ORD MANUSCRIPT COVER SHEET

**Title:** Collaborative Evaluation of *In silico* Predictions for High Throughput Toxicokinetics

**Authors:**

|  |  |  |
| --- | --- | --- |
| John Wambaugh | 1 EPA/ORD/CCTE | [Wambaugh.john@epa.gov](mailto:Wambaugh.john@epa.gov) |
| Nisha Sipes | 1 | [Sipes.nisha@epa.gov](mailto:Sipes.nisha@epa.gov) |
| Kamel Mansouri | 2 Innotiv  3 NICEATM | kamel.mansouri@nih.gov |
| Jon Arnot | 4 Arnot Research & Consulting | jon@arnotresearch.com |
| Trevor Brown | 4 | trevor.n.brown@gmail.com |
| Chris Cook | 1 | cookchristopher732@gmail.com |
| Daniel Dawson | 1 | ddawson@integral-corp.com |
| Sarah Davidson Fritz | 1 | DavidsonFritz.Sarah@epa.gov |
| John DiBella | 5 Simulations Plus | john.dibella@simulations-plus.com |
| Stephen Ferguson | 6 NIEHS | stephen.ferguson@nih.gov |
| Rocky Goldsmith | 1 | rgoldsmith@congruencetx.com |
| Chris Grulke | 1 | [cmgrulke@gmail.com](mailto:cmgrulke@gmail.com) |
| Richard Judson | 1 | Judson.Richard@epa.gov |
| Michael Lawless | 5 | mlawless@simulations-plus.com |
| Gilberto Padilla Mercado | 1 | PadillaMercado.Gilberto@epa.gov |
| Grace Patlewicz | 1 | Patlewicz.Grace@epa.gov |
| Ester Papa | 7 University of Insubria, Varese | ester.papa@uninsubria.it |
| Prachi Pradeep | 8 German Federal Institute for Risk Assessment (BfR) | prachipradeep.15@gmail.com |
| Alessandro Sangion | 2 | alessandro.sangion@mail.utoronto.ca |
| Risa Sayre | 1 | sayre.risa@epa.gov |
| Russell Thomas | 1 | Thomas.Russell@epa.gov |
| Rogelio Tornero-Velez | 1 | tornero-velez.rogelio@epa.gov |
| Barbara Wetmore | 1 | wetmore.barbara@epa.gov |
| Michael Devito | 1 | Devito.Michael@epa.gov |

**Internal Peer Reviewers:** maybe Todd Martin, Dan Chang, or Marina Evans?

**Target Journal:** Not sure, possibly: Computational Toxicology, Journal of Pharmacokinetics and Pharmacodynamics, Drug Metabolism and Disposition, Drug Discovery Today, Environmental Science & Technology (probably too many figures), maybe the special issue from QSAR2021 if Grace can get us in late

**Four bullet point summary:**

* Toxicokinetic (TK) information, such as elimination half-life (thalf, plotted below), is critical for understanding chemical risk
* Here we collected *in silico* (quantitative structure-property relationship, QSPR) predictions of key *in vitro* determinants of TK from several different models
* The models were evaluated for ability to reproduce *in vitro* and *In vivo* measurements of TK
* Overall, high throughput physiologically-based TK (PBTK) model performed similarly when using TK QSPRs as when the actual *in vitro* measured data were used

**One sentence description:** This collaborative trial demonstrates that multiple QSPRs exist that make reasonably accurate chemical structure-based predictions for *in vitro* TK parameters

**Chemicals Involved:** 87 ToxCast chemicals with diverse uses for which high-throughput toxicokinetic (HTTK) data are available

**Collaborative Evaluation of *In Silico* Predictions for High Throughput Toxicokinetics**

John Wambaugh1,\*, Nisha Sipes1, Jon Arnot2, Trevor Brown2, Christopher Cook1,3, Daniel Dawson1, Sarah Davidson1, , John DiBella4, Stephen Ferguson5, Rocky Goldsmith1, Chris Grulke1, Richard Judson1, Michael Lawless4, Kamel Mansouri6, Grace Patlewicz1, Ester Papa7, Prachi Pradeep1,8, Alessandro Sangion2, Risa Sayre1, Russell Thomas1, Rogelio Tornero-Velez1, Barbara Wetmore1, and Michael Devito1

1. Center for Computational Toxicology and Exposure, Office of Research and Development, United States Environmental Protection Agency, Research Triangle Park, North Carolina 28311
2. ARC Arnot Research and Consulting Inc.
3. Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee 38331
4. Simulations Plus, Inc.,
5. Division of the National Toxicology Program (NTP), National Institute of Environmental Health Sciences, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 28317
6. NTP Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 28317
7. Department of Theoretical and Applied Sciences, University of Insubria, Varese
8. German Federal Institute for Risk Assessment (BfR)

\*Corresponding Author: wambaugh.john@epa.gov

# Abstract

To assess public health risks posed by chemicals we need to understand chemical absorption, distribution, metabolism, and elimination by the body (that is, toxicokinetics or TK). Unfortunately, chemical-specific TK data are often unavailable*. In silico* predictions along with high throughput TK (HTTK) methods have the potential to address this gap. This collaborative trial uses a database of *in vivo* measured toxicokinetic data to evaluate *in silico* approaches for HTTK. Six different sets of quantitative structure-property-relationship (QSPR) tools for predicting TK were evaluated. QSPR predictions were evaluated by using the predicted parameter values within a high throughput physiologically based TK (PBTK) model to predict *in vivo* measured plasma concentrations for 92 chemicals, mostly in rats. Root mean squared log10 error (RMSLE) was calculated for toxicologically-relevant dosimetry statistics AUC (time-integrated area under the curve) and Cmax (peak concentration. For Cmax optimal fits to the *in vivo* data indicated a best case RMSLE of 0.5. PBTK using chemical-specific *in vitro* data predicted AUC with RMSLE 0.74, while QSPRs ranged in accuracy from 0.74 to 0.84. For AUC *in vivo* parameters gave RMSLE 0.5. Chemical-specific *in vitro* data had RMSLE 1.3 and QSPRs ranged from 1.3-1.4. *In vivo* parameters for the full concentration time-course data had RMSLE of 0.62. Predictions based on *in vitro* HTTK had RMSLE of 1.18. RMSLE for individual QSPRs ranged from 1.19 to 1.3. QSPRs predict plasma binding well (RMSLE 0.03 – 0.07) but have difficulty predicting metabolic clearance (RMSLE 0.37 – 1.28). However, a consensus prediction using the maximum clearance predicted across all QSPRs predicted AUC with RMSLE 1.03 and full time-course with RMSLE 1.09. The consensus predictions outperformed the *in vitro* measured data for the evaluation chemicals. For novel compounds a consensus QSPR approach may yield a reasonable prediction. This evaluation characterizes the accuracy of HTTK approaches for new chemicals based on both *in vitro* measurement and structure-based *in silico* predictions.

# Introduction

Toxicokinetics (TK) describes the absorption, distribution, metabolism, and excretion (ADME) of a chemical compound in the body as a function of time [1]. Since TK allows the prediction of internal tissue concentrations as a function of chemical exposure it provides critical information for assessing risk posed by a chemical to public health [2]. TK allows interpretation of biomonitoring data [3], dosimetric anchoring of animal toxicity studies [4], and quantitative *in vitro*-*in vivo* extrapolation (or IVIVE) from high throughput bioactivity studies [5]. These *in vitro* bioactivity data are available for thousands of chemicals (for example, the ToxCast [6] and Tox21 [7] screening programs). IVIVE of bioactivity data relies upon TK to relate the concentrations found to be active *in vitro* with doses that could cause these concentrations in tissues [8-12]. Unfortunately, chemical-specific information on TK are often unavailable for thousands of chemicals in commerce and the environment [13]. New approach methodologies (NAMs) are being developed throughput the chemical risk assessment process [14] including new methods for assessing TK [15,16].

For non-therapeutic compounds TK data were traditionally developed using animal studies; for reasons of both ethics and resources these studies are no longer desirable nor practical for the thousands of remaining chemicals [17]. An alternative technology developed by the pharmaceutical industry relies on characterizing certain aspects of TK *in vitro* and then extrapolating to *in vivo* conditions to estimate TK parameters such as AUC, often within a three-fold error [18]. For the past decade, government chemical regulatory agencies working with their collaborators and contractors have been collecting chemical-specific *in vitro* data that allow prediction of TK [8-12,19]. These data currently span approximately one thousand chemicals used for industry, pesticides, diet, therapy, and consumer products. However, thousands more remain uncharacterized.

The U.S. National Academies of Science, Engineering, and Mathematics have recognized that *in vitro* TK data “enabled first-tier risk-based rankings of chemicals on the basis of margins of exposure—the ratio of exposures that cause effects (or bioactivity) to measured or estimated human exposures” [20]. To address the remaining chemicals for which even *in vitro* TK data are unavailable, multiple organizations have developed *in silico* quantitative structure-property relationship (QSPR) models for predicting these values [21-25]. Meanwhile, EPA and other organizations are continuing to accumulate chemical-specific TK data, both *in vivo* – that is, curation of data from the scientific literature as well as a targeted animal studies only when needed [26,27] – as well as *in vitro* [11,19,12].

Here we examine the TK predictions from six different modeling teams. Four modeling teams produced QSPR models for two key toxicokinetic parameters that can be measured *in vitro*: intrinsic hepatic clearance (Clint measured with hepatocyte incubations, [28]) and fraction unbound in plasma (fup, typically measured via rapid equilibrium dialysis [29]). The models are initially evaluated for their ability to reproduce *in vitro* measured values. However, we focus on analysis of the predictions for chemical concentration as a function of time (CvT) that can be made when a generic physiologically-based TK (PBTK) model [30] is used with the predictions from each QSPR. For this study *in vivo* plasma and blood concentration vs. time data for rat and human were initially available for 101 chemicals from the CvTdb [27]. Models are evaluated for ability to reproduce the full CvT curve as well as summary statistics (such as peak plasma concentration) and parameters (such as chemical half-life). Two additional models for chemical half-life are also evaluated. The six modeling teams were provided with chemical identities and some physico-chemical descriptors but were not provided with the actual *in vivo* evaluation data.

# Methods

As summarized in Table 1, three levels of evaluation were made. First (Level 1), QSPR predictions of parameters were compared against *in vitro*-measured values for each chemical where data were available. Then (Level 2) the ability of a PBTK model to predict chemical concentration vs. time behavior was evaluated against *in vivo* measurements for the full time-course observed for each combination of QSPR and chemical. Finally (Level 3) the ability of the QSPRs and PBTK model to predict summary statistics (for example Cmax, AUC, half-life) was evaluated. All analyses were performed in the free, open-source statistical analysis language R [31] v4.4.0. Analyses were performed on a Dell Precision 7560 laptop personal computer. Scripts to perform all analyses are available in RMarkdown [32] format as supplemental material.

## HTTK

R package “httk” [30] v2.3.2 was used for this analysis. “httk” can parameterize a physiologically based toxicokinetic (PBTK) model based on chemical-specific values for fraction unbound in plasma (fup, unitless) and intrinsic hepatic clearance (Clint, µL/min/106 hepatocytes). These experimentally measured values are collected from peer-reviewed literature and provided by “httk”. All the QSPRs analyzed here were trained to human data.

The generic PBTK model used here (model “pbtk”) consists of well-mixed compartments for the gut, kidney, liver, and rest of body. The model is parameterized for a chemical using fup and Clint plus equilibrium tissue:plasma partition coefficients predicted with a modified Schmitt’s method [33,34]. The model simulates both oral and intravenous dosing. Oral dosing is subject to first-pass metabolism by the liver before the compound distributes systemically. Among other species, the model includes physiological information for parameterizing both humans and rats (primarily from [35,33,36]. The model was simulated using command “httk::solve\_pbtk()” with option default.to.human=TRUE – that is, since no rat-specific values are predicted by the models under evaluation, comparisons to data from rats were done using rat physiology but human in vitro TK parameters.

The steady-state ratio of the concentration of chemical in blood and plasma (Rb:p) is an important parameter in the PBTK model and is used for converting between plasma predictions and observations in whole blood. The function httk::get\_rblood2plasma() either retrieves measured values of Rb:p from the literature or predicts the ratio by predicting the red blood cell:plasma equilibrium partition coefficient and then using the hematocrit fraction for the relevant species.

R is an interpreted language (primarily operated by a user from the command-line although scripts are common). The user can alter the values within a table which stores the fup and Clint values for all chemicals (httk::chem.phys\_and\_invitro.data). After alteration, the httk function will proceed using the new values in the table. The HTTK data can be returned to their default values via the command “httk::reset\_httk()”. By default, no QSPR values are included in the table. However, predictions can be loaded with the commands “httk::load\_sipes2017(overwrite=TRUE)”, “httk::load\_pradeep2020(overwrite=TRUE)”, or “httk::load\_dawson2021(overwrite=TRUE)” [23-25]. The argument “overwrite=TRUE” is needed so that *in vitro* measured data are overwritten whenever a chemical-specific prediction is available. To facilitate comparisons, a custom function “clear\_httk()” is included in the supplemental material which deletes all Human Clint and fup values. The OPERA predictions are available as Supplemental Table 1.

## QSPR Models

The QSPR models evaluated are summarized in Table 2. Four different modeling teams previously [23-25] produced quantitative structure-property relationship (QSPR) models for two key toxicokinetic parameters that can be measured *in vitro*: intrinsic hepatic clearance (Clint measured with hepatocyte incubations) and fraction unbound in plasma (fup). The consensus QSPR was constructed for fup by taking the inverse logit of the mean of the logit-transformed fup from each QSPR. For Clint the maximum value predicted by any QSPR was used. Two additional models for chemical half-life were also evaluated. Individual model predictions are available in Supplemental Table 1.

Due to the potential presence of the model evaluation data (that is, measured *in vitro* parameters) in the training sets for some or all the models, we attempted to remove predictions that seemed more like a direct retrieval of the chemical-specific values from a training set. Model predictions were removed for a particular model-chemical combination if both the predictions for Clint and fup were within 1% absolute fold error of the measured values, the predictions were omitted from the evaluations. Omitted predictions are listed in Supplemental Table 6. When a QSPR model prediction was missing for a particular chemical the mean prediction of the other models was used for evaluation purposes.

QSAR training set summary data

* number chemicals, max, min, average values
* will use the averages for missing values

## *In vitro* Data

For 61 of the study chemicals *in vitro* measurements were also available for comparison. These data are collected by the R package “httk” but are drawn from the peer reviewed scientific literature (including [9,5,12]). It is likely that these data were in the training set of some of the QSPR models. The *in vitro* measured values are available in Supplemental Table 1.

## *In vivo* Data

EPA has developed a public database of concentration vs. time data for building, calibrating, and evaluating TK models [27]. Curation and development of the database are ongoing, but when this study began there were 101 chemicals with either rat or human *in vivo* blood or plasma concentration vs. time data. The *in vivo* measured concentration vs. time values are available as Supplemental Table 2.

