Journal of Pharmacokinetics and Pharmacodynamics (2018) 45:659–661
https://doi.org/10.1007/s10928-018-9595-8

PERSPECTIVE

Data standards for model-informed drug development: an ISoP initiative

Andrijana Radivojevic[1] · Brian Corrigan[2] · Nicholas Downie[3] · Robert Fox[4] · Jill Fiedler-Kelly[5] · Huan Liu[4] · Murad Melhem[6] · David Radke[7] · Peter Schaefer[8] · Jing Su[9] · Maciej J. Swat[10] · Nathan S. Teuscher[11] · Neelima Thanneer[12] · Alice Zong[13] · Justin J. Wilkins[14]

Received: 28 May 2018 / Accepted: 31 May 2018 / Published online: 25 July 2018

Introduction

Analysis datasets are fundamental components of pharmacometric analyses and their quality and readiness highly correlate with the efficiency and impact of pharmacometrics deliverables overall [1]. Despite this, structures of datasets vary widely. This article introduces the activities of International Society of Pharmacometrics (ISoP) Data Standards Group towards establishing standards for pharmacometric datasets. The ultimate goal is to reduce the time required to specify, implement and review datasets, and to facilitate portability within and across organizations.

✉ Andrijana Radivojevic
andrijana.radivojevic@intigrowth.com

[1] Pharmacometrics, Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA

[2] Pfizer Global Research, Groton, CT, USA

[3] Bayer AG, Berlin, Germany

[4] AstraZeneca, Waltham, MA, USA

[5] Cognigen Corporation, Buffalo, NY, USA

[6] Amgen, Thousand Oaks, CA, USA

[7] Eli Lilly and Company, Indianapolis, IN, USA

[8] VCA-Plus, Raleigh, NC, USA

[9] Merck & Co., Kenilworth, NJ, USA

[10] Simcyp (A Certara Company), Sheffield, UK

[11] Certara, Princeton, NJ, USA

[12] Bristol-Myers Squibb, Princeton, NJ, USA

[13] Janssen, Horsham, PA, USA

[14] Occams, Amstelveen, The Netherlands

Data in pharmacometrics

The last two decades have seen remarkable changes in the expectations for data used in decision-making, both in drug development and in the clinic, particularly with respect to the speed of collection, analysis, and communication of results. In parallel, sources of data have grown exponentially in type and quantity, challenging our ability to utilize the available information in a cogent and reproducible manner.

Collecting and assembling data that meets minimum agreed standards have become a priority. The idea of open science is becoming central, with health authorities increasingly demanding open access to data [2].

In parallel, reproducible research in pharmacometrics is becoming both more important and straightforward to implement [3], the establishment of good practices in pharmacometrics is gaining momentum [4], production-ready open-source tools are appearing more frequently, and consortia such as DDMoRe [5] and the Critical Path Institute [6] continue to make great contributions to the field. Sharing of model code, workflows and data is being heavily promoted as key in open innovation [7, 8], and data standards will be a critical piece of this surge towards reproducibility and industrialization. Pharmacometric analyses are dependent upon the quality of analysis datasets, which is driven by the quality of reference data sources and preparation processes. The datasets are often difficult to review and share (both internally and externally), as they tend to be developed individually and specifically to cater to the needs of specific analyses, tools, data programmers and pharmacometricians [9]. Consequently, variables are often named and coded inconsistently and without proper documentation, creating significant hurdles in preparing datasets, describing them, sharing them with collaborators and regulators, and reusing them.

Missing data is another major challenge. Missing information can be imputed, and although data imputation should be avoided whenever possible and different methods for doing so exist, standard guidelines on rules for imputation should bring additional quality and efficiency to pharmacometric datasets [10].

Widely used analysis software tools such as NONMEM (ICON Development Solutions, Dublin, Ireland) and Monolix (Lixoft, Antony, France) require input data to be formatted in specific ways, which are often inconsistent with one another, adding an additional layer to the process.

Sources of data are also important, particularly when datasets derived from different studies are being pooled, as is common practice in pharmacometric analyses. Data are usually extracted from Clinical Data Interchange Standards Consortium (CDISC, http://www.cdisc.org) Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) datasets in SAS (SAS Institute, Cary, NC, USA) format, and refactored based on the requirements of the analyst and the tool(s) to be used for the analysis. Both utilize standard domains designed to organize data with consistency, clarity and traceability for regulatory reporting purposes. Unfortunately, concordance across standardized domains is often lacking, owing to study design, data collection, and data cleaning processes that may not consider the utility of data for time-ordered pharmacometric analysis.

Data specification and subsequent preparation is often slow in larger organizations as a result of these factors. Issues unanticipated by analysts may result in imprecise data specifications, making unambiguous interpretation by data programmers difficult. The use of external vendors may further complicate this. Alternatively, when specifications do not exist in an organization, datasets are prepared according to the needs of the analyst and the analysis at hand, and are often poorly documented, creating issues when questions arise after analyses have been completed.

Recognizing these challenges, many organizations are launching internal efforts to improve the processes underlying data preparation, including standardization of methods to define and communicate data requirements.

While this is unarguably useful, more is needed: we propose the development of a common set of data standards for the interchange and analysis of pharmacometric data, flexible enough to fit into established processes, organizational structures, and functional preferences, in order to increase the impact of pharmacometrics on clinical research and drug development.

The benefits of standards

The anticipated benefits of a common, tool-agnostic standard for pharmacometric data are many. Of these, a considerable increase in consistency across the community as a whole is perhaps the most important benefit, allowing data to be prepared, reviewed, used, shared, and reused with higher efficiency within and across organizations. Standards will serve to support and ensure regulatory compliance and audit-readiness. Quality would also be improved: a standardized lexicon and data dictionary would result in fewer errors in the datasets themselves leading to higher quality of the analyses and resulting recommendations. Finally, standards would facilitate the automation or semi-automation of data preparation.

