

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

Ly Ly Pham<sup>1,2</sup>, Katie Paul Friedman<sup>1</sup>, R. Woodrow Setzer<sup>1</sup>, Matthew Martin<sup>1</sup>

<sup>1</sup> National Center for Computational Toxicology, Office of Research and Development, US EPA; RTP, NC

<sup>2</sup> ORISE Post-Doctoral fellow

The large number of chemicals with limited toxicological information for chemical risk decision-making has 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 limit the upper bounds 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. Using the US EPA Toxicity Reference Database (ToxRefDB) systemic toxicity POD values and associated study parameters (e.g., chemical treatment, study type, species, strain, and dose spacing), multiple linear regression and analysis of variance were performed to quantify the explained variance. Unexplained variance, the portion of the total variance not explained by known study parameters, was estimated using mean squared error (MSE). The total variance in the set of  $\log_{10}(\text{POD})$  values was approximately 1, and the residual MSE after adjusting for study parameters was  $\sim 0.33$ . The root mean squared error (RMSE) was  $\sim 0.58$ ; indicating that the best any alternative model prediction of  $\log_{10}(\text{POD})$  would be within  $\pm 0.58 \log_{10}(\text{mg/kg/day})$  units. Chemical treatment appeared to account for a significant fraction of the explained variability [ $0.46 \log_{10}(\text{POD})$ ], suggesting that chemical structure descriptors could be used as a surrogate for chemical identity in predictive modeling. To test this hypothesis, two approaches were used to evaluate the impact of chemical descriptors on explained variance. The first approach stratified the dataset using chemical class, and the second approach used ToxPrint chemotype fingerprints to group chemicals based on structural features. Use of chemical class and the study parameters marginally improved the explained variance over study parameters and use of chemotype fingerprints failed to provide groupings that improved explained variance over individual chemical identity. Both approaches demonstrated the limited capacity for chemical structure to predict complex, systemic POD endpoints for a heterogeneous dataset. This characterization of unexplained variance in *in vivo* data suggests that the upper bound on the residual MSE for predictive models of *in vivo* POD data may approach  $0.33 \log_{10}(\text{mg/kg/day})$ . *This abstract may not reflect U.S. EPA policy.*