## Compartmental Model Fits

For chemicals with CvT data, parameters were estimated for empirical one- and two-compartment toxicokinetic models using R package “invivoPKfit” (<https://github.com/USEPA/CompTox-ExpoCast-invivoPKfit>). Three models were considered: one- and two-compartment empirical TK models and a flat “null hypothesis” where there was no systematic change in concentration vs. time. The model with the lowest Akaike Information Criterion (AIC) value – indicating model parsimony – was selected [37]. Data sets where the flat model was selected were omitted from further analysis. The empirical model fit was then used as a “best case” prediction scenario for comparison with PBTK parameterize by *in vitro* or QSPR predictions.

For both models a half-life was calculated from the terminal elimination rate as thalf = ln(2)/kelim. For the two-compartment model the volume of distribution at steady-state was used as Vd. For both models, clearance was calculated as Cltot = Vd \* kelim. The estimated TK parameters for both models are provided as Supplemental Table 3.

## Evaluation Metrics

Multiple statistics were used to evaluate predictions (*pred*) relative to observed values (*obs*) as appropriate. Relative Predictive Error (RPE) was calculated as RPE =, where if the observed value was 0 then the error was set to zero. We note that if the predicted value is 0 then RPE = -1. Absolute Average Fold Error (AAFE) was calculated as AAFE = , where if pred=0 and obs=0 we assigned = 0. Root Mean Squared Log Error (RMSLE) was calculated as RMSLE = .

Kolmogorov-Smirnov tests were performed using R function *ks.test*.

# Results

## Evaluation Chemicals and Predictions

There were 101 chemicals present in the CvTdb (Sayre, 2020) as of September 2019 that had plasma concentration data following either oral or intravenous doses given to rats or humans. These chemicals included: 57 from the Toxic Substances Control Act (TSCA) active inventory [38], 20 pharmaceuticals, 24 pesticides, 99 that are found in consumer products, 7 per- and poly-fluorinated substances (PFAS) [39], and 64 that are part of the ToxCast screening program. Note that a chemical could be in more than one of these categories.

Of the 101 chemicals, 2 were best described by the “flat” model indicating that the data sets were too noisy to estimate empirical TK parameters. Given that these chemicals were poorly described by basic TK models – potentially indicating problems with chemical analysis sensitivity -- these chemicals were withheld from subsequent analysis. The data for each remaining chemical could be described using either a one- or two-compartment empirical pharmacokinetic models. Separate parameter estimates were made for each combination of compound and species for which there were data. For each remaining chemical the better of the one or two compartment models was used on the basis of model parsimony.

For each QSPR we removed predictions where the predicted values for a given chemical were within 1% for both fup and Clint assuming these values reflected the chemical data present in the training set and the model method allowing for recall of the measurements. This affected 21 chemicals as predicted by OPERA.

We summarize the chemical-specific properties and predictions in Figure 1. In Figure 1 similar chemicals (rows) and properties/predictions (columns) are clustered together based on Euclidean distance. All properties/predictions were centered (mean changed to zero) and scaled (divided by standard deviation) such that the value reflects the number of standard deviations from the mean. Interestingly, the first division between clusters in Figure 1 places all the Clint measurements and predictions on the one side and all the fup measurements and predictions on the other. The physico-chemical properties are divided between those two clusters, with Octanol:Water (partition coefficient, PC), Octanol:Air PC, Molecular Weight, Boiling Point, and Melting Point all clustering with Clint. Water solubility, vapor pressure, and the Henry’s law constant all clustered with fup.

## Heterogeneous Chemical Counts

section on imputations -- difference between regulatory data gaps and statistical perspective (subset of chemicals with overlap between all models more rigorous statistically, superset of chemicals with predictions from most models is more rigorous chemically)

* table of coverage percentages by model/chemical maybe
* provide other figures (maybe for supplemental)
* could use consensu rmse
* how we handle missing observations
* could do both separate rmsle and consensus
* could fill in with average prediction for model
* note how many times consensus used per model
* overlap only evaluation

## Level 1 Analysis

Our first level of evaluation directly compared the predictions of QSPR's with the *in vitro* measured values. We do not have predictions for all the chemicals across all QSPRs – when a prediction is missing, we assign the average prediction across those models that successfully made a prediction for that chemical. if model A does not a prediction for chemical x, we assign mean(model B,C,D) to for chemical x for model A. There were 63 chemicals with *in vitro* measured parameters. We evaluate model performance for Clint in Figure 2. The QSPR with the lowest mean RMSLE (root mean squared log10 error) is OPERA, which uses a nearest neighbors method to retrieve values from similar chemicals in the training set. On the basis of RMSLE the other three models – Dawson (2021), Pradeep (2020), and Simulations Plus ADMET predictor – perform similarly to each other despite the fact that the Dawson model is categorical (that is, predicting only three values: very slow, slow, and fast) while the other models are continuous. Figure 3 shows that all four models perform very well for predicting fup. Most models are within ten-fold and predictions are highly correlated with observed. Mean RMSLE ranged from 0.03 to 0.06 – all very small values. For chemicals where there is not a tight correlation it looks like most models over predict the fup value. OPERA has both over and under predictions in these “outliers". OPERA has the smallest variability in the relative error, ADMET has the most variability, and Dawson and Pradeep are rather on par with one another (between OPERA and SPlus).

We examined the distributions of fold errors between the predictions and the measured data using a Kolmogorov-Smirnov test. For both Clint and fup the only QSPR that differed from the others was OPERA, which had a significant (p-value < 0.05) difference between the distribution of predicted values and the distributions of the three other QSPRs. In Table 3 we summarize the fold errors for the four QSPRs. The median fold error for all for QSPRs is close to zero except for ADMET Predictor, which has a slight bias toward overestimating clearance (log10 fold error of 0.105 in Table 3 corresponding to median predictions being 27% higher than measured). We note that the median fold errors for OPERA are extremely low (effectively zero for fup), even with the obvious training set chemicals (Supplemental Table 6) removed.

Also shown in Table 3 is that the predictions of Clint range from 100x lower than experimental values (log10 folder error of -2) for all four QSPRs to 100x higher (log10 fold error of 2) for three of the four QSPRs. For fup the predictions range from 30x too low to more than a million times overestimated (log10 fold error of 6).

## Level 2 Analysis

For the second level of analysis, we compared predictions based on the QSPR predicted values with actual tissue concentration vs. time data. We evaluated based upon 83 chemicals as described above. All predicted values are used with the HTTK PBTK model to make predictions. We bracket model performance in three ways: First, we use the HTTK PBTK model with the actual in vitro measured values. In subsequent figures this is labeled as “HTTK-InVitro”. Next, for a best-case performance, we use empirical (one or two compartment toxicokinetic model) fits to the *in vivo* data , labelled as “FitsToData”. The one and two compartment models are simpler than the high throughput PBTK model used for all other scenarios, but because they have been optimized to the *in vivo* evaluation data they are expected to outperform the other approaches here. Finally, for worst case performance we use y-randomization so that the measured values for the 83 chemicals are scrambled and assigned to the incorrect chemicals, labelled “HTTK-YRandom”.

Figure 4 shows the full predicted time-courses for each set of model predictions as well as the actual *in vitro* data and empirical model fits. The PBTK model generally underpredicted the *in vivo* data, both when used with *in vitro*-measured parameters (mean RPE -0.71) and with the QSPR-predict values (MRPE for -0.66 for ADMET, -0.74 for Dawson, -0.80 for Pradeep, and -0.67 for OPERA).

In Panel A of Figure 5 we examine the distribution of RMSLE on a per chemical basis, aggregating over all doses, routes, and time points for each chemical. All observed time points are valued equally, without consideration of phase (absorption/distribution/metabolism) and measurement accuracy. First, we observe that the empirical fits to the data yield an average RMSLE of 0.54, corresponding on the arithmetic scale to predictions being on average within a factor of 3.5 times the observed values. The difference between the predictions made with *in vitro* measurements and y-randomized *in vitro* measurements is relatively small, with predictions based upon *in vitro* measurements being within 18x of observations on average and y-randomized values being with 22x of observations on average. All four QSPRs perform similarly to the *in vitro* data, with ADMET predictor being the best with an average prediction within 14x the observations.

Noting that the combination of ADMET predicted values and the generic PBTK model from “httk” produces values more accurate than either the actual measured *in vitro* values or the one compartment model fits, we looked to understand how the ADMET values differed from the rest. As discussed above, the clearance values predicted by ADMET were on average 27% higher than the experimentally measured values. In Panels B and C of Figure 5 we respectively break each time course into two phases, early (that is, time points less than the mean time for a given study) and late (the remaining points). We presume that the early phase will typically include the absorption and distribution phases, included the peak concentration. We note that the same absorption rate is used for all QSPR models (and indeed for all chemicals) based on the mean absorption observed in Wambaugh et al. [26]. All the models perform better for earlier time points than for later time points, and the four QSPRs and *in vitro* measured data all perform effectively the same for later time points. The differences in performance of ADMET predictions seems to be in the early time points.

Early time points are dominated by the ability to correctly predict peak plasma concentration (Cmax). In Figure 6 we examine each methods accuracy in predicting Cmax as determined from the CvT data. As summarized in Table 40 optimal performance is given by the empirical fits to the data, with a coefficient of variation (R2) of 0.94 and a RMSLE of 0.32. Worst case performance is given by the y-randomized measured data with a R2 of 0.42 and RMSLE of .99. The QSPRs perform roughly as well, if not clearly better (ADMET being best) than the *in vitro* measured data, with R2 ranging from 0.48 to 0.59 and RMSLE ranging from 0.83 to 0.93. The superior ability of ADMET to predict Cmax likely is correlated with its better performance at early times.

In Figure 7 we examine the ability to predict time-integrated area under the plasma concentration time course (Area Under the Curve or AUC). Again, the empirical fits give a clear best-case scenario, but here the y-randomization more clearly gives a worst-case scenario. The other various methods are all roughly comparable, although again ADMET outperforms even the in vitro measured values.

In Figure 8 we plot each chemical-specific RMSLE by prediction method with each method as a row and each chemical as a column. These data are also provided in Supplemental Table 7. Chemicals and methods have been clustered based upon Euclidean distance. We see that the largest RMSLE is for the Tamoxifen measured data, potentially influencing the performance of in vitro measured data relative to the predictions of the QSPRs.

## Level 3 Analysis

We then proceed on to the third level of evaluation, in which we use the QSPR predictions to predict toxicokinetic summary parameters – volume of distribution (Vd), half-life for elimination from the body (thalf), and whole-body clearance (Cltot) – and compare the predictions to the values estimated from the empirical fits to the CvT data. The values predicted for each method are provided in Supplemental Table 8.

In Figure 9 we examine predicted vs. observed thalf. As summarized in X none of the models are very successful – the highest coefficient of variation is 0.15 for IFS-QSAR, while QSARINS-Chem and HTTK with ADMET both had an R2 of 0.11. The RMSLE for all models, including HTTK with y-randomized data, was just above 1 (a factor of 10x).

The models were distinctly better than the y-randomization for predicting Vd. As shown in Figure 10 and summarized in X , the HTTK algorithm for predicting Vd [34] when used with y-randomized data had no skill. The models performed similarly to the measured *in vitro* data when used with the Vd algorithm – R2 ranged from 0.10 to 0.16 with the measured data being the worst. For all models the RMSLE again indicated a factor of 10x.

In Figure 11 we examine predictions for CLtot, which depends on both elimination rate (inverse of thalf) and Vd. As summarized in X the y-randomized predictions reassuringly have no skill at predicting *in vivo* clearance, while the combination of HTTK and ADMET predicted values had the most, with a R2 of 0.32 and a RMSLE indicating a factor of 17x. Both IFS-QSAR and QSARINS-Chem had comparable R2 of 0.25 and 0.2 (respectively) and an RMSLE indicating a factor of ~20x. The other QSPRs performed about as well as using the *in vitro* measured data.

# Discussion

TK information, especially elimination Vd, half-life (thalf), and whole-body clearance, is critical for understanding chemical risk. . The NAS has recognized [20] that high throughput (chemical-agnostic) TK models parameterized with chemical-specific in vitro data (that is, in vitro-in vivo extrapolation for toxicokinetics) are a powerful tool for interpreting high throughput screening data for chemical toxicity in terms of chemical risk [2] – *in vitro-in vivo* extrapolation for toxicity. Governments and industry are continuing to accumulate chemical-specific TK data including both *in vivo* concentration vs. time data in key tissues [27,26] and *in vitro* HTTK data [8,9,11,12,10,19]. However, several thousand chemicals remain in need of TK info; the QSPRs evaluated here provide options to fill this gap.

Given a sufficient number of observations for evaluation, the RMSLE can be interpreted as a coefficient of variation for normally distributed errors about the prediction. This means we have 95% confidence that the actual value will occur within +- 2 RMSLE of the prediction.

To assess public health risks posed by chemicals we need to understand chemical absorption, distribution, metabolism, and elimination by the body (that is, toxicokinetics or TK). Unfortunately, chemical-specific TK data are unavailable for thousands of chemicals in commerce and the environment*. In silico* predictions along with high throughput toxicokinetic (HTTK) methods have the potential to address this gap for chemical risk evaluators. This collaborative trial uses a database of *in vivo* measured toxicokinetic data to evaluate *in silico* approaches. In vivo dated were carefully reviewed to include only those data that could be well-described by empirical TK models. Six different sets of quantitative structure-property-relationship (QSPR) tools for predicting TK were evaluated. Four of the QSPR models made predictions for chemical-specific *in vitro* measurements (HTTK data). QSPRs varied in estimated domain of applicability and so no individual QSPR made predictions for all chemicals. An additional consensus prediction was constructed from the various QSPRs. QSPR predictions were evaluated by using the predicted parameter values within a physiologically based TK (PBTK) model to predict *in vivo* measured plasma concentrations for 92 chemicals, mostly in rats with limited human data. This analysis used the generic PBTK model in R package "httk", which is designed to use the *in vitro* parameters predicted by the QSPRs. Across the evaluation chemicals root mean squared log10 error (RMSLE) was calculated for toxicologically-relevant dosimetry statistics AUC (time-integrated area under the curve) and Cmax (peak concentration. For Cmax optimal fits to the *in vivo* data indicated a RMSLE of 0.5 while using *in vitro* values for random (incorrect) chemicals had an AUC RMSLE of 0.89. Using chemical-specific in vitro data predicted AUC with RMSLE 0.74, while QSPRs ranged in accuracy from 0.74 to 0.84. For Cmax in vivo fits gave RMSLE 0.5 and random chemicals gave 1.7. Chemical-specific in vitro data had RMSLE 1.3 and QSPRs ranged from 1.3-1.4. Predictions for the full concentration time-course were also evaluated. Fits to the in vivo data had RMSLE of 0.62, while predictions based on *in vitro* HTTK had RMSLE of 1.18 without using any *in vivo* data. RMSLE for individual QSPRs ranged from 1.19 to 1.3. Any one QSPR performed slightly better than using *in vitro* values for random chemicals (RMSLE 1.35). Most QSPRs predict plasma binding well (RMSLE 0.03 – 0.07) but have difficulty predicting metabolic clearance (RMSLE 0.37 – 1.28). However, a consensus prediction using the maximum clearance predicted across all QSPRs predicted AUC with RMSLE 1.03 had RMSLE 1.09 for the full time course – this is better than using the *in vitro* measured data for the evaluation chemicals. For novel compounds a consensus QSPR approach may yield a reasonable result. This evaluation characterizes the accuracy of HTTK approaches for new chemicals based on both *in vitro* measurement and structure-based *in silico* predictions.

