



Uncertainty Quantification in High Throughput Screening

Applications to Models of Endocrine Disruption, Cytotoxicity, and Zebrafish Development

Eric Dean Watt | NCCT/ORD/EPA | watt.eric@epa.gov | <http://www.ericdwatt.com>

June 26th, 2016

The views expressed in this presentation are those of the author[s] and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Motivation

Motivating Questions

What are the impacts of uncertainty in high throughput screening (HTS)?

How can we quantify uncertainty?

How can we propagate the uncertainty through models and analysis built on HTS results?

ToxCast

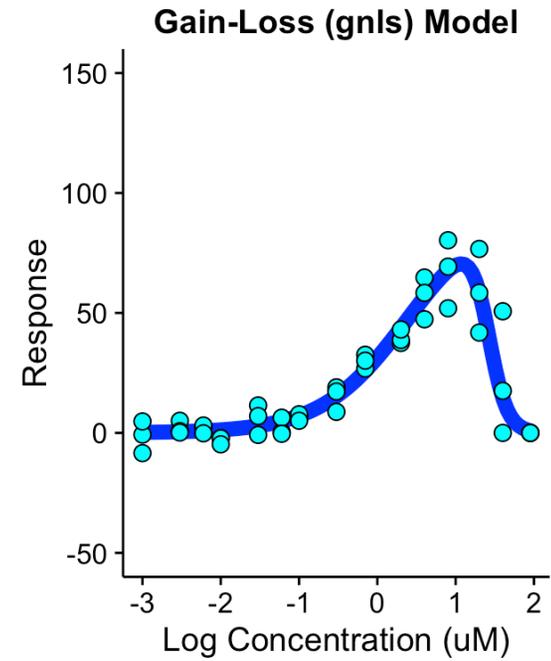
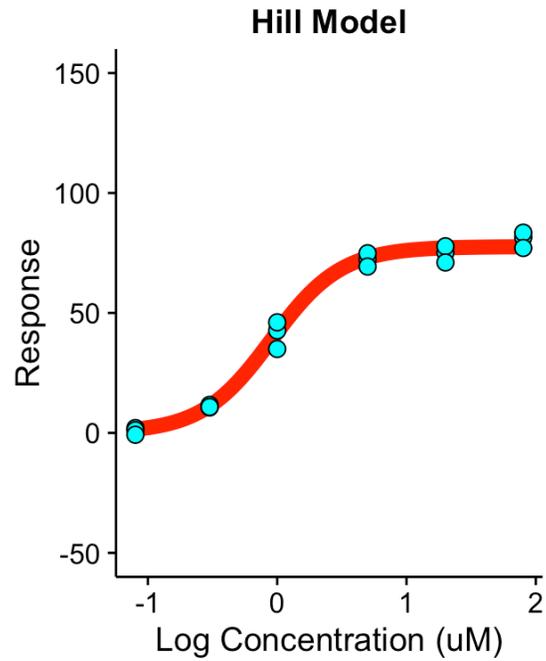
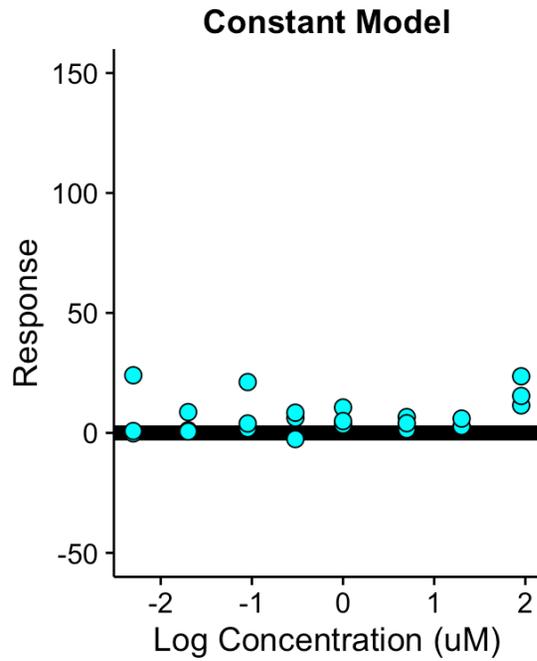
Over 2.6 million concentration response curves from *in vitro* assays

- Dozens of sources with different:
 - technologies
 - concentration spacings
 - response profiles

High throughput analysis requires selection of somewhat arbitrary cutoffs

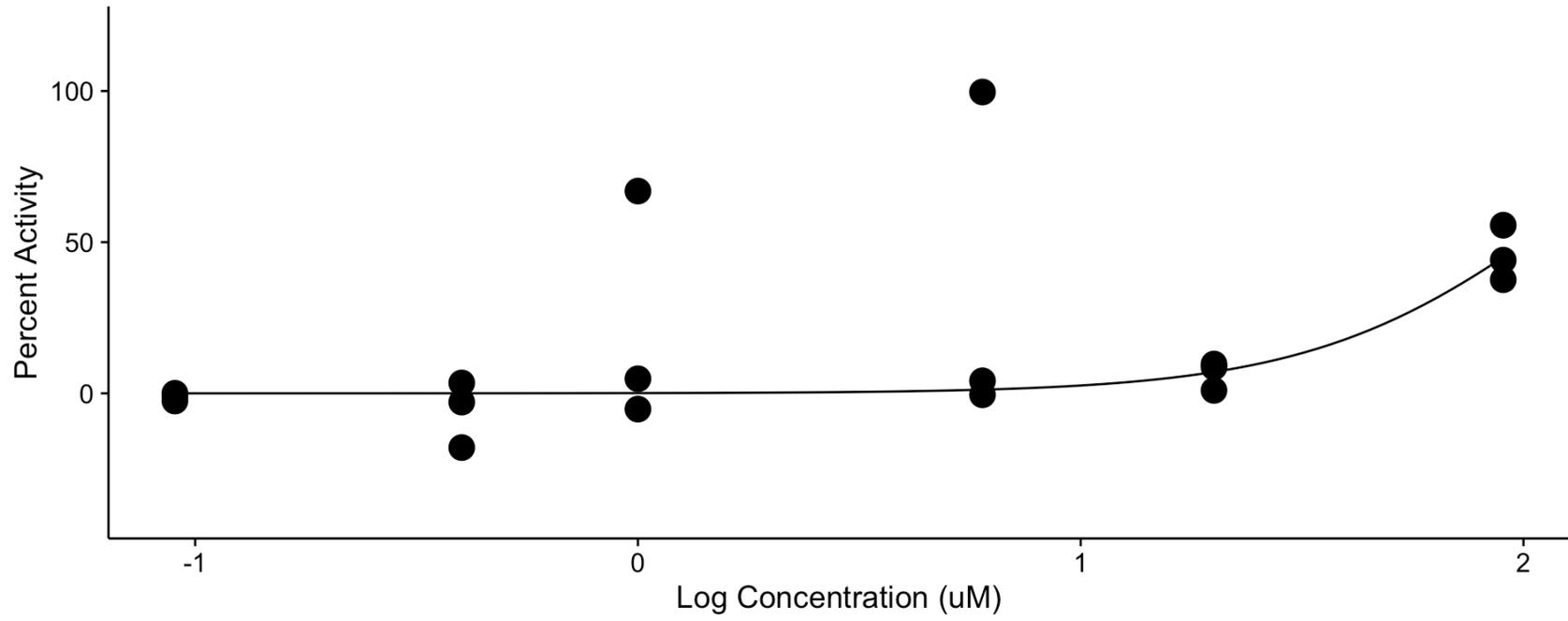
Quantifying uncertainty and confidence intervals helps separate biological activity from assay noise

