Supplemental Appendix with Supplemental Figures

Paul-friedman.katie@epa.gov

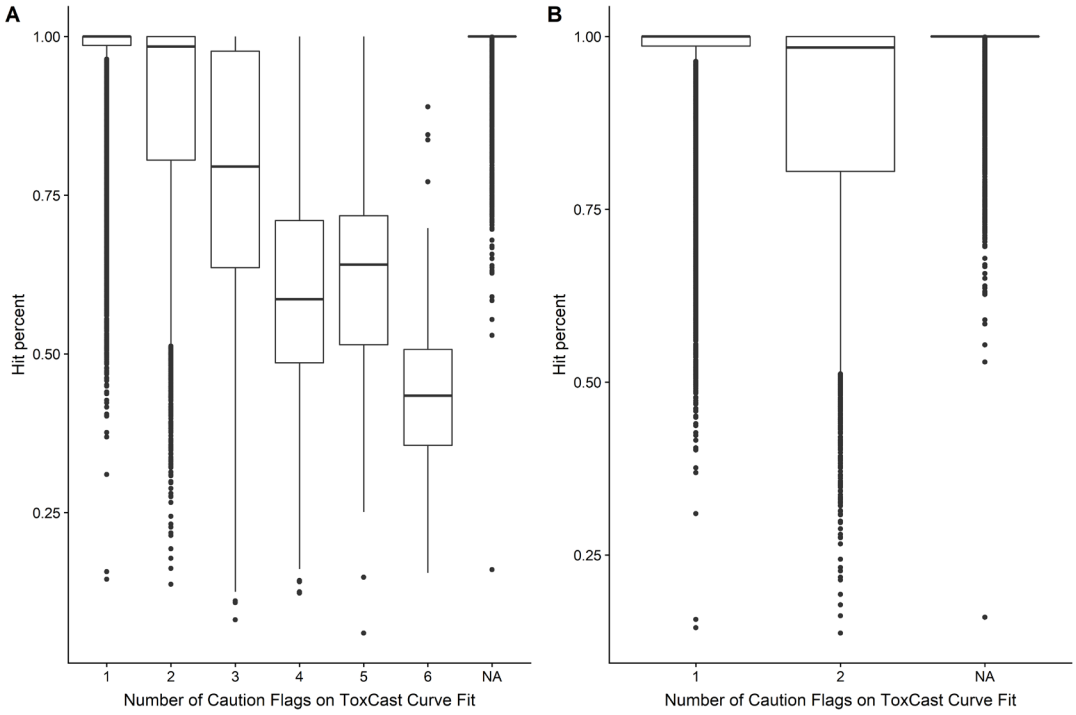
This Supplemental Appendix is intended as additional discussion of some of the details in using *in vitro* bioactivity data, producing administered equivalent doses using high-throughput toxicokinetic (HTTK) information, as well as providing all supplemental figures to support the main text.

1. ToxCast data filtering

The ToxCast data (invitrodbv3.0) (USEPA, 2018) were filtered for this application to remove curve fits that may be less reproducible or reliable for quantitative use (i.e., the modeled 50% activity concentration is less likely to be an informative value). The filtering procedure is discussed in the Methods of the main text. Briefly, information on the uncertainty and caution information associated with curve-fitting were used to filter curve-fits.

First, curves with 3 caution flags (at multi-concentration level 5 in invitrodb) are more frequently associated with with lower curve-fit reproducibility, represented by hit-percent (or the % of 1000 bootstrap resampled curve-fits that would be positive). Thus, the first filtering rule removes perhaps the least reproducible fits by removing curve-fits with greater than or equal to 3 and a hit-percent of 50% or less. In Supplemental Figure 1 below, it is apparent in panel A for the unfiltered dataset that there is a large difference in median hit-percent between 2 and 3 caution flags. In panel B, the hit-percent v. number of flags is illustrated for the filtered dataset.

Supplemental Figure 1: Hit-percent v. Number of flags, for (A) unfiltered ToxCast data and (B) filtered ToxCast data.



