# Best Practices for Preparation of Pharmacometric Analysis Data Sets

**Neelima Thanneer, Amit Roy, Prema Sukumar, Jyothi Bandaru, and Eric Carleen**

*Bristol-Myers Squibb, Princeton, NJ, USA*

## INTRODUCTION

- Timely availability of high quality data sets is essential for accurate and impactful pharmacometric (Pm) analyses.
- Improvements in timeliness and quality are achieved by standardizing data set preparation for Pm analyses, particularly when the data set requires derivations and imputations of variables and data not available in the source data.

## OBJECTIVE

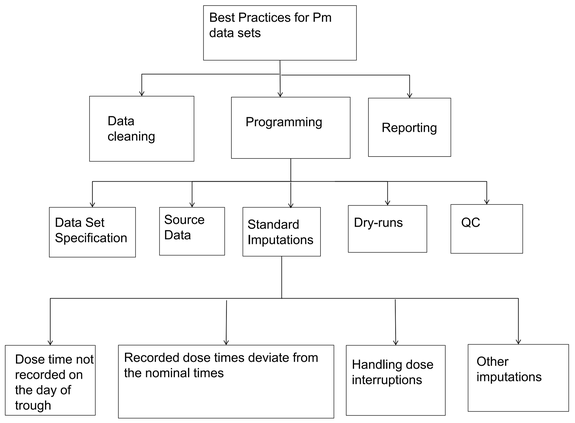
- To describe Best Practices that have enabled high quality data sets to be prepared efficiently.

## METHODS

The three main elements of Best Practices of Pm analysis data set preparation:

- **Data Cleaning**: Standardized edit checks of pharmacokinetic metadata have been developed and implemented within Oracle Clinical (OC), the clinical database, which have enabled cleaning of the PK sample dates and times during the course of study.
- **Programming**: Key attributes of high quality Pm analysis data sets are reproducibility and logical consistency. Developed best practices to program the Pm data sets efficiently while maintaining these key attributes.
  - The key attributes are:
    - Pm data set specification form
    - Standard rules for imputations and derivations
    - Dry-run data set programming
    - Quality Control (QC)
- **Reporting**: The reporting component of Best Practices includes documentation of data set variables, identification of outliers, and informative data summaries.

### Application of Best Practices to Pm data set preparation




## RESULTS

- A typical Pm analysis data set (for analysis using NONMEM) contains subject identification variables, time variables, NONMEM specific event variables, covariates, and flags.

| | |
|---|---|
| ID | This is generated from the unique subject identifier for the study (USUBJID). Standard format is protocol-site-subject. ID is generated sequentially based on USUBJID |
| TIME | Actual time after first dose (ATAFD). For a PK record it is time elapsed from the first dose to the PK measurement |
| ATAPD | Actual time after previous dose. For a PK record it is time elapsed from the last dose to the PK measurement |
| EVID | Event ID: Code for administration or observation, e.g., 1=dose event, 0=observation |
| DV | Dependent variable: Result of the observation |
| CMT | Compartment for observation or dose |
| AMT | Amount of dose administered |
| ADDL | Number of additional doses. Derived for doses without recorded dosing dates/times |
| II | Inter-dose interval |
| OCC | Dosing occasion variable, e.g., occ=1 for 1st occasion of the observation |

- NONMEM data set is built using clinical data from OC and PK concentration data from Watson.

### Pm data set specification form

- The data set structure, variable names, source data and logic are all documented in the specifications.
- Programming algorithms and exclusion flag criteria are documented in a separate section in specs.
- Flags enable exclusion of data from Pm analyses, while retaining the source data in the analysis data set.
- A flag priority is defined in specs to indicate which flag is applied in cases where more than one flag is applicable.

| Variable Name | Variable Label | Units | Type | Notes | Source |
|---|---|---|---|---|---|
| USUBJID | Unique patient identifier: Protocol-Site-Patient | NA | Char | NONMEM ID will be generated from USUBJID | |
| ATAFD | Actual time after first dose | hr | Num | Round to 0.01 | |
| DV | Dependent variable (PK concentration) | ug/mL | Num | Use "." if it's a missing value or below LLQ<br>Round to 0.01 | eToolbox |
| EVID | Event ID | NA | Num | 0 = PK Event<br>1 = Dose Event | Derived according to notes |
| GENDER | Gender | NA | Num | 1 = male<br>2 = female | Level1.dm$SEX |
| BSCR | Baseline serum creatinine | mg/dL | Num | Round to 0.01 | Level1a.lb$LBSTRESN where LBTESTCD=CREAT and LBBLFL=1 |
| FLAG | Exclusion flag | NA | Num | | Derived according to notes in section 4.4 |

### Imputation Rules

- Issues related to data imputation are particularly acute for population pharmacokinetic (PPK) data sets, as detailed dosing history is generally not available in the source data.

## RESULTS (continued)

- Given below are the scenarios in which imputation is typically necessary. All the examples assume BID dosing. ATAFD and ATAPD units are in hours.