ToxCast Models

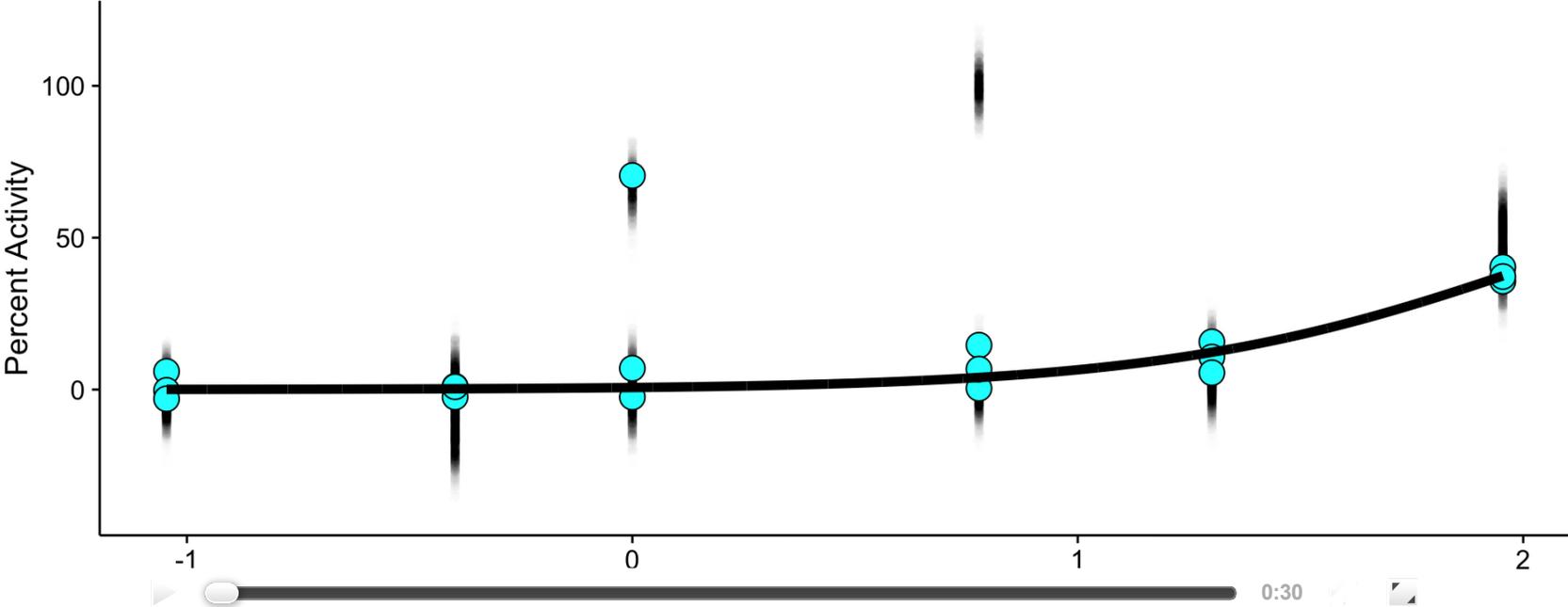


Quantifying Uncertainty

ToxCast Experimental Values and Hill Model

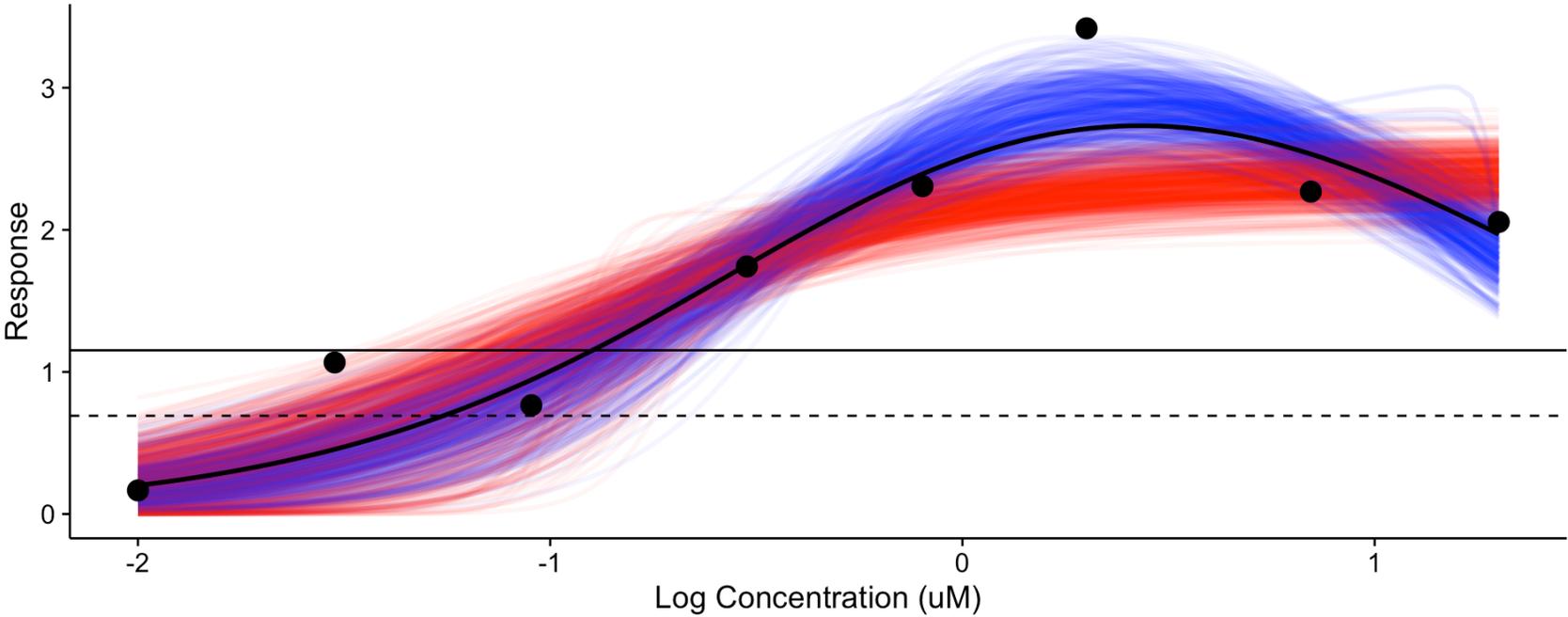