Since the *in vitro* data for HTTK are limited, here we have conducted a collaborative trial of four QSPRs for *in vitro* TK parameters and two additional predictors of *in vivo* TK half-life.

* of how well TK QSPRs would do

The focus of the analysis was using *in vivo* data from rat or human TK studies collected by CvTdb {sayre, 2020}. We report the HTTK RMSLE for the full concentration time course (Cvt) data (“level 2” analysis) for 83 chemicals.

Model performance was closer to y-randomized predictions than to empirical fits to the data



For the TK summary statistics (level 3) we observed Vd had RMSLE 0.89 with R2 0.04. For total clearance CLtot the RMSLE 1.57 and R2 0.03.

* + - For Cmax the RMSLE is 0.84 and the R2 is 0.57. For AUC the RMSLE is 1.11 and the R2 is 0.5.
* Here we have found the HTTK PBTK model performed similarly when using TK QSPRs for Clint and fup as when the actual *in vitro* measured data were used

Comparison with previous evaluation:

Wambaugh et al. [26] found evaluated HTTK-based IVIVE for summary TK endpoints (what we call “level 3” here) for TK using just over forty chemicals. The mean squared error (MSE) observed for Vd was 4.4. For CLtot the MSE was 2.4 for pharma, and 2.93 for non-pharma. However, the fraction of variance explained (R2) was 0.19 for pharma, 0.5 for non-pharma. For Cmax Wambaugh et al. [26] found MSE 5, R2 0.48. Finally, for AUC the MSE was 3.8, with R2 0.62. Why the drop-off?

note that CLtot and AUCinfinity are related

y randomization needs to be clarified

* In Figure 1 describe distribution of chemicals quantitatively – means, upper, lower, distribution of high/medium/low Fup and Clint – this is why y-scrambling doesn't show a large range
* mention phys-chem span (log p, etc)
* Point out that 80-100 chemicals is not "Big data" – need to expand CvT database
* it's a constrained random set of chemicals -- reflecting the correlation and distribuirton (frequency) in this set
* like the cvt data , we can only evaluate and model things that vary across our dataset
* uniform draw would better show true random (no) information case -- call this "Randomized" or something
* We had a very similar result in the Dawson et al. paper for Y-randomization.
* the in vitro probably has a similar spread of values for technical reasons
* What is breakdown rat vs. human in evaluation set?
* We do not expect Fabs to correlate between rat and human – good reason not to worry about Caco2 here
* in HTPBTK here by – using human clint and fup
* In some cases, QSPRs outperformed *in vitro* measurements, indicating value to intra-chemical averaging of data
* more explanation of why in vitro data doesnt do better than the in silico

Limitations of model for in vitro intrinsic clearance

* admet is five cyps -- not an apples to apples comparison (figure 2)
* what are impacts of 3d effects in qsar? is that something we're missing?
* all the models are 2d right now because we don't have data to train qsar models to predict 3d differences (for example chiral pairs)

Discuss chemicals with worst RSMLEs

* Eventually ?

These QSPRs will enable public health risk-based prioritization of many more chemicals in commerce and the environment

Other edits:

* show CvT curve fits and in vitro curve – Supplemental Figures X
* Change label "FitstoData" -> "FitstoInVivoData"
* add mention of versions used
* what error should we care about the most? ten-fold error in clearance or ten-fold or fup?
* fix clint zero point-- clarify that this is only for purposes of plotting
* just one point aove 10^3 for clint, 90% are within two fold of each other (y scramble doesn't do much)
* give some chemical-specific
* add oral vs iv studies to table
* need supplemental table dose regimens (not per point)
* add to table 1 how many models compared

# Acknowledgements

The authors thank Drs. X and Y for their helpful U.S. EPA internal reviews of the manuscript.

# Conflict of Interest

Please declare any COI here

# Funding

Employees of the United States Environmental Protection Agency (EPA) were funded through EPA’s Office of Research and Development (ORD). The Oak Ridge Institute for Science

and Education provided funding for C. Cook. The National Institute of Environmental Health Sciences provided funding for S. Ferguson and K. Mansouri.

Please add your funding here

The views expressed in this publication are those of the authors and do not necessarily represent the views or policies of the U.S. EPA. Reference to commercial products or services does not constituteendorsement.

# References

1. O'Flaherty EJ (1981) Toxicants and drugs: kinetics and dynamics. John Wiley & Sons,

2. Coecke S, Pelkonen O, Leite SB, Bernauer U, Bessems JG, Bois FY, Gundert-Remy U, Loizou G, Testai E, Zaldívar J-M (2013) Toxicokinetics as a key to the integrated toxicity risk assessment based primarily on non-animal approaches. Toxicology in Vitro 27 (5):1570-1577

3. Sobus JR, Tan Y-M, Pleil JD, Sheldon LS (2011) A biomonitoring framework to support exposure and risk assessments. Science of the Total Environment 409 (22):4875-4884

4. National Research Council (1983) Risk Assessment in the Federal Government: Managing the Process. In. National Academies Press, Washington (DC). doi:10.17226/317

5. Wetmore BA (2015) Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. Toxicology 332:94-101

6. Kavlock R, Chandler K, Houck K, Hunter S, Judson R, Kleinstreuer N, Knudsen T, Martin M, Padilla S, Reif D, Richard AM, Rotroff D, Sipes NS, Dix D (2012) Update on EPA’s ToxCast program: providing high throughput decision support tools for chemical risk management. Chemical Research in Toxicology 25 (7):1287-1302

7. Tice RR, Austin CP, Kavlock RJ, Bucher JR (2013) Improving the human hazard characterization of chemicals: a Tox21 update. Environmental Health Perspectives 121 (7):756

8. Rotroff DM, Wetmore BA, Dix DJ, Ferguson SS, Clewell HJ, Houck KA, LeCluyse EL, Andersen ME, Judson RS, Smith CM, Sochaski MA, Kavlock RJ, Boellmann F, Martin MT, Reif DM, Wambaugh JF, Thomas RS (2010) Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. Toxicological Sciences 117 (2):348-358

9. Wetmore BA, Wambaugh JF, Ferguson SS, Sochaski MA, Rotroff DM, Freeman K, Clewell HJ, 3rd, Dix DJ, Andersen ME, Houck KA, Allen B, Judson RS, Singh R, Kavlock RJ, Richard AM, Thomas RS (2012) Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicological Sciences 125 (1):157-174. doi:10.1093/toxsci/kfr254

10. Tonnelier A, Coecke S, Zaldívar J-M (2012) Screening of chemicals for human bioaccumulative potential with a physiologically based toxicokinetic model. Archives of Toxicology 86 (3):393-403

11. Wetmore BA, Wambaugh JF, Allen B, Ferguson SS, Sochaski MA, Setzer RW, Houck KA, Strope CL, Cantwell K, Judson RS, LeCluyse E, Clewell HJ, Thomas RS, Andersen ME (2015) Incorporating High-Throughput Exposure Predictions With Dosimetry-Adjusted In Vitro Bioactivity to Inform Chemical Toxicity Testing. Toxicological Sciences 148 (1):121-136. doi:10.1093/toxsci/kfv171

12. Wambaugh JF, Wetmore BA, Ring CL, Nicolas CI, Pearce R, Honda G, Dinallo R, Angus D, Gilbert J, Sierra T, Badrinarayanan A, Snodgrass B, Brockman A, Strock C, Setzer W, Thomas RS (2019) Assessing Toxicokinetic Uncertainty and Variability in Risk Prioritization.

13. Bell SM, Chang X, Wambaugh JF, Allen DG, Bartels M, Brouwer KLR, Casey WM, Choksi N, Ferguson SS, Fraczkiewicz G, Jarabek AM, Ke A, Lumen A, Lynn SG, Paini A, Price PS, Ring C, Simon TW, Sipes NS, Sprankle CS, Strickland J, Troutman J, Wetmore BA, Kleinstreuer NC (2018) In vitro to in vivo extrapolation for high throughput prioritization and decision making. Toxicology in Vitro 47:213-227. doi:<https://doi.org/10.1016/j.tiv.2017.11.016>

14. Kavlock RJ, Bahadori T, Barton-Maclaren TS, Gwinn MR, Rasenberg M, Thomas RS (2018) Accelerating the Pace of Chemical Risk Assessment. Chemical Research in Toxicology 31 (5):287-290. doi:10.1021/acs.chemrestox.7b00339

15. Breen M, Ring CL, Kreutz A, Goldsmith M-R, Wambaugh JF (2021) High-throughput PBTK models for in vitro to in vivo extrapolation. Expert Opinion on Drug Metabolism & Toxicology (just-accepted)

16. Armitage JM, Hughes L, Sangion A, Arnot JA (2021) Development and intercomparison of single and multicompartment physiologically-based toxicokinetic models: Implications for model selection and tiered modeling frameworks. Environment International 154:106557

17. US EPA (2019) Directive to Prioritize Efforts to Reduce Animal Testing. Washington D.C.

18. Wang Y-H (2010) Confidence Assessment of the Simcyp Time-Based Approach and a Static Mathematical Model in Predicting Clinical Drug-Drug Interactions for Mechanism-Based CYP3A Inhibitors. Drug Metabolism and Disposition 38 (7):1094-1104. doi:10.1124/dmd.110.032177

19. Paini A, Cole T, Meinero M, Carpi D, Deceuninck P, Macko P, Palosaari T, Sund J, Worth A, Whelan M (2020) EURL ECVAM in vitro hepatocyte clearance and blood plasma protein binding dataset for 77 chemicals.

20. National Academies of Sciences E, Medicine (2017) Using 21st century science to improve risk-related evaluations. National Academies Press,

21. Arnot JA, Brown TN, Wania F (2014) Estimating screening-level organic chemical half-lives in humans. Environmental science & technology 48 (1):723-730

22. Papa E, Sangion A, Arnot JA, Gramatica P (2018) Development of human biotransformation QSARs and application for PBT assessment refinement. Food and Chemical Toxicology 112:535-543

23. Sipes NS, Wambaugh JF, Pearce R, Auerbach SS, Wetmore BA, Hsieh J-H, Shapiro AJ, Svoboda D, DeVito MJ, Ferguson SS (2017) An Intuitive Approach for Predicting Potential Human Health Risk with the Tox21 10k Library. Environmental Science & Technology 51 (18):10786-10796. doi:10.1021/acs.est.7b00650

24. Pradeep P, Patlewicz G, Pearce R, Wambaugh J, Wetmore B, Judson R (2020) Using chemical structure information to develop predictive models for in vitro toxicokinetic parameters to inform high-throughput risk-assessment. Computational Toxicology 16:100136

25. Dawson DE, Ingle BL, Phillips KA, Nichols JW, Wambaugh JF, Tornero-Velez R (2021) Designing QSARs for Parameters of High-Throughput Toxicokinetic Models Using Open-Source Descriptors. Environmental science & technology 55 (9):6505-6517

26. Wambaugh JF, Hughes MF, Ring CL, MacMillan DK, Ford J, Fennell TR, Black SR, Snyder RW, Sipes NS, Wetmore BA, Westerhout J, Setzer RW, Pearce RG, Simmons JE, Thomas RS (2018) Evaluating in vitro-in vivo extrapolation of toxicokinetics. Toxicological Sciences 163 (1):152-169

27. Sayre RR, Wambaugh JF, Grulke CM (2020) Database of pharmacokinetic time-series data and parameters for 144 environmental chemicals. Scientific Data

28. Shibata Y, Takahashi H, Chiba M, Ishii Y (2002) Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method. Drug Metabolism and Disposition 30 (8):892-896

29. Waters NJ, Jones R, Williams G, Sohal B (2008) Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding. Journal of Pharmaceutical Sciences 97 (10):4586-4595

30. Pearce RG, Setzer RW, Strope CL, Sipes NS, Wambaugh JF (2017) Httk: R package for high-throughput toxicokinetics. Journal of Statistical Software 79 (1):1-26

31. R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

32. Baumer B, Udwin D (2015) R markdown. Wiley Interdisciplinary Reviews: Computational Statistics 7 (3):167-177

33. Schmitt W (2008) General approach for the calculation of tissue to plasma partition coefficients. Toxicology in Vitro 22 (2):457-467

34. Pearce RG, Setzer RW, Davis JL, Wambaugh JF (2017) Evaluation and calibration of high-throughput predictions of chemical distribution to tissues. Journal of Pharmacokinetics and Pharmacodynamics 44 (6):549-565. doi:10.1007/s10928-017-9548-7

35. Davies B, Morris T (1993) Physiological parameters in laboratory animals and humans. Pharmaceutical Research 10 (7):1093-1095

36. Ruark CD, Hack CE, Robinson PJ, Mahle DA, Gearhart JM (2014) Predicting passive and active tissue: plasma partition coefficients: interindividual and interspecies variability. Journal of pharmaceutical sciences 103 (7):2189-2198

37. Akaike H (1998) Information theory and an extension of the maximum likelihood principle. In: Selected papers of hirotugu akaike. Springer, pp 199-213

38. Schmidt CW (2016) TSCA 2.0: A new era in chemical risk management. National Institute of Environmental Health Sciences,

39. Patlewicz G, Richard AM, Williams AJ, Grulke CM, Sams R, Lambert J, Noyes PD, DeVito MJ, Hines RN, Strynar M (2019) A chemical category-based prioritization approach for selecting 75 per-and polyfluoroalkyl substances (PFAS) for tiered toxicity and toxicokinetic testing. Environmental health perspectives 127 (01):014501

40. U.S. Centers for Disease Control and Prevention (2012) National Health and Nutrition Examination Survey. <http://www.cdc.gov/nchs/nhanes.htm>. Accessed 07/15/2013

41. Wang A, Gerona RR, Schwartz JM, Lin T, Sirota M, Morello-Frosch R, Woodruff TJ (2018) A suspect screening method for characterizing multiple chemical exposures among a demographically diverse population of pregnant women in San Francisco. Environmental Health Perspectives

126 (7):077009

42. Mansouri K, Grulke CM, Judson RS, Williams AJ (2018) OPERA models for predicting physicochemical properties and environmental fate endpoints. Journal of Cheminformatics 10 (1):10

43. Mansouri K, Chang X, Allen D, Judson R, Williams A, Kleinstreuer N (2021) OPERA models for ADME properties and toxicity endpoint. Paper presented at the Society of Toxicology Annual Meeting,

# Figures

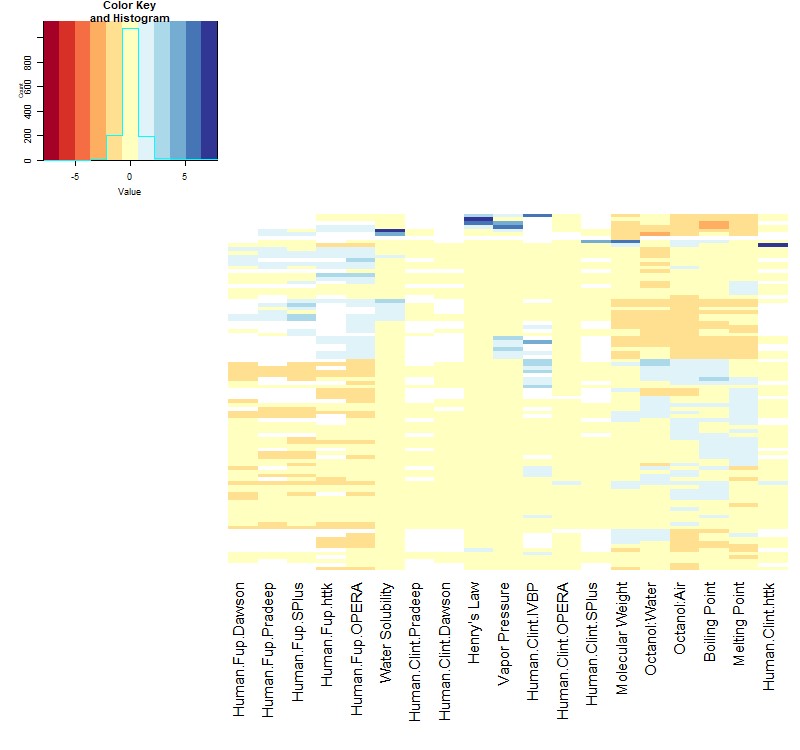


Figure :Columns in this heatmap indicate the physico chemical properties and measured/predicted values for in vitro TK (intrinsic hepatic clearance, Clint, and fraction unbound in plasma, fup). Each row corresponds to one of the 83 chemicals. The in vitro TK measurements (“Human.Clint.InVitro” and “Human.Fup.InvItro”) and predictions for these values from the various QSPRs (Table 2) are indicated by name. Data are normalized on a per column basis by centering (subtracting the mean) and scaling (by standard deviation). Thus, the "Value" of each entry in heatmap indicates the number of standard deviations from the mean. Blank values indicate no prediction.

