Data Standards for Population Pharmacokinetic Analysis (POPPK) and Non-Compartmental Analysis (NCA)

Neelima Thanneer, Bristol-Myers Squibb Company, Princeton, NJ

ABSTRACT

Pharmacometrics (Pm) Programming specializes in creating modeling-ready datasets for Pm analyses. Challenges to create these datasets include the familiar list of accounting for deficiencies in dosing information, standard rules for handling source data issues, and variability in the specifications. Given these challenges the datasets are prepared and documented inconsistently across the companies and sometimes within the companies, making it difficult to review and exchange the datasets. To address this, International society of pharmacometrics (ISoP) has come up with an initiative to standardize the POPPK datasets. This paper will go over the challenges associated with programming POPPK datasets and describes the best practices to handle them. In addition, an update will be provided on the ISoP Data Standards Initiative and ADaM NCA effort from the CDISC PK working group and how that can augment and impact the POPPK and NCA dataset creation process.

INTRODUCTION

POPPK analysis describes the dose-exposure relationship and its variability between and within individuals in the target population in terms of compartmental models. These analyses have become key components in regulatory submissions but the datasets for such analyses are prepared and documented inconsistently. Accurate and transparent preparation and reporting of pharmacometrics analysis datasets is critical factor in ensuring the quality and reproducibility of the Pm analysis, and facilitating its review. The quality of the analyses and results is highly dependent upon the quality of the data.

The POPPK dataset construction process faces many challenges including source data accuracy and completion, merging disparate data sources, and complications with programming logic. This paper will focus on some of the major sources of data issues that pose unique challenges to creating POPPK datasets, and methods to compensate for those issues. The paper also provides an update on the POPPK data standards initiative and describes its application.

SOURCE DATA ISSUES & IMPUTATIONS

Issues related to data imputation are particularly acute for population pharmacokinetic (POPK) datasets, as detailed dosing history is generally not available in the source data[1].

DATA CLEANING

Cleaning of pharmacokinetic (PK) data during the conduct of the study enables recovery of data that may not have otherwise be useable. Standardized edit checks of PK metadata have been developed and implemented within Oracle Clinical (OC), the clinical database, which have enabled cleaning of the dose dates, times and PK sample dates and times during the course of study. The standard checks involve derivation of actual time after previous dose of the samples and comparing it with nominal times to query large deviations between them. Front-loading of data set programming enabled identification of issues that required amendments to the data set specification.

IMPUTATIONS

As much as possible imputations should be avoided, but they may be necessary particularly for POPPK datasets as complete dosing history is generally not available in source data. Below are standard imputation rules for some common scenarios.

Dose date/time imputation typically occurs in below scenarios:

1. Intravenous (IV) dosing where every dose date, time should be recorded

Date of Dose
DD-MMM-YYYY

Time of Dose
HH:MM 24 Hour Clock

If a dose date is available but time is missing:

- If a trough sample was taken on the same day, use the trough time as the dose time (if IV use start time of infusion). One can add 5 min

- If no troughs were taken, use the previous dose time. For BID (twice a day dose)/TID (thrice a day dose) adjust the imputation based on frequency.

- If the first dose time is missing then impute using day 1 lab time or impute based on post first dose sample time. For BID/TID adjust the imputation based on frequency.

- The imputed clock time records should be flagged

If Infusion Duration is missing:

- If infusion stop time is available but infusion start time is missing, the protocol defined duration (e.g. 1 hour or 30 min) is used to determine the start of infusion and vice versa if stop time is not available.

- If both infusion start and stop times are missing on day 1, pre-dose sample time or end of infusion sample time is used along with nominal infusion duration to determine the start of infusion time

- The imputed date/time records should be flagged

2. Oral doses where interval dosing is recorded

Start Date
DD-MMM-YYYY

Stop Date
DD-MMM-YYYY

When Case Report Form (CRF) is designed to capture interval doses with start and stop dates recorded and only dose times relative to PK are recorded

- Variables ADDL and II are derived to capture the not recorded doses for analysis done in NONMEM

- ADDL: Number of additional doses exactly like the current one

- II: Interdose Interval.

- Below example describes how ADDL and II are derived. EVID=0 indicates PK record, EVID=1 indicates dosing record. There are 15 additional doses between 27-AUG-2009 and 12-SEP-2009

every 24 hours.

ID	EVID	DATE	TIME	ADDL	II
1	0	27-AUG-2009	10.30	.	.
1	1	27-AUG-2009	10.45	15	24
1	0	11-SEP-2009	15.05	.	.
1	1	12-SEP-2009	10.45	0	.

- If the dose time relative to PK samples are not recorded then impute using the IV imputation rules
- Programming accounts for missing doses/dose interruptions while applying these imputation rules.

