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

Eric D Watt[1,2] and Richard S Judson[1]

[1]Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee
[2]National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina

Link to CV | Eric Dean Watt, Ph.D. | watt.eric@epa.gov | 919-541-7523

www.epa.gov

Motivating Questions

- What is the impact of uncertainty in outputs from 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 *in vitro* curves
- Many chemicals (> 8,000 unique)
- Many assays (> 800)

Broad assay coverage
- Numerous assay sources (> 10)
- Many biological pathways (> 400)
- Cell free and in cell
- Genes and cell lines from many species including human, rat, mouse, and fish
- Diverse detection methods including fluorescence, colorimetric, radioactive, electronic sensing, and RNA transcription

Broad chemical coverage
- Pesticides, food additives, green alternatives, endocrine reference compounds, water contaminants, fragrances, etc.
- Extensive QC
- Unified chemical library for all assays

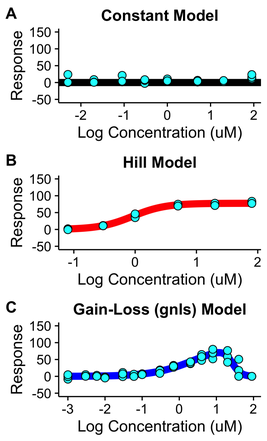

Fig 1 ToxCast models. A) Constant (cnst), B) Hill, and C) Gain-Loss (gnls) models.

ToxCast Pipeline offers consistent analysis
- Multiple models fit to data to determine efficacy (top) and potency (AC50) (Fig 1)
- Model selection based on AIC from model fits, hit call based on efficacy relative to cutoff for winning model

Method: Bootstrap Resampling

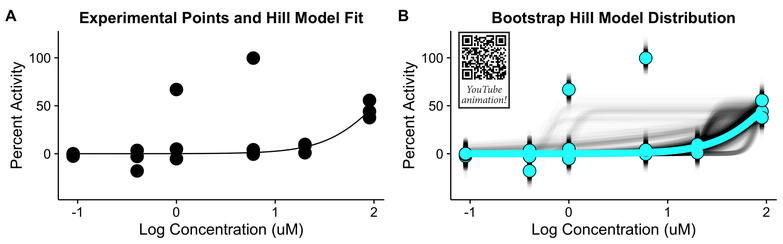

Fig 2 Bootstrap. A) Experimental response values (circles) and hill model fit. B) Uncertainty in response values and fitted model using 1000 bootstrap resamples. Experimental (cyan) and bootstrap resampled (black) response values (circles) and hill model fits (lines).

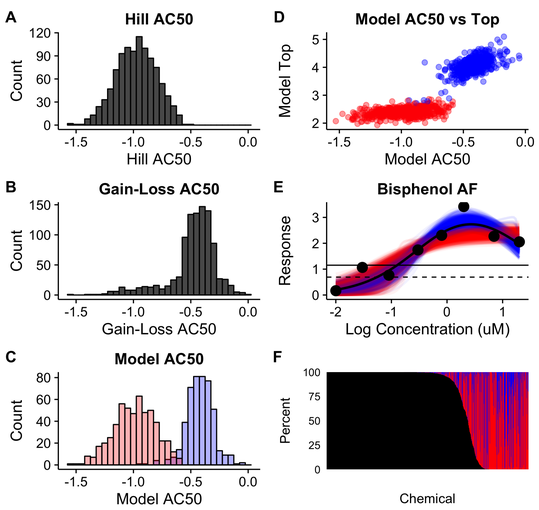

Fig 3 Bisphenol AF ATG_ERa_TRANS_up bootstrap results. The bootstrap results for Bisphenol AF in the ATG ERa TRANS assay are explored. A) Hill model potency (AC50). B) Gnls model potency. C) Winning model potency colored by winning model (hill red, gnls blue). Because each bootstrap sample can have a different winning model, the winning model AC50 distribution can be more variable than those from the individual models. D) Correlation between winning model top and AC50 (hill red, gnls blue). E) Black points and curve are experimental values and fit. Dashed line is 3x baseline MAD. Solid line is assay activity cutoff. Colored curves indicate 1000 bootstrap fits (red hill, blue gnls). The hill model selected 534 and the gnls 466 out of 1000. F) Hit percentage and model selection for all chemicals in the ATG ERa TRANS assay. Black bars indicate chemical not a hit, red is hill model active, blue is gnls active.