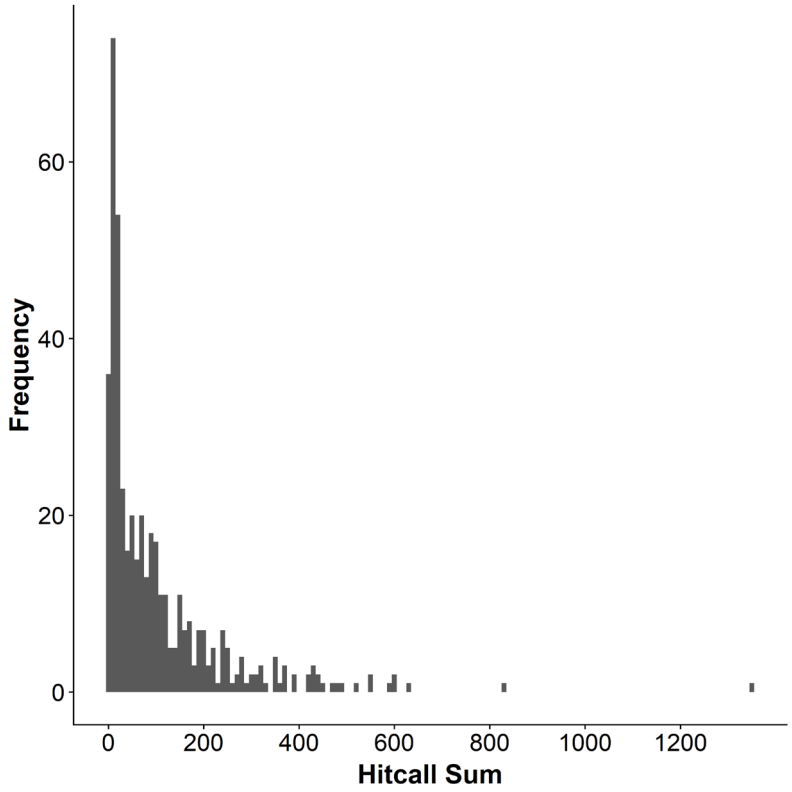
Second, the ToxCast data were also filtered to eliminate curve fits associated with two of the "fit categories," or categories defined in invitrodb that describe the shape of the curvefit and the location of the AC50 (*tcpl* R package vignette https://cran.r-project.org/web/packages/tcpl/vignettes/Data_processing.html) (Filer *et al.*, 2017). Fit categories 36 and 45 were dropped, which are Hill and gain-loss fits with the model top (top of the curve fit, i.e. maximal fitted response) that is less than or equal to 1.2 times the threshold cut-off for a positive response and an AC50 less than or equal to the concentration range screened. These fit categories are thought to be less quantitatively informative because the efficacy is borderline, and the estimated AC50 is in a concentration range where there are no actual data, and thus the true slope of the curve is not defined.

These two sets of filtering steps resulted in a 13.6% reduction in the ToxCast dataset for the 448 substances in this case study, from 54048 curves to 46735 curves.

2. Filtered ToxCast Data

Following filtering, the median hitcall sum was 56, and the range was 0 to 1351. Only The hitcall sum, i.e. the number of positive assay endpoints for a given substance, is shown in a frequency distribution (Supplemental Figure 2).