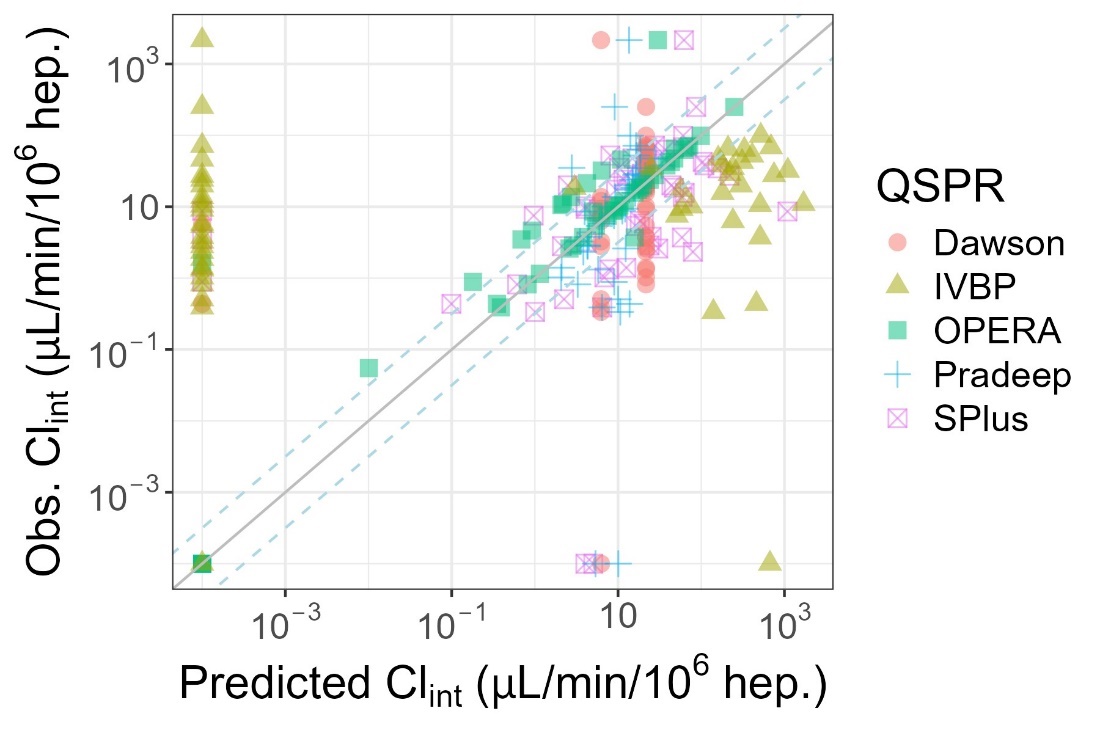
\ 

Figure : Evaluation of Predictions for Intrinsic Hepatic Clearance (Clint). Zero values were plotted at 10-1, the solid line indicates identity (1:1) while the dashed lines indicate 3.2-fold difference.



Figure : Evaluation of Predictions for Fraction Unbound in Plasma (fup). Zero values were plotted at 10-4, the solid line indicates identity (1:1) while the dashed lines indicate 3.2-fold difference

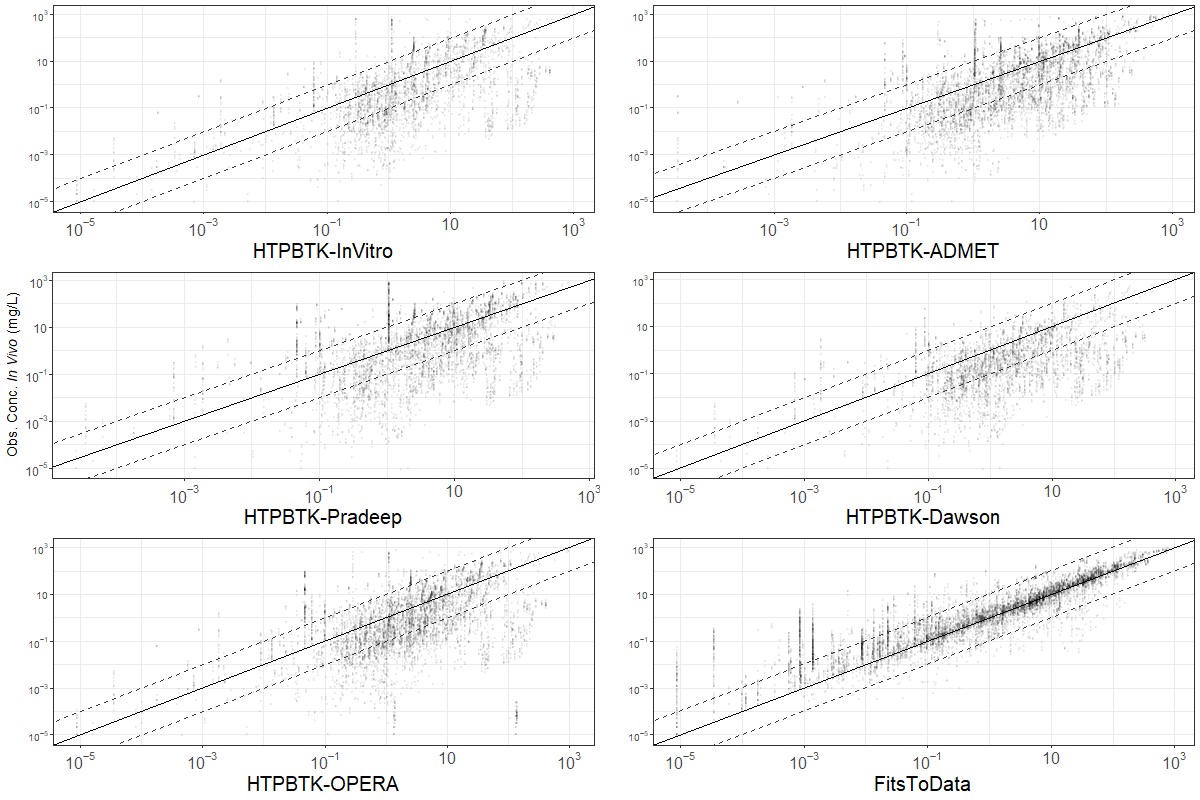


Figure : Comparison of in vivo measured chemical concentration vs. time (CvT) data [27] vs. predictions for empirical models fit to the data (“FitsToData”), and predictions for a PBTK model (“HTTK”) parameterized with chemical specific values either measured in vitro (“HTTK-InVitro”) or predicted with various QSPRs. In each sub-plot the y-axis shows the measured data while the x-axis shows the predictions made using chemical-specific parameters from the various sources. The solid line indicates identity (1:1) while the dashed lines indicate ten-fold difference.

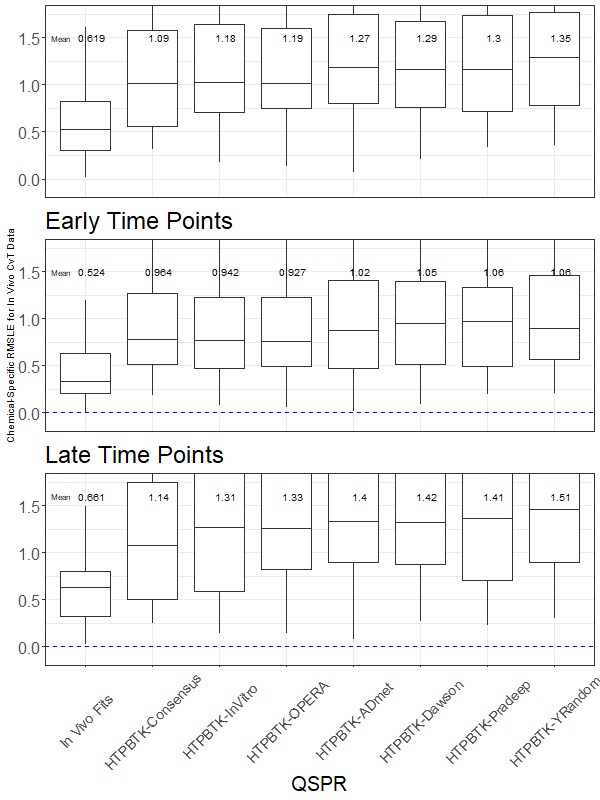


Figure : Chemical-Specific RMSLE for In vivo CvT Data. The upper and lower extent of the box for each model indicates the 25th to 75th quantiles, the mid-line indicates the median (50th quantile) and vertical line indicates 1.5x the range of the box.

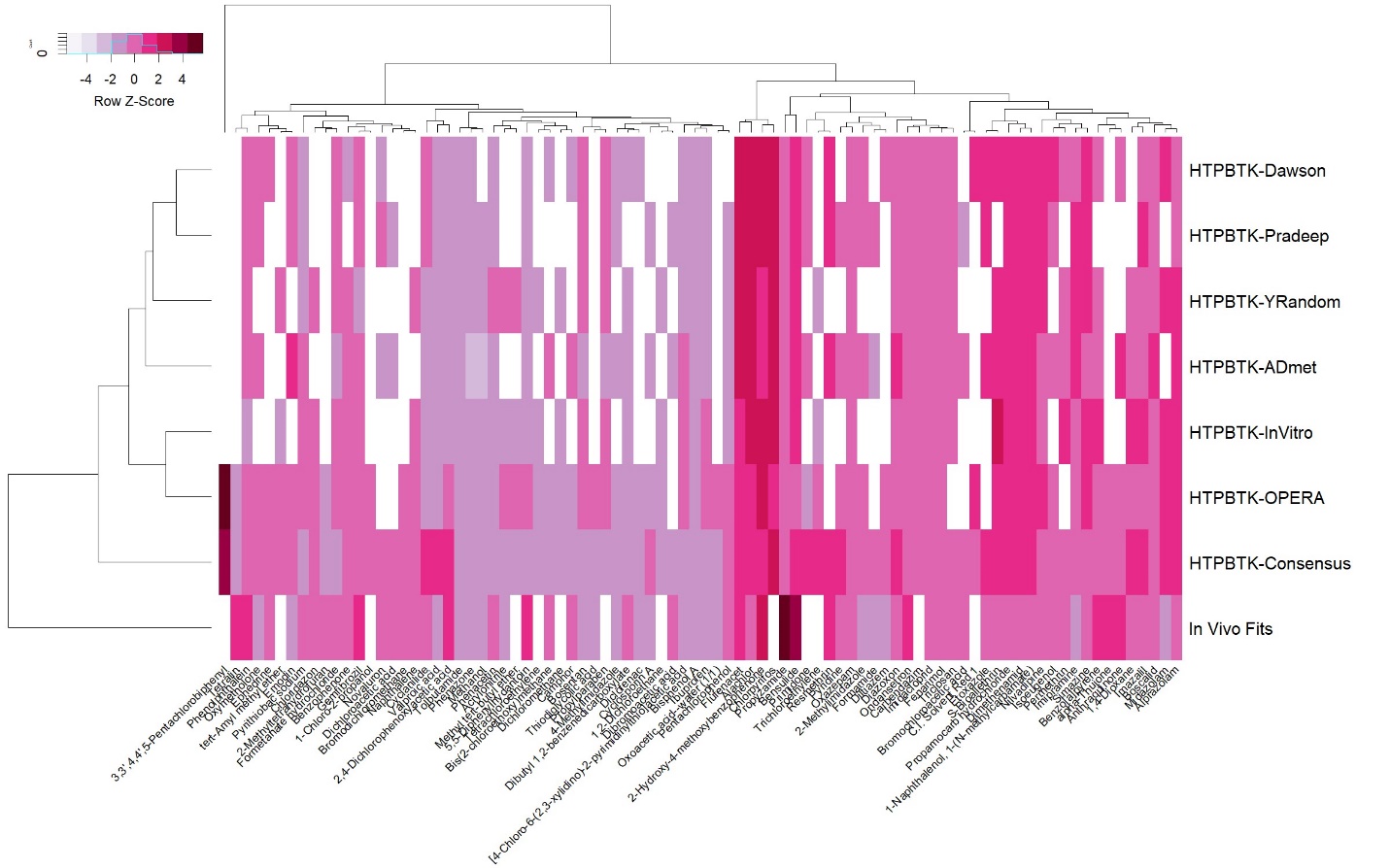
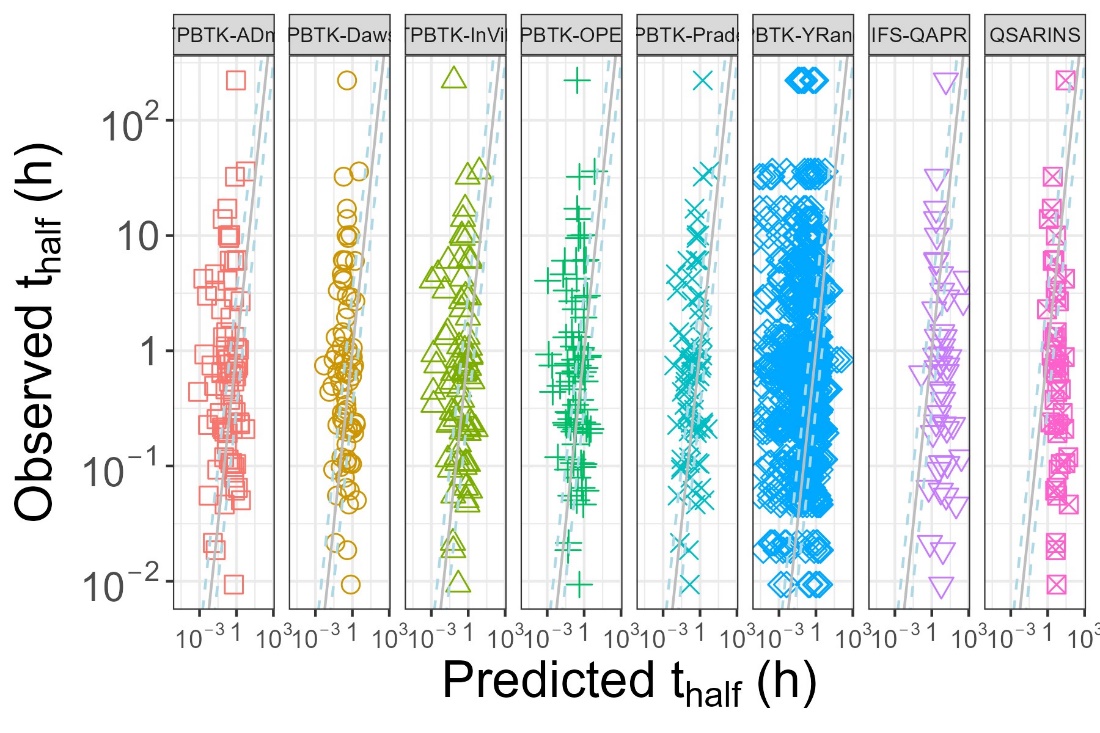


Figure : Values in this heatmap are the root mean squared log10 error (RMSLE) by chemical and predictor (that is, lighter indicates better predictive ability). The columns indicate different evaluation chemicals. The rows give the different prediction methods: The empirical fits to the data are given by “FitsToData”. All other values are calculated using the HTTK PBTK model and either measured values “In vitro”, y-randomized measured values (“Y-Random”) or the various QSPRS.



