Data Collection of Pharmacokinetic Parameters for Chemicals in the EU-ToxRisk Project. Methods for Collecting and Analysing Data Quality.

Lucinda Camidge, Rosemary Porter, Tony Long, Mark Pinches, Nazanin Golbamaki

Sources of pharmacokinetics data

Absorption, distribution, metabolism, and excretion (ADME) are biological processes that determine the concentrations of a chemical in the body. Physiologically Based Pharmacokinetics (PBPK) models are mathematical representations of these processes and provide estimations of the time-dependent chemical disposition. The PBPK model parameters are separated into physiological parameters for the species studied (e.g. alveolar ventilation, blood flow, tissue volumes, and glomerular filtration rate) and compound-specific parameters, which can be measured or estimated computationally. The latter include Plasma Protein Binding (PPB), blood-to-plasma ratio (B/P), Volume of distribution (V_d), intrinsic clearance (CL_{int}), total clearance (CL_{tot}), half life ($t_{1/2}$), and permeability through gastrointestinal membrane.

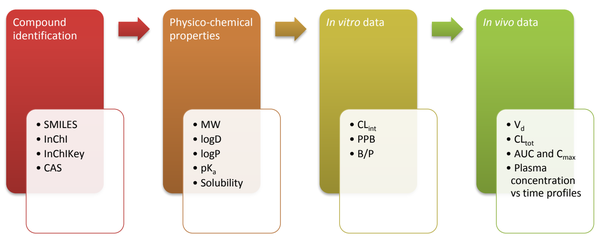


Figure 1. Compound specific data searched in major PBPK databases.

Compound-specific PK parameters were searched in:

- The literature (PubMed and ScienceDirect)
- Public databases: ChEMBL[2], Drugbank[3], PKDB[4], ClinicalTrials.gov[5], httk[6], OChem[7], ADME team[8], PubChem[9], and Physprop[10]
- Reference books: Randall [11] and Goodman and Gilman [12]

The EU-ToxRisk case study chemicals comprised 257 unique compounds. The presence of these chemicals in the identified public PBPK related databases was assessed.

As part of the EU-ToxRisk work package 4 activities, Lhasa Limited has created a PBPK dataset that includes a list of chemical-specific parameters relevant in parameterizing PBPK models. This poster aims to present the research performed on two main areas of the data collection activities:

- identifying sources of compound-specific pharmacokinetics (PK) data
- developing standard guidelines for the data quality assessment (QA)

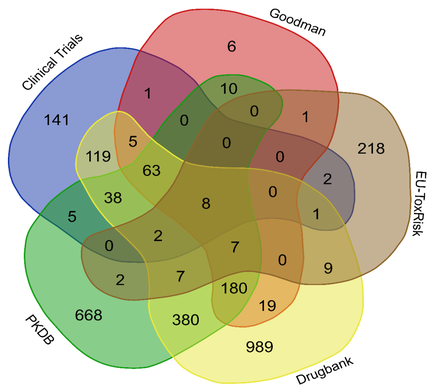


Figure 2. Venn diagram[1] showing availability of kinetics data for the EU-ToxRisk case-study chemicals in individual PBPK databases: PKDB, Drugbank, Goodman and Gilman, and ClinicalTrials.

Data quality assessment

Transparent reporting of research is essential for assessing the validity and quality of results. Nevertheless, whether the data were collected from a publication or a database the following inconsistencies were often present:

- Variations in the domains included in the databases
- Different parameters included in separate databases
- Conflicting data between databases
- Inclusion of both qualitative and quantitative results in domains
- Differences in the reported units
- Differences in the measured parameter

To reduce inconsistencies in the metadata extracted from the sources, the data quality for both *in vitro* and *in vivo* data was assessed by a specific checklist created for each compound-specific PBPK parameter. The checklist was then used to review the metadata before collating the information in the dataset. These lists are divided into two major groups: *in vivo* data and *in vitro* data each including specific guidelines for publications and databases.