Estrogen Receptor Model

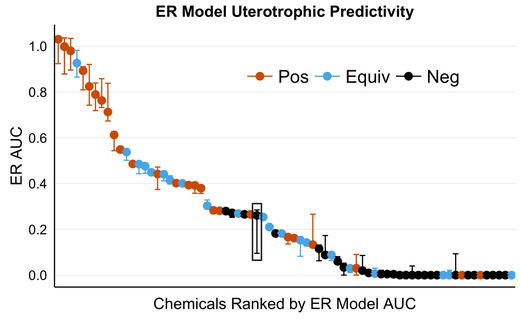

Fig 4 ER model AUC. Chemicals with both an ER model prediction and *in vivo* guideline-like study are plotted, ranked by AUC. Color indicates positive (red), equivocal (blue) or negative (black) activity in *in vivo* assays. Boxed value is explored in Fig 5.

- Recently approved for the Endocrine Disruptor Screening Program (EDSP) Tier 1 battery[1]
- Linear pathway model
- 18 ToxCast assays, 5 sources
- AUC > 0.1 predictive of activity
- Propagate bootstrap to get confidence intervals around AUC values
- **Bootstrap softens cutoff and explains some false positives and negatives**

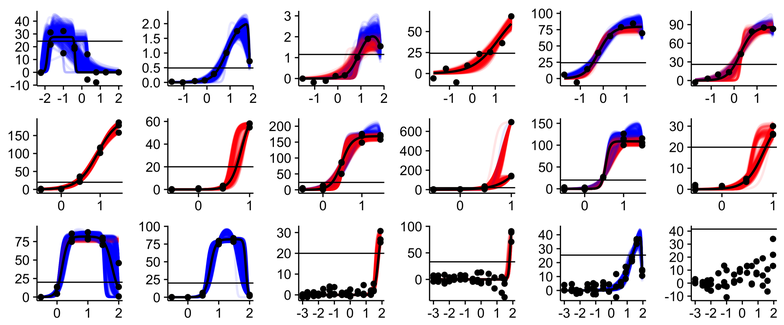

Fig 5 Exploration of source of uncertainty. Results from 18 ER assays with experimental points (circles), fit model (black), cutoff (horizontal line), and bootstrapped curves (hill red, gnls blue) for Nordihydroguaiaretic acid (boxed in Fig 4). ACEA assay (top left panel) is a hit in 652/1000 bootstrap samples. Inactive samples in this assay lower the ER AUC value.

Androgen Antagonism

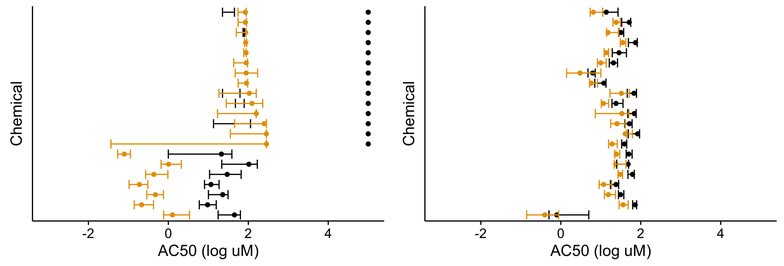

Fig 6 Androgen antagonist activity shift. The androgen antagonist assay was initially run with a concentration of agonist above a biologically relevant amount. A second experiment was performed at a lower, biologically relevant agonist concentration. Chemicals acting as true antagonists are expected to see a potency shift. Using bootstrap confidence intervals, we determine which chemicals have a significant potency shift and are therefore likely to be antagonists. The left panel indicates a region of large shift while the right panel includes chemicals with log shifts on the order of 0.3.

Thyroid Peroxidase Inhibition

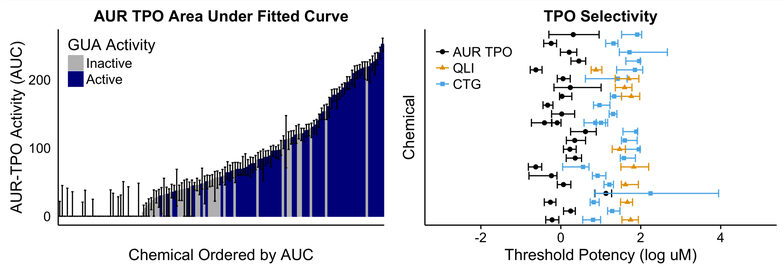

Fig 7 TPO assay comparison. The AUR-TPO assay[3] is compared to the previous GUA-TPO assay. AUR activity is represented as the area under the curve of the fitted model, while historic GUA results are indicated by color. Error bars are the 95% CI in the AUR AUC calculation.

Fig 8 TPO selectivity. To put the AUR-TPO hits into context, two orthogonal assays were run. QLI is a orthogonal enzyme inhibition assay and CTG tests for cytotoxicity. The 95% CIs help to confirm potency shifts between TPO and the other two assays.