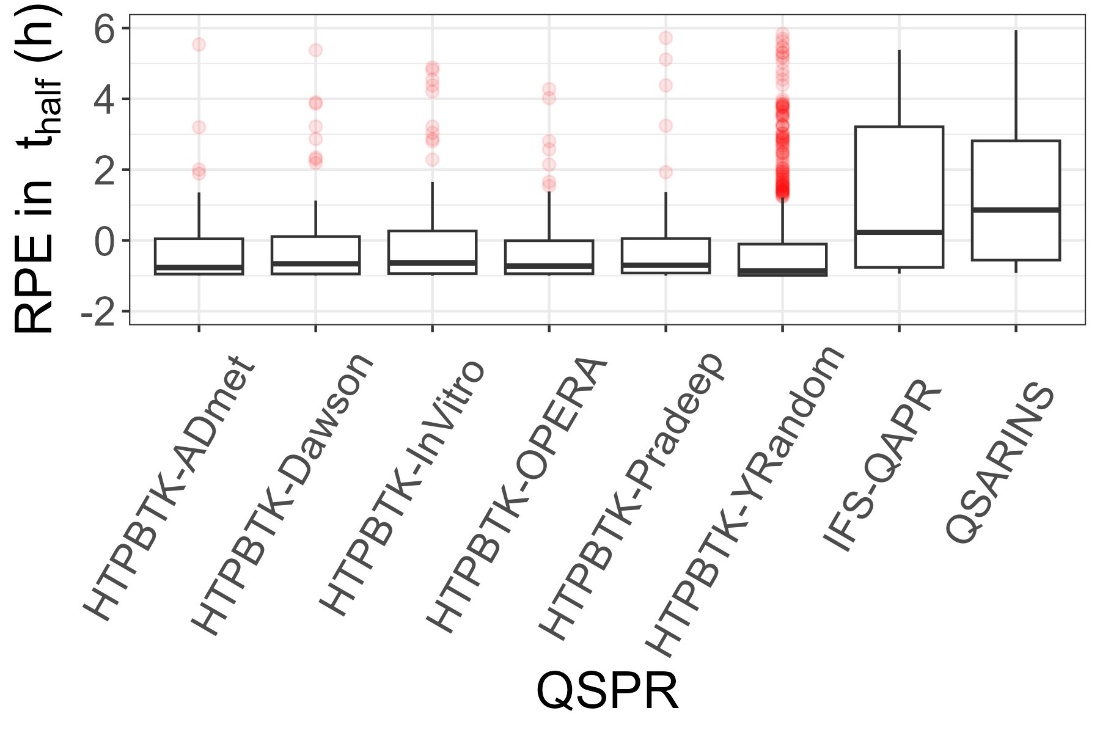
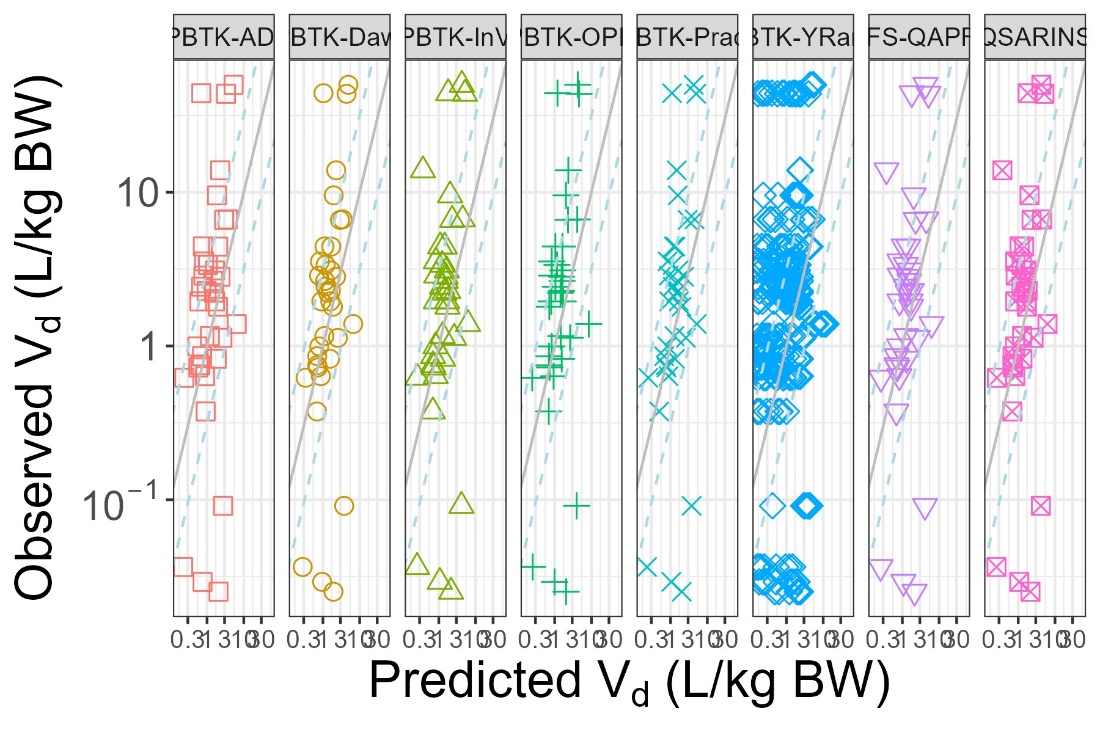


Figure : Comparison of “observed” chemical half-lives based empirical model fits and predictions for chemical half-life based on a PBPK model (“HTTK”) parameterized with chemical specific values either measured *in vitro* (“HTTK-InVitro”) or predicted with various QSPRs. The upper panel shows a scatter plot of predicted vs. observed values, while the lower panel shows the distribution of relative predictive error (RPE). The upper and lower extent of the box for each model indicates the 25th to 75th quantiles, the mid-line indicates the median (50th quantile) and vertical line indicates 1.5x the range of the box.



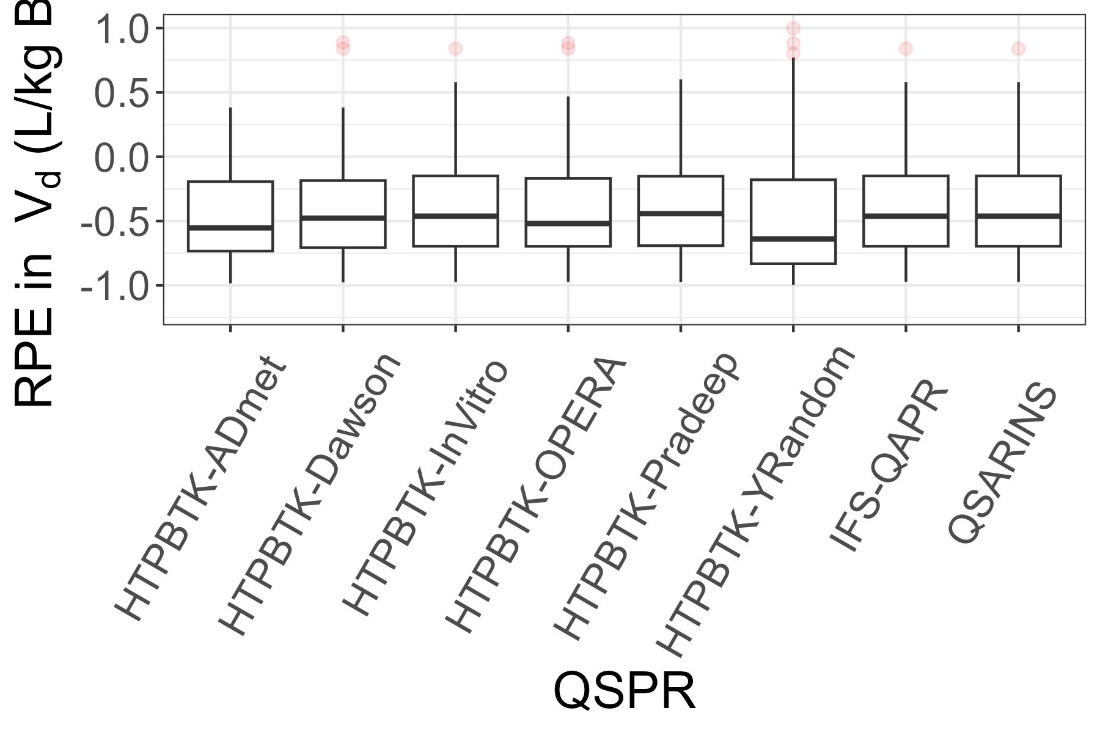
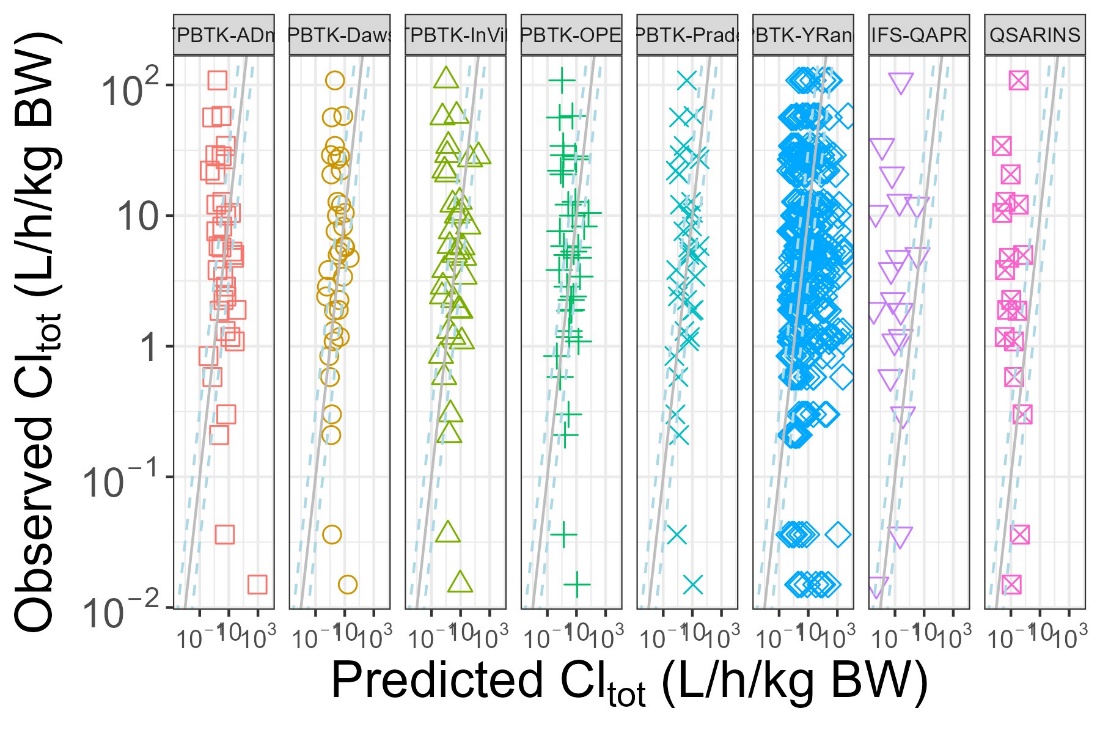


Figure : Comparison of “observed” chemical volumes of distribution based on empirical model fits and predictions for chemical half-life based on a PBPK model (“HTTK”) parameterized with chemical specific values either measured *in vitro* (“HTTK-InVitro”) or predicted with various QSPRs. The upper panel shows a scatter plot of predicted vs. observed values, while the lower panel shows the distribution of relative predictive error (RPE). The upper and lower extent of the box for each model indicates the 25th to 75th quantiles, the mid-line indicates the median (50th quantile) and vertical line indicates 1.5x the range of the box.



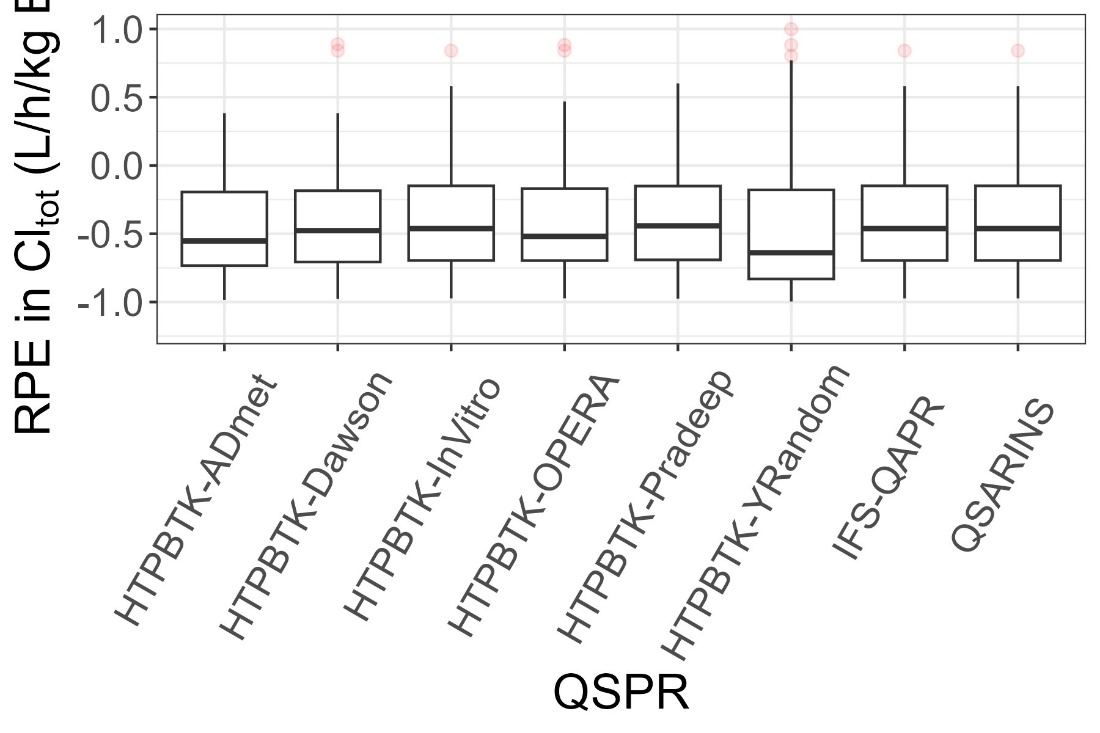


Figure : Comparison of “observed” chemical whole body clearance (Cltot) based on empirical model fits and predictions for chemical half-life based on a PBPK model (“HTTK”) parameterized with chemical specific values either measured *in vitro* (“HTTK-InVitro”) or predicted with various QSPRs. The upper panel shows a scatter plot of predicted vs. observed values, while the lower panel shows the distribution of relative predictive error (RPE). The upper and lower extent of the box for each model indicates the 25th to 75th quantiles, the mid-line indicates the median (50th quantile) and vertical line indicates 1.5x the range of the box.