**Case 1:** Imputation of missing dose time when time of associated trough PK sample (prior to dose) is available

- **Imputation Rule**: Impute the missing time of the subsequent dose to be same as the time of the trough PK sample. For example:

| ID | EVID | DATE | TIME | TROUGH | ATAFD | ADDL | ATAPD |
|---|---|---|---|---|---|---|---|
| 71 | 1 | 10APR12 | 9:15 | . | 1342.42 | 54 | 0.00 |
| 71 | 1 | 07MAY12 | 16:34 | . | 1997.73 | 0(1) | 0.00 |
| 71 | 0 | 08MAY12 | 9:20 | 1 | 2014.50 | . | 16.77(4.77) |
| 71 | 1 | 08MAY12 | 9:20 | . | 2014.50 | 165 | 0.00 |
| 71 | 0 | 08MAY12 | 9:50 | . | 2015.00 | . | 0.50 |

- The value for ADDL and ATAPD in parentheses is wrong as it is incorrect to assume the dose between the recorded dose prior to trough and the trough collection if it is not captured in the CRF.

### Case 2: Derivation of ATAPD when dose is interrupted

- **Derivation Rule**: ATAPD is derived based on date and time of prior nonzero dose. It is important to keep the dose-interrupted records to derive ADDL correctly and can be flagged for exclusion. For example:

| ID | EVID | DATE | TIME | DAY | DOSE | ATAFD | ADDL | ATAPD |
|---|---|---|---|---|---|---|---|---|
| 70 | 1 | 14MAY12 | 10:00 | 83 | 200 | 1968.00 | 1 | 0.00 |
| 70 | 1 | 15MAY12 | 10:00 | 84 | 0 | 1992.00 | 0(1) | 0.00 |
| 70 | 0 | 16MAY12 | 11:18 | 85 | 0 | 2017.30 | . | 37.30(13) |
| 70 | 1 | 16MAY12 | 12:28 | 85 | 200 | 2018.47 | 79 | 0.00 |

### Case 3: Derivation of ADDL when the dose time of the next visit is recorded too far from the current dose time.

- **Derivation Rule**: Derive ADDL taking into account the time difference between two doses. For example:

| ID | DATE | TIME | DAY | EXAMPM | ADDL | ADDL_NEW |
|---|---|---|---|---|---|---|
| 103 | 18MAR13 | 8:40 | 14 | AM | 26 | . |
| 103 | 31MAR13 | 15:30 | 27 | PM | 1 | 0 |
| 103 | 01APR13 | 9:45 | 28 | AM | 54 | . |
| 103 | 28APR13 | 17:20 | 55 | PM | 0 | . |

### Case 4: Deriving ADDL when AM or PM dose is missed in a BID dosing

- **Derivation Rule**: Derive ADDL taking into account the next day missed AM dose. For example:

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

## RESULTS (continued)

### Other dose time imputations

- If there is no PK sample associated with the missing dose time then impute by using next or previous available dosing time.
- Missing first dose time will be imputed from day 1 lab measurement time.

### Quality Control

- Independent Programmer checks the analysis data sets according to the relevent procedural documents (SOP and Work Instructions).
- A QC check list and best practice user guide was developed specifically for checking the preparation of Pm analysis data sets.

### Process Improvements

- Dry-runs:For projects on critical path the programming of Pm data sets is front-loaded by doing dry runs with mock data before the data base lock (DBL).
  - This not only enables delivery of Pm data sets rapidly after DBL but also helps identify any data issues in advance so Pm data set specs can be improved.
- Book of Work: A SharePoint-based book of work was designed to better plan and manage Pm data requests and programming resources. It allows the requestors to enter information ahead of time, for example:

| Project | Title | Number of Studies | Purpose | Requested By | Requested On | Date of Source Data Availability | Due Date |
|---|---|---|---|---|---|---|---|
| A1266 | creation of table for SUSTIVA Pediatrics study | 3 (A1266-922, PACTG 382, PACTG 1021) | Regulatory Filings | Tafoya, Emi | 9/13/2012 | 9/13/2012 | 9/25/2012 |
| A1443 | HCV triple DAA IK (BMS-791325+DCV+ASV) | 1: AI-443-014 | All Other | Tafoya, Emi | 8/21/2012 | 8/22/2012 | 8/31/2012 |
| A1447 | Dataset Preparation PPK for JNDA | 1:(A1447017) | Regulatory Filings | Osawa, Mayu | 9/11/2012 | 1/19/2012 | 9/26/2012 |
| A1447 | DUAL/QUAD eSubmission of PPK, ER and IK datasets | 4: (A1444010, A1447011, A1447016, A1447017) | EOP2 Meetings | Chan, Phyllis | 1/5/2012 | 1/19/2012 | 1/27/2012 |

### Report and eSubmission

Programmers author the data section of the Pm report, including:

- Summary tables for subjects and samples excluded and included from analyses, with reasons for exclusion.
- Graphical and tabular summaries of key covariates and PK sampling schedule. Listings of individual data are presented in the appendices.

Pm data sets are prepared for eSubmission as per FDA guidance http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm180482.htm

- eSubmission data sets and define.doc generation process has been automated.

## CONCLUSIONS

- The implementation of Best Practices has enabled improvements in the efficiency of Pm analysis data set preparation, consistency and documentation of data sets.