Database of pharmacokinetic time-series data and parameters for evaluating the safety of environmental chemicals

Risa R. Sayre[1,2], John F. Wambaugh[1], Christopher M. Grulke[1]

1. U.S. Environmental Protection Agency, National Center for Computational Toxicology, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, USA. 2. Oak Ridge Institute for Science and Education, Oak Ridge, TN, USA

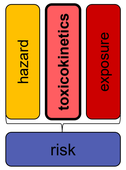
www.epa.gov

Risa R. Sayre | sayre.risa@epa.gov | ORCiD 0000-0002-6173-8020

Christopher Grulke | grulke.chris@epa.gov | ORCiD 0000-0001-5606-7560

1. Background

Time courses of compound concentrations in plasma are used to evaluate the relationship between external administered doses and internal tissue exposures. This type of experimental data is rarely available for the thousands of chemicals to which people may potentially be exposed. An understanding of pharmacokinetics can be developed using *in vitro* assays and *in silico* models, but the quantitative uncertainty cannot be determined without *in vivo* data for external validation. These data were identified as a key area needed for improvement of chemical safety prioritization in a recent review by authors from U.S. government, European government, academia, and industry (Bell et al. 2018). To address this need, we present an extensible public database of pharmacokinetic time-series and parameters.



Summary: Data points describing **changes in amounts of a chemical in an tissue over time after an administered dose** were **extracted** from publications and databases, and **stored** with accompanying contextualizing information in a reproducible way for **usage** in risk characterization or model evaluation.

2. Methods

2.1 Definitions of database terminology

Study: a pharmacokinetic experiment in which a type of *subject*, or organism on which the experiment is performed, is exposed to a single test substance. A study may be described by several series if time-course data is available from several individual subjects or groups of subjects, in several media (tissue, plasma, etc.), or multiple analytes (for example, a study could measure a dosed substance in plasma and its metabolite in urine).

Series: a set of concentration time-series (**CvT**) points measured during a pharmacokinetic experiment, or ***study***. A series may represent data from a single subject, or the mean value from a group of subjects (sometimes with error bars).

Document: a source of experimental data. A study may have more than one document linked to it if it was captured from a paper that was not written by those who performed the experiment.

2.2 Data source identification

Data was sourced from the Chemical Effects in Biological Systems (CEBS) database hosted by the National Toxicology Project, and from individual publications. We used machine learning in Python 3.6 to identify candidate publications from PubMed using Medical Subject Headings (MeSH) terms (F1 score: 0.16) and **abstracts** (F1 score: 0.45).

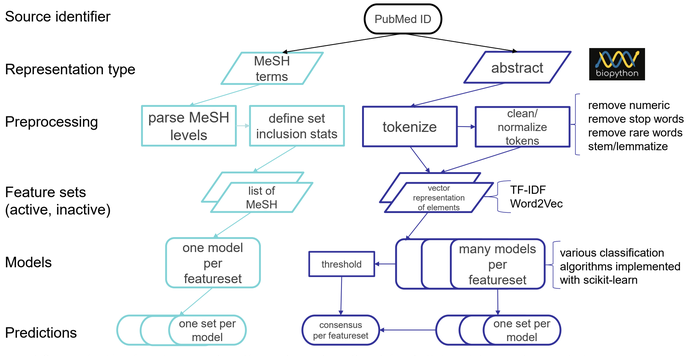


Figure 1. Overview of workflows for machine learning to identify publications containing CvT data from PubMed records *(manuscript in preparation)*.

3. Data Records

3.1 Data record storage

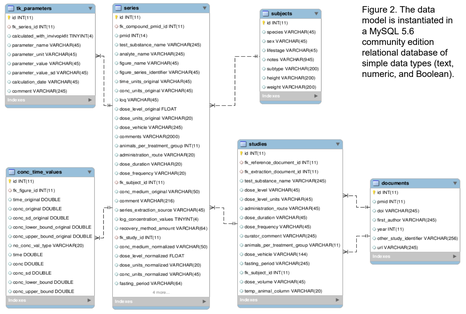


Figure 2. The data model is instantiated in a MySQL 5.6 community edition relational database of simple data types (text, numeric, and Boolean).

3.2 Data collection

CvT data: Efforts at automated data extraction from images were not found to be sufficiently accurate or precise, so we used manual capture.

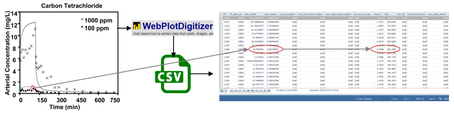


Figure 3. Image files (in this case, from Ng 2007) were uploaded to the webtool *WebPlotDigitizer*. The tool was calibrated to the plot axes, then each data point was clicked (left red oval). Results were exported into comma-separated value transaction files for database import (central red oval). Time units were normalized to *hour*, and concentrations were normalized to the micro scale; all values were rounded to four decimal places (right red oval).

Toxicokinetic parameters: Calculated values summarizing the absorption, distribution, metabolism, and excretion (ADME) represented in a series are included in this dataset, either as presented in original publications and/or as calculated from the CvT data points using the R package *invivopkfit* as described in Wambaugh et al. 2018. It fits the curve as noncompartmental, one-compartment, or two-compartment kinetics, then calculates the corresponding parameters (such as total clearance and concentration at steady-state) and their uncertainty.