# Tables

Table Three Levels of Evaluation were performed

|  |  |  |  |
| --- | --- | --- | --- |
| **Evaluation** | **TK Quantities** | **Chemicals for Evaluation** | **Reference** |
| **Level 1** | *In vitro* TK Measurements  (fup, Clint) | 63 with Measured *In vitro* Data | [30] |
| **Level 2** | TK Concentration vs. Time  (all points, Cmax, time-integral/AUC) | 83 with predictions across multiple QSPRs and empirical model fits | [27] |
| **Level 3** | Summary Statistics  (Vd, thalf, Cltot) | 83 | [27] |

Table QSPR Models Evaluated

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Predictions** | **Mechanism** | **Reference** |
| Simulations Plus ADMET Predictor® | Level 1  (*in vitro* parameters) | Sum of CYP-specific Artificial Neural Network (ANN) | [23] |
| Pradeep 2020 | Level 1 | Random forest and support vectors method | [24] |
| Dawson 2021 | Level 1 | Random forest, clearance organized by categories | [25] |
| OPERA | Level 1 | Nearest-neighbors | [42,43] |
| IFS-QSAR | Level 3  (Half-lives) | Fragment-based Multiple Linear Regressors (MRL) | [21] |
| QSARINS-Chem | Level 3 | Ordinary Least Squares MLR | [22] |

Table : Biases of the QSPRs for predicting in vitro measured values in terms of fold error (FE)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Median Clint AbsFE** | **Median Clint FE** | **Min Clint FE** | **Max Clint FE** | **Median fup AbsFE** | **Median fup FE** | **Min fup FE** | **Max fup FE** |
| SPlus | 0.503 | 0.105 | -2.34 | 2.1 | 0.274 | -0.0179 | -1.47 | 6.27 |
| Dawson | 0.348 | 0.0378 | -2.54 | 1.42 | 0.152 | -0.00384 | -1.23 | 6.38 |
| Pradeep | 0.232 | 0.0194 | -2.2 | 1.5 | 0.154 | -0.0191 | -0.506 | 1.92 |
| OPERA | 0.00141 | -0.00019 | -1.86 | 0.63 | 0.0281 | 0 | -1.11 | 7.1 |

Table 4 Level 2 Analysis of predictions based on empirical model fits and a PBPK model (“HTTK”) parameterized with chemical specific values either measured in vitro (“HTTK-InVitro”) or predicted with various QSPRs. Both the coefficient of determination (R2) and root mean squared error (RMSLE) are calculated on a log10 scale. “HTTK-YRandom” indicates the HTTK PBPK model parameterized with a random permutation of the in vitro data.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Column1** | **HTPBTK-InVitro** | **HTPBTK-InVitro-Caco2Exp** | **HTPBTK-InVitro-Caco2QSPR** | **HTPBTK-ADmet** | **HTPBTK-Dawson** | **HTPBTK-Pradeep** | **HTPBTK-OPERA** | **FitsToData** | **HTPBTK-YRandom** | **HTPBTK-InVitro-Measured** |
| AAFE | 8.94 | 8.87 | 9.05 | 8.87 | 8.93 | 9.38 | 9.52 | 3.28 | 11.88 | NA |
| RMSLE | 1.26 | 1.25 | 1.26 | 1.24 | 1.25 | 1.28 | 1.27 | 0.82 | 1.39 | NA |
| MRPE | 1.51 | 1.33 | 1.37 | 2.64 | 1.45 | 1.7 | 2.1 | -0.23 | 2.84 | NA |
| RMSLE.early | 1.13 | 1.12 | 1.13 | 1.1 | 1.12 | 1.15 | 1.15 | 0.79 | 1.22 | NA |
| RMSLE.late | 1.55 | 1.53 | 1.55 | 1.55 | 1.55 | 1.56 | 1.55 | 0.9 | 1.76 | NA |
| AAFE.early | 6.77 | 6.78 | 6.9 | 6.36 | 6.7 | 7.22 | 7.23 | 2.88 | 8.45 | NA |
| AAFE.late | 19.2 | 18.52 | 19.05 | 22.07 | 19.68 | 19.26 | 20.26 | 4.76 | 30.13 | NA |
| MRPE.early | 0.8 | 0.68 | 0.7 | 1.46 | 0.76 | 1.04 | 1.23 | -0.13 | 1.85 | NA |
| MRPE.late | 6.29 | 5.68 | 5.97 | 16.74 | 6.69 | 6.85 | 9.53 | -0.59 | 13.52 | NA |
| RMSLE.bychem | 0.968 | NA | NA | 1.12 | 0.978 | 1.04 | 0.987 | 0.521 | 1.22 | 0.941 |
| RMSLE.bychem.early | 0.907 | NA | NA | 0.909 | 0.847 | 0.907 | 0.895 | 0.372 | 1.02 | 0.878 |
| RMSLE.bychem.late | 1.23 | NA | NA | 1.41 | 1.16 | 1.18 | 1.29 | 0.58 | 1.52 | 1.24 |

Table Level 3 Analysis of predictions based on empirical model fits and a PBPK model (“HTTK”) parameterized with chemical specific values either measured in vitro (“HTTK-InVitro”) or predicted with various QSPRs. Both the coefficient of determination (R2) and root mean squared error (RMSLE) are calculated on a log10 scale. “HTTK-YRandom” indicates the HTTK PBPK model parameterized with a random permutation of the in vitro data.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Cmax** | | | **AUC** | | | | |
| **Predictor** | **R2** | **RMSLE** | **RPE** | **R2** | **RMSLE** | **RPE** | **RPE Low** | **RPE High** | |
| Empirical Fits | 0.95 | 0.95 | -0.09 | 0.96 | 0.31 | 0.32 | 0.41 | 0.06 | |
| HTTK-InVitro | 0.57 | 0.84 | -0.02 | 0.50 | 1.11 | 3.44 | 6.72 | -0.46 | |
| HTTK-ADmet | 0.62 | 0.78 | 0.15 | 0.62 | 0.97 | 5.88 | 10.50 | -0.44 | |
| HTTK-Dawson | 0.60 | 0.81 | 0.02 | 0.58 | 1.01 | 3.32 | 5.00 | -0.04 | |
| HTTK-Pradeep | 0.57 | 0.83 | -0.11 | 0.40 | 1.22 | 2.29 | 5.99 | -0.80 | |
| HTTK-OPERA | 0.58 | 0.83 | 0.11 | 0.56 | 1.04 | 4.67 | 7.93 | 0.39 | |
| HTTK-YRandom | 0.46 | 0.93 | 0.53 | 0.10 | 1.48 | 5.10 | 12.10 | -0.88 | |

Table Analysis of predictions based on empirical model fits and a PBPK model (“HTTK”) parameterized with chemical specific values either measured in vitro (“HTTK-InVitro”) or predicted with various QSPRs. Both the coefficient of determination (R2) and root mean squared error (RMSLE) are calculated on a log10 scale. “HTTK-YRandom” indicates the HTTK PBPK model parameterized with a random permutation of the in vitro data.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Level 3** | | | | | |
|  | **thalf** | | **Vd** | | **Cltot** | |
| **Predictor** | **R2** | **RMSE** | **R22** | **R222** | **R23** | **RMSE3** |
| HTTK-InVitro | 0.04 | 1.43 | 0.04 | 0.89 | 0.03 | 1.57 |
| HTTK-ADmet | 0.02 | 1.45 | 0.13 | 0.85 | 0.20 | 1.42 |
| HTTK-Dawson | 0.09 | 1.39 | 0.11 | 0.86 | 0.20 | 1.42 |
| HTTK-Pradeep | 0.00 | 1.46 | 0.13 | 0.85 | 0.02 | 1.57 |
| HTTK-OPERA | 0.09 | 1.39 | 0.05 | 0.89 | 0.09 | 1.52 |
| HTTK-YRandom | 0.01 | 1.45 | 0.00 | 0.91 | 0.02 | 1.57 |
| QSARINS | 0.10 | 1.00 | 0.04 | 0.89 | 0.19 | 1.17 |
| IFS-QAPR | 0.15 | 0.97 | 0.04 | 0.89 | 0.25 | 1.12 |

# Supplemental Figures

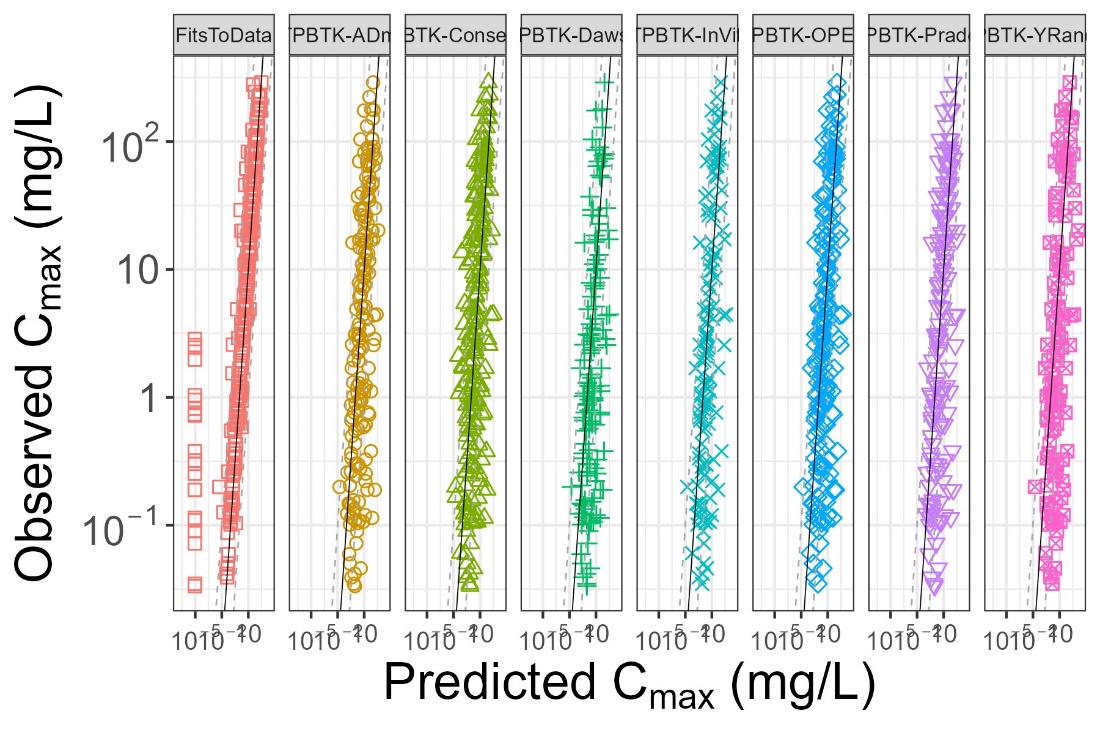


Figure 6: Evaluation of Predictions for Cmax based on empirical model fits (“1CompFits”), and a PBPK model (“HTTK”) parameterized with chemical specific values either measured in vitro (“HTTK-InVitro”) or predicted with various QSPRs.

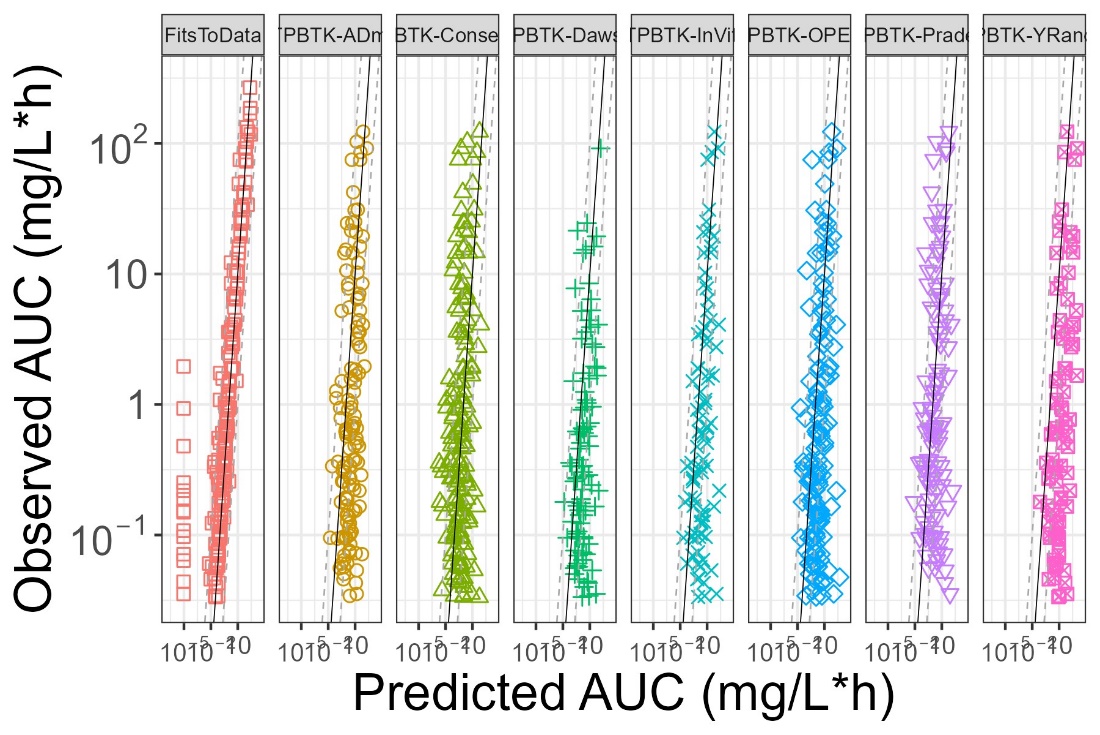


Figure 7: Evaluation of Predictions for AUC based on empirical model fits (“1CompFits”), and a PBPK model (“HTTK”) parameterized with chemical specific values either measured in vitro (“HTTK-InVitro”) or predicted with various QSPRs.