Bootstrap Sampling



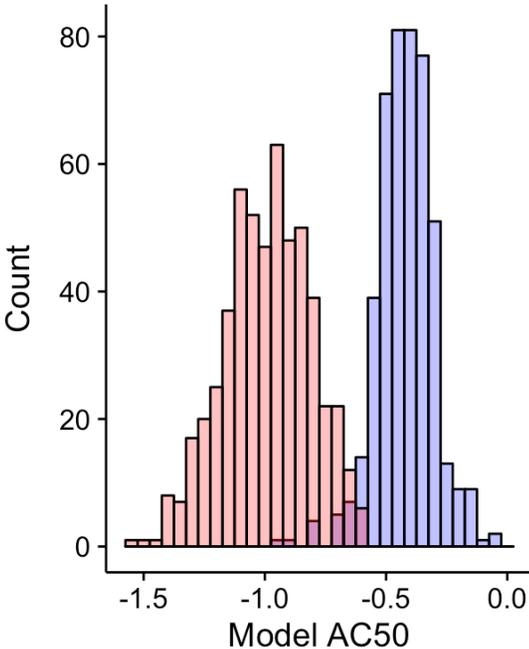
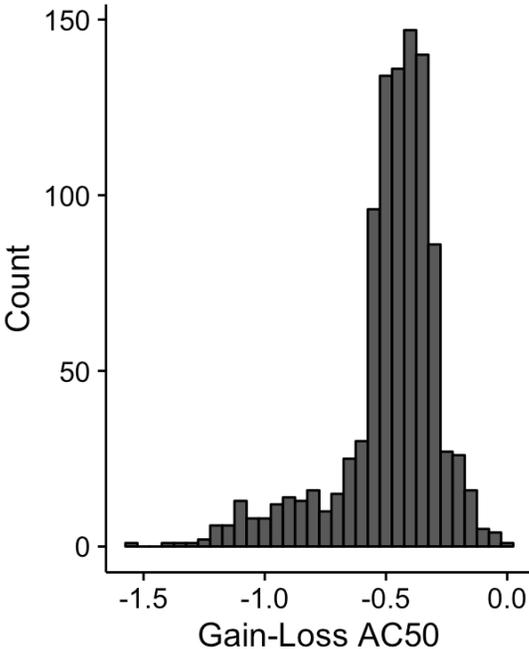
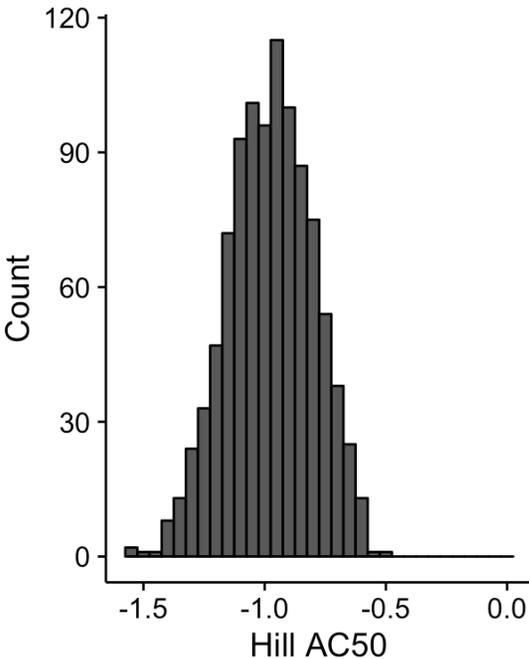
Model Selection