Metadata: Manually captured study details based on the test guideline for metabolism and pharmacokinetics released by the U.S. EPA Office of Prevention, Pesticides, and Toxic Substances (1998). Some metadata apply to an entire study, while some are specific to a series. Inclusion in this set does not imply that studies meet OPPTS testing requirements.

3.3 Data record content

The total set consists of 16542 series from 601 **studies** for 154 **test substances** and 196 analytes (test substances or their metabolites). Of the test substances, 25 are pharmaceuticals, 28 are pesticides, and the remaining 101 are other environmental chemicals, which includes ingredients from personal care products, industrial chemicals, disinfection or combustion byproducts, and other diverse sources.

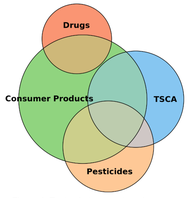


Figure 4. Test substance categories determined through list overlaps

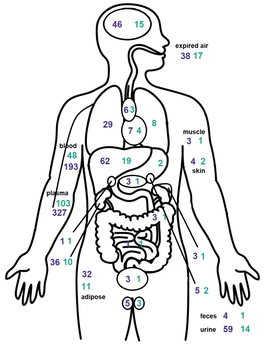


Figure 5. Count of studies and test substances with CvT results in different media (for any species, but represented on a human body).

Figure 6. Count of studies and test substances by species type.

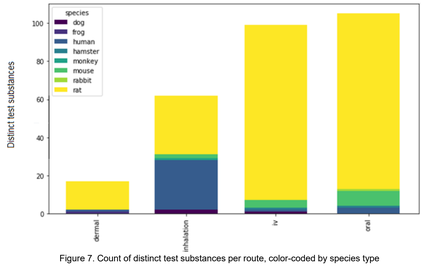


Figure 7. Count of distinct test substances per route, color-coded by species type

5. Data quality

Reporting error: We could not know this value and did not evaluate it

Measurement error in chemical quantification: Recovery amounts will vary based on method and method execution

Capture error due to graphical presentation: Some figures were blurry. Some concentration points were below the datum, which implies a printing error.

Chemical identification validity: Chemicals in unstructured data were usually only identified by their name in the source documents. This leaves some room for ambiguity (Richard & Williams, 2002). Names were mapped to unique substances (designated by DSSTox substance ID, and when possible, structure ID) through expert curation.

Intra- and inter-curator extraction variability: Variability was measured using the fractional difference ($\frac{\sigma_{AUC_{1,\ldots,n}}}{A_{plot}}$) between results collected at different times by the same curator, and between results from different curators.

The fractional difference in AUCs from 13 tested documents was 0.06% for intra-curator, and 0.44% for inter-curator.

Random error due to uncontrollable experimental factors: One document described several trials of the same experiment to serve as controls for different treatment groups. The fractional difference across these trials was 5%.

6. Usage examples

Comparison of compound distribution in different species and media at different doses

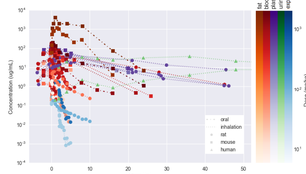


Figure 7. Comparison of all series results for a single test substance from the database.

Evaluation & calibration of physiologically-based pharmacokinetic (PBPK) model results

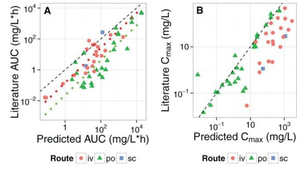


Figure 8 (from Wambaugh et al. 2015) See an example of this usage at "Development of a Generalized Inhalation Model for Use with the High-Throughput Toxicokinetics (*httk*) Package in R", poster P167 in session Computational Toxicology II.

Other ideas:

- Compare ADME of different compounds, and investigate relationship to chemical structure (QSAR)
- Meta-analyses on pharmacokinetic studies
- Calculation of other pharmacokinetic parameters

7. Acknowledgements

The authors would like to acknowledge the many scientists who generated the original *in vivo* data.

Matthew Linakis contributed 65 series of inhalation data.

Bell, S.M. et al. In vitro to in vivo extrapolation for high throughput prioritization and decision making. Toxicol In Vitro. 2018 Mar;47:213-227.
Ng, L.J. et al. Incorporation of acute dynamic ventilation changes into a standardized physiologically based pharmacokinetic model. Inhal Toxicol. 2007 Mar;19(3):247-63.
Richard, A.M. & Williams, C.R. Distributed structure-searchable toxicity (DSSTox) public database network: a proposal. Mutat Res. 2002 Jan 29;499(1):27-52.
Rohatgi, A. WebPlotDigitizer, Version 4.1 https://automeris.io/WebPlotDigitizer
Wambaugh, J.F. et al. Evaluating In Vitro-In Vivo Extrapolation of Toxicokinetics. Toxicol Sci. 2018 May 1;163(1):152-169. doi: 10.1093/toxsci/kfy020.
Wambaugh, J.F. et al. Toxicokinetic Triage for Environmental Chemicals. Toxicol Sci. 2015 Sep;147(1):55-67. doi: 10.1093/toxsci/kfv118.

This project was supported in part by an appointment to the Internship/Research Participation Program at the National Center for Computational Toxicology, U.S. Environmental Protection Agency, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and EPA.

Want to contribute data? The development and population of this database is ongoing. Interested parties should contact the lead or corresponding author.

This poster does not necessarily reflect U.S. EPA policy.