# Supplemental Tables

Supplemental Table 1: List of 102 chemicals, physico-chemical descriptors, and *in vitro* measured values, and QSPR predictions

*SupTable-QSPRPredsandInVitroData.xlsx*

Supplemental Table 2: Concentration vs. time data (Sayre et al., 2020)

*SupTable-CvTData.xlsx*

Supplemental Table 3: Empirical (one and two compartment model) toxicokinetic parameter estimates

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Compound | CAS | Species | Reference | AIC.1comp | AIC.2comp | Model | Vdist | kelim | halflife |
| 1,2-dichloroethane | 107-06-2 | rat | 18 | 286.7 | 290.7 | 1Comp | 3.65 | 0.3431 | 2.02 |
| 1,4-dioxane | 123-91-1 | rat | 24, 6 | 1373 | 1373 | 2Comp | 0.4464 | 0.01508 | 45.97 |
| 1-chloro-2-propanol | 127-00-4 | rat | 176 | 5.553 | 10.39 | 1Comp | 2.835 | 1.344 | 0.5157 |
| 2,3,7,8-tetrachlorodibenzo-p-dioxin | 1746-01-6 | rat | 177 | -49.13 | -45.42 | 1Comp | 0.2341 | 0.001496 | 463.4 |
| 2,4-dichlorophenoxyacetic acid | 94-75-7 | rat | 164, 192 | 238.4 | 242.4 | 1Comp | 0.3759 | 0.2186 | 3.171 |
| 2-hydroxy-4-methoxybenzophenone | 131-57-7 | rat | 129 | -82.97 | -112 | 2Comp | 11.6 | 3.304 | 0.2098 |
| 2-methylimidazole | 693-98-1 | rat | 150 | 29.44 | 9.339 | 2Comp | 1.999 | 0.02467 | 28.1 |
| 2-methyltetrahydrofuran | 96-47-9 | rat | 175 | 420.5 | 417.6 | 2Comp | 1.072 | 0.9591 | 0.7227 |
| 4-methylimidazole | 822-36-6 | rat | 151 | 28.6 | 0.162 | 2Comp | 1.218 | 0.8444 | 0.8209 |
| acrylonitrile | 107-13-1 | rat | 43 | 195.4 | 170.3 | 2Comp | 0.7476 | 9.897 | 0.07003 |
| alachlor | 15972-60-8 | rat | 192 | -139 | -135.5 | 1Comp | 130.9 | 0.05883 | 11.78 |
| alpha-thujone | 546-80-5 | rat | 130 | -69.3 | -80.52 | 2Comp | 13.94 | 0.5856 | 1.184 |
| alprazolam | 28981-97-7 | rat | 192 | -87 | -116.9 | 2Comp | 9.42 | 3.702 | 0.1872 |
| anthraquinone | 84-65-1 | rat | 166 | 1.042 | NA | 1Comp | 4.74 | 0.09222 | 7.517 |
| antipyrine | 60-80-0 | rat | 192 | -4.438 | -2.125 | 1Comp | 1.133 | 0.5825 | 1.19 |
| benzophenone | 119-61-9 | rat | 171 | -27.96 | -52.25 | 2Comp | 12.76 | 0.2639 | 2.627 |
| bisphenol a | 80-05-7 | rat | 192 | -13.29 | -87.26 | 2Comp | 1.787 | 4.052 | 0.1711 |
| boscalid | 188425-85-6 | rat | 192 | -44.32 | -73.71 | 2Comp | 10.82 | 1.346 | 0.5151 |
| bosentan | 147536-97-8 | rat | 192 | -18.22 | -14.22 | 1Comp | 3.156 | 0.2439 | 2.841 |
| bromochloroacetic acid | 5589-96-8 | rat | 132 | 192.1 | NA | 1Comp | 0.3796 | 2.259 | 0.3069 |
| bromodichloromethane | 75-27-4 | rat | 154 | -7.809 | -45.51 | 2Comp | 22.72 | 3.576 | 0.1938 |
| carbaryl | 63-25-2 | rat | 192 | -55.78 | -24.19 | 1Comp | 34.44 | 6.131 | 0.1131 |
| carbendazim | 10605-21-7 | rat | 192 | 28.13 | 32.09 | 1Comp | 33.15 | 0.245 | 2.829 |
| chloridazon | 1698-60-8 | rat | 192 | -62.76 | NA | 1Comp | 9.044 | 0.07509 | 9.23 |
| chlorpyrifos | 2921-88-2 | rat | 192 | -88.14 | -84.14 | 1Comp | 100.3 | 0.1788 | 3.877 |
| cyclanilide | 113136-77-9 | rat | 192 | -30.3 | -66.03 | 2Comp | 0.3674 | 0.3297 | 2.102 |
| cyclosporin a | 59865-13-3 | rat | 192 | 113 | 76.75 | 2Comp | 1.593 | 0.3693 | 1.877 |
| di-n-butyl phthalate | 84-74-2 | rat | 135 | -3.349 | -35.07 | 2Comp | 1.587 | 8.08 | 0.08578 |
| diazinon-o-analog | 962-58-3 | rat | 192 | -58.93 | -68.18 | 2Comp | 177.3 | 0.000139 | 5001 |
| dibromoacetic acid | 631-64-1 | rat | 133 | 498.9 | 793 | 1Comp | 0.4029 | 1.692 | 0.4096 |
| dichloroacetic acid | 79-43-6 | rat | 134 | 230.2 | 240.1 | 1Comp | 0.3354 | 2.837 | 0.2443 |
| diltiazem | 34933-06-7 | rat | 192 | 12.39 | -9.811 | 2Comp | 2.747 | 3.384 | 0.2048 |
| dimethenamid | 87674-68-8 | rat | 192 | -120.1 | -142.7 | 2Comp | 246.7 | 2.019 | 0.3434 |
| dl-camphor | 76-22-2 | rat | 165 | 4.603 | -22.82 | 2Comp | 7.166 | 2.286 | 0.3032 |
| emodin | 518-82-1 | rat | 160 | 4.277 | 7.361 | 1Comp | 0.6335 | 2.886 | 0.2402 |
| etoxazole | 153233-91-1 | rat | 192 | -148.9 | -173 | 2Comp | 34.3 | 0.6847 | 1.012 |
| fenarimol | 60168-88-9 | rat | 192 | -68.23 | -91.56 | 2Comp | 12.17 | 0.3854 | 1.799 |
| flufenacet | 142459-58-3 | rat | 192 | -197.7 | -195.1 | 1Comp | 58.71 | 0.0321 | 21.59 |
| formamide | 75-12-7 | rat | 172 | 802.6 | NA | 1Comp | 0.5236 | 0.03136 | 22.1 |
| formetanate hydrochloride | 23422-53-9 | rat | 192 | -17.95 | -13.95 | 1Comp | 29.4 | 1.07E-08 | 64830000 |
| free carbon disulfide | 75-15-0 | rat | 167 | 52.7 | 33.58 | 2Comp | 3.469 | 6.702 | 0.1034 |
| gemfibrozil | 25812-30-0 | rat | 163 | 226.2 | NA | 1Comp | 0.7988 | 0.06215 | 11.15 |
| glyoxylic acid monohydrate | 563-96-2 | rat | 156 | 59.3 | 18.78 | 2Comp | 0.5298 | 19.34 | 0.03584 |
| hexachlorobenzene | 118-74-1 | rat | 174, 180 | NA | -242.2 | 2Comp | 7.088 | 0.002409 | 287.7 |
| hexobarbital | 15307-86-5 | rat | 192 | 222.2 | 492.2 | 1Comp | 0.4199 | 4.003 | 0.1732 |
| ibuprofen | 15687-27-1 | rat | 192 | 118.4 | 92.01 | 2Comp | 0.687 | 0.8539 | 0.8118 |
| imazalil | 35554-44-0 | rat | 192 | -54.53 | -91.25 | 2Comp | 10.4 | 5.226 | 0.1326 |
| imipramine | 50-49-7 | rat | 192 | -0.5092 | -20.95 | 2Comp | 45.44 | 0.2763 | 2.509 |
| isoeugenol | 97-54-1 | rat | 157 | 28.35 | -56.77 | 2Comp | 27.95 | 2.59 | 0.2676 |
| l-ephedrine | 299-42-3 | rat | 136 | -36.84 | -65.5 | 2Comp | 11.6 | 0.9385 | 0.7385 |
| methanol | 67-56-1 | rat | 17 | 46.39 | 47.83 | 1Comp | 0.992 | 0.3046 | 2.275 |
| methyl tert-butyl ether | 1634-04-4 | human | 51 | -134.9 | -141.9 | 2Comp | 0.8526 | 1.356 | 0.511 |
| methylene chloride | 75-09-2 | rat | 18 | 312.7 | 313.7 | 1Comp | 1.883 | 0.4908 | 1.412 |
| methyleugenol | 93-15-2 | rat | 158, 170 | NA | -5.438 | 2Comp | 8.303 | 3.048 | 0.2274 |
| midazolam | 59467-70-8 | rat | 192 | -36.58 | -49.41 | 2Comp | 3.487 | 3.998 | 0.1734 |
| nilvadipine | 75530-68-6 | rat | 192 | -93.6 | -113.4 | 2Comp | 10.74 | 1.03 | 0.673 |
| nitrite | 14797-65-0 | rat | 137 | 121.8 | 124.2 | 1Comp | 2.528 | 0.3789 | 1.829 |
| novaluron | 116714-46-6 | rat | 192 | -82.71 | -102.7 | 2Comp | 7.068 | 0.07823 | 8.861 |
| octylphenol | 140-66-9 | rat | 60 | -207.3 | -299.5 | 2Comp | 49.92 | 0.3106 | 2.231 |
| oxazepam | 604-75-1 | rat | 159 | 146 | 102.8 | 2Comp | 16.9 | 0.05402 | 12.83 |
| oxymetholone | 434-07-1 | rat | 168 | 67.48 | 30.16 | 2Comp | 8.674 | 3.082 | 0.2249 |
| pentachlorophenol, purified | 87-86-5 | rat | 161 | 519.8 | 522.2 | 1Comp | 0.0887 | 0.08031 | 8.631 |
| perfluorodecanoic acid | 335-76-2 | rat | 147 | 293.3 | NA | 1Comp | 0.2088 | 0.001265 | 548 |
| perfluorohexane-1-sulphonic acid – potassium salt | 3871-99-6 | rat | 143 | 464.4 | NA | 1Comp | 0.2362 | 0.001256 | 552 |
| perfluorooctane sulfonate | 45298-90-6 | rat | 148 | 173.9 | NA | 1Comp | 0.2185 | 0.000885 | 782.9 |
| perfluorooctanoic acid | 335-67-1 | rat | 146, 192 | 1075 | 1012 | 2Comp | 46.58 | 1.43E-07 | 4843000 |
| phenacetin | 62-44-2 | rat | 1, 192, 62, 63, 67 | 265.4 | 215.2 | 2Comp | 1.599 | 0.7362 | 0.9415 |
| phenytoin | 57-41-0 | rat | 192 | 228.4 | NA | 1Comp | 1.273 | 0.2112 | 3.282 |
| potassium perfluorobutane sulfonate | 29420-49-3 | rat | 144 | 330.5 | 318 | 2Comp | 0.2398 | 0.1955 | 3.545 |
| primidone | 125-33-7 | rat | 140 | 125.7 | 129.7 | 1Comp | 1.107 | 0.1883 | 3.68 |
| propamocarb hydrochloride | 25606-41-1 | rat | 192 | -68.01 | -95.77 | 2Comp | 8.131 | 3.207 | 0.2161 |
| propylparaben | 94-13-3 | rat | 64 | 139.9 | NA | 1Comp | 0.03969 | 333.2 | 0.00208 |
| propyzamide | 23950-58-5 | rat | 192 | -113.1 | NA | 1Comp | 11.09 | 0.09099 | 7.618 |
| pyridine | 110-86-1 | rat | 149, 162 | 516.2 | 516.4 | 1Comp | 1.241 | 0.07107 | 9.753 |
| pyrithiobac sodium | 123343-16-8 | rat | 192 | -6.596 | -2.142 | 1Comp | 1.09 | 0.08863 | 7.821 |
| resmethrin | 10453-86-8 | rat | 192 | -70.11 | NA | 1Comp | 59.95 | 0.2174 | 3.189 |
| s-bioallethrin | 28434-00-6 | rat | 192 | -77.36 | -83.84 | 2Comp | 47.54 | 0.831 | 0.8341 |
| simazine | 122-34-9 | rat | 192 | -64.68 | -62.89 | 1Comp | 3.532 | 1.992 | 0.348 |
| solvent red1 | 1229-55-6 | rat | 66 | -54.87 | -60 | 2Comp | 1.875 | 1.066 | 0.6502 |
| tamoxifen | 10540-29-1 | rat | 141 | -26.92 | -26.42 | 1Comp | 37050 | 0.07607 | 9.112 |
| tert-amyl methyl ether | 994-05-8 | human | 51 | -149.4 | -172.1 | 2Comp | 2.608 | 1.342 | 0.5167 |
| tetrachloroethylene | 127-18-4 | rat | 4 | 70.07 | 72.5 | 1Comp | 12.73 | 0.08628 | 8.034 |
| thiodiglycolic acid | 123-93-3 | rat | 155 | 43.13 | 20.29 | 2Comp | 0.7874 | 5.729 | 0.121 |
| trichloroethylene | 79-01-6 | rat | 18, 29 | 80.68 | 1508 | 1Comp | 2.565 | 0.8173 | 0.8481 |
| valproic acid | 99-66-1 | rat | 192 | 432.8 | 430.5 | 2Comp | 0.574 | 0.7342 | 0.9441 |
| wyeth-14643 | 50892-23-4 | rat | 173 | 113.1 | 102.8 | 2Comp | 0.4481 | 0.6937 | 0.9992 |
| bis 2-chloroethoxy methane | 111-91-1 | rat | 155 | NA | -0.4358 | 2Comp | 2.621 | 3.149 | 0.2202 |
| fluorotelomer alcohol 8+2 | 678-39-7 | rat | 131 | NA | -74.51 | 2Comp | 12.96 | 1.419 | 0.4884 |
| naphthalene | 91-20-3 | rat | 169 | NA | -29.5 | 2Comp | 5.786 | 2.63 | 0.2635 |
| perfluorohexanoic acid | 307-24-4 | rat | 145 | NA | 429.3 | 2Comp | 1.783 | 0.1387 | 4.996 |
| tetralin | 119-64-2 | rat | 152 | NA | 0.244 | 2Comp | 5.271 | 1.94 | 0.3573 |
| triclosan | 3380-34-5 | rat | 192 | NA | 27.86 | 2Comp | 0.3428 | 0.2387 | 2.904 |

Supplemental Table 4: Chemicals that could not be fit by either a one- or two-compartment model using R package invivoPKfit

|  |  |  |
| --- | --- | --- |
| **DTXSID** | **PREFERRED\_NAME** | **CASRN** |
| DTXSID0021125 | Phenolphthalein | 77-09-8 |
| DTXSID2020139 | Benzo(a)pyrene | 50-32-8 |
| DTXSID2021103 | Pentachloroanisole | 1825-21-4 |
| DTXSID5032442 | Imidacloprid | 138261-41-3 |
| DTXSID8021359 | Tolbutamide | 64-77-7 |
| DTXSID8022292 | Permethrin | 52645-53-1 |
| DTXSID8023393 | Ondansetron | 99614-02-5 |
| DTXSID9032329 | Bensulide | 741-58-2 |