Select the winning model for each bootstrap sample



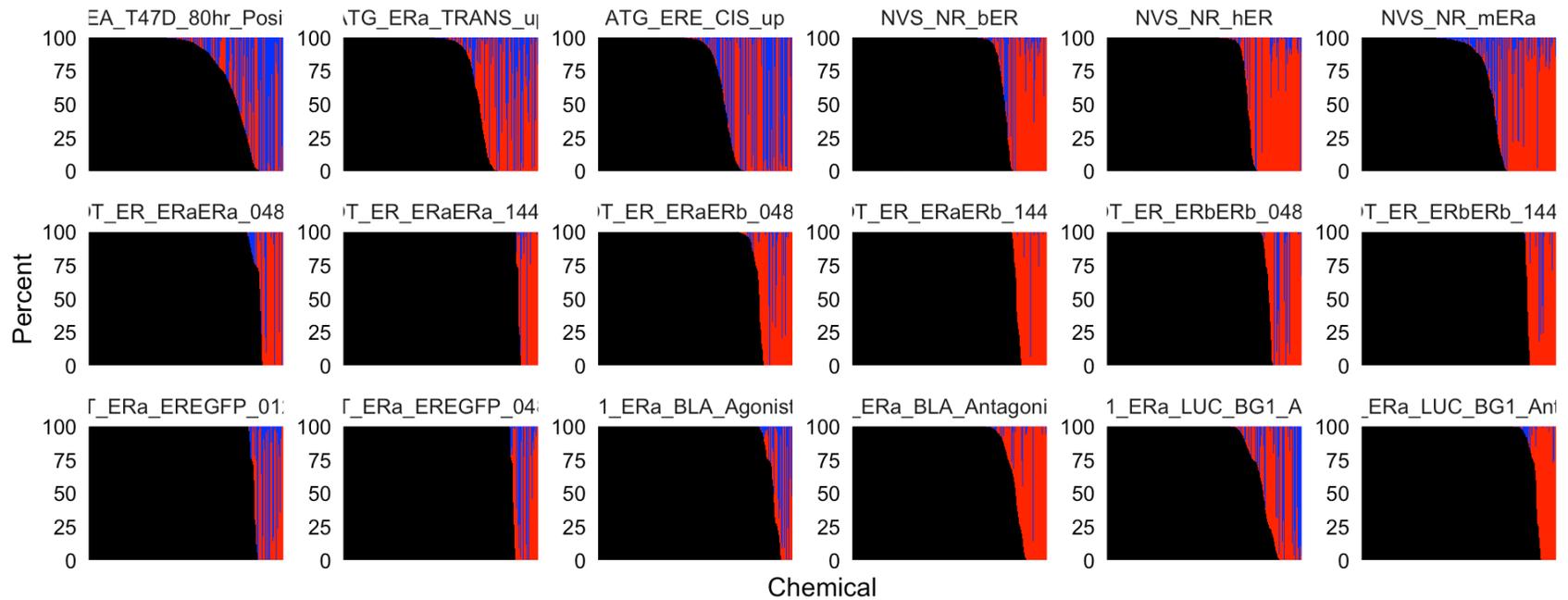
Model Selection

Distribution of model parameters



Hit Call

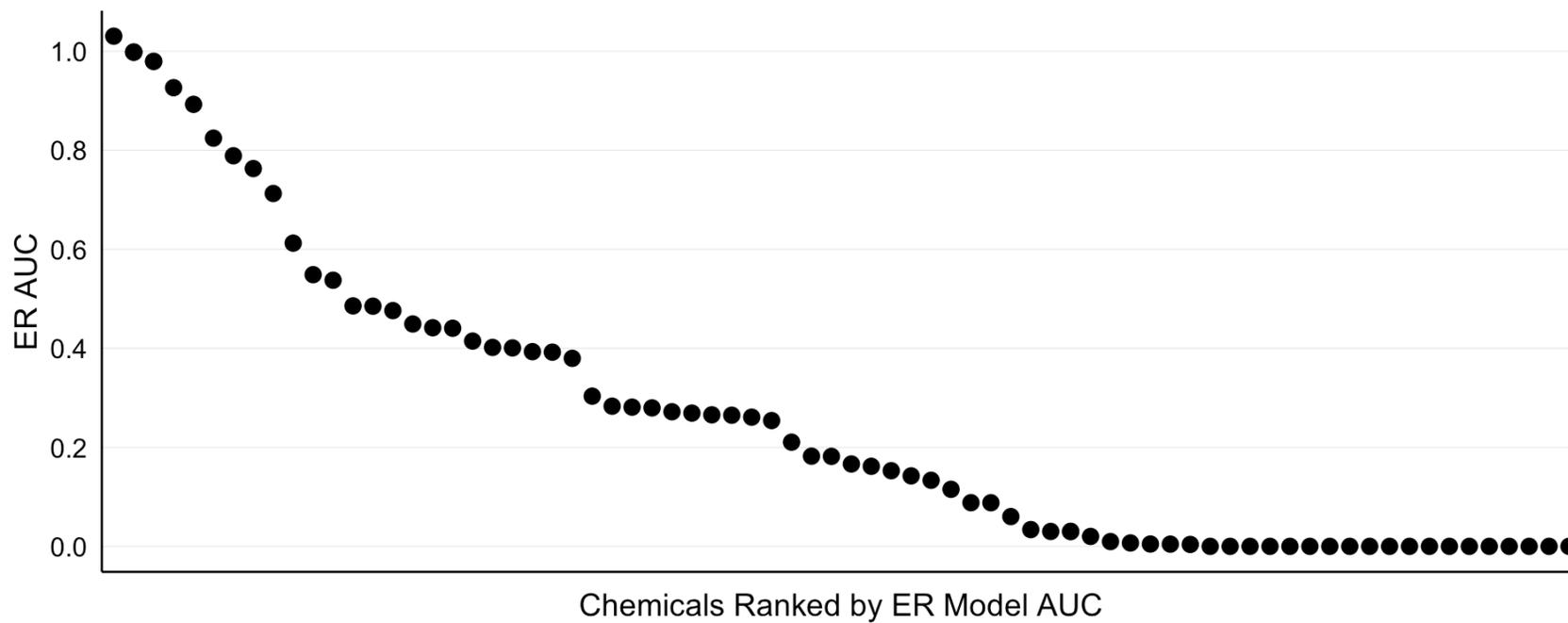
Convert activity determination from binary to continuous probability value