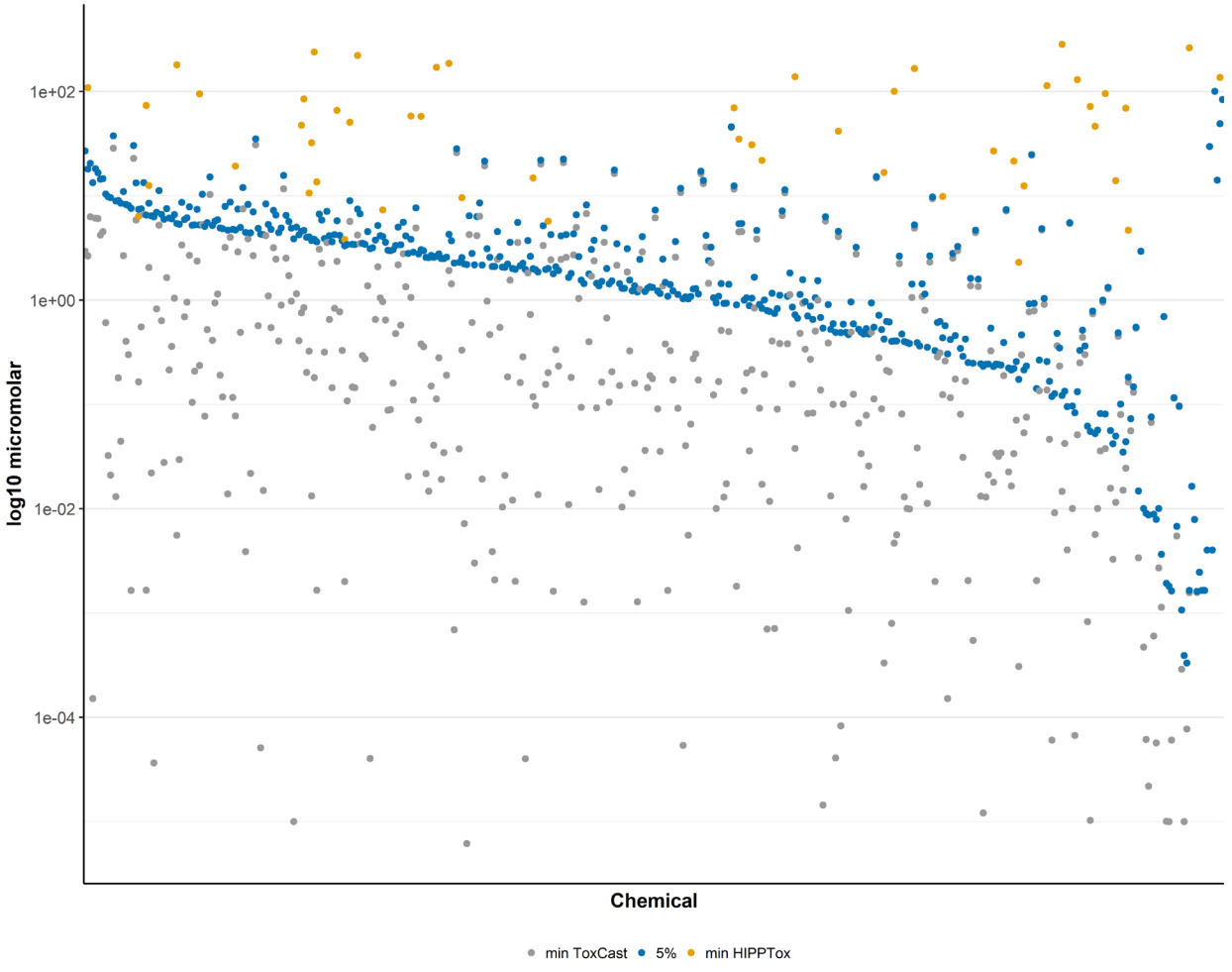
Supplemental Figure 2: Frequency distribution of the hitcall sum for the filtered ToxCast data.



