Developing Generic Toxicokinetic Models with R Package "httk" for Enhanced Reporting Accuracy and Statistical Evaluation

John F. Wambaugh[1], Sarah E. Davidson[1], Robert G. Pearce[1,2], Michael Devito[1], Gregory S. Honda[1,2], Matthew W. Linakis[3], Elaina Kenyon[1], Annabel Meade[1,2], Tom Purucker[1], Caroline L. Ring[1], Mark A. Sfeir[1,2], and R. Woodrow Setzer[1]

1) Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, 27711
2) Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee 37831
3) Ramboll U.S. Consulting, Inc.

Poster number: P751
Abstract Number: 4052

John Wambaugh | wambaugh.john@epa.gov
https://orcid.org/0000-0002-4024-534X

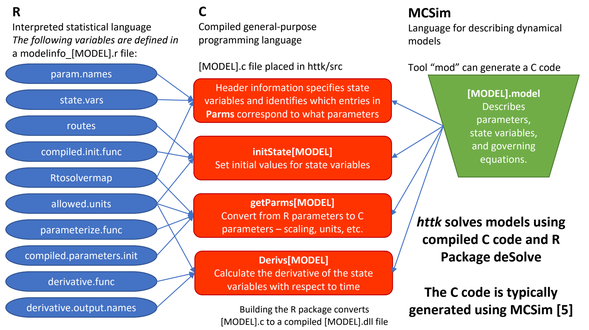
Introduction

- Decisions about the risk posed by a chemical to the public health involve models for toxicity, exposure, and the dose-response relationship. Toxicokinetics (TK) informs the dose-response relationship by describing the absorption, distribution, metabolism, and excretion (ADME) of a chemical dose
- The "*httk*" R package (https://cran.r-project.org/package=httk) [1] informs *in vitro-in vivo* extrapolation (IVIVE) for chemical risk assessment by providing both chemical-specific *in vitro* data for predicting TK and a suite of "generic" models (including PBTK) that use those data to predict ADME.
- The ***high throughput toxicokinetics (httk)*** environment addresses three barriers to the incorporation of PBTK models into risk assessment in non-pharmaceutical regulatory settings [2]:
  1) **Data availability** (that is, lack of clinical trials): *httk* provides access to chemical-specific data and predictions for thousands of chemicals
  2) **Model reproducibility**: Despite the deterministic nature of scientific computing, it is often difficult to reproduce computational results [3]
  3) **Statistical evaluation**: New generic TK models can be evaluated across the growing library of chemicals with both *in vitro* and *in vivo* data

Here we discuss how new models may be created and interlaced with the *httk* environment, so that previously evaluated tools and data may be incorporated alongside new approaches

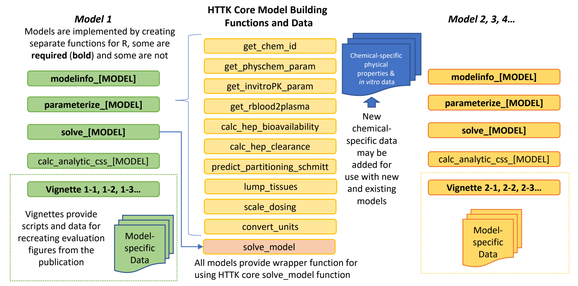
Integrating New Models

- ***httk* makes use of three open-source computer programming languages: R, MCSim, and C**
- Typical users and the *httk* design team use R [4], an interpreted statistical analysis language
- TK modelers can use MCSim [5] to describe new models and the MCSim function "mod" is then used to convert MCSim to C code that can be compiled for rapid execution by R
- Here we indicate the name of a new model (such as "gas_pbtk" or "fish_pbtk") with "[MODEL]"
- *httk* simulates the compiled C models with R package deSolve [6], using instructions provided by modelinfo_[MODEL].R
- **Once integrated a model can be used for TK ("calc_mc_tk()") and IVIVE ("calc_mc_css()")**



Generic Functions for TK Models

- ***httk* is intended to permit flexibility with respect to the description of toxicokinetics but allows the reuse of data and tools for tasks such as parameterization, unit conversion, and Monte Carlo simulation**
- By using pre-existing, peer reviewed, and publicly vetted data and code we aim to minimize errors in physiology, chemistry, calculation, and implementation – "Do things once"
- For example, we have a single function for conversion of units (such as µM to mg/L) called "convert_units()" and reuse it wherever needed to minimize the chances of making a mistake



Parameter Tables

- **Models developed for the *httk* suite are designed to be used to make chemical-, species-, and tissue-specific predictions, drawing from data tables provided with the R package**

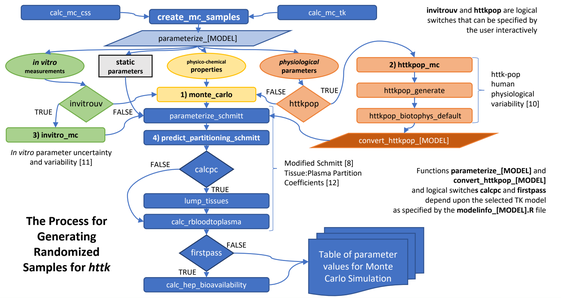
Tables available within *httk*

chem.physical_and_invitro.data	Chemical-specific values for physico-chemical properties and *in vitro* measurements
chem.invivo.PK.data	Chemical concentration vs. time data (CvTdb) for evaluating model predictions [7]
physiology.data	Species-specific physiological parameters, currently covering human, monkey, rat, rabbit, and mouse
tissue.data	Tissue composition data for Schmitt's method of partition coefficient prediction [8]
Tables.Rdata.stamp	Time stamp identifying when the other tables were created

