ORD MANUSCRIPT COVER SHEET

**Title:** Great Title about *invivoPKfit* and the CvTdb

**Authors:** Christopher Cook, Lucas Albrecht, Grace Cary, Brenda Edwards, Derik Haggard, Nancy Hanley, Michael F. Hughes, Anna Jarnagin, Tirumala D. Kodavanti, Evgenia Koral-Bexell, Anna Kreutz, Mayla Ngo, Caroline Ring, Caitlyn Patullo, Risa Sayre, Bhaskar Sharma, Jonathan Wall, Hiroshi Yamazaki, and John Wambaugh

Labs/Centers: CCTE

**External collaborators:**

* Showa Pharmaceutical University

**Internal Peer Reviewers:** TBD

**Target Journal:** Possibly: Computational Toxicology, Journal of Pharmacokinetics and Pharmacodynamics, Journal of Statistical Software

**Four bullet point summary:**

* words
* words
* words
* words

**One sentence description:** This collaborative trial demonstrates that …

**Chemicals Involved:** ~100 CvTdb chemicals with oral and/or intravenous administration routes

**Great Title about *invivoPKfit* and the CvTdb**

Christopher Cook1,2, Lucas Albrecht1, Grace Cary1,2, Brenda Edwards1, Derik Haggard1, Nancy Hanley1, Michael F. Hughes1, Anna Jarnagin1,2, Tirumala D. Kodavanti1,2, Evgenia Koral-Bexell1,2, Anna Kreutz1,2, Mayla Ngo1,2, Caroline Ring1, Caitlyn Patullo1,2, Risa Sayre1,3, Bhaskar Sharma1, Jonathan Wall1, Hiroshi Yamazaki4, and John Wambaugh1

1. Center for Computational Toxicology and Exposure, Office of Research and Development, United States Environmental Protection Agency, Durham, North Carolina 28311
2. Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee 38331
3. Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
4. Showa Pharmaceutical University, Machida, Tokyo 194-8543, Japan

\*Corresponding Author: cook.christopher@epa.gov

# Abstract

The Concentration versus Time Database (CvTdb) is a public repository of toxicokinetic (TK) data developed by the Center for Computational Toxicology and Exposure at the U.S. Environmental Protection Agency (EPA). The CvTdb contains manually curated experimental time-course data of chemical compound concentrations in body fluids and tissues, along with descriptive metadata, from hundreds of publications. The CvTdb is a “living” repository, to which new data can be added as they become available. Recent efforts have been made to 1) Expand the scope of in-house data curation and 2) Integrate sets of CvT data shared by other institutions. To standardize the format of data extracted from numerous publications, by several curators, we developed an R Shiny application that checks completed template curation files for both format and data correctness. New data sets include those from collaborators of Showa Pharmaceutical University with more than 200 compounds. Standardized data reporting allows for analysis of TK trends across studies. For example, we observe that among replicate observations (those with common chemical/reference/dose/route/timepoint data), 85.9% are within two-fold of the mean concentration. When the data are subset by route, we observe that 84.0% of oral administration observations, and 89.7% of intravenous administration observations, are within two-fold of the mean concentration. Data are systematically analyzed using invivoPKfit, an R package that fits standardized 1- and 2- compartmental TK models to all data associated with a particular compound, including data that spans multiple references. We used invivoPKfit to fit models and estimate TK parameters, such as volume of distribution and elimination half-life. Overall, the CvTdb serves as a platform for assessing TK trends across a large and standardized set of data, as well as for calibrating and validating TK models.

# Introduction

## 1a) What is the CvTdb?

The Concentration versus Time Database (CvTdb) is a database of standardized time-course data of chemical compound concentrations measured in vivo from hundreds of diverse publications. It serves as a platform for assessing PK/TK models on a large set of data. (Sayre, Grulke, and Wambaugh, 2020).

The CvTdb contains 389 unique chemicals, including 70 pharmaceuticals (~30%). 101 analytes are present in EPA’s ToxCast dataset (~53%).

1b) Integration of Data Shared by Other Institutions

Showa Pharmaceutical University shared CvT data consisting of over 200 unique compounds (Kamiya et al., 2020; Kamiya et al., 2021).

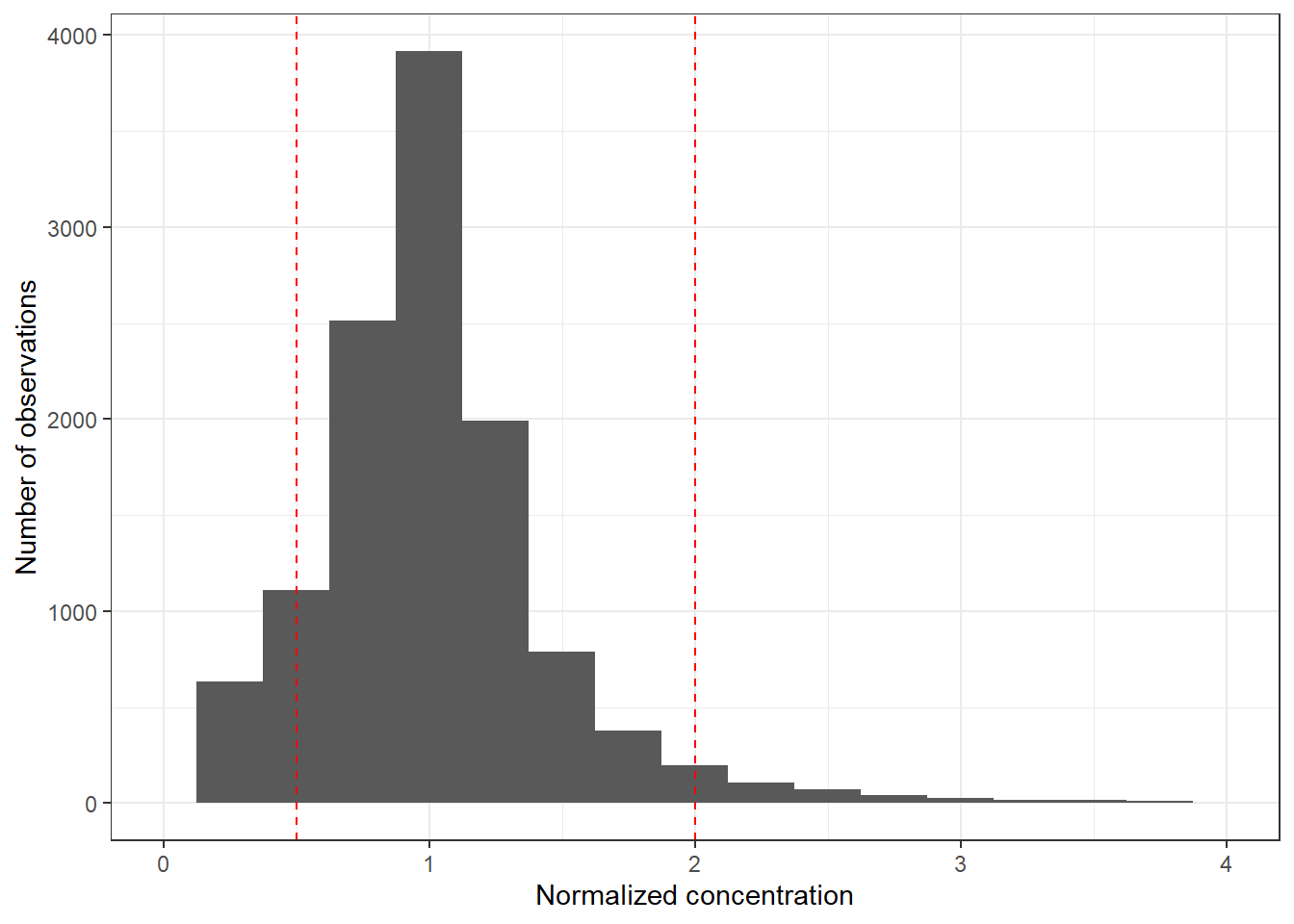


Figure 1 The CvTdb data was grouped into unique combinations of compound/reference/route/dose/timepoint. Groups without replicate timepoints were removed. If all replicate timepoints corresponded to non-detect concentration values, those data were removed. If some, but not all replicate timepoints corresponded to non-detect concentration values, those values were set to equal 0.45 of the minimum value. Each concentration was normalized by the mean concentration of each group.