3. Selection of *In Vitro* data for Calculation of Administered Equivalent Dose

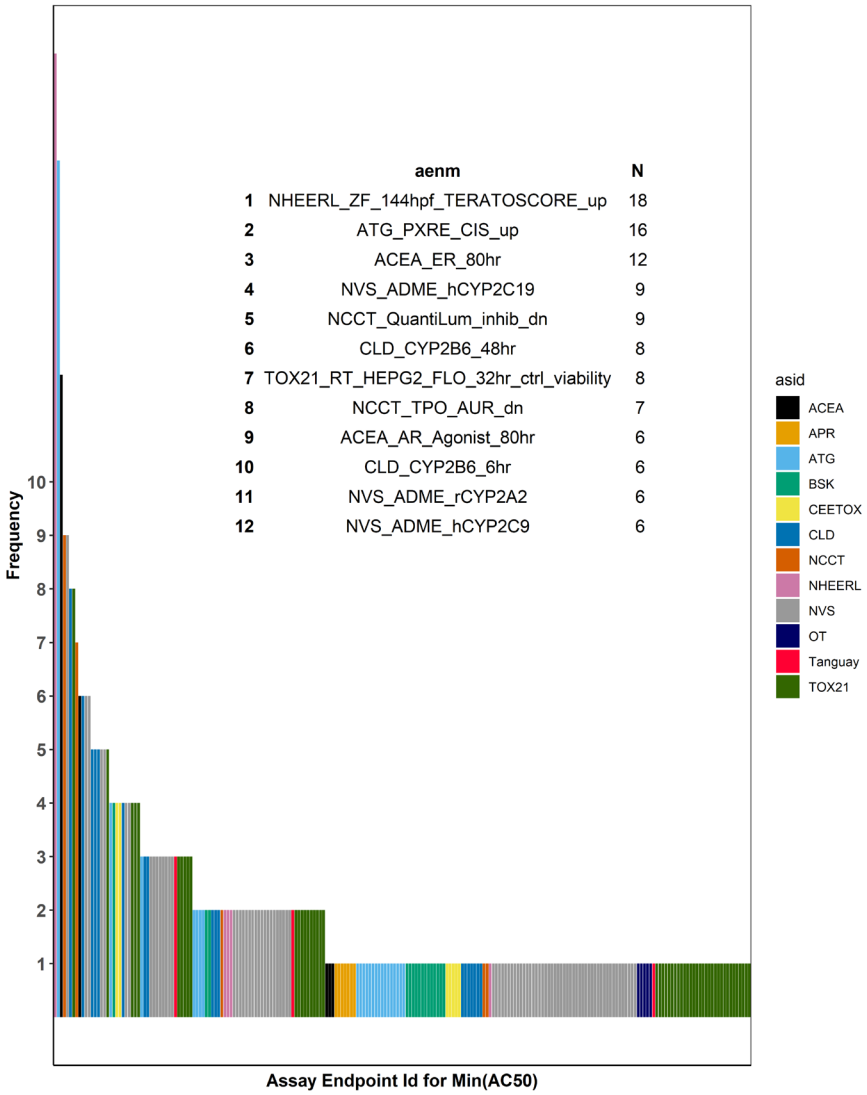
A high-throughput approach to interpretation of the ToxCast data for the purposes of this case study identifies a threshold concentration at which we expect bioactivity. We chose to use the minimum of either the 5th percentile from the distribution of AC50 values or the minimum effect concentration value (EC10) from the HIPPTox models available for each substance. The logic associated with the decision was multi-factorial - 1) the 5th percentile of AC50 values was intended to protect against extreme AC50 values that may have resulted from spurious fits or perhaps minute changes in bioactivity that were not broadly representative of a threshold of bioactivity; and 2) when HIPPTox EC10 values were available, the minimum of the HIPPTox value or the 5th percentile AC50 from ToxCast was used because the HIPPTox value was thought to be more indicative of adversity rather than a conservative threshold for bioactivity; however, as ToxCast does not have complete biological coverage of all tissues, the HIPPTox model could provide information on lung, kidney, and liver that might not be indicated by ToxCast.

Supplemental Figure 3: Comparison of ToxCast minimum and 5th percentile AC50 and HIPPTox EC10.



In using the 5th percentile of the ToxCast AC50 distribution by substance, the value is no longer distinctly connected to a specific assay endpoint in invitrodb. However, it was important to examine the potential for disproportionate influence of any one ToxCast assay endpoint on the minimum AC50 value, as this minimum value greatly influences in the 5th percentile of the AC50 distribution. In Supplemental Figure 4, the assay endpoints that drove the minimum AC50 for more than 5 substances are displayed. There did not appear to be a distinct influence of assay source (which is highly connected to assay platform) on the frequency. The assay endpoint with the highest frequency in Supplemental Figure 4, is zebrafish teratogenicity scoring. This assay endpoint may be particularly susceptible to certain substances such as organophosphate insecticides or highly cytotoxic/systemically toxic substances that were present in the case study set.

