PharmaSUG 2022 - Paper AP-146

# Merging Pharmacokinetic (PK) Sample Collection and Dosing Information for PK Analysis Based on Study Design and Lessons Learned

Durga Kalyani Paturi, Pharmaceutical Product Development

## ABSTRACT

The PK merge dataset is a vital source dataset for PK statistical analysis. PK concentrations, sample collection, and dosing variables are merged and subsequently used as input for creating PK parameters. Because of the increasing complexity of clinical trial protocols, data issues, and the possible need of adding exclusion flags, one must carefully understand the study design, Statistical Analysis Plan (SAP), and specifications before building a merge dataset.

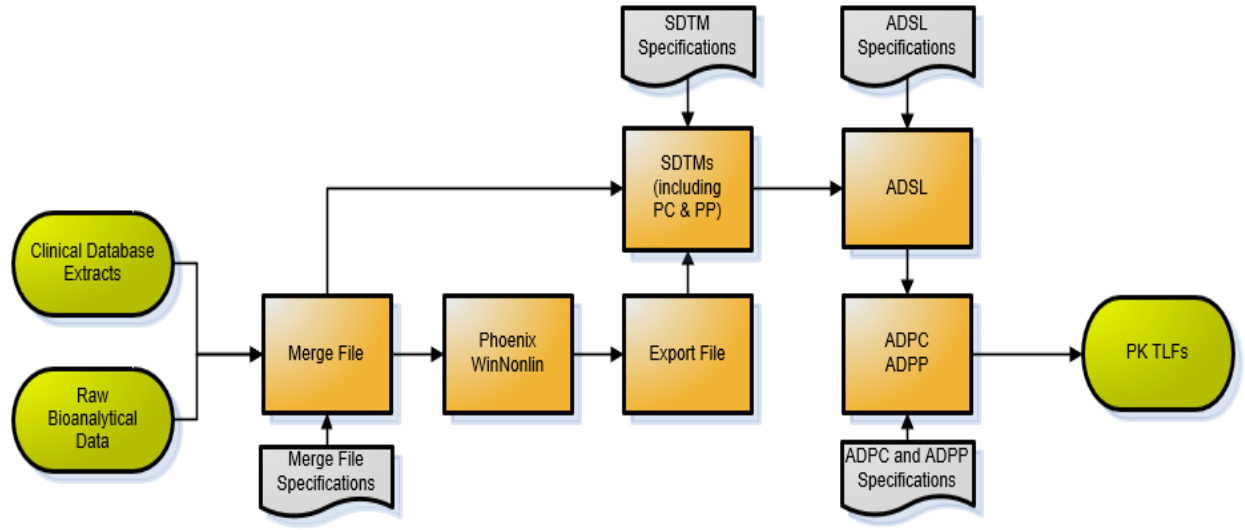
This paper explains the merge data set structure and presents examples of the merge data set in different study designs and various scenarios of data exclusions. Finally, we discuss the interesting lessons learned from the programmer's point of view.

## INTRODUCTION

Most Phase 1 or Phase 2 clinical studies are designed primarily to investigate the safety/tolerability, pharmacokinetics (PK), and pharmacodynamics (PD) of a compound. Statistical inferential analyses of PK parameters play a pivotal role in determining the primary/secondary endpoints of a clinical trial, aid in dose optimization, and thereby influence the decision-making process for subsequent clinical studies. PK merge dataset comprises PK concentrations from bioanalytical data, dosing, sample collection, and timing information collected from the clinical database. Depending on the study design, programmers create a SAS® program that pulls in the various variables from different source datasets. Pharmacokineticists use PK merge dataset as an input file in the Phoenix® WinNonlin® software to generate PK/PD parameters. PK parameters are calculated using non-compartmental (NCA) PK analysis in Phoenix® WinNonlin® (Certara USA, Inc., Princeton, NJ, USA). Once the PK/PD parameters are generated, Clinical Data Interchange Standards Consortium (CDISC) compliant SDTM pharmacokinetic parameters (PP) dataset and Analysis Data Model (ADaM) standard ADPP dataset are created to produce PK tables, listings, and figures.

