Uncertainty Quantification in High Throughput Screening: Applications to Models of Endocrine Disruption, Cytotoxicity, and Zebrafish Development

Eric D. Watt [1,2], Richard Judson [2]

[1] ORISE, Oak Ridge, TN, [2] National Center for Computational Toxicology, ORD, US EPA, RTP, NC

Using uncertainty quantification, we aim to improve the quality of modeling data from high throughput screening assays for use in risk assessment. ToxCast is a large-scale screening program that analyzes thousands of chemicals using over 800 assays representing hundreds of biochemical and cellular processes, including endocrine disruption, cytotoxicity, and zebrafish development. Over 2.6 million concentration response curves are fit to models to extract parameters related to potency and efficacy. Models built on ToxCast results are being used to rank and prioritize the toxicological risk of tested chemicals and to predict the toxicity of tens of thousands of chemicals not yet tested in vivo. However, the data size also presents challenges. When fitting the data, the choice of models, model selection strategy, and hit call criteria must reflect the need for computational efficiency and robustness, requiring hard and somewhat arbitrary cutoffs. When coupled with unavoidable noise in the experimental concentration response data, these hard cutoffs cause uncertainty in model parameters and the hit call itself. The uncertainty will then propagate through all of the models built on the data. Left unquantified, this uncertainty makes it difficult to fully interpret the data for risk assessment.

We used bootstrap resampling methods to quantify the uncertainty in fitting models to the concentration response data. Bootstrap resampling determines confidence intervals for both the potency and efficacy parameters. In addition, resampling allowed us to shift model selection and hit calling from a binary determination to a probabilistic quantification. These uncertainty estimates were then propagated through mathematical models built to predict bioactivity from ToxCast data. We explored the utility of our method by quantifying uncertainty in numerous examples: estrogen and androgen model agonist scores; selectivity scoring of in vitro thyroid peroxidase potency relative to cytotoxicity and non-specific enzyme inhibition; bioactivity assay potencies relative to global cytotoxicity; and developmental toxicity results from zebrafish assays. By quantifying the uncertainty in model predictions, we separated biological activity from assay noise, improved the quality of model output, and thus increased confidence in model predictions for use in risk assessments.