PREDICTIVE MODEL OF SYSTEMIC TOXICITY

Matt Martin[1], LyLy Pham[1], Jacques Clouzeau[2], Sophie Loisel-Joubert[2], Delphine Blanchet[2], Hicham Noçairi[2], Gladys Ouedraogo[2]

[1]US EPA- ORD – NCCT, RTP, NC [2]L'Oréal, Paris, France

Matt Martin | martin.matt@epa.gov | 919-541-4104

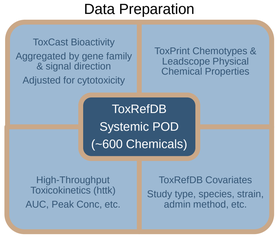
Abstract & Objective

In an effort to ensure chemical safety in light of regulatory advances away from reliance on animal testing, USEPA and L'Oréal have collaborated to develop a quantitative systemic toxicity prediction model. Prediction of human systemic toxicity has proved difficult and remains a gap in chemical safety assessment using alternative approaches. By leveraging multiple data sources including high-throughput screening (ToxCast Bioactivity) and chemical descriptor and property (ToxPrint chemotypes and Leadscope Chemical Properties), and high-throughput toxicokinetic (httk) data, a predictive model of systemic point-of-departures (POD) was developed. The model specifically predicts the chemical-level POD using data from roughly 3000 studies across nearly 600 chemicals. Systemic POD were curated in ToxRefDB from numerous study types, across multiple species and dose administration methods. Rather than attempt to adjust all POD to a single species or strain or dose duration or study type or administration method, these *in vivo* study parameters were included as covariates in the modeling process. Using Random Forest modeling, *in vivo* covariates alone accounted for roughly 17.5% of the variance in the dataset. A model developed using a combination of *in vivo* covariates, biological, chemical, and kinetic parameters explained at total of 35% of the variance. The combination of covariates and features explain more variance in the data than either do individually, demonstrating the advantage of incorporating *in vivo* covariates into the modeling process instead of adjusting POD a priori. The final resulting model was also enriched for features measuring xenobiotic metabolism gene expression as well oxidative stress markers demonstrating the importance for accounting for kinetics and non-specific bioactivity in predicting systemic toxicity. Herein, we have generated an externally predictive model of systemic toxicity capable of being used as a safety assessment tool.

Objective: Develop a predictive model of systemic toxicity point-of-departures incorporating inherent differences in *in vivo* study parameters, chemical structure and properties, bioactivity and kinetics for use as a safety assessment tool providing appropriate performance baselines, benchmarks, and uncertainty estimates.

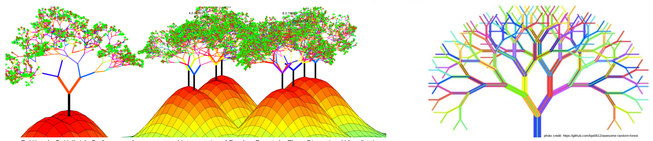
Methods

Data Preparation



1. Derive chemical-level points-of-departure (POD)
2. Characterize POD to serve as Covariates:
   - LOEL vs LOAEL & NOEL vs NOAEL
   - No. of Studies
   - Study type, species, admin method, etc.
   - Dose spacing
3. Merge input datasets (ToxCast, Toxprint, RTK, etc)
   - Impute missing data (median)
4. Generate random forest models
   - Sample across NO(A)EL and LO(A)EL
   - Combine trees and forests for final model
5. Evaluate models
   - Performance compared to baseline & benchmarks
   - Biological plausibility

Random Forest Model Generation

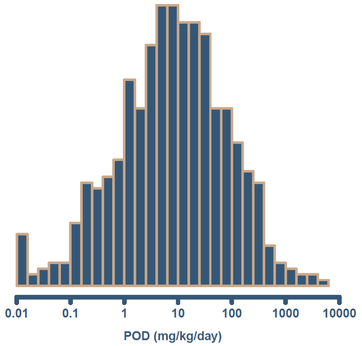


R. Hänsch, O. Hellwich, Performance Assessment and Interpretation of Random Forests by Three-Dimensional Visualizations, IVAPP 2015 - Proceedings of the International Conference on Information Visualization Theory and Application, March 2015