Supplemental Figure 4: Frequency of ToxCast assay driving minimum AC50



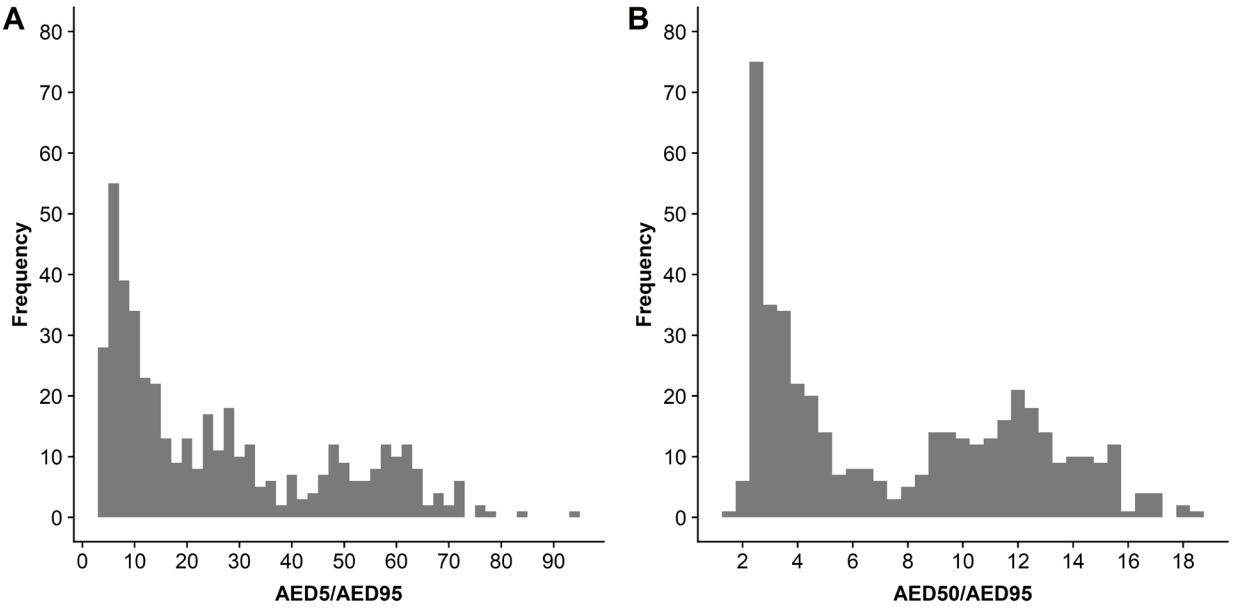
4. Assumptions in the application of high-throughput kinetics

Inter-individual variability in toxicokinetics may be accounted for using the httk R package (Pearce *et al.*, 2017). In this case study, we considered the selection of the 5th, 50th, and 95th percentiles of the distribution of administered dose equivalent (AED) values to represent the full range of dose values an adult population might receive in order to achieve the steady state concentration (micromolar) used as input. The AED produced from the 95th percentile is lower and thus more conservative.

In Supplemental Figure 5, the details on the difference between the 5th, 50th, and 95th percentile for inter-individual variability in AED computation are demonstrated. Selection of the AED50 could mean selection of an AED that is 1.7- to 19-times larger than the AED95; for many chemicals, the AED50 is 2-5 times larger.

Supplemental Figure 5: Width of the interval produced by the httk Monte Carlo simulation.

(A) The fold-differences between the 5th and 95th percentiles of the Monte Carlo simulation (AED5/AED95). (B) The fold-differences between the 50th and 95th percentiles of the Monte Carlo simulation (AED50/AED95). The frequency distribution of AED50/AED95 for the 448 chemicals in this case study demonstrated that for many chemicals AED50 was 2-5 times larger than AED95, though in some cases the differences was much greater.