If YOU have CvT data, please consider collaborating with us! Visit https://github.com/USEPA/CompTox-PK-CvTdb or email Wambaugh.John@epa.gov

## General

* Toxicokinetics (TK) describes the absorption, distribution, metabolism, and excretion (ADME) of a chemical compound in the body as a function of time
* TK allows for the prediction of internal tissue concentrations as a function of chemical exposure and provides critical information for the assessment of a risk posed by a chemical to public health
* This data is referred to as “concentration *versus* time” (CvT) data
* TK involves mathematically describing CvT data such that extrapolation is possible

## CvTdb and *invivoPKfit* can Help

* There is a need for a standardized and reproducible approach to fitting pharmacokinetic models
* A standardized approach would make one data set for a chemical intercomparable with other data sets
* We address this problem with two tools – the Concentration *versus* Time Database (CvTdb) and the R package, *invivoPKfit*
* The CvTdb is a public repository of standardized TK data developed by the Center for Computational Toxicology and Exposure (CCTE) at the U.S. Environmental Protection Agency (EPA)
* *invivoPKfit* is an R package developed by CCTE to fit generic 1- and 2-compartment models to the diverse set of data within the CvTdb
* We estimate model parameters for the CvTdb using *invivoPKfit* and compare them to those reported in Lombardo, which compiles parameters from models tailored to specific sets of data

In Figure 2, 85.1% of replicate timepoint data in the CvTdb is within a factor of two of their respective mean. The “factor of two” rule may not be appropriate if the data itself is spread beyond a factor of two.

## Review on other packages

* There are other software packages to fit models to CvT data, notably *PK*, *pmxTools*, and *PKconverter*
* *invivoPKfit* differentiates itself from these other tools by allowing for meta-analyses between data of the same chemical but from different studies…
* Maybe a table describing them?

# Methods

What is invivoPKfit?

R package that fits 1- and 2- compartment models to CvT data and extrapolates TK parameter values (e.g., half-life, elimination rate, and volume of distribution)

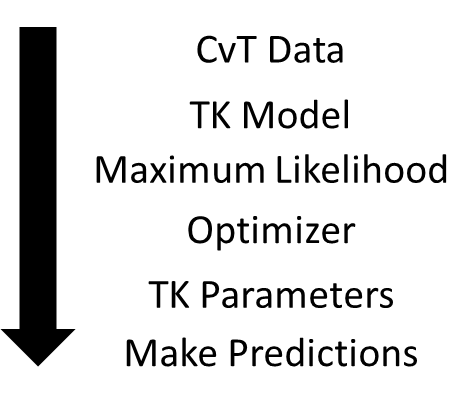


Figure 2 invivoPKfit workflow.

## *invivoPKfit* Models

**1-Compartment Model**

)Civ(dose, t) = doseVde(−kelim ∗ t)

Coral(dose, t) = Fbio ∗ doseVd ∗ kabs ∗ (e−kelim ∗ t − e−kabs ∗ t)kabs − kelimCoral(dose, t) = Fbio ∗ doseVd ∗ kabs ∗ (e−kelim ∗ t − e−kabs ∗ t)kabs − kelim

Parameters include VdVd (volume of distribution), kelimkelim (elimination rate), kabskabs (absorption rate; oral only), and FbioFbio (fraction bioavailable; oral only).

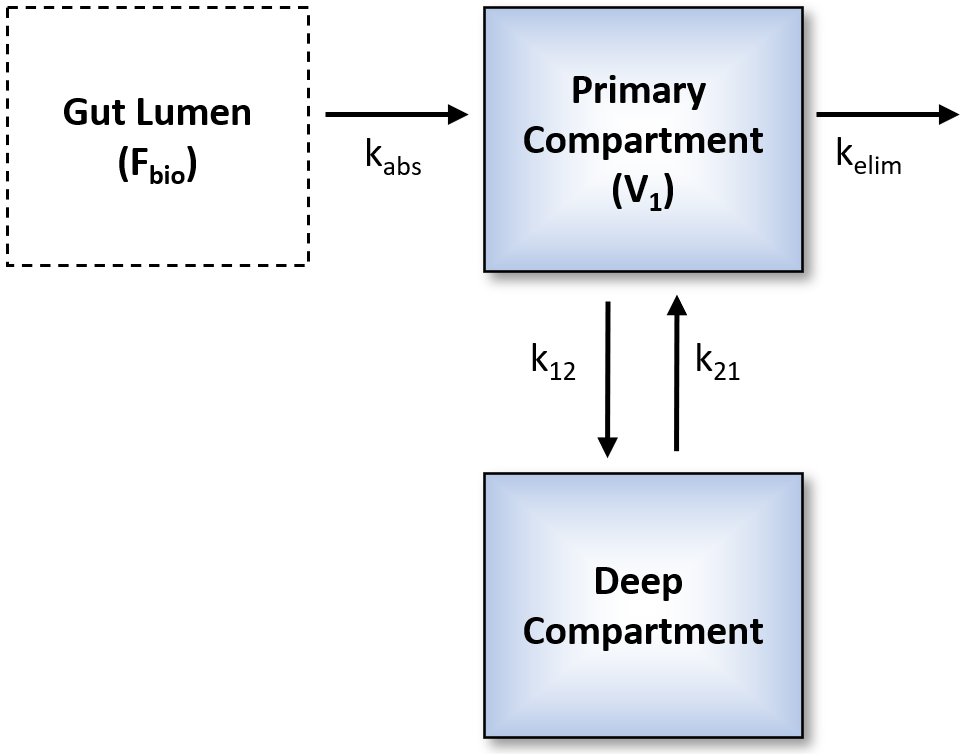
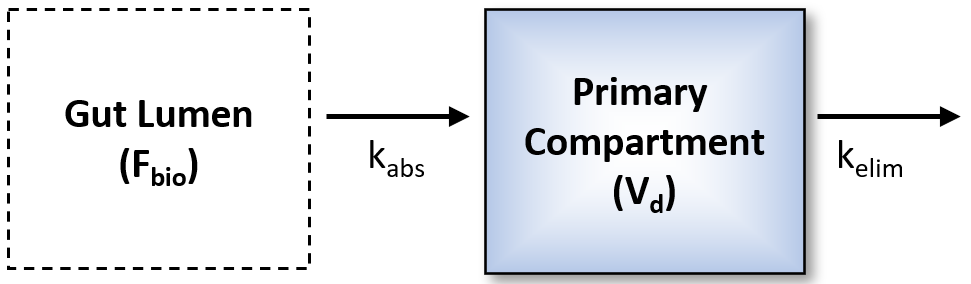


Figure 3 A) 1-Compartment Model. Adapted from Wambaugh et al. (2018). B) 2-Compartment Model. Adapted from Wambaugh et al. (2018).

* *invivoPKfit* fits generic 1- and 2- compartment models to all data
* These were copied from the Boomer page, but can be cited from

**2-Compartment Model**

* *Parameters include V1V1 (volume of primary compartment), kelimkelim, k12k12 (distribution rate to deep compartment), k21k21 distribution rate from deep compartment), kabskabs (oral only), FbioFbio (oral only).*

## Statistical Model Description

## invivoPKfit calculates the probability of Cobs given a PK model M(p, t) and a statistical model S(sigma), where P = L(Cobs, M. S), and assumes that Cobs are log-normally distributed around theconcentrations predicted by M(p,t) (Cox and Hinkley, 1979).

