U.S. Environmental Protection Agency
Office of Research and Development
Center for Computation Toxicology & Exposure
Chemical Characterization & Exposure Division
Computational Exposure and Toxicokinetics Branch

Quality Assurance Project Plan

Title: ExpoCast Project: R Package "httk" for High Throughput Toxicokinetic Modeling

QA Category: ☐ A ☒ B

ORD National Program Project/Task ID: CSS 2.6

QAPP was Developed: ☒ Intramurally ☐ Extramurally: Click here to enter text.

QAPP Accessibility: QAPPs will be made internally accessible via the ORD QAPP intranet site upon final approval *unless the following statement is selected*.
☐ I do NOT want this QAPP internally shared and accessible on the ORD intranet site.

Project Type(s) (check all that apply):
☐ Environmental Measurements ☐ Environmental Technology ☐ Decision Support Tool ☒ Existing Data ☒ Informatics ☐Geospatial ☐ Method Development ☒ Model Application ☒ Model Development ☒ Software and Data Management ☐Remote Sensing ☐ Technical Assessment ☐ Other

Approvals

Prepared by: John Wambaugh	Digitally signed by JOHN WAMBAUGH Date: 2021.04.09 11:53:03 -04'00'	
	Signature	*Date*
Supervisor: Peter Egeghy	PETER EGEGHY Digitally signed by PETER EGEGHY Date: 2021.04.09 16:56:33 -04'00'	
	Signature	*Date*
QA Manager: Adam Swank	ADAM SWANK Digitally signed by ADAM SWANK Date: 2021.03.16 13:14:27 -04'00'	
	Signature	*Date*

Table of Contents

Part A. Project Management
A1. Distribution List

Name	Affiliation	Title/Role	Email	Products
Wambaugh, John	Center for Computational Toxicology and Exposure (CCTE)	Technical Lead	wambaugh.john@epa.gov	1-9
Swank, Adam	CCTE	Quality Assurance Manager (QAM)	swank.adam@epa.gov	1-9
Williams, Antony	CCTE	Principal Investigator / CompTox Chemicals Dashboard Architect	Williams.antony@epa.gov	1-9
Ring, Caroline	CCTE	Principal Investigator/ Mathematical Modeler	ring.caroline@epa.gov	7,8
Wetmore, Barbara	CCTE	Principal Investigator/ Biologist	wetmore.barbara@epa.gov	4
Dawson, Dan	CCTE	Post-doctoral researcher	Dawson.dan@epa.gov	5
Tornero-Velez, Rogelio	CCTE	Principal Investigator/ Mathematical Modeler	tornero-velez.rogelio@epa.gov	5
Evans, Marina	CCTE	Principal Investigator/ Mathematical Modeler	evans.Marina@epa.gov	3
Breen, Miyuki	CCTE	Post-doctoral researcher	Breen.miyuki@epa.gov	7
Kenyon, Elaina	CCTE	Principal Investigator/ Biologist	Kenyon.marine@epa.gov	1
Purucker, Tom	CCTE	Principal Investigator	Purucker.Tom@epa.gov	9
Davidson, Sarah	CCTE	Statistician	Davidson.Sarah.E@epa.gov	1-9
Goldsmith, Rocky	CCTE	Principal Investigator / Computational Chemist	Goldsmith.michael@epa.gov	5
Dustin Kapraun	Center for Public Health and Environmental Assessment (CPHEA)	Principal Investigator/ Mathematical Modeler	kapraun.dustin@epa.gov	1

A3. Project/Task Organization and Responsibilities

This Quality Assurance Project Plan (QAPP) was developed with reference to Guidance for Quality Assurance Project Plans for Modeling EPA QA/G-5M (U.S. Environmental Protection Agency, 2002).

This QAPP covers a project that supports R package "httk". These software tools use high throughput methods to describe the absorption, distribution, metabolism, and excretion of chemicals by the body (that is, toxicokinetics). httk allows systematic, reproducible, and transparent *in vitro-in vivo* extrapolation (IVIVE) of high throughput bioactivity screening data to estimate human-relevant exposures. The following list provides the project participants along with their respective responsibilities.

John Wambaugh is a principal investigator in the CCTE and Project Co-Lead of the ExpoCast (Exposure Forecasting) project. He is responsible for designing the structure of the httk package; overseeing the development of new models; writing technical manuscripts; reviewing and debugging code; and performing statistical evaluation to characterize the model accuracy and certainty.

Adam Swank is the Quality Assurance Manager (QAM) in the CCED and will receive copies of all QA and QC related information. He will review the QAPP and any pertinent QA related documents to ensure the quality of the research is upheld. The QAM will perform, at their or the technical lead's discretion, an audit of the process.

Caroline Ring is a principal investigator in CCTE. She is the original developer of the httk-pop Monte Carlo human variability simulator. She is leading projects to improve variability simulations and better model sensitive lifestages. She writes technical manuscripts describing her work.

Evgenia Korol-Bexell is a Federal post-doctoral researcher in the CCTE. She develops models for and analyzes data better characterizing oral absorption of chemicals. She writes technical manuscripts describing her work.

Barbara Wetmore is a principal investigator in CCTE and an expert in high throughput *in vitro* characterization of TK. She provides consulting support and she leads an EPA laboratory that generates new data that contributes to the httk package. She writes technical manuscripts describing her work.

Marina Evans is a principal investigator in CCTE. She is leading a project to develop a high throughput physiologically-based toxicokinetic (PBTK) model with a dermal

route of exposure. She develops and revises the TK models and performs statistical analysis. She writes technical manuscripts describing her work.

Elaina Kenyon is a principal investigator in CCTE. She oversees projects augmenting HTTK's abilities to describe sensitive (pediatric) populations. She develops and revises the TK models and performs statistical analysis. She writes technical manuscripts describing her work.