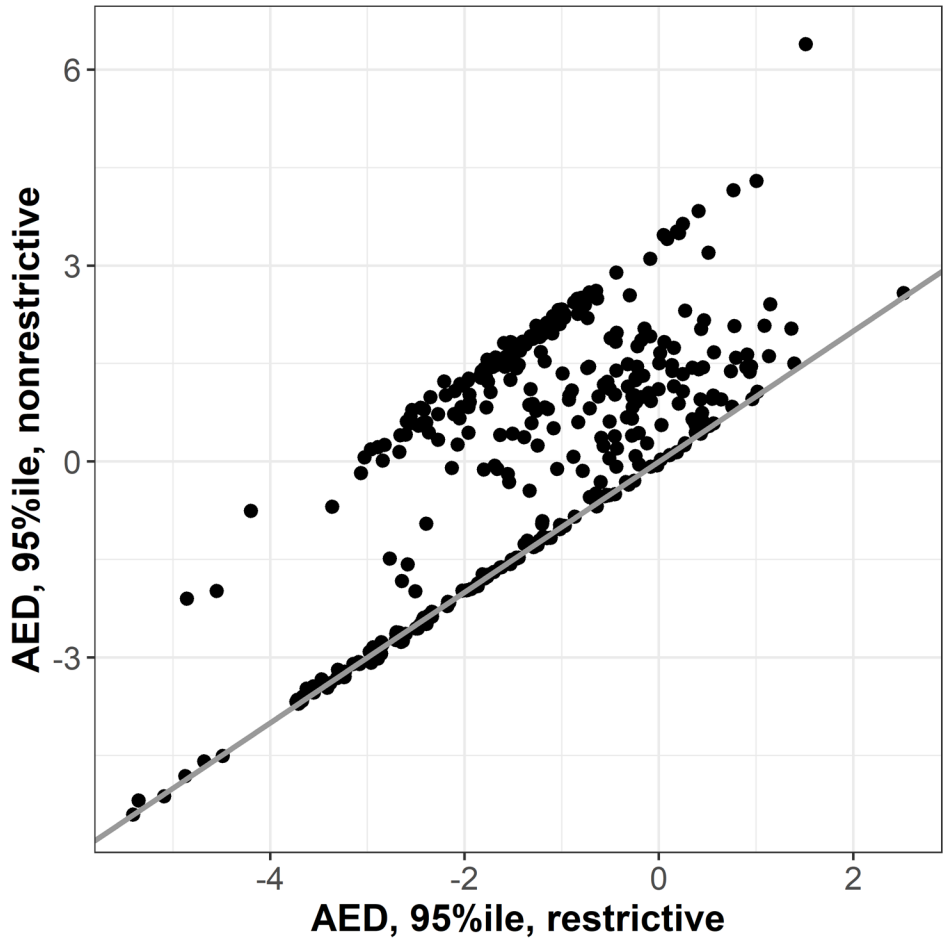


An additional decision that affects the estimated AED is assumption of restrictive clearance. Restrictive clearance indicates that chemical bound to protein is relatively unavailable for hepatic metabolism or renal excretion (whereas non-restrictive clearance assumes that chemical bound to protein rapidly

disassociates from that protein for metabolism and excretion). The degree to which a protein bound chemical is available for metabolism and excretion is likely chemical specific and a continuous function (i.e., not binary). Currently, there is no way to predict or measure this property for a chemical. Restrictive clearance was used in this case study as a conservative assumption. In Supplemental Figure 6, the estimated AED values based on non-restrictive and restrictive clearance are shown. Because the amount of chemical bound to protein can vary from 0-100%, the AEDs produced using a non-restrictive clearance assumption may be as much as two to three orders of magnitude greater than those produced using a restrictive clearance assumption (on a log10-mg/kg/day scale and based on current measurement ability). The amount of difference observed depends on how much of the chemical is thought to be protein-bound; the more highly protein-bound the chemical, the greater the shift observed. For the chemicals in the case study not thought to be highly protein-bound, the AEDs produced using non-restrictive and restrictive clearance assumptions are generally equivalent.

Supplemental Figure 6: Administered equivalent doses (AEDs) estimated using non-restrictive and restrictive clearance assumptions.

The AED values were estimated for the 448 case study chemicals using non-restrictive and restrictive clearance with no other changes to the other httk parameters.