In vitro data quality assessment

❑ Publications

To distinguish which *in vitro* data source is of the best quality, different requirements from the data sources were considered. Publications that reported the following were preferred: species, original references (where applicable), standardised units, pooled samples to reduce variability, assay type and methodology, metabolite information/data (where applicable), and protein binding data within the assay.

A further checklist for each compound-specific parameter was used to aid in data selection. For example, with data collection for CL_{int} the following criteria were set for publications:

- Hepatic microsomal assays (E.g. PubChem AID[13]: 89954; CHEMBL697440)
- Standardised units of mL/min/g of protein
- V_{max} and K_m reported
- Enzyme involvement if known

Those publications which met the general requirements, as well as the parameter specific requirements were selected over those which did not report all the above requirements.

❑ Databases

Databases were also assessed on a parameter-by-parameter basis. Each database was evaluated for: chemical space, number of compounds, assay type and methodology, format of the data, and reported chemical identifiers. For the example of PPB, data are reported in ChEMBL[2], DrugBank[3], PKDB[4], httk[6], and OChem[7]. There is an overlap of compounds between these datasets and using the above criteria and the parameter specific checklist, data was selected from the most appropriate source. It was common for the databases to use common references, in which case the original publication was sourced and assessed.

The PPB parameter specific checklist was as follows:

- A single data point or absolute value (rather than a range)
- Consistent with other datasets
- Assay methodology included Rapid Equilibrium Dialysis (E.g. PubChem AID[13]: 269936; ChEMBL869833)
- Human data

Figure 3 demonstrates the process in which conflicted data sources were assessed and subsequent data selected on quality criteria. In Figure 3, a range of data was selected for data inputting as it met more of the QA criteria. In this case, an average from the range would be taken forward: 5.65 % PPB for the compound Busulfan.

Compound Name	PPB(%)	Assay	Reference	PPB(%)	Assay	Reference
BUSULFAN	32	Not Specified	DrugBank	2.7-14	HPLC (Human plasma sample)	Brunton L et al. (2006) Goodman & Gillman's The Pharmacological Basis of Therapeutics. 11th Edition. The McGraw-Hill Companies, Inc.

✓ Absolute Value

↓

x Original Reference: Wikipedia, no further reference provided.

x Range of Values
✓ Original Reference
✓ Methodology detailed
✓ Second reference correlates with value (Obach et al. 2008)

Figure 3. Example of how conflicting data from different sources are assessed for quality, and how selections for data extraction are decided. Example table shows data that does not correlate for the chemical Busulfan. The corresponding circles detail the reasons behind the data selection. Red correlates to the reference excluded. Green correlates to the reference included.

In vivo data quality assessment

❑ Publications

In vivo data are affected by additional demographics and baseline characteristics; therefore, it is crucial for publications to report additional information. To assess the quality of publications, an internal QA tool was developed, Figure 4. This was based on existing reporting guidelines and literature around *in vivo* studies, such as The ClinPK[14] statement.

The relevant variables explaining inter and intra-patient variability can include, but are not limited to:

- Age
- Weight
- Sex
- Race
- Metabolism (extensive vs poor metabolisers for example)
- Health status (especially for renal studies)

One form of *in vivo* data used in PBPK modelling are the plasma concentration-time profiles. Most literature reported these as graphs, from which the AUC was calculated. To gather these individual data points the software GetData Graph Digitiser[15] was used.

