root residual mean square error (RMSE) for study-level lowest effect level (LEL) values (on a $\log_{10}$-mg/kg/day basis) approaches 0.5 $\log_{10}$-mg/kg/day.

Table 1. Questions on animal study reproducibility

3 main questions	What is the range of possible systemic effect values (mg/kg/day) in replicate studies?	What is the maximal accuracy of a model that attempts to predict a systemic effect values for an unknown chemical?	What is the probability that an effect in adult animals will be observed in replicate studies?
Statistical approach to the question	• Residual root mean square error (RMSE) is an estimate of variance in the same units as the systemic effect values. • The RMSE can also be used to define a minimum prediction interval, or estimate range, for a model.	• The mean square error (MSE) is used to approximate the unexplained variance (not explained by study descriptors). • This unexplained variance limits the R-squared on a new model.	• Understand the reproducibility of treatment-related changes in specific endpoint targets (e.g., any effect on liver).

Previous work suggests that the variance (estimated by RMSE) in study-level effect values from repeat dose studies in animals approaches 0.5 log10-mg/kg/day.

- Total variance in systemic toxicity effect values likely approaches 0.75-1 (units of $(\log_{10}\text{-mg/kg/day})^2$)
- MSE (unexplained variance) is 0.2 – 0.4 (units of $(\log_{10}\text{-mg/kg/day})^2$)
- RMSE is 0.45-0.60 $\log_{10}$-mg/kg/day
- RMSE is used to define a 95% minimum prediction interval (i.e., based on the standard deviation or spread of the residuals)

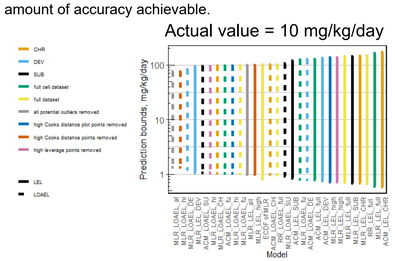
Calculating the minimum prediction interval width based on results of multi-linear regression modeling

$$lower\ bound\ on\ prediction = 10^{qnorm(0.025*RMSE)}$$

$$upper\ bound\ on\ prediction = 10^{qnorm(0.075*RMSE)}$$

Figure 1. Estimating minimum prediction intervals based on animal study variance (based on Pham *et al.*, 2020).

If attempting to use a NAM-based predictive model for prediction of a reference systemic effect level value of 10 mg/kg/day, it is likely that given the variability in reference data of this kind, that a model prediction of somewhere between 1 and 100 mg/kg/day would be the greatest amount of accuracy achievable.

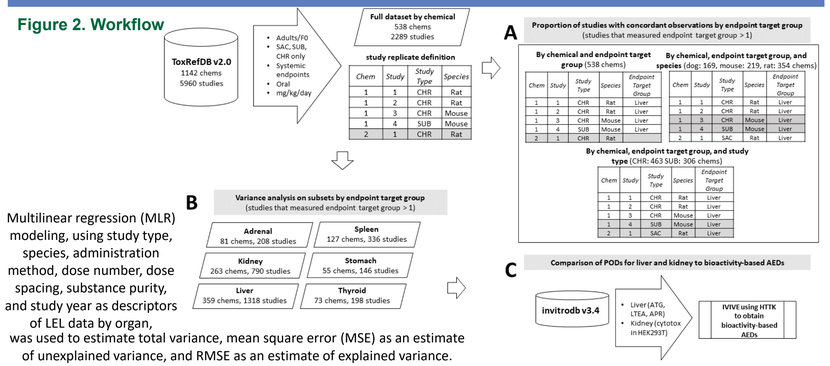


Based on Pham LL, Watford S, Pradeep P, Martin MT, Thomas RS, Judson RS, Setzer RW, Paul Friedman K. *Accepted.* "Variability in in vivo studies: Defining the upper limit of performance for predictions of systemic effect levels." Computational Toxicology.
https://doi.org/10.1016/j.comtox.2020.100126

U.S. Environmental Protection Agency
Office of Research and Development

Observations of liver, kidney, stomach, spleen, thyroid and adrenal gland from the Toxicity Reference database (v2.0) were included in this analysis to understand the reproducibility of organ-level effects.

Figure 2. Workflow



Multilinear regression (MLR) modeling, using study type, species, administration method, dose number, dose spacing, substance purity, and study year as descriptors of LEL data by organ, was used to estimate total variance, mean square error (MSE) as an estimate of unexplained variance, and RMSE as an estimate of explained variance.

A: What is the qualitative reproducibility of organ-level effect observations in repeat dose studies of adult animals?