## 

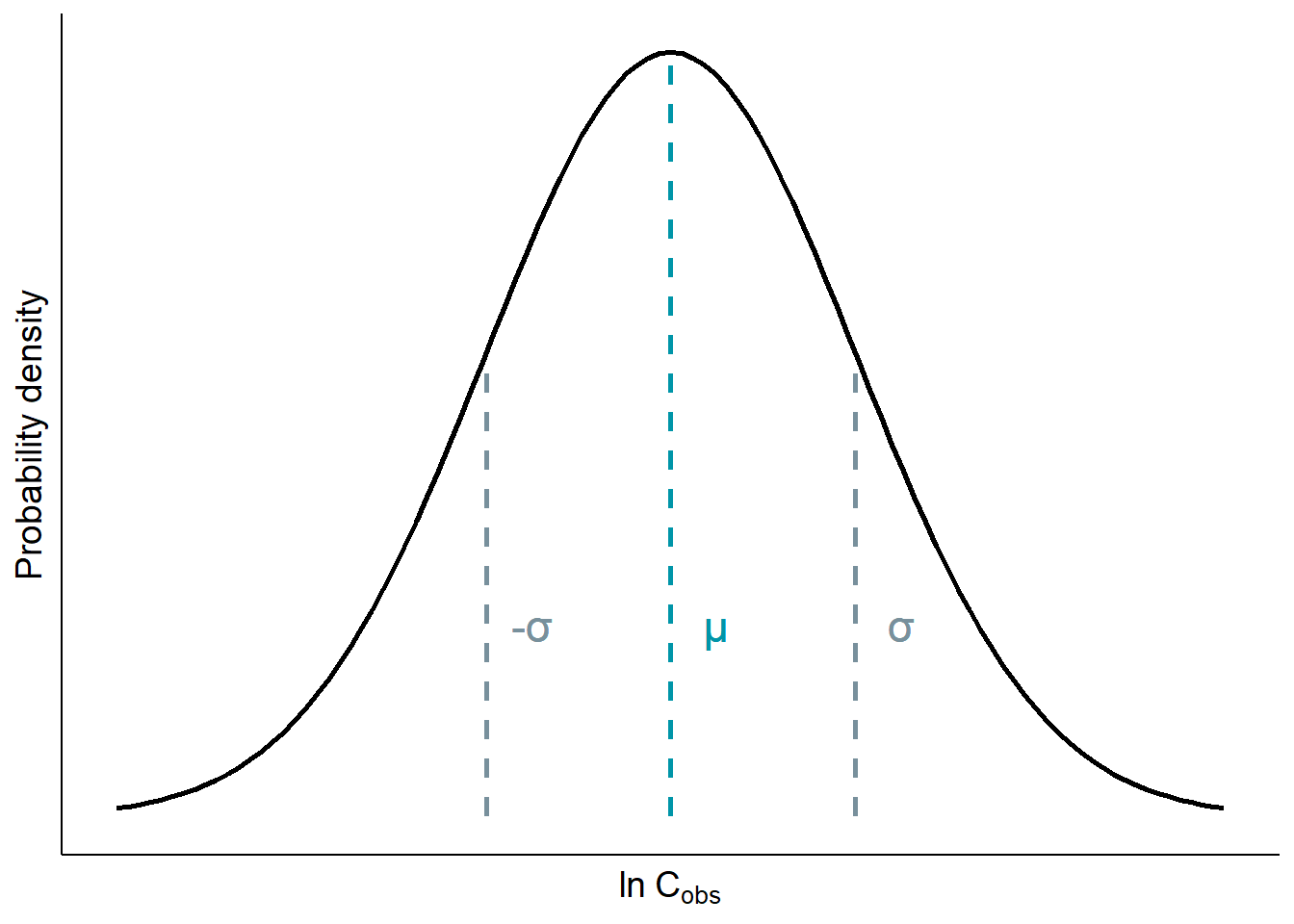


Figure 4 Log-Normal Distribution

## How *invivoPKfit* Finds Solutions to Models

* *invivoPKfit* estimates the most likely parameter values for generic 1- and 2- compartment models describing the TK of a compound given CvT observations
* We use the R package *optimx* to optimize the log-likelihood equation and estimates the parameters and sigma
* Confidence intervals on the estimated parameters were calculated using the Hessian of the likelihood function to estimate standard deviations of the log parameter (Bartlett, 1953a,b).
* The R package, optimx (Nash and Varadhan, 2011; Nash, 2014) optimizes the log-likelihood equation and estimates pp and σσ.
* We used the Hessian of the likelihood function to estimate standard deviations of the log parameter (Dovi, 1991).

## *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 each chemical 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>). Between the one- and two-compartment models, the one with the lower Akaike Information Criterion (AIC) value – indicating model parsimony – was selected [37]. 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.

# Results

## Evaluation Chemicals and Predictions

There were 102 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.

Of the 102 chemicals, there were 10 chemicals that could only be predicted by OPERA and were omitted from the rest of the analysis (Supplemental Table 5). Two chemicals, Oxoacetic acid--water (1/1) and Nitrite have CvT data but do not have either measured or predicted values for both Clint and fup from any model to date.

The data for each chemical was fit using maximum likelihood estimation to one and two compartment empirical pharmacokinetic models, with separate fits for each combination of compound and species for which there were data. Maximum likelihood estimates could not be obtained for either model for 8 chemicals (listed in Supplemental Table 4). These chemicals were withheld from subsequent analysis. For each remaining chemical the better of the one or two compartment models was used on the basis of model parsimony.

Eliminating chemicals where the CvT data could not be described by an empirical model or for which there were only one QSPR that could make predictions left 83 chemicals with *in vivo* CvT data and 68 chemicals with *in vitro* measure fup and Clint.

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 only 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.

## Toxicokinetic Trends Across Chemicals

We can explore trends across chemicals and references using the CvTdb.

“In PBPK modeling, predictions that are, on average, within a factor of 2 of the experimental data have frequently been considered adequate” (WHO, 2010).

## 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 Table 4, 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 Table 4, 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 Table 4 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.

## **Predicted Concentrations**

63.3% of 16,675 concentrations predicted by the 1-compartment model fit within a factor of two of the observed concentrations.

69.1% of 13,097 concentrations predicted by the 2-compartment model fit within a factor of two of the observed concentrations.