Tom Purucker is a principal investigator in CCTE. He oversees a project developing an aquatic species model for HTTK. She develops and revises the TK models and performs statistical analysis. She writes technical manuscripts describing her work.

Dustin Kapraun is a principal investigator in CPHEA. He is leading a project to develop a high throughput PBTK model for human gestational exposure that includes fetal growth. He develops and revises the TK models and performs statistical analysis. He writes technical manuscripts describing his work.

Antony Williams is a principal investigator in CCTE and is architect of the public-facing CompTox Chemicals Dashboard (https://comptox.epa.gov/). When provided with updated chemical-specific predictions of TK, he works to provide those predictions to the public.

Sarah Davidson is a statistician in CCTE and is the liaison of the Scientific Computing and Data Curation division to the HTTK project.

A4. Problem Background

Thousands of chemicals have been profiled by high-throughput screening programs such as ToxCast and Tox21 (Dix, Houck et al. 2007, Schmidt 2009); these chemicals are tested in part because most of them have limited or no data on hazard, exposure, or toxicokinetics. Toxicokinetic models describe the absorption, distribution, metabolism, and excretion of chemicals by the body.

Linking hazard and exposure predictions to estimate risk requires toxicokinetics. Since most chemicals lack toxicokinetic data obtained by traditional (that is, low throughput *in vivo* assays), the use of high-throughput, *in vitro* toxicokinetic methods (HTTK) is required (Wetmore, Wambaugh et al. 2015). HTTK measures key determinants of toxicokinetics that, when used with toxicokinetic models, allow prediction of internal tissue concentrations resulting from chemical exposure.

EPA has created multiple, different HTTK models that are included within the httk software package. These models are designed to be parameterized using high-throughput *in vitro* toxicokinetic data (plasma protein binding and hepatic clearance), as well as structure-derived physicochemical properties and species-specific physiological data (Pearce, et al., 2017b). New models are continuously under development while older models continue to be evaluated and refined (Wambaugh, et al., 2018). These various models represent different levels of complexity, routes of exposure, and susceptible populations.

HTTK data and models allow translation of bioactive concentrations identified by high throughput screening (for example, ToxCast) to doses (mg/kg body weight/day) that would cause tissue concentrations equal to the bioactive concentration. These doses can then be compared with estimated chemical intake rates (that is, exposure) to estimate the margin between bioactive and exposure doses. This margin allows for chemical priority setting.

A5. Project Description, Objectives, and Schedule

This project supports the development, refinement, and evaluation of the R package "httk" (Pearce, et al. 2017a). Package httk enables the inclusion of toxicokinetics in the statistical analysis of chemicals undergoing high-throughput screening. The httk tools are made publicly available as an R package distributed on the Comprehensive R Archive Network (https://cran.r-project.org/) supporting the freely-available statistical programming language R. httk allows systematic, reproducible, and transparent *in vitro-in vivo* extrapolation (IVIVE) (Coecke, Pelkonen et al. 2013, Wetmore 2015, Bell, Chang et al. 2018) of high throughput bioactivity screening data to estimate human-relevant exposures. A "reverse

dosimetry" (Tan, Liao et al. 2006) approach can be used to predict exposure doses sufficient to cause tissue concentrations that have been identified as bioactive by high-throughput screening. The package contains tools for Monte Carlo sampling (Ring et al., 2017) and reverse dosimetry along with functions for the analysis of concentration vs. time simulations. This package is structured to be augmented with additional chemical data as they become available.

The objective of this project is to further develop, refine, and evaluate the R package "httk" in support of the Chemical Safety for Sustainability (CSS) program's Rapid Exposure and Dosimetry (RED). Contributing products are listed below:

Product 1

Development and evaluation of a HT-PBTK model that includes human gestational growth and exposure. (**FY21**)
Lead: Dustin Kapraun

Product 2

Development and evaluation of a HT-PBTK model that includes an aerosol inhalation (lung) route of exposure. (**FY21**)
Lead: Matt Linakis (USAF)

Product 3

Development and evaluation of a HT-PBTK model that includes a dermal (skin) route of exposure. (**FY21**)
Lead: Marina Evans

Product 4

Incorporation of new human *in vitro* TK data from EPA, Health Canada, and peer-reviewed literature. (**FY21**)
Lead: John Wambaugh

Product 5

Incorporation of quantitative structure-property relationships (QSPRs) for predicting plasma protein binding and intrinsic hepatic clearance from chemical structure. (**FY21**)
Lead: Rocky Goldsmith

Product 6

Development and evaluation of a HT-PBTK model that includes allows configurable tracking of five compounds that may be metabolites of each other. (**FY23**)
Lead: John Wambaugh

Product 7

Incorporation of new population biometric data (NHANES) and expansion of HTTK-pop Monte Carlo simulator to include inhalation exposure (**FY22**)
Lead: Caroline Ring

Product 8

Development and evaluation of a HT-PBTK model that allows rapid lumping and splitting of model compartments (that is, grouping and ungrouping of tissues). (**FY23**)
Lead: Caroline Ring

Product 9

Development and evaluation of a HT-PBTK model for aquatic species. (**FY23**)
Lead: Tom Purucker

A6. Quality Objectives and Criteria

Mathematical modeling is a descriptive science, and there are no absolute, quantitative rules for judging quality (Oreskes 1998, Pilkey and Pilkey-Jarvis 2007, McLanahan, El-Masri et al. 2012). However, certain evaluations should always be performed to allow assessment of key aspects including accuracy, documentation, uncertainty, and applicability of the code, models, and data. Quality objectives for models included in the HTTK package are quantification of bias, uncertainty, and fit to existing data. These should be evaluated using appropriate statistical methods, such as root mean squared error (RMSE) or relative predictive error (RPE) when comparing models across differing data sets. When comparing models using the same data, information criteria (such as Akaike Information Criterion (Yamaoka, Nakagawa et al. 1978) or Deviance Information Criterion (Spiegelhalter, Best et al. 2002)) might be considered appropriate. Garcia, Ibrahim et al. (2015) found that parameter identifiability (Cobelli and Distefano 3rd 1980) may only be proven using a full Bayesian analysis, but this is often beyond the resources available. Quality objectives for HTTK package coding include accuracy of model implementation in computer code, documentation of coding decisions and code structure, and reproducible performance of the code. Quality objective for data sets included in the HTTK package – both data for model parameterization and data for model evaluation – are completeness (for example, missing values), correctness (for example, numbers, decimal places, and units), and documentation (for example, references).