photo credit: https://github.com/bpb962/awesome-random-forest

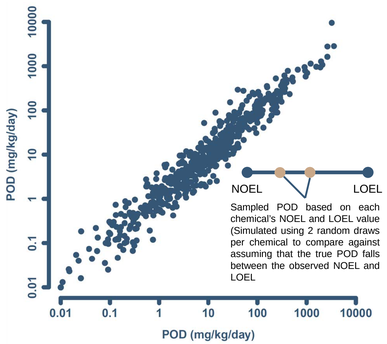
Model Baseline & Benchmark

Baseline: Predict the population mean



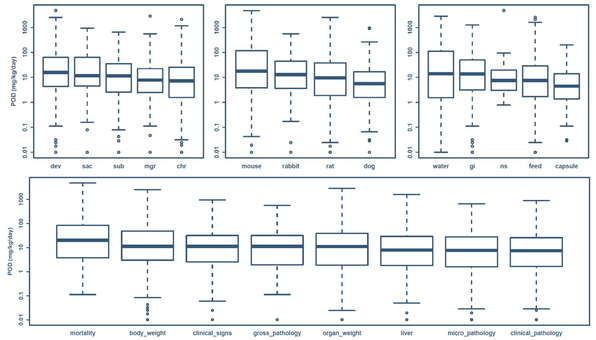
SD ≈ 1 (log mg/kg/day) ≈ RMSE
0 % Variance Explained

Benchmark:



RMSE ≈ 0.3 (log mg/kg/day)
92% Variance Explained

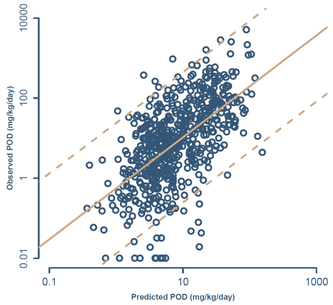
Incorporating *In Vivo* Covariates



- Boxplots illustrate the relationship between *in vivo* variables (corvariates) and chemical-level PODs across the 596 chemicals
- EXAMPLE: A chemical where its POD is defined by critical effects in a CHR (chronic) study will have, on average, a lower POD than a chemical where its POD was defined by a DEV (developmental) study

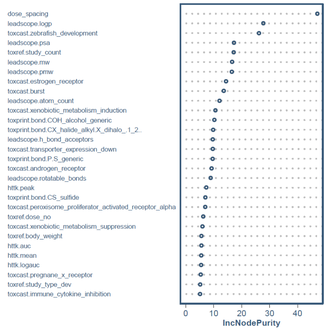
- 7% variance explained by dose spacing alone (RMSE = 0.98)
- 17.5% variance explained by all *in vivo* covariates (RMSE = 0.93)
- Dose spacing, study count, number of doses, pod type (LEL vs LOAEL), body weight were most important contributors to variance

Model Results & Performance



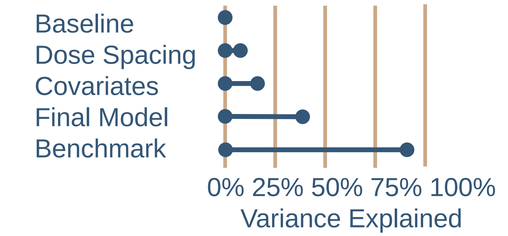
RMSE ≈ 0.8 (log mg/kg/day)

- 35% variance explained by all input variables, including *in vivo* covariates
- 17.5% of the 35% variance was explained by biological, chemical, and kinetic parameters



Biologically and chemically plausible variables selected:
Zebrafish development, xenobiotic metabolism, oxidative stress, cytotoxicity, gene induction, PXR, ER, LogP, OPs, Peak concentration

Summary Results, Conclusions & Future Directions



FUTURE DIRECTIONS

- Include completed high-throughput toxicokinetics (httk) datasets as well as exploration of further curated chemical structure and property datasets
- Benchmark dose modeling of *in vivo* endpoints for improved POD estimates
- Improve understanding of *in vivo* study variability and uncertainty for improved benchmarking

CONCLUSIONS

- Dose spacing is the single largest contributor of explained variance
- *In vivo* covariates provide context around individual chemical POD
- Final model explains additional variance with biologically and chemically plausible parameters being of highest importance
- Unexplained variance not a product of model development failure, but that of specific factors and uncertainty inherent to *in vivo* POD derivations themselves