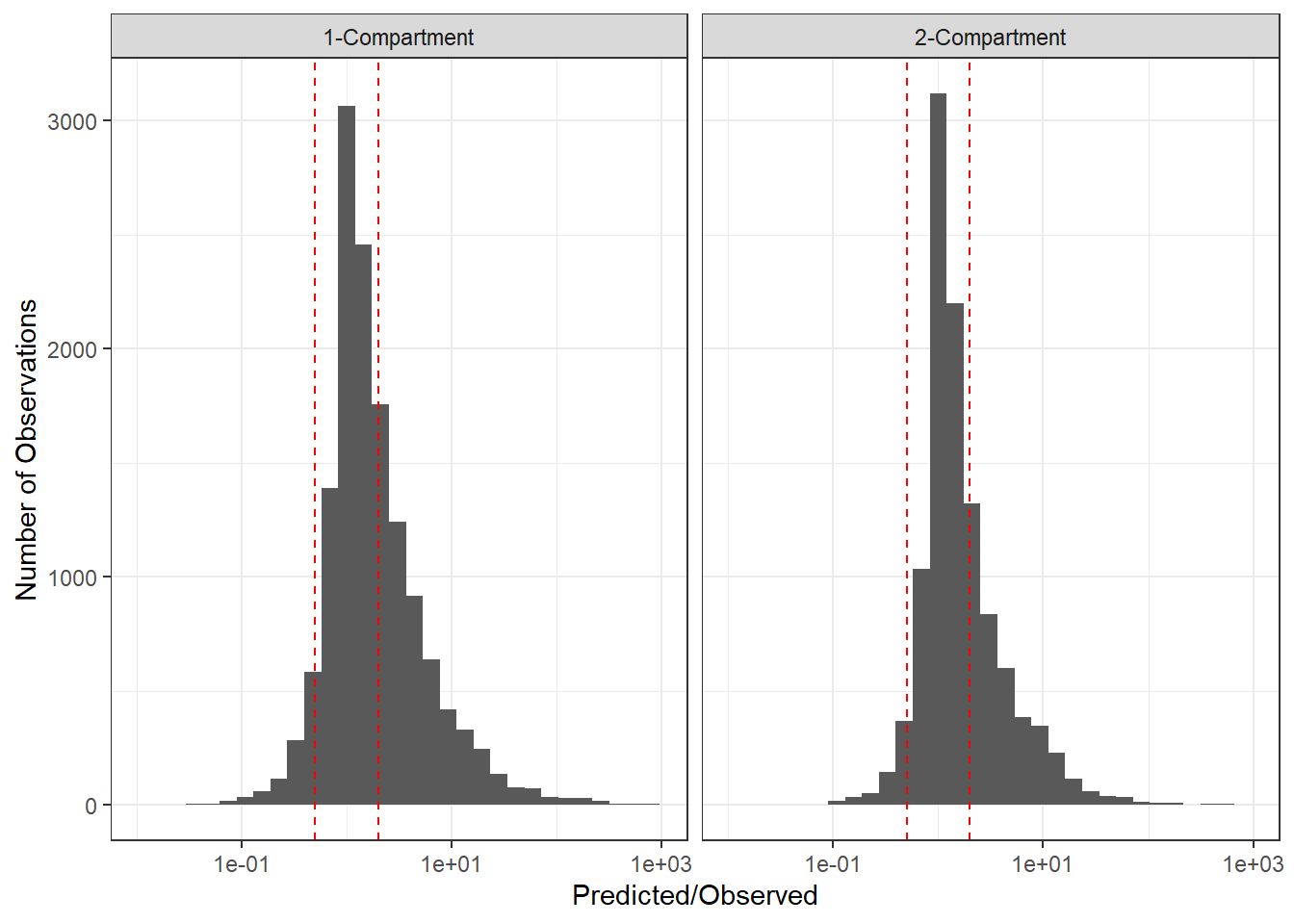


Figure 5 Concentrations were predicted using the estimated model parameters. Predicted concentrations were normalized by observed concentrations, such that Cpred / Cobs . The red dashed lines mark the ‘factor of two’ bounds. This plot does not include ‘Joint Analysis’ fits.

## invivoPKfit**Performance**versus**Data Variability**

Can start to gauge how data variability influences the performance of a model

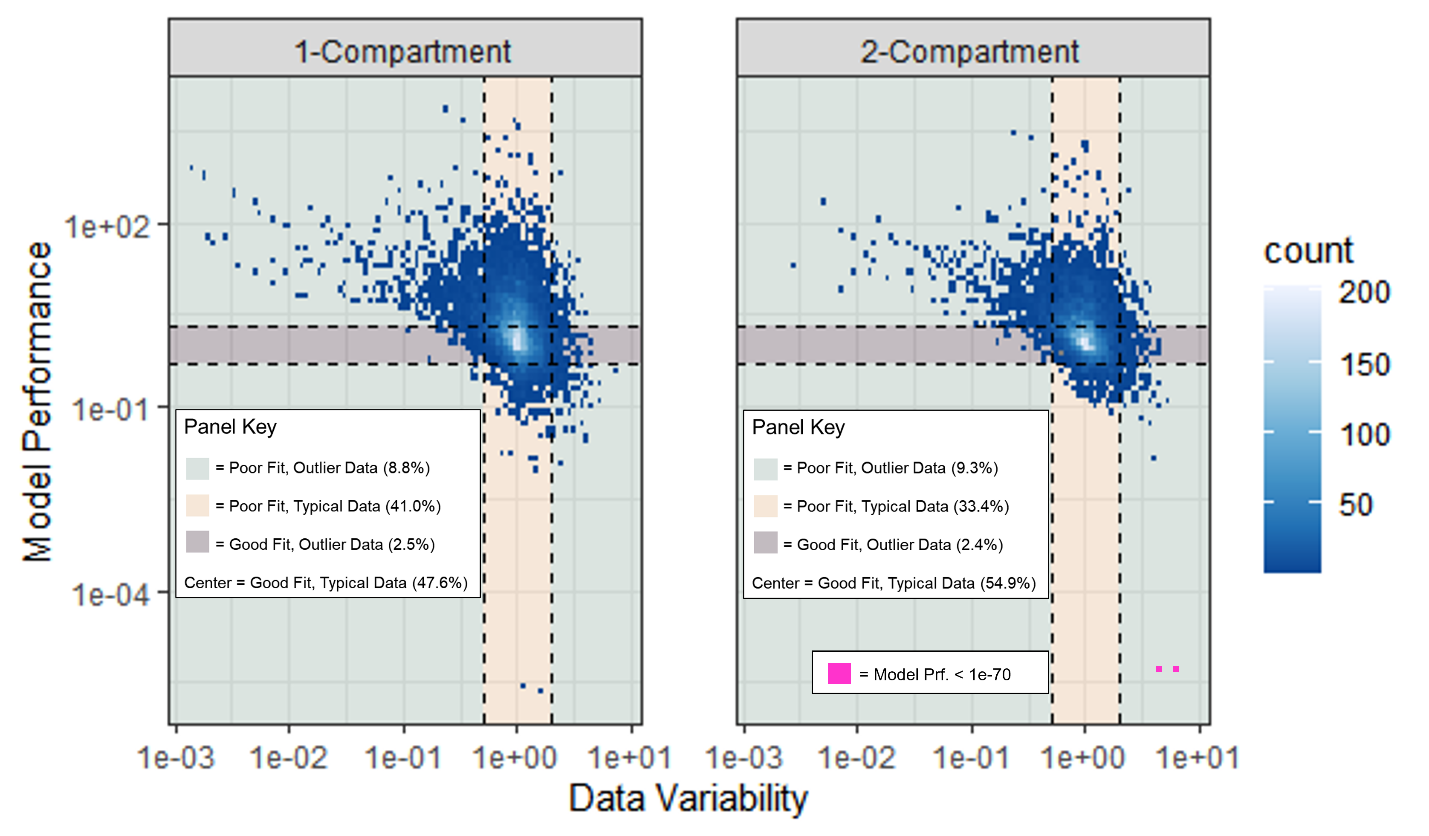


Figure 6 Predicted concentrations were normalized by observed concentrations (see Figure 7), and plotted against observed concentrations normalized by experiment-timpoint specific replicates (see Figure 2). The dashed lines mark the ‘factor of 2’ bounds for each axis. Sections were assigned based on both the ability of a model to accurately predict an observed concentration and the spread of the observed data. Points colored pink represent concentrations significantly underpredicted by the model.

## **Diagnosing Poor Fits**

Plotting model predictions against observed data can help determine why a model might fit poorly. A poor fit could indicate 1) a deficiency in the model, 2) an inaccurate curation of data, and/or 3) a deficiency in the experimental collection of data.