The R package "httk" exists to allow model evaluation. All models, documentation, and evaluation data can and should be "packaged" together. Maintenance and expansion of the package will be described by the Standard Operating Procedure (SOP) "HTTK R Package Support Standard Operating Procedures".

A7. Specialized Training Requirements/Certification

Specialized expertise and qualifications needed to participate in this project include proficiency with developing mathematical models of biological systems (especially toxicokinetics) in modern computer programming languages. Team members must be knowledgeable in at least one of the following fields: biomedical engineering, mathematical physiology, applied mathematics, applied statistics, physics, mathematics, chemistry, chemical engineering, or computer science. Team members should be fluent programmer in a modern programming language (for example, R, Python, Julia). This research project involves data science, mathematical modeling, toxicokinetics, analytical science, data management, and computational toxicology.

Team, members must become comfortable with the R programming language, physiologically-based toxicokinetic (PBTK) modeling, and the MCSim modeling language. Team members need to be comfortable with the use of version control software. Team members must become familiar with the development, revision, and release of R packages for public use.

For trainees working on the project, a basic knowledge of programming and mathematics and the ability to identify and extract relevant data and information from articles and databases will be required. Additional training will be provided, if needed. Reference to this training will be documented in an electronic lab notebook that is easily accessible to all with the initials of the trainer and trainee and date of completion. If expert workshops or meetings are required to advance the working knowledge of individuals involved in the projects, the name of these workshops, the attendees, and their dates of attendance will also be placed into the electronic lab notebook. If certificates were issued for these workshops, they will be placed into a separate binder containing relevant lab documents, or, if digital, will be placed into a folder labeled "Certificates of Training" that will be located on the Principal Investigator's computer desktop.

A8. Documentation and Records

When a new individual becomes involved in the project, this QAPP will be sent to her or him by John Wambaugh. The individual will be encouraged to ask questions should any part of this document be unclear, and upon reading their name(s) and dates of completion will be entered into an open lab notebook or electronic record of training. Each project contributor should be given access to L:/ExpoCast. Each project contributor should be given a Confluence and Bitbucket account with access to the "HTTK" git repository.

Part B. Data Acquisition and Measurement

B1. Model Development

Model Design

A key component of this project is the building of generic TK models (Van der Molen, Kooijman et al. 1996, Schmitt and Willmann 2004, Brightman, Leahy et al. 2006, Edginton, Schmitt et al. 2006, Jongeneelen and Berge 2011) that can be parameterized from *in vitro* data and physico-chemical properties (Wetmore, Wambaugh et al. 2012, Wetmore, Wambaugh et al. 2015). A generic TK model uses the same structure – simulating the same physiological and chemical processes – for all chemicals. The generic TK modeling approach offers the advantage of consistent use of an open-source, transparent tool, which hopefully reduces the likelihood of programming errors (McLanahan, El-Masri et al. 2012). However, there are many TK processes that are important for only some classes of chemicals. So, we expect a generic model to have larger uncertainty than a "bespoke" chemical-specific model, despite the greater confidence in model implementation.

R packages are assigned three version numbers: [MAJOR].[MINOR].[PATCH]. While public releases of httk are numbered sequentially (for example, 1.1.0, 1.2.0, 1.3.0, …) within ORD there may be many different versions. Versions can differ based on the TK models included, the data included, or the functioning of the httk code. Different versions of the package are managed collaboratively using the "git" version control software as implemented with Bitbucket. Internal versions should always have major version number 99. Each project should have a unique minor number. Versions will be renumbered with the next sequential public number when it is cleared for release. Internal packages can be built and shared within ORD and with Material Transfer Agreement partners. Use "R CMD build [PATH]" to get a distributable version. While there is a separate Github repository, this is only for public dissemination and is used for distributing code that has been cleared for external release by EPA management.

Each upcoming version should have a branch on Bitbucket: https://ncct-bitbucket.epa.gov/projects/HTTK/repos/httk/branches

Model Coding

Rather than create a single, "super model" we add individual models to the package containing different features. Each of these models should have been evaluated against measured data (Sayre, Wambaugh et al. 2019) as in Figure 1b.

Initial model development can proceed in any programming language, but for inclusion in the httk R package it must be transformed to C and an interface must be written in R. This interface is specified in the modelinfo_[MODELNAME].R file. A parameterize-[MODELNAME] function must be indicated by the modelinfo file and can be included either as a separate ".R" file or in the modelinfo file. If an analytic steady-state solution can be obtained this should be indicated in the modelinfo file and may either be included in that file or as a separate file. The MCSim (https://www.gnu.org/software/mcsim/) provides basic functionality for model description and compilation to C code. The MCSim "mod" function converts MCSim models into C code that can be compiled. We then run these compiled models with HTTK

All functions and data sets should be documents using roxygen2 (https://CRAN.R-project.org/package=roxygen2). Data files are documented in data.R file in /R directory. Need to add a \describe{ \item{colname}{}} with an item for each column. If all code is properly documented using roxygen2, then documentation files can be created by going into the "httk" package directory (setwd) within R and running "roxygenize()". Documentation of parameter descriptions can be reused with @inherit, @inheritParams, and @inheritSections.