Applications

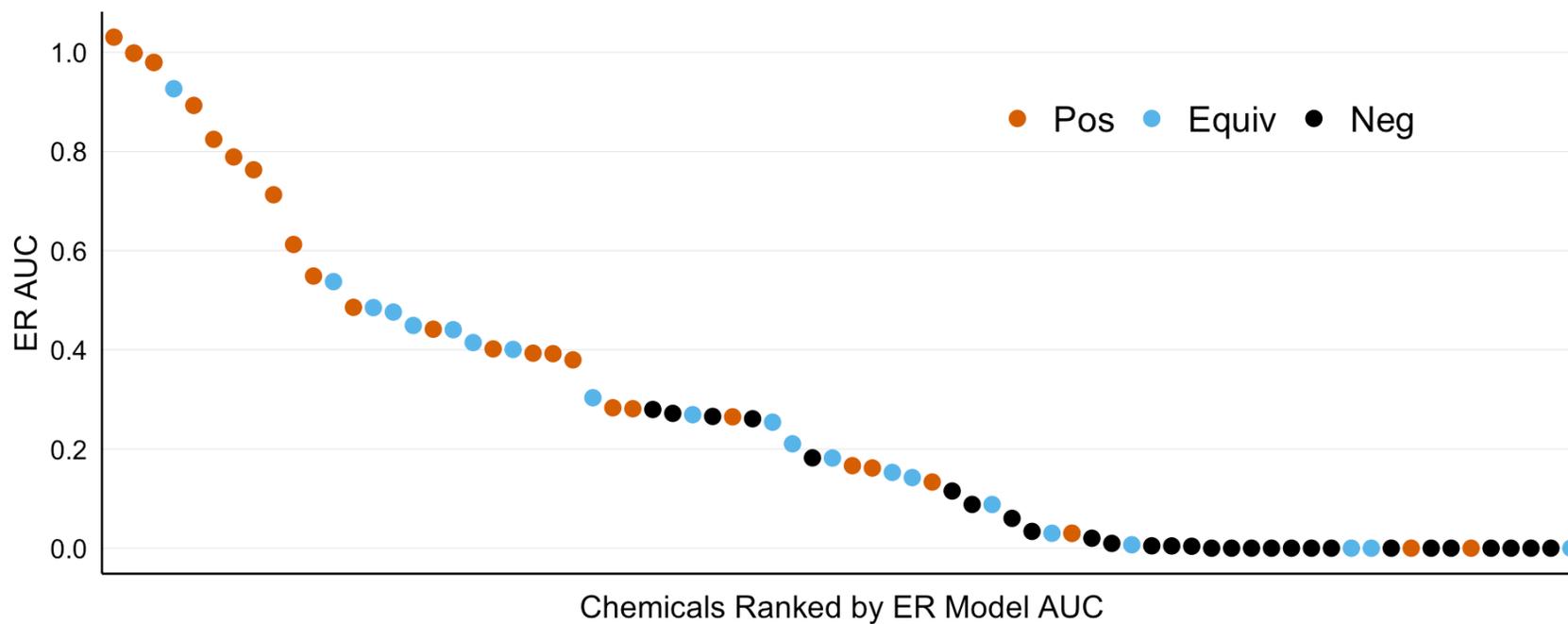
ER Model

Activity cutoff = 0.1 AUC, what is the uncertainty around this number?



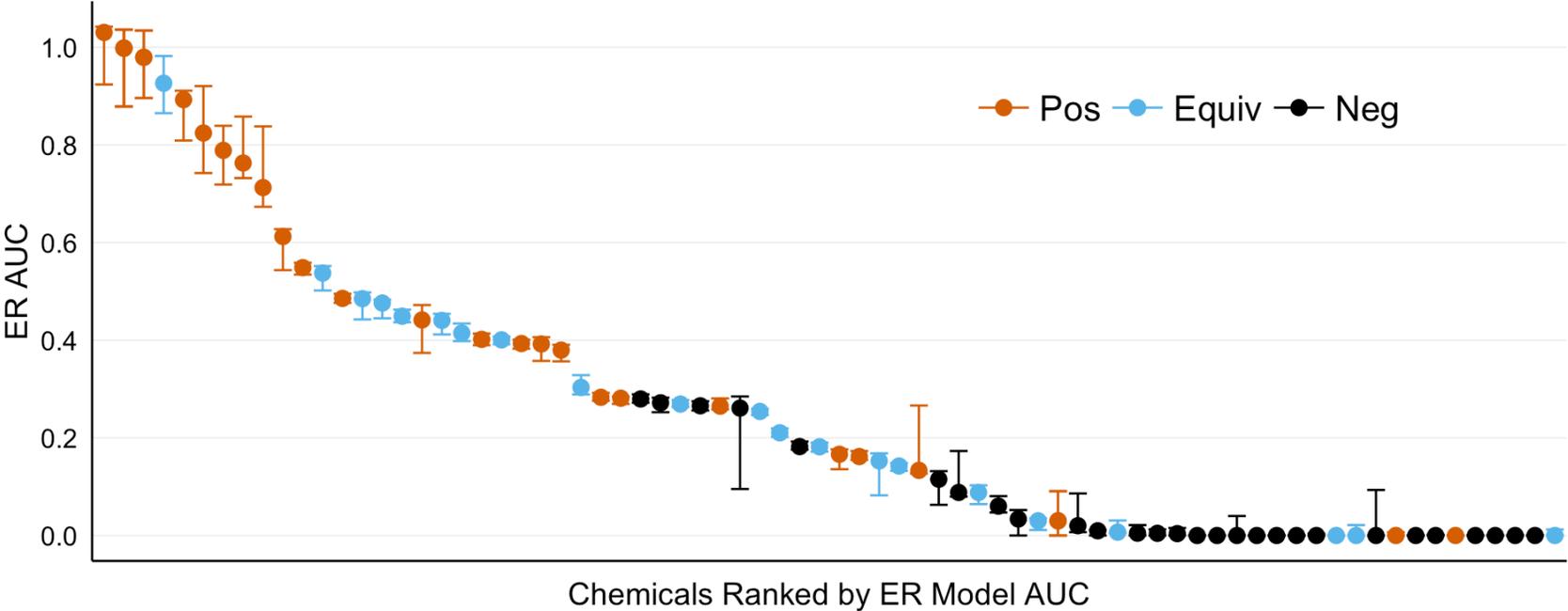
ER Model

Activity cutoff = 0.1 AUC, what is the uncertainty around this number?