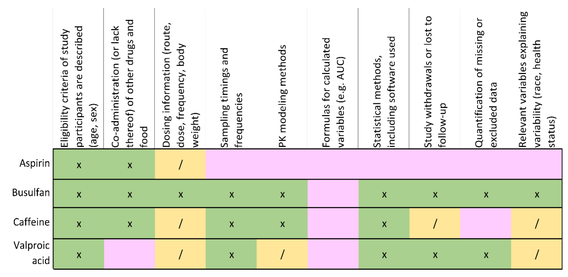
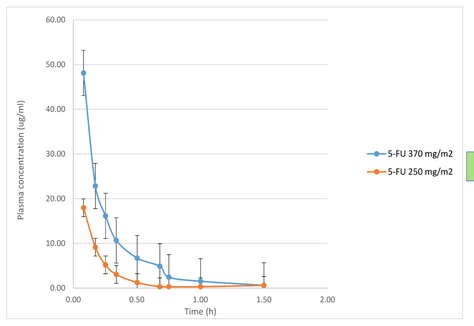


Figure 4. Example of QA tool used in data gathering projects created for *in vivo* studies. Green shows the publication had reported the item correctly; yellow indicated the need for additional information and items in pink were missing from publications.

Figure 5 shows a graph used from a PBPK publication[16] for 5-Fluorouracil (5-FU). An image of the graph was loaded onto the software, which then allowed the user to set maximum and minimum values for both axis. Due to the quality of the image being low, 'Point capture mode' had to be used instead of 'Auto trace lines'. Once all points had been selected, data was then exported into an excel file. From this, further information was added, such as compound identifiers and administration route. Intravenous bolus (IVB) administration required no further information. For per oral data, oral bioavailability and absorption rate were necessary for modelling purposes.



Compound	CAS	Dose	Units	Admin	Time	Units	Value	Units
Fluorouracil	51-21-8	250	mg/m2	IVB	0.08	h	17.97	ug/ml
Fluorouracil	51-21-8	250	mg/m2	IVB	0.17	h	9.14	ug/ml
Fluorouracil	51-21-8	250	mg/m2	IVB	0.25	h	5.18	ug/ml
Fluorouracil	51-21-8	250	mg/m2	IVB	0.34	h	3.05	ug/ml
Fluorouracil	51-21-8	250	mg/m2	IVB	0.50	h	1.22	ug/ml
Fluorouracil	51-21-8	250	mg/m2	IVB	0.68	h	0.30	ug/ml
Fluorouracil	51-21-8	250	mg/m2	IVB	0.75	h	0.30	ug/ml
Fluorouracil	51-21-8	250	mg/m2	IVB	1.00	h	0.30	ug/ml
Fluorouracil	51-21-8	370	mg/m2	IVB	0.08	h	48.12	ug/ml
Fluorouracil	51-21-8	370	mg/m2	IVB	0.17	h	22.84	ug/ml
Fluorouracil	51-21-8	370	mg/m2	IVB	0.25	h	16.14	ug/ml
Fluorouracil	51-21-8	370	mg/m2	IVB	0.34	h	10.66	ug/ml
Fluorouracil	51-21-8	370	mg/m2	IVB	0.50	h	6.70	ug/ml
Fluorouracil	51-21-8	370	mg/m2	IVB	0.68	h	4.87	ug/ml
Fluorouracil	51-21-8	370	mg/m2	IVB	0.75	h	2.44	ug/ml
Fluorouracil	51-21-8	370	mg/m2	IVB	1.00	h	1.52	ug/ml
Fluorouracil	51-21-8	370	mg/m2	IVB	1.50	h	0.61	ug/ml

Figure 5. 5-FU concentration-time graph from literature[16] converted by GetData to individual time points.

❑ Databases

To ensure the inconsistencies found in the data QA are reduced it is preferred that *in vivo* databases comply with the following:

- Homogeneity between parameters including parameter name and units
- Homogeneity between compound identifiers
- Baseline characteristics presented clearly in tabular format. If these are not present, the original source of the data should be provided so these can be found

ClinicalTrials.gov[17] held the greatest source of data for parameters such as AUC and C_{max}. However, extraction proved difficult and time-consuming due to the differences in parameter names. For example, 'Area under the concentration-time curve' and 'AUC' were both used amongst others, meaning there was no common name resulting in manual extraction being carried out, increasing the chance of human error.