Programmers are encouraged to include the author tag in the code they create and to add themselves to the authors list if modifying pre-existing code. CRAN allows the recognition of different levels of contribution – the two most likely to be relevant are "aut" – "Full authors who have made substantial contributions to the package and should show up in the package citation" – and "ctb" – "Authors who have made smaller contributions (such as code patches etc.) but should not show up in the package citation". If you have created a new model you are a full author.

When you add a new feature, add something to httk/tests to test it

Model Evaluation

In the context of the httk R package model evaluation involves two critical parts 1) evaluating the performance of the new model (as in Figure 1b) and 2) making sure the new changes do not break other aspects of the package.

We should maintain the ability to calculate a root mean squared error (RMSE) for each model, and then recalculate it every time we are done revising the package to make sure that a change made for one model affects the other models in the intended way (often, not at all).

An advantage that is distinct to generic TK models is that we can evaluate the model predictions in the absence of *in vivo* data for a specific chemical. We do this by evaluating the model predictions using as much *in vivo* data as we can find for other chemicals. We can estimate bias and uncertainty and try to correlate these with chemical-specific properties (Pearce et al., 2017b, Wambaugh et al., 2015). We can then consider whether the generic model can be extrapolated to other situations (chemicals without *in vivo* data).

The takeaway is that the trick is not building a new model, but rather evaluating that new model. Each new addition to HTTK should include an evaluation of model performance using *in vivo* data, and a discussion of how generalizable the new model (or revision) is expected to be.

Checking the impact of changes.

1) *Go to DOS shell, run "R CMD INSTALL newversionpath" – fix any errors that prevent the install*
2) *Once you have your new version installed, load it in r "library(httk)" and check out that the new functions work right.*
3) *Go to DOS shell, run "R CMD build newversionpath" – fix any errors that come up.*

To install a previous version use "R CMD INSTALL otherversionpath"

To install official last CRAN release from within R type install.package("httk")

Using the R "tests" Directory Functionality

"Testing is a vital part of package development. It ensures that your code does what you want it to do" (http://r-pkgs.had.co.nz/tests.html) R includes automated testing functionality that runs whenever you build ("R CMD build") a package. Test scripts are stored in the "httk/tests" directory along with saved output.

Within the "httk/tests" folder:

1) The code in each ".R" file should run

2) The outputs in the ".Rout.save" files should match (unless you changed something that should impact it)

To create a new ".Rout.save" file use "R CMD BATCH - myTest.R myTest.Rout.save" from the command prompt.

If you calculate a RMSE you can always use a statement like "mymodel.RMSE <= X" so that if the RMSE improves no warning is thrown.

When you add a new feature, add something to httk/tests to test it

Random Number Seeds

Since many of the applications of httk involve Monte Carlo simulation it is incredibly important to use the command "set.seed(SEED)" before running any code that uses random numbers. The value of SEED is immaterial if it is properly recorded. This should be used for all manuscript figures, analyses, and tests.

B2. R Package Maintenance

Curating *in vitro* TK Data

Most of the data in the HTTK package initially came from the Rotroff, Wetmore et al. (2010), Wetmore, Wambaugh et al. (2012), and Wetmore, Wambaugh et al. (2015) publications, with a new bolus from EPA HTTK contract research. However, there are many examples of papers in the peer reviewed literature that provide relevant information. The data of particular interest are:

- fraction unbound in plasma
- intrinsic hepatic clearance (hepatocyte incubation assay)
- fraction bioavailable
- blood:plasma ratio
- volume of distribution

As we expand the model, we may eventually also want:

- blood:air ratio
- metabolic K_M, V_{max} for a Michaelis-Menten equation

Often a paper will only provide some of these values. One of the things the function get_cheminfo() checks is whether the combination of data from all sources is sufficient to run a given model.

Checking Units

In some cases, units are not a problem since the relevant parameters are fractions/ratios. However, we always need to be careful to translate:

- intrinsic hepatic clearance units should be μL/min/10^6 hepatocytes
- volume of distribution units should be L/kg body weight
- K_M should probably be μM
- V_{max} depends on the process

Data Storage

Until we have a production version of a database, we rely on a script to compile HTTK data (including *in vitro* measurements, *in vivo* measurements, and physico-chemical properties) from an ever-growing list of various papers. The script and supporting data are stored in their own Git repository:

https://ncct-bitbucket.epa.gov/scm/httkdat/httkdatatables.git

The script load_package_data_tables.R is used to add each tables data using the function add_datatable. The script generates a Tables.RData file (to be placed within the "httk/data" directory), that contains most of the relevant chemical-specific data. In particular, the table chem.physical_and_invitro.data contains the physico-chemical properties *and in vitro* TK data for each chemical. The data from each source should be maintained in the Git repository. The script also creates a sysdata.rda file (to be placed within the "httk/R" directory) contains data used internally by the R functions but that is not exported to the user. Finally, the script generates text versions of key tables to allow Git to more easily track the changes (the binary files are backed up but change tracking is unavailable).

Uploading to CRAN

New versions of the package are cleared as supplemental material for peer-reviewed scientific manuscripts. Once cleared, the package is made public by uploading to CRAN. CRAN is a public source for exchanging R code and datasets. In part because they are providing a public service, they are quite picky about exactly what is in a package that is uploaded to CRAN. We use the R CMD check function to make sure our package is up to snuff. The standards evolve over time, so first you must install the latest (development) versions of R and RTools. Then you use these to build a "tarball" for upload to CRAN:

R CMD build --resave-data [PATH]

This will create a file named httk-X.tar.gz Then you must check the tarball. The "—as-cran" argument makes the checking more stringent and matches what is done when it is uploaded:

R CMD check --as-cran httk_1.9.8.tar.gz

The goal is to have no NOTES or ERRORS when check is done. However, we seem to always get a NOTE about the size of the package, which has been acceptable to CRAN so far.