- Over the past decade EPA, its collaborators, and its contractors have made chemical-specific measurements of *in vitro* TK determinants (largely plasma protein binding and intrinsic hepatic clearance) for over 1,200 chemicals in humans and more than 200 chemicals in rats
- Additional chemical-specific data as well as physiological information and tissue-descriptors have been curated from the scientific literature
- Physico-chemical properties are predicted with the OPERA quantitative-structure-activity relationship (QSAR) suite [9]

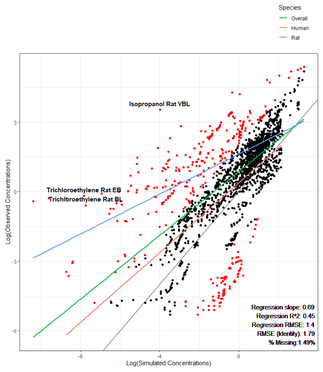
Monte Carlo

- ***httk* propagates both uncertainty and variability with Monte Carlo (MC) simulation**
- "create_mc_samples()" is the primary function and creates a table of samples with a column for each parameter and a row for each "draw" of the MC simulation
- Population variability in physiological parameters is controlled by function "httkpop_mc()" [10], while measurement uncertainty and population variability in chemical-specific parameters is controlled by "invitro_mc()" [11]
- Since httk-pop predicts correlated human population variation in more than 50 biometrics [10], modelers may include a custom function "convert_httkpop_[MODEL]()" to relate those parameters to their PBTK model
- For rapid Monte Carlo it is recommended that modelers develop an analytical steady-state solution for their model ("calc_analytic_steadystate_[MODEL]()"



Model Evaluation Data

- **The most challenging task is not building a new model, but evaluating that model**
- Each new addition to *httk* should include an evaluation of model predictions relative to *in vivo* data where available (as in the table "httk::chem.invivo.PK.data" from CvTdb [7])
- Based on the outcome of your evaluation, it may be possible to estimate how generalizable the new model is expected to be
- Linakis et al. (2020) [13] added a generic gas inhalation PBTK model to *httk*, included in version 2.0.0. They estimated the root mean squared error (RMSE) for model predictions for 41 volatile organic chemicals in 143 inhalation exposure experiments drawn from the CvTdb [7] including exhaled breath (EB), blood (BL), and venous blood (VBL) data



Community Involvement

- Version control software is used by EPA and a public GitHub repository is maintained at: https://github.com/USEPA/CompTox-ExpoCast-httk
- Users of *httk* can use GitHub to identify bugs with functions, problems with data, and even submit suggested fixes
- Each release follows the R version numbering convention, with three numbers indicating: "Major Releases", "Minor Releases", and "Patches". The current version of is 2.0.4
  - Patches include relatively minor alterations of the functions and data to address problems identified by the scientific community
  - Minor releases always accompany peer-reviewed publications, and typically cover new models or data sets
  - A major release would involve restructuring most or all the underlying R code and none is currently planned

Conclusions

- All users can access the peer-reviewed data and functions of *httk*, however model developers can incorporate new models by developing a TK model description in the language MCSim and then using a modelinfo_[MODEL].R file to describe that model to *httk*
- An upcoming manuscript will describe this procedure in detail
- *httk* models allow transparency, portability, statistical evaluation, and community feedback

https://www.epa.gov/research/

References:

1) 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
2) Clark LH, Setzer RW, Barton HA (2004) Framework for evaluation of physiologically-based pharmacokinetic models for use in safety or risk assessment. Risk Analysis 24 (6):1697-1717
3) Piccolo, S. R., & Frampton, M. B. (2016). Tools and techniques for computational reproducibility. Gigascience, 5(1), s13742-016.
4) R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
5) Bois FY, Maszle DR (1997) MCSim: a Monte Carlo simulation program. Journal of Statistical Software 2 (1):1-60
6) Soetaert KE, Petzoldt T, Setzer RW (2010) Solving differential equations in R: package deSolve. Journal of statistical software 33
7) Sayre, R. R., Wambaugh, J. F., & Grulke, C. M. (2020). Database of pharmacokinetic time-series data and parameters for 144 environmental chemicals. Scientific data, 7(1), 1-10.
8) Schmitt W (2008) General approach for the calculation of tissue to plasma partition coefficients. Toxicology in Vitro 22 (2):457-467
9) Mansouri, K., Grulke, C. M., Judson, R. S., & Williams, A. J. (2018). OPERA models for predicting physicochemical properties and environmental fate endpoints. Journal of chemoinformatics, 10(1), 1-19.
10) Ring CL, Pearce RG, Setzer RW, Wetmore BA, Wambaugh JF (2017) Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability. Environment International 106:105-118
11) Wambaugh, J. F., Wetmore, B. A., Ring, C. L., Nicolas, C. I., Pearce, R. G., Honda, G. S., ... & Thomas, R. S. (2019). Assessing toxicokinetic uncertainty and variability in risk prioritization. Toxicological Sciences, 172(2), 235-251.
12) Pearce, R. G., Setzer, R. W., Davis, J. L., & Wambaugh, J. F. (2017). Evaluation and calibration of high-throughput predictions of chemical distribution to tissues. Journal of pharmacokinetics and pharmacodynamics, 44(6), 549-565.
13) Linakis MW, Sayre RR, Pearce RG, Sfeir MA, Sipes NS, Pangburn HA, Gearhart JM, Wambaugh JF (2020) Development and Evaluation of a High Throughput Inhalation Model for Organic Chemicals Journal of Exposure Science & Environmental Epidemiology