INFLUENCING CRF DESIGN

- CRF's are not consistently designed to capture the information needed for Pharmacometrics analysis
- First dose time should be collected for all interval dosing collection designs
- For BID and TID dosing, CRF should capture the prior 2 and 3 dose times relative to PK.
- Dose interruptions for BID and TID doses should indicate which doses are missing
- Below example shows ADDL=2 in the highlighted record but it should show 1 taking into account the missed AM dose on 21APR14 as total dose indicates 1 capsule. So ADDL derivation should include dose amount in the logic which further adds to the complexity.

ID	DATE	TIME	DAY	EXAMPM	TOTAL_ DOSE	ADDL
1	26MAR14	9:21	58	AM	2	49
1	20APR14	9:21	83	AM	2	2
1	21APR14	23:34	84	PM	1	0
1	22APR14	10:08	85	AM	1	0

These standards help minimize the variability on how different modelers handle imputations and missing data. They also help bring efficiency in programming and consistency across studies and compounds as POPPK analysis is done on a pooled dataset from several studies and it's important to use consistent rules and algorithms.

To meet aggressive timelines, a rigorous, systematic process to meet these challenges can be instrumental and can be most efficiently performed by a dedicated Pm Programming group who understand source data issues, modeling concepts and tool specific requirements. Most companies now have their own Pm programming groups to support this work for efficiency, faster turnaround, better quality and consistency. However, there is still a lot of variability on how different companies prepare their POPPK datasets and in their internal standards. Thus an industry wide standard will provide multitude of benefits.

POPPK DATA STANDARDS

With the sponsorship of International Society of Pharmacometrics (ISoP), the standards group envisions the development of common POPPK data standards for interchange and analysis that would further move the field of pharmacometrics and its impact on drug development. The draft standards have completed open review process. Currently the team is addressing the comments and intends to hand over the final draft of the standards to CDISC for feasibility assessment.

APPLICATION OF THE STANDARD

Described below are some of the benefits of applying the standards. Internally at BMS, POPPK data standards similar to the one described above have been developed and implemented. This has led to the development of:

Web-based spec: Given the standard structure, naming conventions, controlled terminology, rules for imputations and derivations for common core variables, it was possible to develop web-based data specification form with a back end server to store the information. All this information is pre-filled in the standard POPPK template specification. User just needs to select the template and add additional variables specific to study. This cuts down the variability across the specifications from different pharmacometricians. Minimizes the iterations of the specifications and improves efficiency and consistency.

Programming: Developed SAS® macros to semi-automate the programming for PPK datasets. There are still some non-standard issues related to source data and study specific information that need to be accounted for but large part of the programming is automated by applying the standard structure, naming conventions and rules for imputations and derivations.

Quality Control: Automated quality control (QC) to improve the QC process of the POPPK datasets[2].

Developed a checklist to capture the data issues specific to the pharmacometrics datasets. The tool reads the standard PPK dataset and performs below checks listed in the checklist. Some of them are listed below:

- Generates bar, scatter and histogram plots of continuous and categorical variables

- Outputs frequency table of all variables in the dataset

- Provides summary statistics for all numeric variables (Table 1)

- Other checks in the QC Checklist (Table 2)

- Scatter plots of nominal and actual times (Figure 1), nominal and actual doses, dosing over time and concentrations over time

- Bar charts of lab results and physical measurements (Figure 2).

- Normal ranges from the National Library of Medicine MedlinePlus for various lab tests are integrated into the bar charts to compare how the values with normal range values.

- Ensures the dataset and variable attributes are same between specs and the dataset.

	n	max	min	mean	median	missing
BSTUDYN	8698	40.0	40.0	40.0	40.0	0
NMID	8698	4070678.0	4026.0	2445350.577	4018118.0	0
ATAFD	8697	30287.9	-16858.47	4214.178	2736.72	1
NTAFD	8444	30288.0	0.0	4233.402	2688.0	254
NTAPD	8444	336.0	0.0	67.146	0.0	254
ATAPD	8297	7056.17	0.0	130.909	0.0	401
EVID	8698	1.0	0.0	0.69	1.0	0
MDV	8698	1.0	0.0	0.741	1.0	0
AMT	6002	1905.0	4.87	218.729	215.0	2696
RATE	6002	2526.32	4.62	272.298	234.0	2696
DUR	6002	2.0	0.13	0.878	1.0	2696
CONC	2255	429.0	0.25	39.813	27.1	6443
CONCSTAT	2696	2.0	0.0	0.164	0.0	6002
LOGC	2255	6.061	-1.386	3.062	3.3	6443
NDOSE	8229	257.0	0.1	22.168	3.0	469
DOSENUM	8297	79.0	1.0	12.832	9.0	401
OCC	8698	27.0	0.0	4.426	3.0	0
AGE	8698	83.0	18.0	61.573	63.0	0
SEXN	8698	2.0	1.0	1.197	1.0	0
BBWT	8698	148.1	38.3	71.997	70.7	0
BALB	8698	4.9	2.6	3.897	3.9	0