NHANES Biometrics

In Ring, Pearce et al. (2017) we used the McNally, Cotton et al. (2014) regressions to build a Monte Carlo population variability simulator that used biometrics from the National Health and Nutrition Examination Survey (NHANES) to estimate physiology relevant to toxicokinetics. NHANES is a rolling survey that includes properties such as height, weight, age, ethnicity, and gender for roughly ten thousand individuals that can be statistically weighted to reproduce the modern U.S. population. By drawing these actual individuals, we preserve correlations that occur between these properties. Eventually we will need to develop a standard operating procedure for updating using the latest NHANES data, however this methodology does not yet exist.

B3. Non-direct Measurements (Data Acquisition Requirements)

The model inputs (for example., chemical specific *in vitro* measures of toxicokinetics or species-specific measures of physiology) are one of:

1) Existing EPA HTTK information included in previous releases of the R package "httk" (https://CRAN.R-project.org/package=httk)
2) New chemical-specific *in vitro* measures of toxicokinetics from EPA contractors, EPA laboratories, or other government organizations
3) New *in vivo* measures of chemical tissue concentration vs. time
4) Curated literature *in vivo* measures of chemical tissue concentration vs. time from an EPA database (Sayre, Wambaugh et al. 2019)
5) Information newly extracted by project team members from peer-reviewed scientific literature
6) New *in silico* predictions of toxicokinetic properties (for example, Sipes, Wambaugh et al. (2017))
7) Chemical structure and use information from EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard)
8) Chemical property estimates derived from existing EPA/CCTE or EPA models (for example, OPERA https://github.com/kmansouri/OPERA or EPI-Suite https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface)

9) Survey data obtained by the U.S. Centers for Disease Control and Prevention (CDC) (https://www.cdc.gov/nchs/nhanes/index.htm)

Many of these efforts have existing data collection QAPPs; for example, those associated with the contractor work assignments that originally collected the data or have gone through extensive peer review processes and are used by EPA or other regulatory bodies (for example, EPI-Suite; CDC data). In other cases, the data are generated by other Office of Research and Development research activities with their own approved QAPP. In both these cases, the original QAPP or data quality process will be recorded and cited where necessary. Quality objectives and criteria for data specifically generated for this project (for example obtained from the literature, publicly available databases, or other online sources) are discussed below.

When using new data quality checks will be put in place to protect against inadvertent errors. All original documents (for example html files, data files, PDFs) etc. will be archived by the PI and retained in case of the need for further analysis. Data will initially be visually assessed to ensure that formatting is consistent (for example, identification numbers are not incorrectly reformatted into dates) and that a name and link to the database are also provided. Where possible, QA scripts will be written to check downloaded data for incorrect or inconsistent values (for example, weight fractions not greater than one). These data quality scripts will also be kept in the same location as the original data.

As values obtained from literature are entered into the appropriate storage locations (for example, spreadsheets or word documents), they will be spot checked to make sure that numbers and decimal places are correct In addition, the parameter or input name, along with the appropriate units, for that will be listed. Finally, the original literature source for that data, along with a link to the website or model, will be included for the value. This will allow for rechecking of the value during model development if model output appears erroneous. Spot checks of numbers, input names, units, and literature sources will be carried out on ten percent of the collected values.

B4. Documentation, Data Management, and Hardware/Software Configuration

All information related to a project will be stored in a directory named:

L:\Lab\NCCT_ExpoCast\ExpoCast[YEAR]\[ProductName]

The data will be stored in standard machine-readable format (typically a comma separated value, CSV, file) that is easily read by different programming languages

and can be easily disseminated to others via required EPA systems (for example, ScienceHub). All data will ultimately be distributed as a new version of the httk R package.

Manuscript Documentation

Each manuscript should have an archival directory named:

L:/ExpoCast/ExpoCast[YEAR]/[PROJECTNAME]

A final working data processing script should be stored along with the raw data on the L:/ drive in a directory called:

L:/ExpoCast/ExpoCast[YEAR]/[PROJECTNAME]/rawdata/

This data processing code should take raw data and convert it to the form used by the paper to make figures. A final working script to make all figures associated with the paper and any necessary processed data should be stored in

L:/ExpoCast/ExpoCast[YEAR]/[PROJECTNAME]/figures/

Multiple copies of manuscript milestones should be stored:

Milestone	Sub-Directory
Initial drafts, figures, scripts, and other material	EarlyDrafts
Version submitted to tech review and edits	TechReview
Version submitted to clearance review and edits	ClearanceReview
Version submitted to journal	Submission
Any revisions submitted to journal and edits	Revision[#]
Any submissions to new journal	Submission[#]

Each manuscript should be accompanied by a new vignette (http://r-pkgs.had.co.nz/vignettes.html) that demonstrates how to create the appropriate figures from the paper using the new version of the httk package.

Part C. Assessment and Oversight
C1. Internal Assessment and Response Actions

Models

New coding is captured using Git, and documented both via Roxygen (Wickham, Danenberg et al. 2020). Scripts for generating figures evaluating models (and, potentially for inclusion in manuscripts) should be documented in a manner allowing reproducibility, for example through RMarkdown (Xie, Allaire et al. 2018). These scripts may subsequently be turned into vignettes for distribution with the R package. Scripts should be documented at a sufficient level of detail to allow understanding of how it is functioning. Accuracy of model representation and documentation in new coding are reviewed by a member of the team other than the author who is familiar with the programming language and can identify mistakes. R automated testing functionality, "R CMD build", is used to confirm proper function of new code.