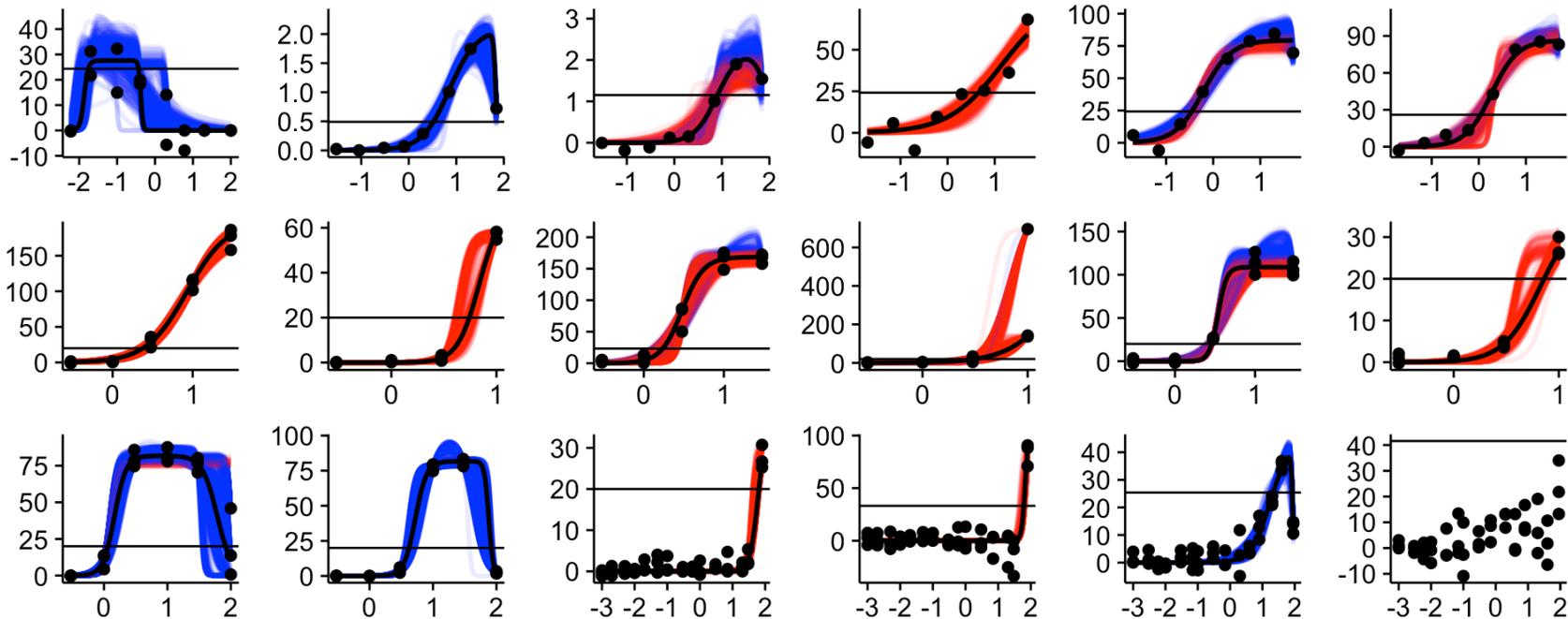
ER Model

Activity cutoff = 0.1 AUC, what is the uncertainty around this number?



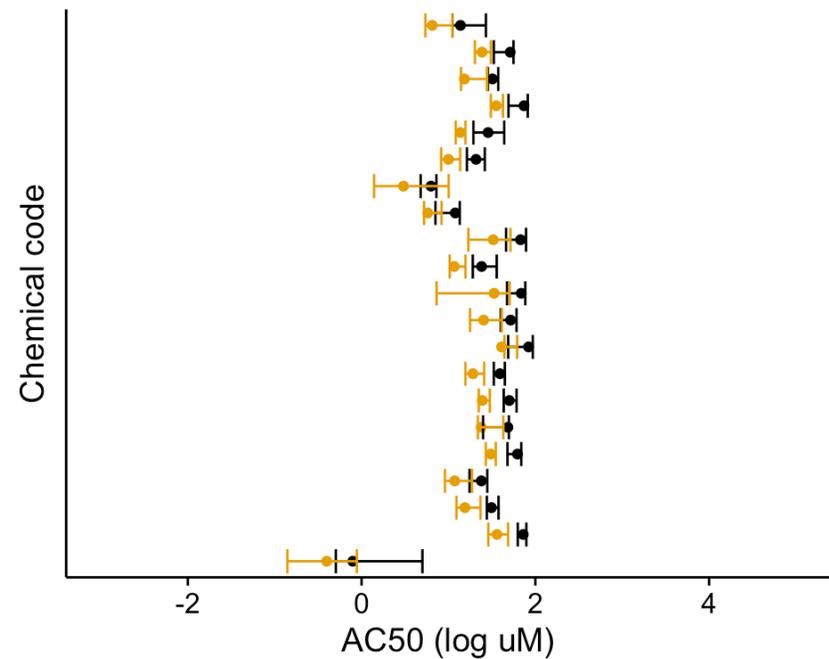
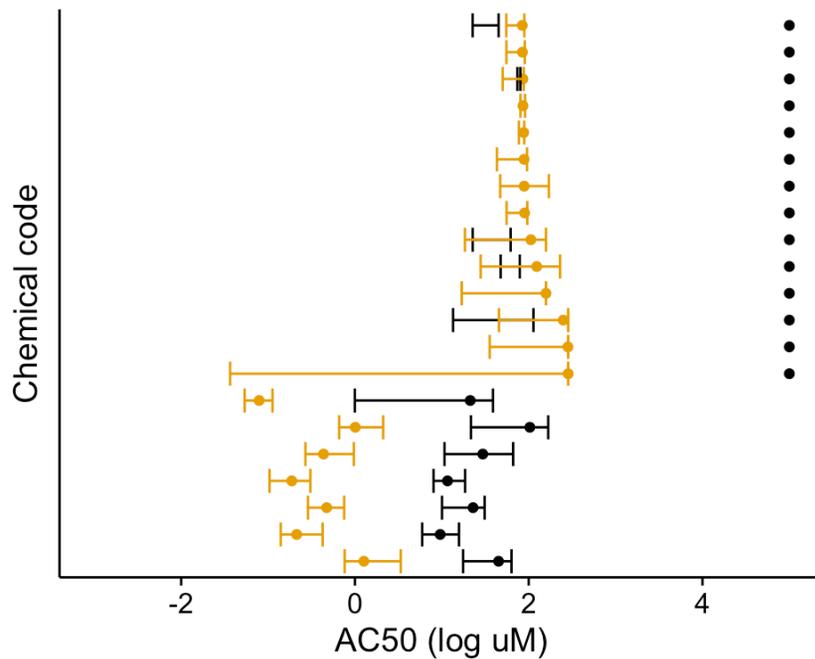
ER Model