Many of these activities could be supported by freely-available and standardized software tools, which would further increase efficiency. Standards will make training of programmers and pharmacometricians both faster and easier. They will provide a knowledge sharing platform, help to close communication gaps among different function groups, both inside and between institutions, and facilitate stable and streamlined processes—which will in turn increase quality and reduce cycle time. All of these are anticipated to result in more time for scientists and programmers to focus on their mission: understanding and making decisions based on data.

A data standard for population PK analysis

Population PK analysis is used to describe factors influencing variability in the dose-exposure relationship between and within individuals in the target population, and is often used as a basis for simulations to inform dose selection. Such analyses are performed regularly throughout development, and are key components of regulatory submissions. While relatively complex, population PK datasets nonetheless represent a relatively straightforward starting point for standardization: all such analyses rely on data items corresponding to time, drug concentrations, drug dosing, and individual identifiers.

We propose that a standard for population PK datasets should be tool-agnostic, and should include the following aspects:

- *A standard core specification* Standard variable names, labels, descriptions, and units for commonly used variables will be specified. The standard will include naming conventions for different classes of variables (e.g. whether it was collected at baseline or varies with time). The core will not change between studies and analyses, and provides traceability and consistency.

- *Source data* Source SDTM/ADaM data fields will be identified for all standard core variables, ensuring consistency within and across organizations.
- *Derived data* Guidance for creating common core derived variables (such as observation occasion) in a standardized and consistent manner will be provided.
- *Missing data* Guidance on how to handle missing data and information gaps, and standardized, consistent rules for imputation, flagging and derivations for common scenarios will be provided.
- *Controlled terminology and code list* The standard will support CDISC controlled terminology for character values and introduce standard code lists for numeric representations (e.g. sex, race, country, yes/no).
- *An implementation guide* will provide a detailed description of how to apply the standards in practice.

Outlook

The first version of our draft PK data standard will be released shortly. The next step will be to harmonize the draft standard with existing CDISC standards, and we hope software vendors will include native support for these standards in the medium to long term.

The proposed standard only covers input data. A standard for output data—population and individual model parameters, metadata, and simulation output—is needed as well. Some of the work needed for this has already been accomplished by the DDMoRe consortium (as part of standard output, SO [http://www.ddmore.eu/projects/so-standard-output]), but some adaptation and extension will be needed.

Ultimately, we wish to be able to extend this approach to more complex datasets for more complex problems (including longitudinal PK-PD, as well as problems involving odd-type data).

We envisage these to be evolving standards, and welcome collaboration from any and all interested parties.

Acknowledgements The authors thank and acknowledge the contributions of ISoP Data Standards initiative group members: Ted Grasela, Jeffry Florian, Timothy Bergsma, Warwick Benger, Christian Rasmussen, Henning Schmidt, Liping Zhang, Scott Pivirotto, Stuart Pearce, Pragathi Kotha Venkata, Vishak Subramoney and Ana Henry, as well as René Bruno for his feedback on the manuscript.

Compliance with ethical standards

Conflict of interest A.R. is an employee of IntiGrowth LLC, B.C. of Pfizer Global Research, N.D. of Bayer AG, R.F. and H.L. of AstraZeneca, J.F.-K. of Cognigen Corporation (a SimulationsPlus company), M.M. of Amgen Inc., D.R. of Eli Lilly and Company, P.S. of VCA-Plus, J.S. of Merck, M.J.S. of Simcyp (a Certara company), N.S.T. of Certara, N.T. of Bristol-Myers Squibb, A.Z. of Janssen Pharmaceutical Companies of Johnson & Johnson, J.J.W. of Occams.

References

1. Grasela T, Fiedler-Kelly J, Hitchcock D (2009) Improving the efficiency and ensuring the quality of data assembly for pharmacometric analysis. PAGE 18 Abstr 1552. www.page-meeting.org/?abstract=1552
2. European Medicines Agency policy on publication of clinical data for medicinal products for human use. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2014/10/WC500174796.pdf
3. Wilkins JJ, Jonsson EN (2013) Reproducible pharmacometrics [tutorial]. PAGE 22 Abstr 2774. www.page-meeting.org/?abstract=2774
4. Marshall SF et al (2016) Good practices in model-informed drug discovery and development: practice, application, and documentation. CPT Pharmacomet Syst Pharmacol 5(3):93–122
5. Harnisch L, Matthews I, Chard J, Karlsson MO (2013) Drug and disease model resources: a consortium to create standards and tools to enhance model-based drug development. CPT Pharmacomet Syst Pharmacol 2:e34
6. Woosley RL, Myers RT, Goodsaid F (2010) The critical path institute's approach to precompetitive sharing and advancing regulatory science. Clin Pharmacol Ther 87(5):530–533
7. Romero K et al (2010) Pharmacometrics as a discipline is entering the "industrialization" phase: standards, automation, knowledge sharing, and training are critical for future success. J Clin Pharmacol 50(9 Suppl):9S–19S
8. Karlsson MO, Conrado D, Romero K, Sarr C, Wilkins JJ (2017) Open innovation: towards sharing of data, models and workflows. Eur J Pharm Sci 109S:S65–S71
9. Radivojevic A, Schmidt H (2015) Datasets for pharmacometric analyses: internal review and standardization efforts. J Pharmacokinet Pharmacodyn 42:S81–S82
10. Thanneer N, Roy A (2014) Best practices for preparation of pharmacometric analysis data sets. ACoP5. http://www.acop7.org/assets/Legacy_ACOPs/ACOP5/Poster_Abstracts/m-058.pdf