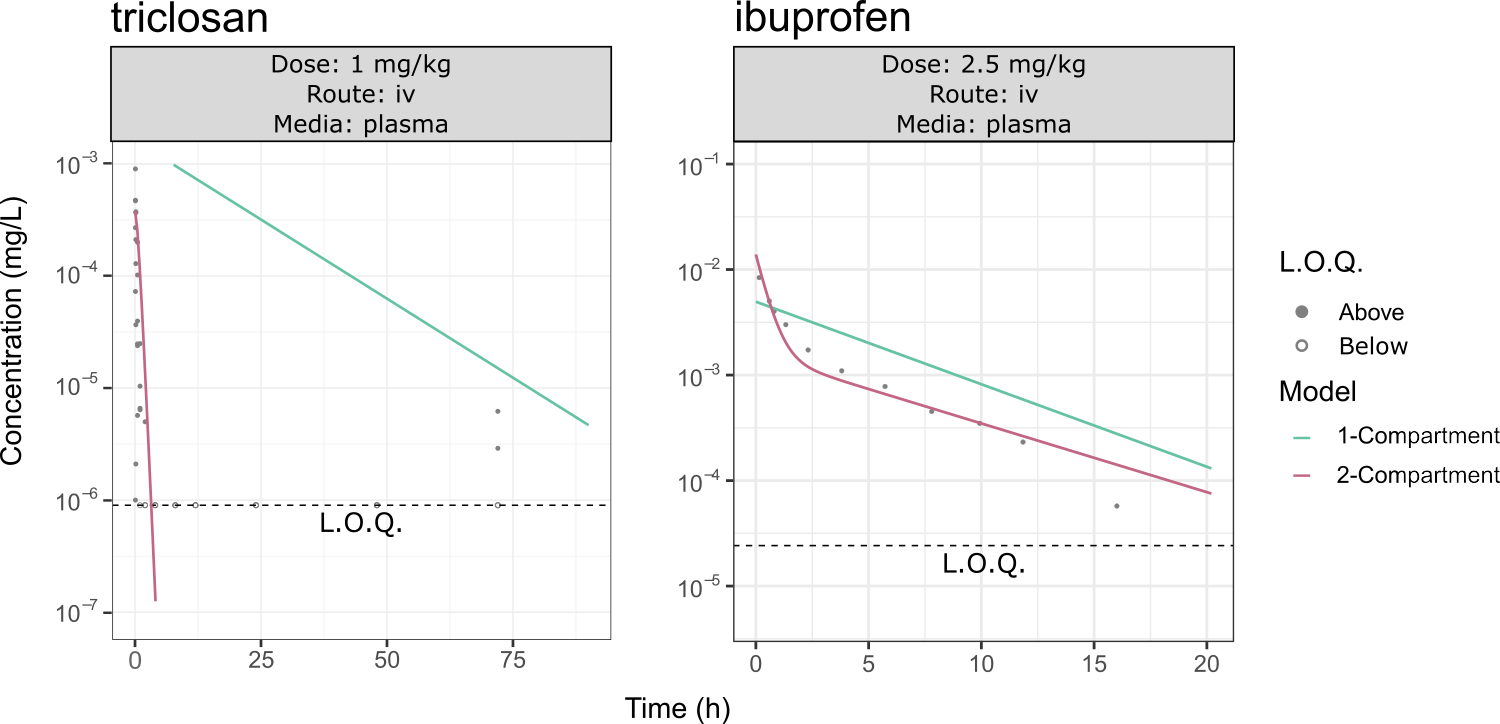


Figure 7 1-compartment model (green line) and 2-compartment model (red line) fit through CvT data (gray points) Open points represent non-detects that were below the limit of quantitation (L.O.Q.).

# Discussion

# References

Boxenbaum, Harold G., Sidney Riegelman, and Robert M. Elashoff. 1974. “Statistical Estimations in Pharmacokinetics.” Journal of Pharmacokinetics and Biopharmaceutics 2 (2): 123–48. https://doi.org/10.1007/BF01061504.

Cox, D. R., and D. V. Hinkley. 1979. Theoretical Statistics. CRC Press.

Dovi, V. G., O. Paladino, and A. P. Reverberi. 1991. “Some Remarks on the Use of the Inverse Hessian Matrix of the Likelihood Function in the Estimation of Statistical Properties of Parameters.” Applied Mathematics Letters 4 (1): 87–90. https://doi.org/https://doi.org/10.1016/0893-9659(91)90129-J.

Kamiya, Yusuke, Kentaro Handa, Tomonori Miura, Junya Ohori, Makiko Shimizu, Masato Kitajima, Fumiaki Shono, Kimito Funatsu, and Hiroshi Yamazaki. 2021. “An Updated in Silico Prediction Method for Volumes of Systemic Circulation of 323 Disparate Chemicals for Use in Physiologically Based Pharmacokinetic Models to Estimate Plasma and Tissue Concentrations After Oral Doses in Rats.” Chemical Research in Toxicology 34 (10): 2180–3. https://doi.org/10.1021/acs.chemrestox.1c00249.

Kamiya, Yusuke, Shohei Otsuka, Tomonori Miura, Manae Yoshizawa, Ayane Nakano, Miyu Iwasaki, Yui Kobayashi, et al. 2020. “Physiologically Based Pharmacokinetic Models Predicting Renal and Hepatic Concentrations of Industrial Chemicals After Virtual Oral Doses in Rats.” Chemical Research in Toxicology 33 (7): 1736–51. https://doi.org/10.1021/acs.chemrestox.0c00009.

Kudo, Naomi, Masanori Katakura, Yasunori Sato, and Yoichi Kawashima. 2002. “Sex Hormone-Regulated Renal Transport of Perfluorooctanoic Acid.” Chemico-Biological Interactions 139 (3): 301–16. https://doi.org/10.1016/s0009-2797(02)00006-6.

Nash, John C. 2014. “On Best Practice Optimization Methods in R.” Journal of Statistical Software 60 (2): 1–14. https://doi.org/10.18637/jss.v060.i02.

Nash, John C., and Ravi Varadhan. 2011. “Unifying Optimization Algorithms to Aid Software System Users: optimx for R.” Journal of Statistical Software 43 (9): 1–14. https://doi.org/10.18637/jss.v043.i09.

O’Flaherty, E. J. 1981. Toxicants and Drugs: Kinetics and Dynamics. John Wiley & Sons.

Sayre, Risa R., John F. Wambaugh, and Christopher M. Grulke. 2020. “Database of Pharmacokinetic Time-Series Data and Parameters for 144 Environmental Chemicals.” Scientific Data 7 (1): 122. https://doi.org/10.1038/s41597-020-0455-1.

Sheiner, L. B. 1985. “Analysis of Pharmacokinetic Data Using Parametric Models. II. Point Estimates of an Individual’s Parameters.” Journal of Pharmacokinetics and Biopharmaceutics 13 (5): 515–40. https://doi.org/10.1007/BF01059333.

Thorne, W. Brent. 2019. Posterdown: An R Package Built to Generate Reproducible Conference Posters for the Academic and Professional World Where Powerpoint and Pages Just Won’t Cut It. https://github.com/brentthorne/posterdown.

Wakefield, Jon, Leon Aarons, and Amy Racine. 1999. “The Bayesian Approach to Population Pharmacokinetic/Pharmacodynamic Modeling.” https://doi.org/10.1007/978-1-4612-1502-8\_4.

Wambaugh, John F., Michael F. Hughes, Caroline L. Ring, Denise K. MacMillan, Jermaine Ford, Timothy R. Fennell, Sherry R. Black, et al. 2018. “Evaluating in Vitro-in Vivo Extrapolation of Toxicokinetics.” Toxicological Sciences : An Official Journal of the Society of Toxicology 163 (1): 152–69. https://doi.org/10.1093/toxsci/kfy020.

World Health Organization. 2010. “Characterization and Application of Physiologically Based Pharmacokinetic Models in Risk Assessment.” WHO Document Production Services, Geneva, Switzerland.