Table 1

| QC Checks	Pass	Detail/Note
All variable names do not exceed 8 characters.	Yes	N/A
The dataset is sorted as described in the specification.	N/A	Program error or not enough information.
All variable label do not exceed 40 characters.	Yes	N/A
Unique count of USUBJID equal to unique count of NMID.	Yes	Unique count of USUBJID: 406 Unique count of NMID: 406
Formats are not associated with variables in the dataset.	Yes	N/A
All variables defined in the dataset specification are present in the dataset.	No	Following variables are not in the dataset: CROSSOVR
Check variable type matches the specification.	No	Check QC Function "Check Variable Type" for more info
No subject with PK records but no dosing records.	No	Following subject have only pk records:
Variables are ordered in the dataset as defined in the specification	Yes	N/A

Table 2
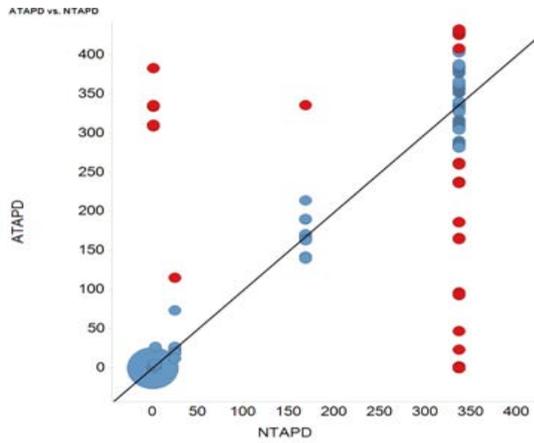
Figure 1
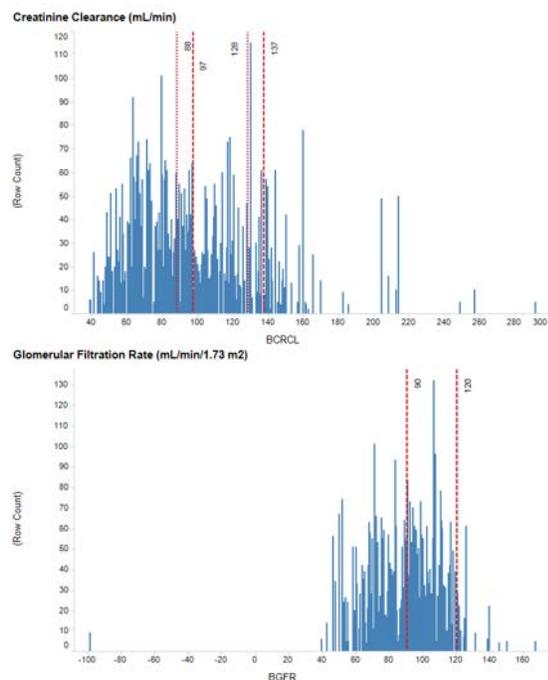
Figure 2

Electronic Submission: Generation of e-sub complaint POPPK datasets and define.pdf has been fully automated. SAS® macros have been developed to read the standard POPPK dataset, the dataset specification form and export the e-sub compliant SAS® transport file and define.pdf.

NCA DATA STANDARDS

The process of creating and coming to consensus on a standardized way of assembling and presenting data takes time, and when decades have passed with each group solving the problem of standardization their own way, bringing those groups together can be a difficult task. With Non-compartmental Analysis (NCA), however, the endeavor brings a little more potential for success. While some companies still use customized solutions, a large majority of the academia and the industry rely on a handful of tools to perform NCA and those tools handle data in a very similar fashion. Given the similarity of input formats, an ADaM sub-team has been working on the development of a new ADaM standard for NCA. The team's goal in creating this new standard, is to create a format that has the least impact on the way the majority of companies perform NCA analysis. This standard addresses not only the basic time from dose calculations and concentration data handling, but also patient cofactors often used to subset and present results for many of the different types of early phase studies. This developing standard should allow the format to be used not only for analysis, but also reporting of the basic tables, figures, and listings associated with early phase Pharmacokinetic analysis. The draft standards have been through one public review and a first review by the larger ADaM standards group and are currently undergoing final refinement before release of the draft guide for a final round of public feedback. The ADaM sub-team is hopeful that the establishment of standards for NCA will not only allow for simplification of analysis, but also regulatory submissions, reporting, and even data collection.

CONCLUSION

The implementation of standard rules of imputation and standards for POPPK datasets has enabled improvements in the quality, efficiency, consistency of Pm analysis dataset preparation and documentation of these datasets. It enabled development of automated QC tools and semi-automated web based specifications. Influencing CRF designs to collect key data will reduce programming challenges. Once the standards get published there will be a considerable improvement in the variability of the datasets prepared across the industry and makes data interchange and collaboration efficient.

REFERENCES

[1] Thanneer, N, Roy, A. Best Practices for Preparation of Pharmacometric Analysis Data Sets. ACoP5 (2014). [Abstract]

[2] Dombrowsky, E, Luo, Y, Thanneer, N Use of Python and Spotfire for Quality Control of Pharmacometric Analysis Dataset Preparation. ACoP8 (2017). [Poster]

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Neelima Thanneer
Bristol-Myers Squibb Company
Neelima.thanneer@bms.com