Results

❑ Data availability assessment

The total number of records identified for the EU-ToxRisk chemicals in work package 4 are: CL_{int} 30, and PPB 75. These values were obtained from examining references detailed above. The ClinicalTrials database contained AUC, and Cmax data for 13 chemicals in the EU-ToxRisk case-studies. These examples demonstrate the scope of data gaps for parameters within the EU-ToxRisk dataset.

❑Benefits of guidelines for metadata QA

Global and parameter specific quality assessment guidelines:

- Allow one to assess the quality of databases in general, in terms of accessibility and format
- Assess the quality of data sources for specific parameters needed in PBPK modelling
- Provide standard reference lists to compare different datasets and data sources more easily when parameters, units and identifiers are common between them

Conclusions

Lhasa Limited has fostered several efforts to collect PK data from various, scattered sources and compile these data into a single database. Throughout these attempts the need to establish standard guidelines lead to the creation of a set of rules to apply to the identified metadata and prevent populating the dataset with inconsistent information and thereby developing a dataset suitable for use by modellers. In addition, identification of the data gaps is essential for prioritising *in vitro* tests that can overcome the lack of data for specific chemicals of the project. These benefits have yet to be realised and are dependent on the awareness of the available data sources. Hence, the issue of data availability was addressed and presented in the current poster.

References

1.Venn Diagram tool, http://bioinformatics.psb.ugent.be/webtools/Venn/.
2.CHEMBL database release 22.1. (EMBL-EBI, 2016).
3.Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 46, D1074–D1082 (2018).
4.Moda, T. L., Torres, L. G., Carrara, A. E. & Andricopulo, A. D. PK/DB: database for pharmacokinetic properties and predictive in silico ADME models. Bioinformatics 24, 2270–2271 (2008).
5.Zarin, D. A., Tse, T., Williams, R. J., Califf, R. M. & Ide, N. C. The ClinicalTrials.gov Results Database — Update and Key Issues. N. Engl. J. Med. 364, 852–860 (2011).
6.Pearce, R. G., Setzer, R. W., Strope, C. L., Sipes, N. S. & Wambaugh, J. F. httk : R Package for High-Throughput Toxicokinetics. J. Stat. Softw. 79, (2017).
7.Sushko, I. et al. Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. J. Comput. Aided Mol. Des. 25, 533–554 (2011).
8.Legehar, A., Xhaard, H. & Ghemtio, L. IDAAPM: integrated database of ADMET and adverse effects of predictive modeling based on FDA approved drug data. J. Cheminformatics 8, (2016).
9.Kim, S. et al. PubChem Substance and Compound databases. Nucleic Acids Res. 44, D1202–D1213 (2016).
10.Bloch, D. Computer Software Review. Review of PHYSPROP Database (Version 1.0). J. Chem. Inf. Model. 35, 328–329 (1995).
11.Baselt, R. C. Disposition of Toxic Drugs and Chemicals in Man. (Biomedical Publications, 2017).
12.Brunton, L. L., Lazo, J. S. & Parker, K. L. Goodman and Gilman's The Pharmacological Basis of Therapeutics. (The McGraw-Hill Companies, 2006).
13.Wang, Y. et al. PubChem BioAssay: 2017 update. Nucleic Acids Res. 45, D955–D963 (2017).
14.Kanji, S. et al. Reporting Guidelines for Clinical Pharmacokinetic Studies: The ClinPK Statement. Clin. Pharmacokinet. 54, 783–795 (2015).
15.Fedorov, S. GetData Graph Digitizer. (2013).
16.Bocci, G. et al. Comparative pharmacokinetic analysis of 5-fluorouracil and its major metabolite 5-fluoro-5,6-dihydrouracil after conventional and reduced test dose in cancer patients. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 6, 3032–3037 (2000).
17.Zarin, D. A., Tse, T., Williams, R. J. & Carr, S. Trial Reporting in ClinicalTrials.gov — The Final Rule. N. Engl. J. Med. 375, 1998–2004 (2016).

The EU-ToxRisk project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 681002

http://www.eu-toxrisk.eu/