Understand sources of false positives and negatives



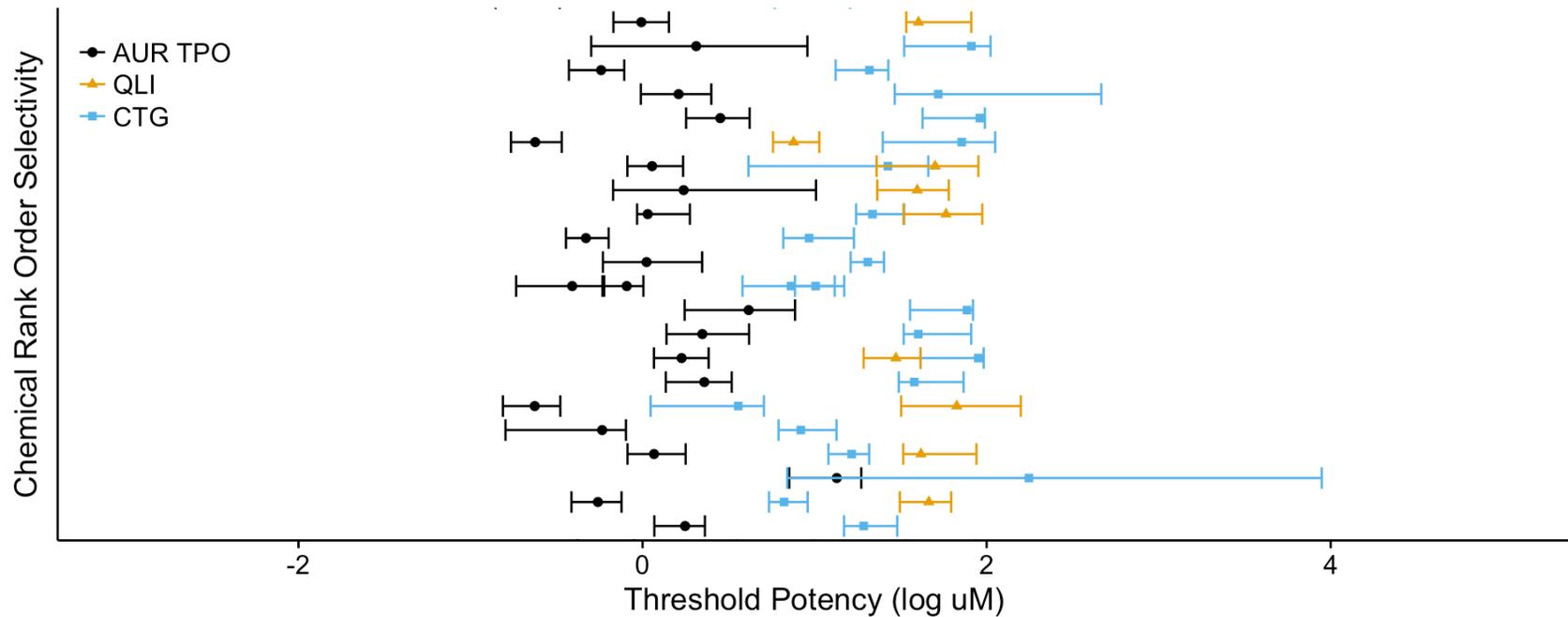
Androgen Receptor Antagonism Potency Shift

Shift from high (black) to low (orange) agonist conc statistically significant?



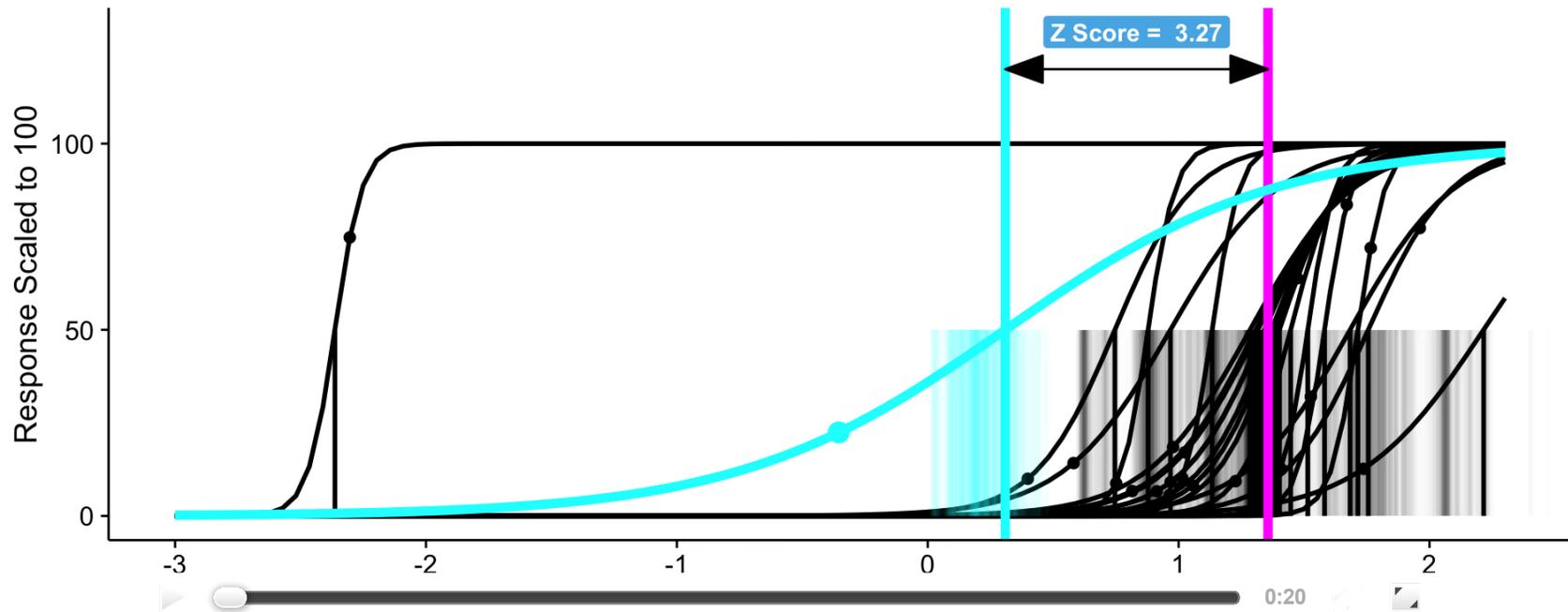
Thyroid Peroxidase

Are thyroid (black circles) potencies separate from orthogonal assays?