Toxicokinetic models may be assessed by comparing predicted tissues concentrations as a function of time against *in vivo* data wherever available. Analysis of the residuals (difference between predictions and observations) allow quantitative estimation of both bias and uncertainty. (Barton, Chiu et al. 2007, Wambaugh, Wetmore et al. 2015, Cohen Hubal, Wetmore et al. 2018, Linakis, Sayre et al. 2020) For a given data set, models are evaluated for fit to existing data sets using statistics such as root mean squared error (RMSE) and relative predictive error (RPE). The use of generic models allows a wider array of data (that is, across multiple chemicals) to be used for evaluation provided that the differences in how the data were collected across multiple experiments can be appropriately captured. Once performance is characterized statistically, this performance can be used as a diagnostic to confirm functionality following modification of the httk R package – that is, we expect an unchanged model to continue to perform the same for unchanged data and a change in performance may indicate an unintended error.

Secondary Data Sets

Secondary data is visually assessed to ensure consistent formatting and that a name and link to the database are also provided. Where possible, QA scripts are written to check downloaded data for incorrect or inconsistent values. Data sets obtained from literature are spot checked to make sure that numbers, decimal places, and units are correct. Original literature source for that data, along with a link to the website or model, will be included for the value. Spot checks are carried out on ten percent of the collected values. All original documents are retained by the PI.

C2. External Assessment and Response Actions

At the discretion of the researchers as well as the QAM a mid-project technical system audit may be requested by the project QAM to assess the quality assurance process of the project. Any findings, research best practices, or improvements will be compiled in a report and filed with other QA documents. Any corrective actions and responses will be documented and filed with the Assessment Report. These findings and corrective measures will be in the office of the Principal Investigator.

C3. Reports to Management

The director and the QAM will be apprised of progress on a semi-annual basis and at the end of the study, or as and when the data will be presented during work-in-progress meetings or at scientific meetings. The types of progress reports may include brief (that is, 1-5 page) progress reports, presentations, scientific posters, meeting abstracts, and/or manuscript drafts. Dr. Wambaugh will be responsible for preparing and distributing the reports.

Upon EPA clearance products will be published as part of a peer-reviewed manuscript in which the data are included (potentially as supplemental material). A new version of the R package "httk" (https://CRAN.R-project.org/package=httk) will be publicly released containing the data. The processed chemical-specific data will be provided to the Comptox Chemicals Dashboard team (https://comptox.epa.gov/dashboard) as "ADME" (absorption, distribution, metabolism, excretion) information for public release. The data will also be archived with EPA's Science Inventory.

Part D. Data Evaluation and Usability

D1. Data Review and Verification

Model coding is reviewed by a second member of the team and tested using R automated testing functionality. Secondary data is visually assessed initially and spot for numbers, decimal places, and units on ten percent of the collected values. Refer to Section C1. Internal Assessments and Response Actions.

D.2 Evaluation

Models are assessed for bias and uncertainty and evaluated for fit to existing data sets using appropriate statistical methods. We expect a generic model to be less accurate with respect to reproducing *in vivo* ADME measurements, but more likely to be accurately reported, reproducible, and statistically-evaluated.

When *in vivo* measured TK data (that is, chemical concentration vs. time) are available, these data allow evaluation of TK models. The predictions of the chemical specific model may be assessed for both bias and uncertainty (Chiu, Barton et al. 2007), as illustrated in Fig. Figure 1a. When no *in vivo* TK data are available, a generic PBTK model can instead be parameterized (Leahy, 2006; Pearce, et al., 2017b; Poulin & Theil, 2002; Schmitt, 2008). To evaluate that model, overall predictions can be compared to *in vivo* TK data for those chemicals with data, as illustrated in Figure 1b. Although predictions generated for any one chemical using a generic model can be expected to have larger uncertainty than those from a chemical-specific model, there can be greater confidence that the model structure has been appropriately implemented since it has been evaluated against a larger amount of data (Clark, Setzer et al. 2004, Barton, Chiu et al. 2007, Wambaugh, Wetmore et al. 2015).

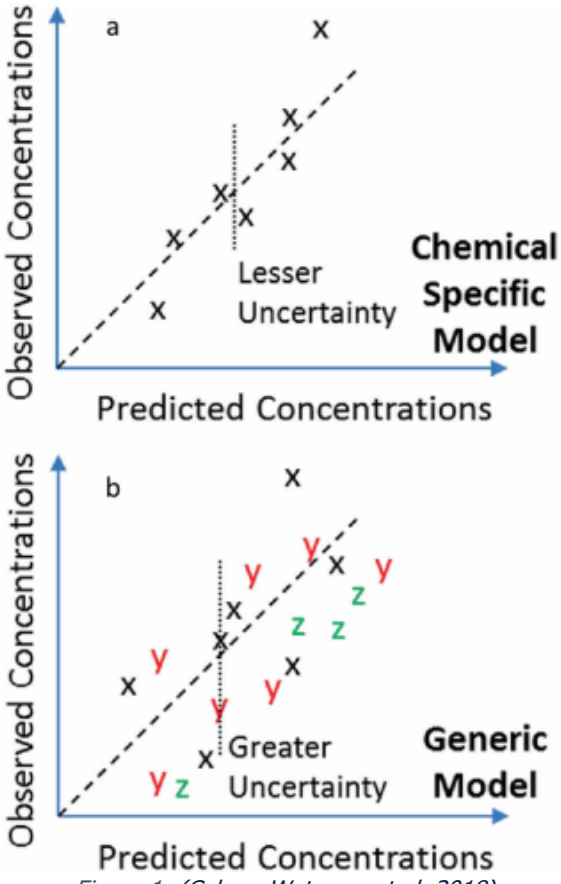

Figure 1: (Cohen, Wetmore et al. 2018)

Clark (Clark, Setzer et al. 2004) identified six steps for the evaluation of PBTK models, which are also applicable to the empirical models used in R package "httk", these are the assessment of 1) model purpose, 2) model structure and biological appropriateness, 3) mathematical description, 4) computer implementation, 5) model fitness and parameter values, 6) specialized analyses. The use of generic models in a standardized environment helps address the first four points. Issues of parameter appropriateness and model fitness require further evaluation using tools such as statistical regression and machine learning (Pilkey and Pilkey-Jarvis 2007). The R statistical programming environment provides thousands of approaches for this evaluation, while databases like CvTdb (Sayre, Wambaugh et al. 2020) provide observations to which predictions may be compared.