# 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

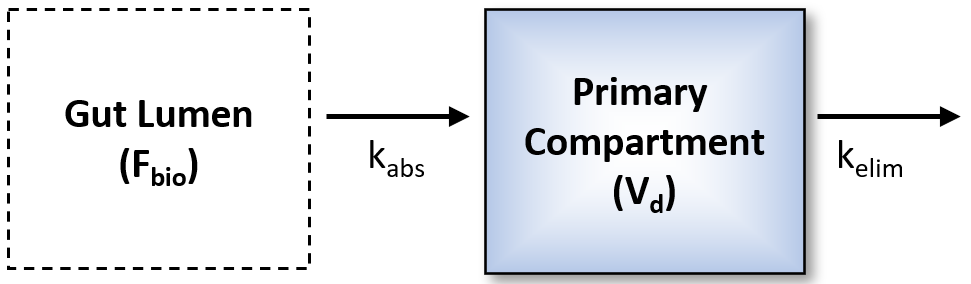


Figure 8: test

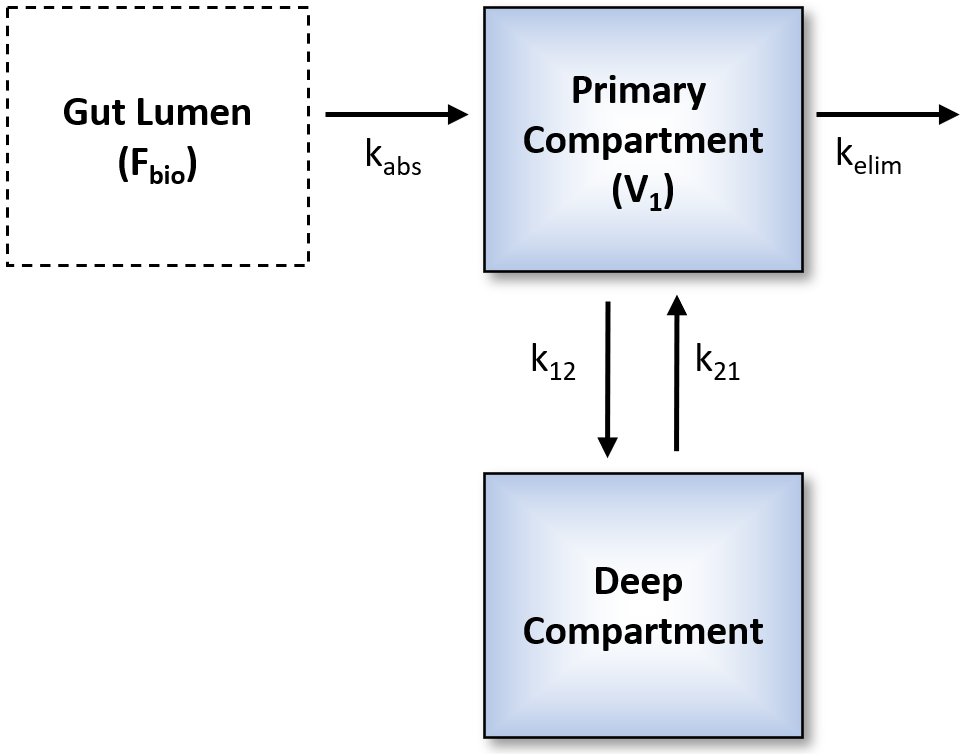


Figure 9: test

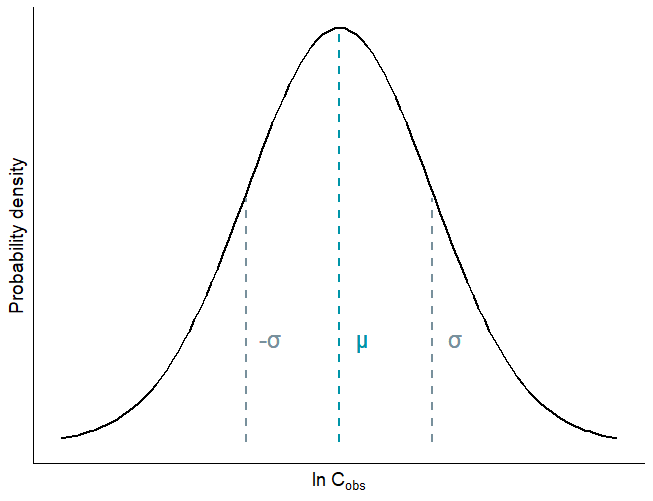


Figure 10: Log-normal distribution. Here are more details.

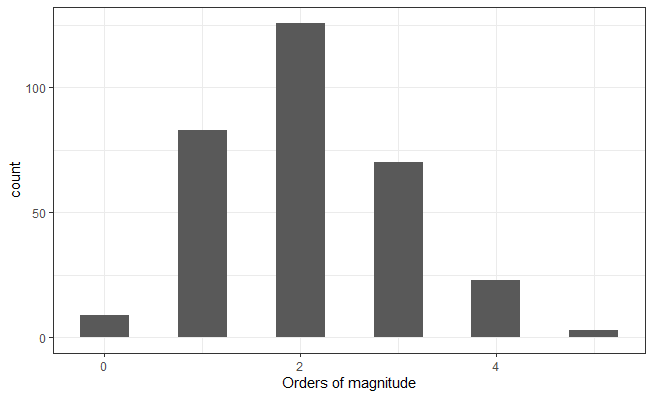


Figure 11: Orders of magnitude difference for individual series. The "2" average is a case for the taking the log-normal distribution, as otherwise there would be a bias toward higher numbers.

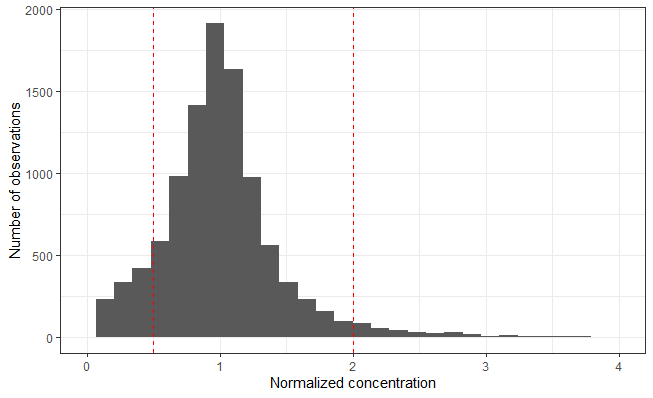


Figure 12: The CvTdb data was grouped into unique combinations of reference/route/dose/timepoint. Groups without replicate timepoints were removed. If all replicate timepoints corresponded to NA (non-detect) concentration values, those data were removed. If some, but not all replicate timepoints corresponded to NA concentration values, those NA concentration values were set equal to the limit of quantitation (LOQ). If the LOQ was unknown, it was assumed to equal 0.45 of minimum concentration value present. Each concentration value was normalized by the mean concentration value of each group. The red dashed lines mark the upper and lower bounds of data within two-fold of the mean (0.5 and 2).

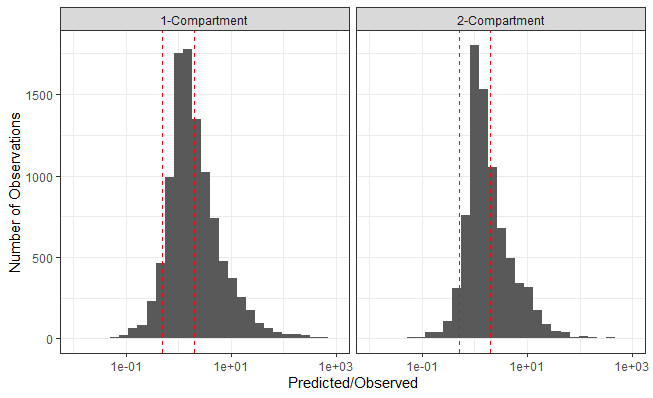


Figure 13: Concentrations were predicted by applying estimated model parameters to the timepoints present in the data. Predicted concentrations were normalized by observed concentrations at their corresponding timepoints. The red dashed lines mark the lower and upper bounds of predicted data within two-fold of observed data.

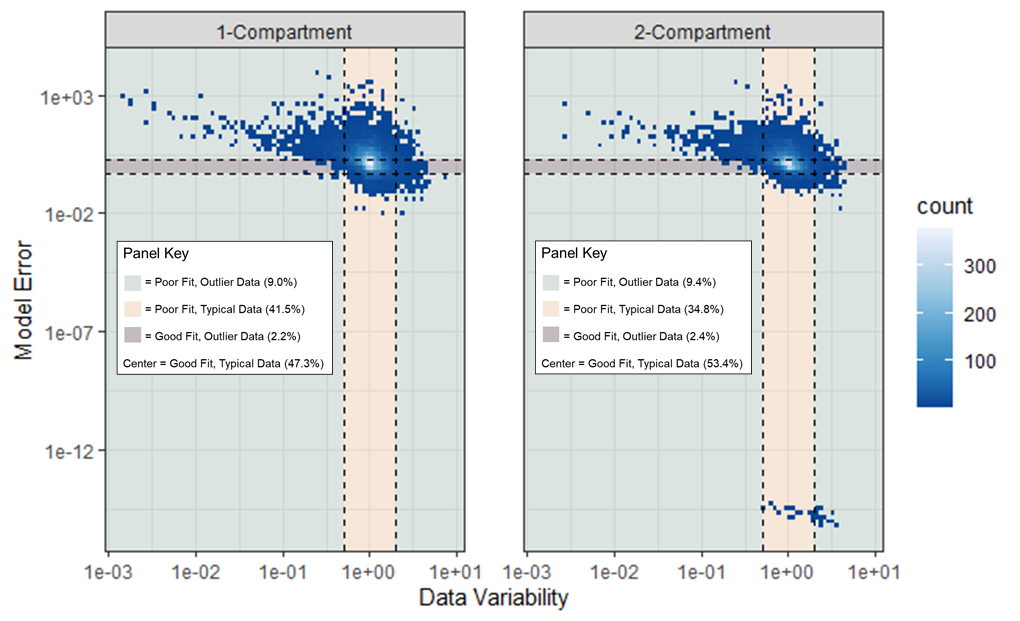
––– 

Figure 14: Density plot of Model Error (predicted concentration / observed concentration) versus Data Variability (concentration normalized by mean concentration of replicate timepoints; see Figure 3). Colored panels define the ‘goodness’ of both the data and the model, based on whether those points are within two-fold of the mean (x-axis) and two-fold of the observed concentration (y-axis). The “Center” panel (the densest area) represents data where points correspond to both a Data Variability within two-fold of the mean and a Model Error within two-fold of the observed concentrations. Under the 1-Compartment Model, 47.3% of points lie in the “Center” area. Under the 2-Compartment Model, 53.4% of points lie in the “Center” area.