Cytotoxicity

- Cytotoxicity, cell stress, and other non-specific responses can interfere with assay results
- 35 'burst' assays to measure cytotoxicity
- Z score reflects potency shift between assay and onset of cytotoxicity
- Z score < 3 indicates assay activity confounded by cytotoxicity
- **Propagate bootstrap through Z score calculation to determine significance of potency shift and soften cutoff**

$$zscore = -\frac{modl\_ga - cyto\_pt}{global\_mad}$$

$$cyto\_pt = \begin{cases} median(modl\_ga[\text{burst assays}]) & \text{if } \geq 2 \text{ burst assays are active} \\ 3 & \text{if } < 2 \text{ burst assays are active} \end{cases}$$

$$global\_mad = median(mad[(nhit > 1)])$$

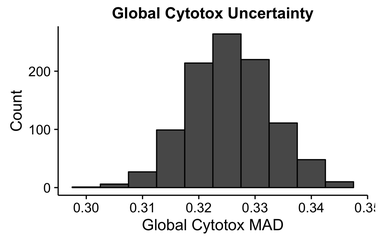

Fig 9 Global cytotoxicity MAD. The global MAD is used as a measure of variability in the cytotox potency and is calculated by aggregating all chemicals over all cytotoxicity assays. Using the bootstrap results we determine a distribution of the global cytotox MAD for use in calculating the Z score distribution (Fig 10).

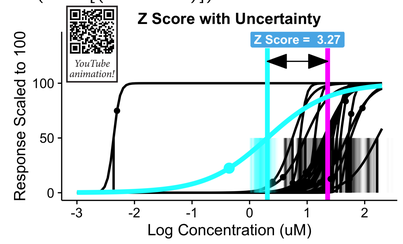

Fig 10 Z score calculation. The z score is calculated as the potency (vertical lines) difference between the assay of interest (cyan) and the median potency in cytotox assays (magenta), divided by the global MAD (Fig 9). We incorporate uncertainty quantification in each parameter. The top of all curves is scaled to 100, with circles representing the cutoff scaled into the same units. Bars of half height represent bootstrapped AC50 values for each assay.

Zebrafish Development

- Chemical treatment of Zebrafish (*Danio rerio*) embryos from 0.25 to 5 days post fertilization[2]
- Response values: 100 (dead), 50 (unhatched), and 0-50 degree of malformation in 20 discrete steps
- Used to screen for developmental toxicants
- Random deaths and unhatched embryos
- cause outlier spikes well outside of normal range
- Outliers can cause difficulty when fitting the ToxCast models, representing a 'worst case' for model fitting and uncertainty quantification
- **Confidence interval width can detect chemicals for examination and retesting**

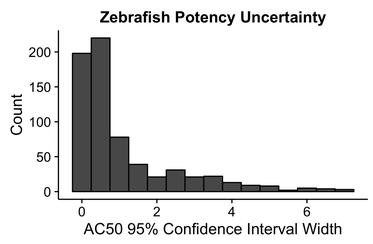

Fig 11 Zebrafish uncertainty. The 95% confidence intervals (CIs) of the AC50 for all chemicals in the assay are indicated. While many chemicals have a 95% width of < 1.5 log uM, there are more than 100 with large CIs.

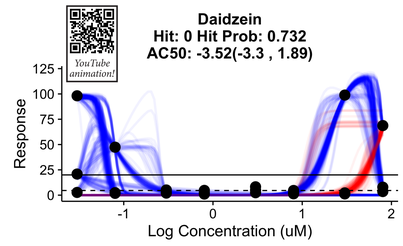

Fig 12 Exploration of high CI width. One of the zebrafish chemicals with wide confidence intervals, Daidzein, is plotted above. The dashed horizontal line indicates 3x baseline MAD, the solid horizontal line is the assay cutoff, circles indicate the experimental concentration-response values, and the curves indicate the bootstrapped hill (red) and gnls (blue) winning models. While this chemical is not found active in the pipeline, bootstrapping categorizes this curve as active 73.2% of the time. The winning (inactive) model for the experimental points found the AC50 at -3.52, while the bootstrapped values for the active samples ranged between -3.3 and 1.89.

References

1. Federal Register. Use of High Throughput Assays and Computational Tools; Endocrine Disruptor Screening Program; Notice of Availability and Opportunity for Comment. 2015 Jun pp. 35350–35355. Report No.: 80 FR 35350.
2. Padilla S, Corum D, Padnos B, Hunter DL, Beam A, Houck KA, et al. Zebrafish developmental screening of the ToxCast™ Phase I chemical library. Reproductive Toxicology. 2012 Apr;33(2):174–87.
3. Paul Friedman K, Watt ED, Hornung MW, Hedge JM, Judson RS, Crofton KM, et al. Tiered High-Throughput Screening Approach to Identify Thyroperoxidase Inhibitors Within the ToxCast Phase I and II Chemical Libraries. Toxicological Sciences. 2016 May;151(1):160–80.