D.3 Reconciliation with User Requirements

The products being generated expand the models and data available for predicting toxicokinetics from in vitro data. By quantitatively evaluating uncertainty across a range of chemicals, and correlating errors with chemical features, it is possible to estimate the predictive accuracy of these tools to support chemical risk decision

making. The R package "httk" provides a documented collection of tools and data, but ongoing expansion and curation will lead to the identification, and in some cases unintentionally cause the creation, of bugs in the code.

The existence of a bug or suggestion for a new feature is first documented (currently via Google Docs but potentially with Kira) on the "HTTK Bugs and Ideas List":

https://docs.google.com/document/d/1ejkTuFmh46UIYpb495bKaNs2FtI1HILEgl_lT4G6gFM/edit

When addressing an issue from the list a Git branch containing code and/or data addressing the bug will be used. Once the issues is confirmed to be properly addressed, the NEWS file will be updated and the issue will be marked as resolved. The new code will be bundled into either a patch or feature branch for new release of HTTK.

When modifying R package "httk" care should be taken to maintain consistency (as in, names of parameters and function arguments), keeping documentation current, providing relevant examples, ensuring all necessary functions and data re included, identifying and including relative literature references, and most importantly paying careful attention to units on data, parameters, and predictions.

Part E. Supporting Information

E.1 References

Barton, H. A., W. A. Chiu, R. W. Setzer, M. E. Andersen, A. J. Bailer, F. Y. Bois, R. S. DeWoskin, S. Hays, G. Johanson and N. Jones (2007). "Characterizing uncertainty and variability in physiologically based pharmacokinetic models: state of the science and needs for research and implementation." Toxicological Sciences **99**(2): 395-402.

Bell, S. M., X. Chang, J. F. Wambaugh, D. G. Allen, M. Bartels, K. L. R. Brouwer, W. M. Casey, N. Choksi, S. S. Ferguson, G. Fraczkiewicz, A. M. Jarabek, A. Ke, A. Lumen, S. G. Lynn, A. Paini, P. S. Price, C. Ring, T. W. Simon, N. S. Sipes, C. S. Sprankle, J. Strickland, J. Troutman, B. A. Wetmore and N. C. Kleinstreuer (2018). "In vitro to in vivo extrapolation for high throughput prioritization and decision making." Toxicology in Vitro **47**: 213-227.

Brightman, F. A., D. E. Leahy, G. E. Searle and S. Thomas (2006). "Application of a generic physiologically based pharmacokinetic model to the estimation of xenobiotic levels in human plasma." Drug Metabolism and Disposition **34**(1): 94-101.

Chiu, W. A., H. A. Barton, R. S. DeWoskin, P. Schlosser, C. M. Thompson, B. Sonawane, J. C. Lipscomb and K. Krishnan (2007). "Evaluation of physiologically based pharmacokinetic models for use in risk assessment." Journal of Applied Toxicology **27**(3): 218-237.

Clark, L. H., R. W. Setzer and H. A. Barton (2004). "Framework for evaluation of physiologically-based pharmacokinetic models for use in safety or risk assessment." Risk Analysis **24**(6): 1697-1717.

Cobelli, C. and J. J. Distefano 3rd (1980). "Parameter and structural identifiability concepts and ambiguities: a critical review and analysis." American Journal of Physiology-Regulatory, Integrative and Comparative Physiology **239**(1): R7-R24.

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

Cohen, E. H., B. Wetmore, J. Wambaugh, H. El-Masri, J. Sobus and T. Bahadori (2018). "Advancing internal exposure and physiologically-based toxicokinetic modeling for 21st-century risk assessments." Journal of exposure science & environmental epidemiology.

Cohen Hubal, E. A., B. Wetmore, J. Wambaugh, H. El-Masri, J. Sobus and T. Bahadori (2018). "Advancing internal exposure and physiologically-based toxicokinetic modeling for 21st-century risk assessments." Journal of Exposure Science and Environmental Epidemiology.

Dix, D. J., K. A. Houck, M. T. Martin, A. M. Richard, R. W. Setzer and R. J. Kavlock (2007). "The ToxCast program for prioritizing toxicity testing of environmental chemicals." Toxicological Sciences **95**(1): 5-12.
Edginton, A. N., W. Schmitt and S. Willmann (2006). "Development and evaluation of a generic physiologically based pharmacokinetic model for children." Clinical pharmacokinetics **45**(10): 1013-1034.
Garcia, R. I., J. G. Ibrahim, J. F. Wambaugh, E. M. Kenyon and R. W. Setzer (2015). "Identifiability of PBPK models with applications to dimethylarsinic acid exposure." Journal of pharmacokinetics and pharmacodynamics **42**(6): 591-609.
Jongeneelen, F. J. and W. F. T. Berge (2011). "A generic, cross-chemical predictive PBTK model with multiple entry routes running as application in MS Excel; design of the model and comparison of predictions with experimental results." Annals of Occupational hygiene **55**(8): 841-864.
Linakis, M. W., R. R. Sayre, R. G. Pearce, M. A. Sfeir, N. S. Sipes, H. A. Pangburn, J. M. Gearhart and J. F. Wambaugh (2020). "Development and Evaluation of a High Throughput Inhalation Model for Organic Chemicals " Journal of Exposure Science & Environmental Epidemiology.
McLanahan, E. D., H. A. El-Masri, L. M. Sweeney, L. Y. Kopylev, H. J. Clewell, J. F. Wambaugh and P. Schlosser (2012). "Physiologically based pharmacokinetic model use in risk assessment—why being published is not enough." Toxicological Sciences **126**(1): 5-15.
McNally, K., R. Cotton, A. Hogg and G. Loizou (2014). "PopGen: a virtual human population generator." Toxicology **315**: 70-85.
Oreskes, N. (1998). "Evaluation (not validation) of quantitative models." Environmental Health Perspectives **106**(Suppl 6): 1453.
Pilkey, O. H. and L. Pilkey-Jarvis (2007). Useless arithmetic: why environmental scientists can't predict the future, Columbia University Press.
Ring, C. L., R. G. Pearce, R. W. Setzer, B. A. Wetmore and J. F. Wambaugh (2017). "Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability." Environment International **106**: 105-118.
Rotroff, D. M., B. A. Wetmore, D. J. Dix, S. S. Ferguson, H. J. Clewell, K. A. Houck, E. L. LeCluyse, M. E. Andersen, R. S. Judson, C. M. Smith, M. A. Sochaski, R. J. Kavlock, F. Boellmann, M. T. Martin, D. M. Reif, J. F. Wambaugh and R. S. Thomas (2010). "Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening." Toxicological Sciences **117**(2): 348-358.
Sayre, R. R., J. F. Wambaugh and C. M. Grulke (2019). Database of pharmacokinetic time-series data and parameters for XX environmental chemicals. U. S. E. P. Agency.
Sayre, R. R., J. F. Wambaugh and C. M. Grulke (2020). "Database of pharmacokinetic time-series data and parameters for 144 environmental chemicals." Scientific Data.
Schmidt, C. W. (2009). TOX 21: new dimensions of toxicity testing, National Institute of Environmental Health Sciences.