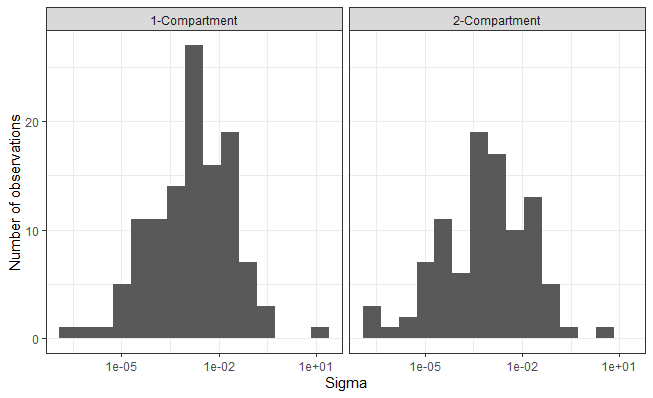


Figure 15:

# Tables

# Supplemental Tables

Supplemental Table 1: Empirical (one and two compartment model) toxicokinetic parameter estimates

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 | -279.8 | -275.9 | 1Comp | 3650 | 0.3431 | 2.02 |
| 1,4-dioxane | 123-91-1 | rat | 24, 6 | -367.4 | -368.1 | 2Comp | 658.8 | 0.02368 | 29.27 |
| 1-chloro-2-propanol | 127-00-4 | rat | 176 | -1391 | -1392 | 2Comp | 2382 | 1.993 | 0.3479 |
| 2,3,7,8-tetrachlorodibenzo-p-dioxin | 1746-01-6 | rat | 177 | -108.1 | -107 | 1Comp | 1452 | 0.001888 | 367.2 |
| 2,4-dichlorophenoxyacetic acid | 94-75-7 | rat | 164, 192 | -1670 | -1733 | 2Comp | 339.4 | 1.738 | 0.3989 |
| 2-hydroxy-4-methoxybenzophenone | 131-57-7 | rat | 129 | -4667 | -4683 | 2Comp | 14350 | 0.3336 | 2.078 |
| 2-methylimidazole | 693-98-1 | rat | 150 | -565.2 | -593 | 2Comp | 2440 | 0.02029 | 34.17 |
| 2-methyltetrahydrofuran | 96-47-9 | rat | 175 | -3240 | -3240 | 1Comp | 812.7 | 0.3836 | 1.807 |
| 4-methylimidazole | 822-36-6 | rat | 151 | -716.5 | -755.5 | 2Comp | 1274 | 0.7412 | 0.9351 |
| acrylonitrile | 107-13-1 | rat | 43 | -260.5 | -285.6 | 2Comp | 747.6 | 9.896 | 0.07004 |
| alachlor | 15972-60-8 | rat | 192 | -965.6 | -962.4 | 1Comp | 173200 | 0.07371 | 9.404 |
| alpha-thujone | 546-80-5 | rat | 130 | -2351 | -2390 | 2Comp | 10510 | 0.7298 | 0.9498 |
| alprazolam | 28981-97-7 | rat | 193 | -529.1 | NA | 1Comp | 2671 | 2.273 | 0.305 |
| anthraquinone | 84-65-1 | rat | 166 | -2502 | NA | 1Comp | 3786 | 0.133 | 5.21 |
| antipyrine | 60-80-0 | rat | 197 | -101.1 | -98.84 | 1Comp | 1133 | 0.5825 | 1.19 |
| benzo(a)pyrene | 50-32-8 | rat | 36, 38, 46, 59 | -777.5 | -884.6 | 2Comp | 3258 | 0.4873 | 1.422 |
| benzophenone | 119-61-9 | rat | 171 | -5427 | -5565 | 2Comp | 12870 | 0.2211 | 3.135 |
| bis 2-chloroethoxy methane | 111-91-1 | rat | 155 | -1090 | -1172 | 2Comp | 2323 | 3.444 | 0.2013 |
| bisphenol a | 80-05-7 | rat | 192 | -201.9 | -204.4 | 2Comp | 1929 | 4.372 | 0.1585 |
| boscalid | 188425-85-6 | rat | 192 | -798.1 | -855 | 2Comp | 11250 | 1.098 | 0.631 |
| bosentan | 147536-97-8 | rat | 198 | -225.5 | NA | 1Comp | 3156 | 0.2439 | 2.841 |
| bromochloroacetic acid | 5589-96-8 | rat | 132 | NA | -2067 | 2Comp | 3939 | 0.3587 | 1.932 |
| bromodichloromethane | 75-27-4 | rat | 154 | -923.6 | -977.7 | 2Comp | 19190 | 2.019 | 0.3433 |
| carbaryl | 63-25-2 | rat | 192 | -465.9 | -260.3 | 1Comp | 41640 | 4.485 | 0.1545 |
| carbendazim | 10605-21-7 | rat | 201 | -68.57 | -64.57 | 1Comp | 33150 | 0.245 | 2.829 |
| chloridazon | 1698-60-8 | rat | 192 | -893.7 | NA | 1Comp | 6923 | 0.08218 | 8.434 |
| chlorpyrifos | 2921-88-2 | rat | 202 | -364.5 | -366.8 | 2Comp | 43220 | 0.5533 | 1.253 |
| cyclanilide | 113136-77-9 | rat | 192 | -1223 | -1277 | 2Comp | 361.7 | 0.2194 | 3.159 |
| cyclosporin a | 59865-13-3 | rat | 203, 204 | -271.8 | -308.3 | 2Comp | 1605 | 0.3603 | 1.924 |
| diazinon-o-analog | 962-58-3 | rat | 192 | -884.7 | -904.1 | 2Comp | 271100 | 8.04E-05 | 8619 |
| dichloroacetic acid | 79-43-6 | rat | 134 | -699.9 | -699.6 | 1Comp | 528.2 | 1.767 | 0.3923 |
| diltiazem | 34933-06-7 | rat | 207 | -236.3 | -258.5 | 2Comp | 2745 | 3.392 | 0.2044 |
| dimethenamid | 87674-68-8 | rat | 192 | -911.5 | -964 | 2Comp | 159600 | 0.8059 | 0.8601 |
| di-n-butyl phthalate | 84-74-2 | rat | 135 | -49.46 | -45.46 | 1Comp | 1916 | 5.349 | 0.1296 |
| dl-camphor | 76-22-2 | rat | 165 | -851.9 | -895.5 | 2Comp | 6283 | 3.609 | 0.192 |
| emodin | 518-82-1 | rat | 160 | -713.7 | -668.6 | 1Comp | 715.3 | 2.152 | 0.3221 |
| etoxazole | 153233-91-1 | rat | 192 | -1539 | -1626 | 2Comp | 31930 | 0.711 | 0.9749 |
| fenarimol | 60168-88-9 | rat | 192 | -806 | -815 | 2Comp | 8003 | 0.2633 | 2.632 |
| flufenacet | 142459-58-3 | rat | 192 | -1414 | NA | 1Comp | 62440 | 0.03117 | 22.24 |
| fluorotelomer alcohol 8+2 | 678-39-7 | rat | 131 | -3425 | NA | 1Comp | 11420 | 0.1106 | 6.266 |
| formamide | 75-12-7 | rat | 172 | -1936 | -1934 | 1Comp | 460.5 | 0.03309 | 20.95 |
| formetanate hydrochloride | 23422-53-9 | rat | 192 | -101 | -97.04 | 1Comp | 25240 | 8.31E-08 | 8341000 |
| free carbon disulfide | 75-15-0 | rat | 167 | -268.3 | -280.3 | 2Comp | 3352 | 7.072 | 0.09802 |
| gemfibrozil | 25812-30-0 | rat | 163 | -1367 | -1393 | 2Comp | 1266 | 0.2384 | 2.907 |
| glyoxylic acid monohydrate | 563-96-2 | rat | 156 | -325.9 | -368.1 | 2Comp | 898.4 | 13.42 | 0.05164 |
| hexachlorobenzene | 118-74-1 | rat | 180 | -8917 | -3018 | 1Comp | 4790 | 0.000306 | 2265 |