Table 2. Repeated concordance of organ-level findings.

$$\%\ Concord = \frac{chemical\ with\ positive\ finding\ in\ all\ studies + chemicals\ with\ negative\ finding\ in\ all\ studies}{total\ chemicals\ tested}$$

Endpoint target group	% Concord	Chem	+Pos	-Neg	Mixed
adrenal	60.2	538	8	316	214
Kidney	38.8	538	54	155	329
Liver	42.4	538	149	79	310
Spleen	56.5	538	17	287	234
Stomach	71.7	538	14	372	152
Thyroid	66.2	538	11	345	182

Endpoint target group	Study Type	% Concord	Chem	+Pos	-Neg	Mixed
adrenal	CHR	67.8	463	8	306	149
kidney	CHR	49	463	58	169	236
liver	CHR	54.6	463	160	93	210
spleen	CHR	67.8	463	16	298	149
stomach	CHR	79	463	22	344	97
thyroid	CHR	70	463	10	314	139
adrenal	SUB	73.5	306	10	215	81
kidney	SUB	52.6	306	65	96	145
liver	SUB	66	306	143	59	104
spleen	SUB	68	306	24	184	98
stomach	SUB	85	306	10	250	46
thyroid	SUB	81	306	11	237	58

% Concord = percent concordant chemicals; Chem = total # chemicals tested at the endpoint target group; +Pos = # chemicals with positive observations in all available studies; -Neg = # chemicals with negative observations in all available studies; Mixed = chemicals with at least 1 study that was not positive

Endpoint target group	Species	% Concord	Chem	+Pos	-Neg	Mixed
adrenal	dog	84.6	169	8	135	26
	mouse	84	219	6	178	35
	rat	66.9	354	17	220	117
kidney	dog	67.5	169	20	94	55
	mouse	63.5	219	43	96	80
	rat	57.6	354	106	98	150
liver	dog	71	169	86	34	49
	mouse	67.1	219	96	51	72
	rat	61.3	354	157	60	137
spleen	dog	78.1	169	9	123	37
	mouse	74	219	16	146	57
	rat	65.5	354	31	201	122
stomach	dog	87.6	169	2	146	21
	mouse	80.4	219	7	169	43
	rat	79.9	354	11	272	71
thyroid	dog	78.7	169	8	125	36
	mouse	90.4	219	3	195	21
	rat	77.4	354	28	246	80

Replicate studies were defined by chemical only, chemical and species, and chemical and study type to estimate concordance in observed organ-level effects (repeated presence/absence of weight, gross or histopathological changes) for a chemical. Total concordance (% chemicals with positive or negative agreement across replicates), depending on the organ and replicate definition, ranged from 39 - 88%, with slightly greater average within-species concordance. Organs associated with more negative chemicals (stomach, thyroid, adrenal) had slightly higher rates of concordance in this range.

B: What is the variance in organ-level effects in repeat dose studies, and is it smaller than study-level variance?

Table 3. Results of MLR to estimate unexplained and explained variance in organ LELs.

$$organLEL \sim b_0 + chemical * b_1 + species * b_2 + study\ type * b_3 + administration\ method * b_4 + dose\ spacing * b_5 + number\ of\ dose\ levels * b_6 + study\ year * b_7 + \%\ substance\ purity * b_8$$

Chems = # chemicals; N = number of studies; Var = total variance; MSE = mean square error on the model; RMSE = root residual mean square error; % var explained = % of total variance explained by study descriptors

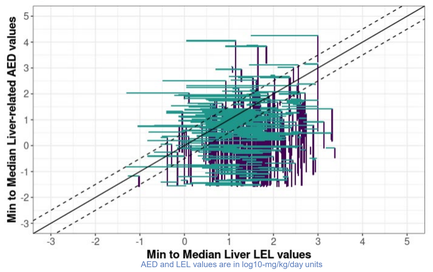
Endpoint Target Group	Chem	N	Var	MSE	RMSE	% var explained
adrenal	81	208	0.756	0.349	0.591	53.8
kidney	263	790	0.765	0.316	0.562	58.7
liver	359	1318	0.745	0.355	0.596	52.3
spleen	127	336	0.671	0.318	0.564	52.6
stomach	55	146	0.553	0.173	0.416	68.7
thyroid	73	198	0.721	0.378	0.615	47.6

Total variance at the organ level is generally less than or equal to total variance at the study-level. The RMSE at the organ level is similar to the study level RMSE in Pham *et al.* The % variance explained is similar to the lower estimate of % variance explained at the study level in Pham *et al.*

C. How well do liver-related bioactivity-derived administered equivalent doses approximate liver-related LEL values?

A 3 compartment steady-state model (R library(httk)) was used in the *in vitro* to *in vivo* extrapolation of administered equivalent doses in order to include the largest number of chemicals in this comparison. The plasma steady state concentration for the median individual based on Monte Carlo simulation of human physiological parameters was used.

Figure 3. Liver-related AEDs vs. Liver LEL values.



The linear relationship between liver LELs and liver-related AEDs is poor, but the AED values generally provide a conservative estimate of liver LEL.

- Wilcoxson tests of the LEL and AED values by DSSTox chemical identifier suggest that mean values are different ($p<0.05$) for 163/306 chemicals in this comparison (53% of the time).
- However, the median liver-related AED was less than the minimum liver LEL value for 175/306 chemicals (57%) and the minimum liver-related AED was less than the minimum liver LEL value for 267/306 (87%). 286/306 chemicals (93%) demonstrated a minimum liver-related AED less than the minimum liver LEL + 0.5 log10-,mg/kg/day.

With MSE used to indicate unexplained variance, results suggest study descriptors accounted for 52-69% of the total variance in organ-level LELs. A NAM would be unlikely to explain more than the variance explained by study level descriptors, or 70%, of the variance in these data. The RMSE for these organ-level statistical models ranged 0.4 – 0.6 $\log_{10}$ mg/kg/day, suggesting organ-level variance in LEL values was similar to overall study LEL variance. Therefore, a good NAM might predict organ-level LELs within ± 1 $\log_{10}$-mg/kg/day. This work suggests thresholds on NAM accuracy for repeat dose, organ-level effects in adult animals.