Schmitt, W. and S. Willmann (2004). "Physiology-based pharmacokinetic modeling: ready to be used." Drug Discovery Today: Technologies **1**(4): 449-456.
Sipes, N. S., J. F. Wambaugh, R. Pearce, S. S. Auerbach, B. A. Wetmore, J.-H. Hsieh, A. J. Shapiro, D. Svoboda, M. J. DeVito and S. S. Ferguson (2017). "An Intuitive Approach for Predicting Potential Human Health Risk with the Tox21 10k Library." Environmental Science & Technology **51**(18): 10786-10796.
Spiegelhalter, D. J., N. G. Best, B. P. Carlin and A. Van Der Linde (2002). "Bayesian measures of model complexity and fit." Journal of the royal statistical society: Series b (statistical methodology) **64**(4): 583-639.
Tan, Y.-M., K. H. Liao, R. B. Conolly, B. C. Blount, A. M. Mason and H. J. Clewell (2006). "Use of a physiologically based pharmacokinetic model to identify exposures consistent with human biomonitoring data for chloroform." Journal of Toxicology and Environmental Health, Part A **69**(18): 1727-1756.
Van der Molen, G., S. Kooijman and W. Slob (1996). "A generic toxicokinetic model for persistent lipophilic compounds in humans: an application to TCDD." Fundamental and applied Toxicology **31**(1): 83-94.
Wambaugh, J. F., B. A. Wetmore, R. Pearce, C. Strope, R. Goldsmith, J. P. Sluka, A. Sedykh, A. Tropsha, S. Bosgra and I. Shah (2015). "Toxicokinetic triage for environmental chemicals." Toxicological Sciences **147**(1): 55-67.
Wambaugh, J. F., B. A. Wetmore, R. Pearce, C. Strope, R. Goldsmith, J. P. Sluka, A. Sedykh, A. Tropsha, S. Bosgra, I. Shah, R. Judson, R. S. Thomas and R. W. Setzer (2015). "Toxicokinetic Triage for Environmental Chemicals." Toxicological Sciences **147**(1): 55-67.
Wetmore, B. A. (2015). "Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment." Toxicology **332**: 94-101.
Wetmore, B. A., J. F. Wambaugh, B. Allen, S. S. Ferguson, M. A. Sochaski, R. W. Setzer, K. A. Houck, C. L. Strope, K. Cantwell, R. S. Judson, E. LeCluyse, H. J. Clewell, R. S. Thomas and M. E. Andersen (2015). "Incorporating High-Throughput Exposure Predictions With Dosimetry-Adjusted In Vitro Bioactivity to Inform Chemical Toxicity Testing." Toxicological Sciences **148**(1): 121-136.
Wetmore, B. A., J. F. Wambaugh, S. S. Ferguson, M. A. Sochaski, D. M. Rotroff, K. Freeman, H. J. Clewell, 3rd, D. J. Dix, M. E. Andersen, K. A. Houck, B. Allen, R. S. Judson, R. Singh, R. J. Kavlock, A. M. Richard and R. S. Thomas (2012). "Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment." Toxicological Sciences **125**(1): 157-174.
Wickham, H., P. Danenberg, G. Csárdi and M. Eugster (2020). roxygen2: In-Line Documentation for R.
Xie, Y., J. J. Allaire and G. Grolemund (2018). R markdown: The definitive guide, CRC Press.
Yamaoka, K., T. Nakagawa and T. Uno (1978). "Application of Akaike's information criterion (AIC) in the evaluation of linear pharmacokinetic equations." Journal of pharmacokinetics and biopharmaceutics **6**(2): 165-175.

E.2 Acronyms and Abbreviations

ADME – absorption, distribution, metabolism, and excretion
CDC -- U.S. Centers for Disease Control and Prevention
CCTE – Center for Computational Toxicology and Exposure
CPHEA – Center for Public Health and Environmental Assessment
CRAN – the Comprehensive R Archive Network
CSV – Commas separated values file format
HTTK – High Throughput Toxicokinetics
IVIVE -- *in vitro-in vivo* extrapolation
ORISE – Oak Ridge Institute for Science Education
PBTK -- physiologically-based toxicokinetic
QA – Quality Assurance
QAM – Quality Assurance Manager
QAPP – Quality Assurance Project Plan
RMSE -- root mean squared error
RPE – relative predictive error
SOP – standard operating procedure
TK -- toxicokinetic