| hexobarbital | 15307-86-5 | rat | 206, 209 | NA | -189.5 | 2Comp | 2152 | 0.3759 | 1.844 |
| ibuprofen | 15687-27-1 | rat | 210 | -185.5 | -211.9 | 2Comp | 687 | 0.8539 | 0.8118 |
| imazalil | 35554-44-0 | rat | 192 | -405.6 | -436.5 | 2Comp | 16370 | 3.427 | 0.2023 |
| imipramine | 50-49-7 | rat | 211, 212 | -215 | -228.6 | 2Comp | 39970 | 0.3045 | 2.276 |
| isoeugenol | 97-54-1 | rat | 157 | -4214 | -4493 | 2Comp | 25090 | 1.949 | 0.3556 |
| l-ephedrine | 299-42-3 | rat | 136 | -1884 | -1891 | 2Comp | 16250 | 0.4206 | 1.648 |
| methanol | 67-56-1 | rat | 17 | -64.13 | -60.13 | 1Comp | 992 | 0.3046 | 2.275 |
| methyl tert-butyl ether | 1634-04-4 | human | 51 | -355.9 | -363 | 2Comp | 893.1 | 1.343 | 0.5161 |
| methylene chloride | 75-09-2 | rat | 18 | -198.5 | -197.5 | 1Comp | 1883 | 0.4908 | 1.412 |
| methyleugenol | 93-15-2 | rat | 158, 170 | -5782 | -5836 | 2Comp | 3780 | 0.3344 | 2.073 |
| midazolam | 59467-70-8 | rat | 217 | -285.3 | NA | 1Comp | 3287 | 1.18 | 0.5876 |
| naphthalene | 91-20-3 | rat | 169 | -1721 | -1820 | 2Comp | 4317 | 1.104 | 0.6277 |
| nilvadipine | 75530-68-6 | rat | 219 | -570.8 | -572.6 | 2Comp | 5910 | 0.3737 | 1.855 |
| nitrite | 14797-65-0 | rat | 137 | -916.3 | -916.5 | 2Comp | 2766 | 0.4414 | 1.57 |
| novaluron | 116714-46-6 | rat | 192 | -1258 | -1282 | 2Comp | 7105 | 0.05488 | 12.63 |
| octylphenol | 140-66-9 | rat | 60 | -1465 | -1557 | 2Comp | 49920 | 0.3106 | 2.231 |
| ondansetron | 99614-02-5 | rat | 220 | -850.4 | -894.2 | 2Comp | 617.2 | 30.69 | 0.02259 |
| oxazepam | 604-75-1 | rat | 159 | -1789 | -1941 | 2Comp | 13120 | 0.1889 | 3.67 |
| oxymetholone | 434-07-1 | rat | 168 | -688.5 | NA | 1Comp | 5639 | 0.1631 | 4.251 |
| pentachlorophenol, purified | 87-86-5 | rat | 161 | -1585 | NA | 1Comp | 82.06 | 0.07683 | 9.021 |
| perfluorodecanoic acid | 335-76-2 | rat | 147 | -2278 | -2361 | 2Comp | 306.2 | 0.000804 | 862.6 |
| perfluorohexane-1-sulphonic acid â€“ potassium salt | 3871-99-6 | rat | 143 | -1153 | NA | 1Comp | 350.9 | 0.001978 | 350.5 |
| perfluorooctane sulfonate | 45298-90-6 | rat | 148 | -1869 | -1907 | 2Comp | 278.6 | 0.000965 | 718 |
| permethrin | 52645-53-1 | rat | 192 | -723.3 | NA | 1Comp | 64800 | 0.1253 | 5.53 |
| phenacetin | 62-44-2 | human | 14 | -715.9 | NA | 1Comp | 1502 | 0.4745 | 1.461 |
| phenacetin | 62-44-2 | rat | 1, 221, 62, 63, 67 | -355.9 | -385 | 2Comp | 1865 | 0.6633 | 1.045 |
| phenolphthalein | 77-09-8 | rat | 139 | -834.6 | -886.5 | 2Comp | 1408 | 1.153 | 0.601 |
| phenytoin | 57-41-0 | rat | 222, 223 | -256.1 | -262.5 | 2Comp | 3687 | 0.1736 | 3.993 |
| potassium perfluorobutane sulfonate | 29420-49-3 | rat | 144 | -1413 | -1452 | 2Comp | 204.1 | 0.2538 | 2.731 |
| primidone | 125-33-7 | rat | 140 | -463 | -459.9 | 1Comp | 1109 | 0.2159 | 3.211 |
| propamocarb hydrochloride | 25606-41-1 | rat | 192 | -595.7 | -603.1 | 2Comp | 8878 | 2.974 | 0.2331 |
| propyzamide | 23950-58-5 | rat | 192 | -1635 | NA | 1Comp | 9560 | 0.1086 | 6.383 |
| pyridine | 110-86-1 | rat | 149, 162 | -2362 | -2369 | 2Comp | 316.2 | 0.02765 | 25.07 |
| pyrithiobac sodium | 123343-16-8 | rat | 192 | -604 | NA | 1Comp | 678.3 | 0.13 | 5.331 |
| resmethrin | 10453-86-8 | rat | 192 | -794.5 | -823.1 | 2Comp | 48620 | 1.48 | 0.4683 |
| s-bioallethrin | 28434-00-6 | rat | 192 | -553.5 | -555.1 | 2Comp | 33980 | 0.7573 | 0.9152 |
| simazine | 122-34-9 | rat | 192 | -651.3 | -650 | 1Comp | 3592 | 1.939 | 0.3575 |
| solvent red1 | 1229-55-6 | rat | 66 | -275.9 | -281 | 2Comp | 2154 | 1.058 | 0.6552 |
| tamoxifen | 10540-29-1 | rat | 141 | -357.8 | 282 | 1Comp | 605100 | 14.33 | 0.04837 |
| tert-amyl methyl ether | 994-05-8 | human | 51 | -342.8 | -365.5 | 2Comp | 5393 | 1.04 | 0.6662 |
| tetrachloroethylene | 127-18-4 | rat | 4 | -81.9 | -79.47 | 1Comp | 12730 | 0.08628 | 8.034 |
| tetralin | 119-64-2 | rat | 152 | -1825 | -2004 | 2Comp | 6321 | 1.725 | 0.4019 |
| thiodiglycolic acid | 123-93-3 | rat | 155 | -442.9 | -489 | 2Comp | 976.6 | 4.789 | 0.1447 |
| tolbutamide | 64-77-7 | rat | 225, 226 | -139.7 | NA | 1Comp | 182.9 | 0.1618 | 4.284 |
| trichloroethylene | 79-01-6 | rat | 18, 29 | -250.9 | -313 | 2Comp | 318.8 | 11.58 | 0.05988 |
| triclosan | 3380-34-5 | rat | 192 | -334.7 | -326.2 | 1Comp | 617.9 | 0.06498 | 10.67 |
| valproic acid | 99-66-1 | rat | 227, 228 | -201.8 | NA | 1Comp | 220.9 | 1.101 | 0.6294 |
| wyeth-14643 | 50892-23-4 | rat | 173 | -1050 | NA | 1Comp | 448.8 | 0.3273 | 2.118 |

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 |