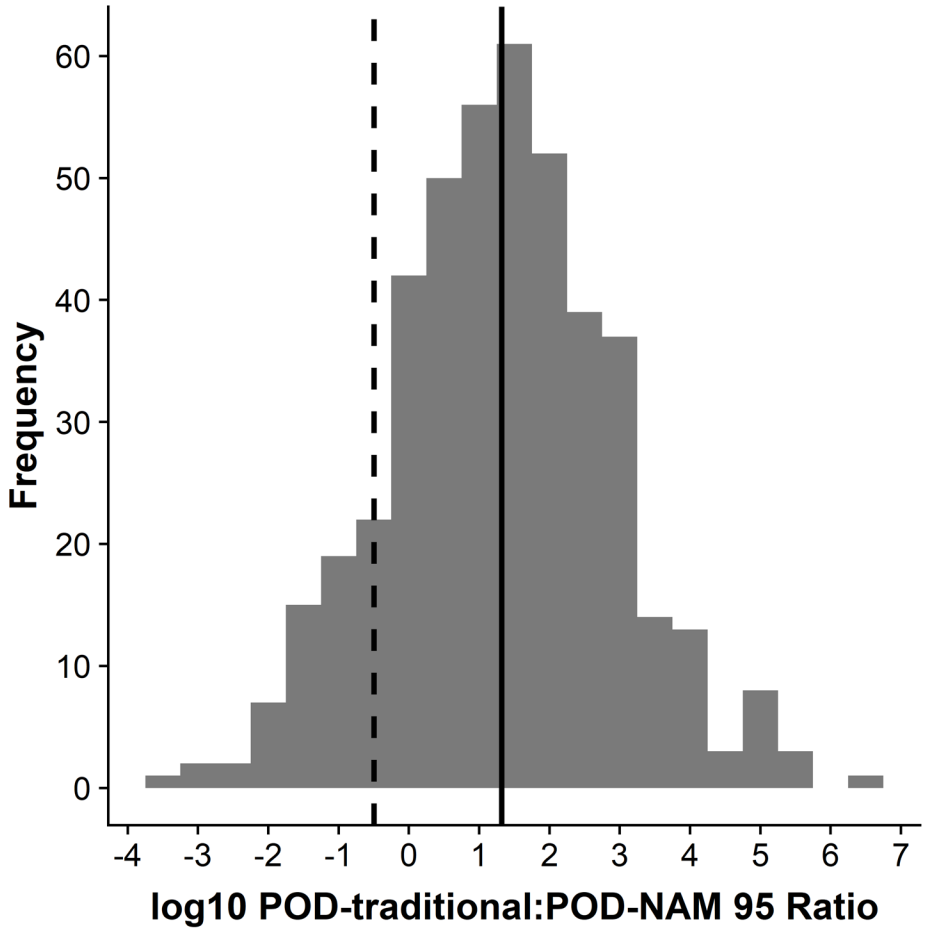


5. Considering inter-species differences in dose

Inter-species differences were considered in the context of dosimetry by performing allometric scaling as described in the main text.

Supplemental Figure 7. Distribution of the POD ratio for allometrically scaled $POD_{traditional}$ data

The human equivalent dose (HED) $\log_{10}POD_{NAM,95}$: $POD_{traditional}$ ratio is illustrated for 447 of the 448 case study chemicals. The dashed line indicates a ratio = 0 and the solid line indicates the median ratio equal to 1.3.



6. References for the Supplemental Appendix

Filer, D. L., Kothiya, P., Setzer, R. W., Judson, R. S., and Martin, M. T. (2017). tcpl: the ToxCast pipeline for high-throughput screening data. *Bioinformatics* **33**(4), 618-620.

Pearce, R. G., Setzer, R. W., Strope, C. L., Wambaugh, J. F., and Sipes, N. S. (2017). httk: R Package for High-Throughput Toxicokinetics. *J Stat Softw* **79**(4), 1-26.

USEPA (2018). ToxCast & Tox21 Summary Files from invitrodb_v3. . In (https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data.