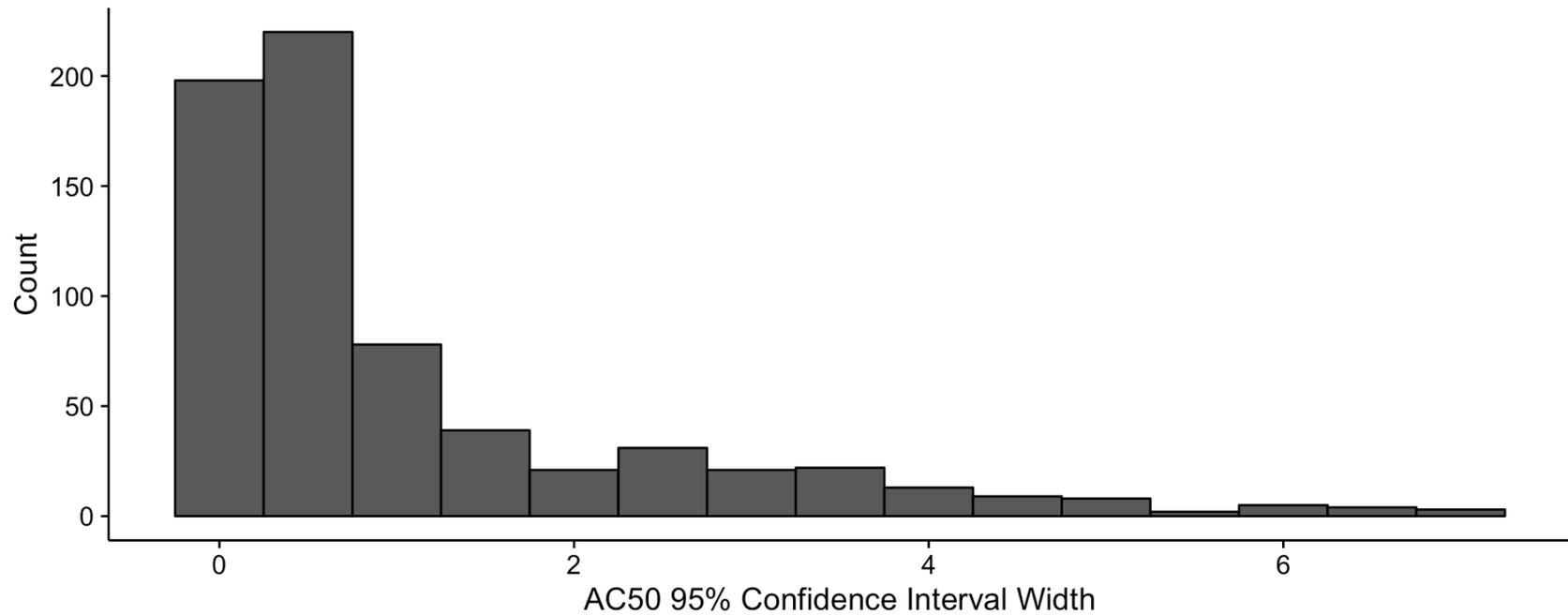
Cytotoxicity

Z score with uncertainty from all components



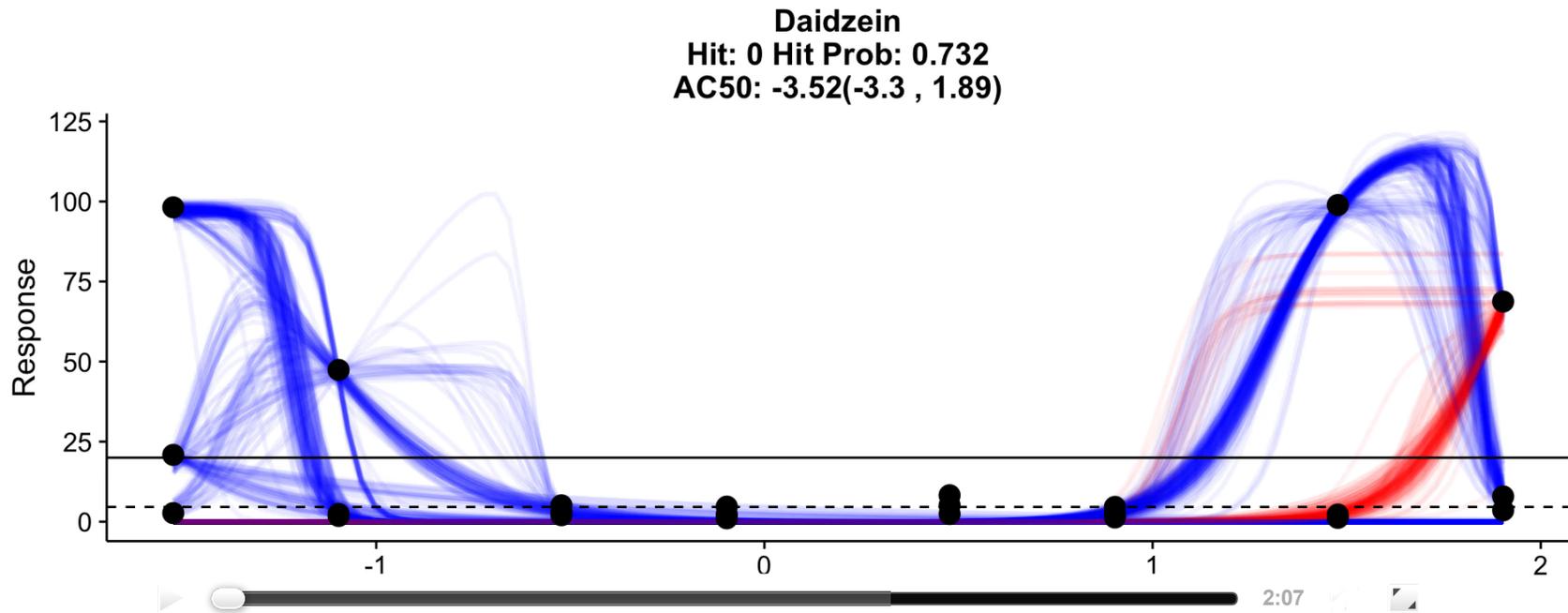
Zebrafish

Over 100 chemicals with potency uncertainty $> 2 \log \mu\text{M}$



Zebrafish

Bootstrap analysis flags chemicals with wide potency uncertainty



Motivating Questions

- What is the impact of uncertainty in outputs from high throughput screening (HTS)?
 - Potentially large shifts in model parameters, model selection, and even activity
- How can we quantify uncertainty?
 - Bootstrap resampling
- How can we propagate the uncertainty through models and analysis built on HTS results?
 - Apply model and analysis to each bootstrap sample, then aggregate the results

Conclusions

- Confidence intervals and distributions now calculated for model parameters
- Provide model selection and hit call probabilities
- Establish statistical basis for cutoffs and comparisons
- Detect and understand false positives and negatives
- Flag samples for manual inspection and retesting
- Method is applicable for diverse assays

Thanks!