Supplemental Table 5: Chemicals that could only be predicted by OPERA

|  |  |  |
| --- | --- | --- |
| DTXSID | PREFERRED\_NAME | CASRN |
| DTXSID0022985 | Ephedrine | 299-42-3 |
| DTXSID1051432 | Thiodiglycolic acid | 123-93-3 |
| DTXSID3032179 | 3,3',4,4',5-Pentachlorobiphenyl | 57465-28-8 |
| DTXSID30575892 | Oxoacetic acid--water (1/1) | 563-96-2 |
| DTXSID3061635 | 1-((2-Methoxyphenyl)azo)-2-naphthol | 1229-55-6 |
| DTXSID5020285 | 1-Chloro-2-propanol | 127-00-4 |
| DTXSID5024219 | Nitrite | 14797-65-0 |
| DTXSID50881104 | Diltiazem | 34933-06-7 |
| DTXSID7030066 | 2,3,4,7,8-Pentachlorodibenzofuran | 57117-31-4 |
| DTXSID80108992 | Perfluorooctanesulfonate | 45298-90-6 |

Supplemental Table 6: Chemicals whose measured values were potentially retrieved “as is” from model training sets and were therefore removed from the evaluation:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| DTXSID | PREFERRED\_NAME | CASRN | Human.Clint.httk | Human.Fup.httk | Human.Clint.pred | Human.Fup.pred | QSPR | Human.Clint.AbsFE | Human.fup.AbsFE |
| DTXSID0020652 | Gemfibrozil | 25812-30-0 | 52.2 | 0.03 | 52 | 0.03 | OPERA | 0.00175 | 0 |
| DTXSID1021087 | Oxazepam | 604-75-1 | 3.27 | 0.04 | 3.27 | 0.04 | OPERA | 0.000133 | 0 |
| DTXSID1021116 | Phenacetin | 62-44-2 | 9.35 | 0.6 | 9.35 | 0.6 | OPERA | 0.000186 | 0 |
| DTXSID1022265 | Alachlor | 15972-60-8 | 62.9 | 0.133 | 62.9 | 0.13 | OPERA | 0.000276 | 0.00991 |
| DTXSID1043881 | Imipramine | 50-49-7 | 35 | 0.1 | 35 | 0.1 | OPERA | 0.00062 | 0 |
| DTXSID2032552 | Flufenacet | 142459-58-3 | 29 | 0.137 | 29 | 0.14 | OPERA | 0.00015 | 0.00941 |
| DTXSID3034872 | Chloridazon | 1698-60-8 | 1.87 | 0.427 | 1.87 | 0.43 | OPERA | 0 | 0.00304 |
| DTXSID4021268 | Simazine | 122-34-9 | 6.33 | 0.284 | 6.34 | 0.28 | OPERA | 0.000686 | 0.00616 |
| DTXSID4032376 | Dimethenamid | 87674-68-8 | 19.9 | 0.232 | 20 | 0.23 | OPERA | 0.000654 | 0.00376 |
| DTXSID4032405 | Formetanate hydrochloride | 23422-53-9 | 13.7 | 0.885 | 13.7 | 0.89 | OPERA | 0.000317 | 0.00245 |
| DTXSID5025607 | Methyleugenol | 93-15-2 | 26.6 | 0.142 | 26.5 | 0.14 | OPERA | 0.000818 | 0.00494 |
| DTXSID5032442 | Imidacloprid | 138261-41-3 | 2.81 | 0.656 | 2.81 | 0.66 | OPERA | 0 | 0.00264 |
| DTXSID5032600 | Cyclanilide | 113136-83-9 | 0.34 | 0.984 | 0.34 | 0.98 | OPERA | 0 | 0.00183 |
| DTXSID5037523 | Diazoxon | 962-58-3 | 34.6 | 0.327 | 34.6 | 0.32 | OPERA | 0.000125 | 0.0094 |
| DTXSID6021117 | Phenazone | 60-80-0 | 0.506 | 0.97 | 0.51 | 0.96 | OPERA | 0.00325 | 0.0045 |
| DTXSID6034849 | Propamocarb hydrochloride | 25606-41-1 | 5.81 | 0.843 | 5.81 | 0.84 | OPERA | 0 | 0.00155 |
| DTXSID7046627 | Bosentan | 147536-97-8 | 2.34 | 0.02 | 2.34 | 0.02 | OPERA | 0.000743 | 0 |
| DTXSID8021359 | Tolbutamide | 64-83-7 | 1.32 | 0.04 | 1.32 | 0.04 | OPERA | 0.00131 | 0 |
| DTXSID8023393 | Ondansetron | 99614-02-5 | 1.4 | 0.25 | 1.41 | 0.25 | OPERA | 0.00247 | 0 |
| DTXSID8024151 | Imazalil | 35554-44-0 | 1.02 | 0.03 | 1.02 | 0.03 | OPERA | 0 | 0 |
| DTXSID9020247 | Carbaryl | 63-25-2 | 27.3 | 0.692 | 27.3 | 0.69 | OPERA | 0 | 0.00126 |

Supplemental Table 7: Chemical-specific root mean square log10 errors for the full TK concentration time course data by QSPRs

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | FitsToData | ADMET | Dawson | InVitro | OPERA | Pradeep | Y-Random |
| 1,2-Dichloroethane | 0.406 | 0.576 | 1.05 | 1.11 | 0.957 | 1.06 | 0.65 |
| 1,4-Dioxane | 0.598 | 1.79 | 1.82 | 1.82 | 1.8 | 1.83 | 1.84 |
| 2-Hydroxy-4-methoxybenzophenone | 0.751 | 1.64 | 1.73 | 2 | 1.98 | 1.87 | 1.71 |
| 2-methylimidazole | 0.157 | 1.36 | 1.36 | 1.36 | 1.36 | 1.36 | 1.36 |
| 2-Methyltetrahydrofuran | 0.731 | 1.25 | 1.89 | 1.89 | 1.94 | 1.99 | 1.91 |
| 2,3,7,8-Tetrachlorodibenzo-p-dioxin | 0.417 | 0.417 | 0.417 | 0.417 | 0.417 | 0.417 | 0.417 |
| 2,4-D | 0.476 | 1.16 | 0.696 | 0.624 | 0.624 | 1.14 | 1.68 |
| 4-methylimidazole | 0.21 | 1.13 | 1.31 | 1.31 | 0.879 | 1.44 | 1.32 |
| acrylonitrile | 0.421 | 0.272 | 1.24 | 1.38 | 1.17 | 1.2 | 1.4 |
| Alachlor | 0.901 | 1.28 | 1.43 | 1.09 | 1.29 | 1.68 | 2.89 |
| alpha-Thujone | 0.584 | 1.08 | 1.12 | 1.11 | 1.12 | 1.11 | 1.17 |
| Alprazolam | 0.396 | 1.3 | 1.28 | 1.17 | 1.18 | 1.28 | 1.17 |
| Anthraquinone | 0.319 | 1.19 | 1.02 | 1.05 | 1.01 | 1.09 | 0.948 |
| Antipyrine | 0.0183 | 0.626 | 0.626 | 0.626 | 0.626 | 0.626 | 0.626 |
| Benzophenone | 0.383 | 0.637 | 0.673 | 0.663 | 0.659 | 0.655 | 0.622 |
| Bis 2-Chloroethoxy Methane | 0.483 | 1.44 | 1.68 | 1.68 | 1.74 | 1.7 | 1.72 |
| Bisphenol A | 0.127 | 0.614 | 0.401 | 0.419 | 0.348 | 0.469 | 0.36 |
| Boscalid | 0.247 | 0.902 | 0.671 | 0.695 | 0.7 | 0.694 | 0.652 |
| Bosentan | 0.396 | 0.766 | 0.666 | 0.432 | 0.669 | 0.445 | 0.746 |
| Bromochloroacetic acid | 0.632 | 1.08 | 1.24 | 1.24 | 1.2 | 1.33 | 1.3 |
| Bromodichloromethane | 0.634 | 0.643 | 0.768 | 0.768 | 0.797 | 0.917 | 1.02 |
| Carbaryl | 0.141 | 0.627 | 0.608 | 0.488 | 0.46 | 0.71 | 0.553 |
| Carbendazim | 0.303 | 2.06 | 2.12 | 2.15 | 2.15 | 2.07 | 1.49 |
| Chloridazon | 1.24 | 2.44 | 2.6 | 2.32 | 2.6 | 2.8 | 1.17 |
| Chlorpyrifos | 0.803 | 2.09 | 2.09 | 2.46 | 2.43 | 2.11 | 2.15 |
| Cyclanilide | 0.552 | 1.5 | 2.3 | 2.03 | 2.26 | 2.39 | 1.82 |
| Cyclosporin A | 0.154 | 0.894 | 0.879 | 0.688 | 0.596 | 0.655 | 0.811 |
| Di-n-butyl phthalate | 0.32 | 0.368 | 0.35 | 0.34 | 0.341 | 0.36 | 0.429 |
| Diazinon-o-analog | 0.737 | 3.19 | 2.98 | 3.02 | 3.07 | 2.93 | 0.933 |
| dibromoacetic acid | 0.57 | 1.02 | 1.43 | 1.43 | 1.39 | 1.48 | 1.54 |
| dichloroacetic acid | 0.604 | 0.929 | 0.939 | 0.939 | 0.924 | 0.953 | 0.947 |
| Dimethenamid | 0.756 | 0.983 | 0.994 | 0.994 | 0.994 | 0.996 | 1.2 |
| DL-Camphor | 0.396 | 0.56 | 1.24 | 1.34 | 1.39 | 1.19 | 0.998 |
| Emodin | 0.426 | 0.342 | 0.691 | 0.608 | 0.705 | 0.434 | 0.856 |
| Etoxazole | 0.525 | 1.12 | 0.891 | 0.874 | 0.874 | 0.849 | 0.837 |
| Fenarimol | 0.305 | 0.69 | 0.684 | 0.875 | 0.862 | 0.85 | 0.674 |
| Flufenacet | 1.27 | 1.49 | 1.34 | 1.29 | 1.35 | 1.34 | 1.23 |
| Formamide | 0.326 | 1.2 | 1.31 | 1.31 | 1.31 | 1.33 | 1.27 |
| Free Carbon disulfide | 0.213 | 0.247 | 0.325 | 0.325 | 0.227 | 0.446 | 1.15 |
| Gemfibrozil | 0.683 | 0.715 | 0.82 | 0.942 | 0.808 | 0.734 | 0.761 |
| Hexachlorobenzene | 0.315 | 0.333 | 0.647 | 0.643 | 0.643 | 0.7 | 0.743 |
| Ibuprofen | 0.272 | 0.722 | 1.18 | 0.858 | 0.993 | 0.801 | 1.44 |
| Imazalil | 0.245 | 0.961 | 0.659 | 1.44 | 1 | 1.4 | 1.23 |
| Imipramine | 0.307 | 0.649 | 0.638 | 0.626 | 0.642 | 0.753 | 0.627 |
| Isoeugenol | 0.419 | 0.837 | 0.924 | 0.905 | 0.9 | 0.844 | 1.29 |
| methanol | 0.0488 | 0.689 | 0.689 | 0.689 | 0.689 | 0.689 | 0.689 |
| methyl tert-butyl ether | 0.307 | 0.427 | 0.768 | 0.768 | 0.702 | 0.768 | 0.768 |
| methylene chloride | 0.274 | 0.564 | 1.04 | 1.1 | 0.935 | 1.07 | 1.13 |
| Midazolam | 0.373 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 2.42 |
| Nilvadipine | 0.633 | 1.13 | 0.842 | 1.33 | 0.91 | 0.857 | 0.899 |
| Novaluron | 0.281 | 1.12 | 1.12 | 0.807 | 0.807 | 0.855 | 1.25 |
| Oxazepam | 0.249 | 0.795 | 0.813 | 0.88 | 0.815 | 0.823 | 1.13 |
| Oxymetholone | 0.262 | 0.759 | 0.622 | 0.617 | 0.636 | 0.609 | 0.614 |
| Pentachlorophenol, purified | 0.332 | 0.565 | 0.825 | 0.885 | 0.868 | 0.892 | 0.867 |
| Perfluorodecanoic acid | 0.313 | 0.919 | 1.02 | 0.858 | 0.903 | 1.11 | 1.22 |
| perfluorohexane-1-sulphonic acid – potassium salt | 0.834 | 2.04 | 2.35 | 1.35 | 2.23 | 2.51 | 1.5 |
| Perfluorohexanoic Acid | 0.937 | 1.15 | 1.34 | 0.834 | 0.799 | 1.66 | 2.08 |
| Permethrin | 0.352 | 1.01 | 1.12 | 1.08 | 1 | 1.09 | 1.43 |
| phenacetin | 0.7 | 0.861 | 0.9 | 0.97 | 0.94 | 0.995 | 0.893 |
| Phenytoin | 0.576 | 0.395 | 0.841 | 0.481 | 0.462 | 0.67 | 0.494 |
| Potassium Perfluorobutane Sulfonate | 0.56 | 1.61 | 1.86 | 1.31 | 1.32 | 2.3 | 2.25 |
| Primidone | 0.318 | 1.96 | 1.94 | 1.96 | 1.66 | 1.94 | 1.61 |
| Propamocarb hydrochloride | 0.414 | 1.13 | 1.93 | 1.01 | 1.01 | 1.02 | 1.06 |
| propylparaben | 0.602 | 0.891 | 0.645 | 0.752 | 0.757 | 0.713 | 0.465 |
| Propyzamide | 0.62 | 1.42 | 1.52 | 1.05 | 1.06 | 1.08 | 2.09 |
| Pyridine | 0.445 | 1.28 | 1.35 | 1.35 | 1.37 | 1.33 | 0.958 |
| Pyrithiobac sodium | 0.937 | 1.25 | 2.27 | 0.75 | 0.647 | 2.58 | 2.66 |
| Resmethrin | 0.316 | 0.833 | 1.06 | 0.694 | 0.694 | 1.18 | 1.49 |
| S-Bioallethrin | 0.393 | 0.928 | 1.07 | 2.29 | 1.22 | 1.25 | 1.01 |
| Simazine | 0.455 | 0.836 | 0.636 | 0.639 | 0.717 | 0.637 | 0.717 |
| Tamoxifen | 0.347 | 0.938 | 1.83 | 4.24 | 3.99 | 2.15 | 2.68 |
| tert-amyl methyl ether | 0.0947 | 0.348 | 0.847 | 0.847 | 0.7 | 0.847 | 0.847 |
| tetrachloroethylene | 0.183 | 0.369 | 0.876 | 0.38 | 0.357 | 1.27 | 1.39 |
| Tetralin | 0.689 | 0.575 | 0.981 | 1.29 | 1.58 | 0.802 | 0.496 |
| trichloroethylene | 0.476 | 0.436 | 0.689 | 0.848 | 0.427 | 0.805 | 0.983 |
| Valproic acid | 0.434 | 0.881 | 0.971 | 0.969 | 0.843 | 1.37 | 1.6 |
| Wyeth-14643 | 0.458 | 0.503 | 0.764 | 0.555 | 0.628 | 0.875 | 0.95 |

Supplemental Table 8: Level 3 Predictions

*SupTable-Level3.xlsx*