## PK STUDY WORKFLOW

As shown below in figure 1, for PK data analysis, CDISC compliant SDTM PC and SDTM PP are primarily created: SDTM PC contains PK concentrations whereas SDTM PP is created from PK parameters. ADPP dataset is then derived from the SDTM PP domain with additional variables added from ADSL or derived, as needed. Similarly, the ADPC dataset is created from SDTM PC and SDTM EX domains with ADSL and additional derived variables, as needed. Analysis ready ADPP dataset is the primary source for generating PK statistical summary tables that support the outcome of PK analysis.




ADPC = Analysis Dataset Pharmacokinetic Concentration; ADPP = Analysis Dataset Pharmacokinetic Parameter; ADSL = Analysis Dataset Subject Level; PC = Pharmacokinetic Concentration; PK = Pharmacokinetic; PP = Pharmacokinetic Parameter; SDTM = Study Data Tabulation Model; TLFs = Tables, Listings and Figures.

**Figure 1: Example PK Analysis Workflow from Raw Data to PK Tables, Listings, and Figures**

## VARIABLES NEEDED FOR MERGE DATASET

An example of PK merge data structure is shown below in table 1. The merge dataset consists of three basic data types: dosing records, PK/PD concentrations, and actual time variables.

### PC DATASET

SUBJECT - Subject Identifier
GRPID - Period/Day
PCTPTNUM - Nominal Time Point in Numeric Format
PCTATNUM - Actual Time in Numeric Format (PCDTC-EXDTC), where PCDTC is Collection Date/Time of PK Sample Collection and EXDTC is Dosing Date/Time.

### BIOANALYTICAL DATA

ANALYTE - Name of Compound being Measured (Typically Parent Drug or Metabolite)
PCORRES - Original Concentration in Character Format
PCORRESU - Original Concentration Units in Character Format
PCSTRESN - Original Concentration in Numeric Format
PCSTRESA - Concentration with BLQ Data Handling Rules Applied in Numeric Format
SPEC - Specimen (Matrix) Type

### EX (DOSING) DATASET

EXDOSE - Dose
EXDOSEU - Units for Dose
EXTM - Time of Dosing
DOSETYPE - Type of Dose (Extravascular or Intravascular)
TAU - Dosing Interval (Only applicable for Multiple Dosing Studies)

| SUBJECT | GRPID | PCTPTNUM | PCTATNUM | PCORRES | PCORRESU | PCSTRESN | PCSTRESA | SPEC | DOSETYPE | EXDOSE | EXDOSEU | EXTM | TAU |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 101 | Day 1 | 0 | 0 | BLQ | ng/mL | 0 | 0 | PLASMA | Extravascular | 6.083 | mg | 0 | |
| 101 | Day 1 | 0.25 | 0.25 | 25.23 | ng/mL | 25.23 | 25.23 | PLASMA | | | | | |
| 101 | Day 1 | 0.5 | 0.5 | 50.15 | ng/mL | 50.15 | 50.15 | PLASMA | | | | | |
| 101 | Day 1 | 1 | 1 | 100.34 | ng/mL | 100.34 | 100.34 | PLASMA | | | | | |
| 101 | Day 1 | 2 | 2 | 150.88 | ng/mL | 150.88 | 150.88 | PLASMA | | | | | |
| 101 | Day 1 | 4 | 4 | 60.04 | ng/mL | 60.24 | 60.24 | PLASMA | | | | | |
| 101 | Day 1 | 6 | 6 | 20.53 | ng/mL | 20.53 | 20.53 | PLASMA | | | | | |
| 101 | Day 1 | 8 | 8 | BLQ | ng/mL | 0 | | PLASMA | | | | | |
| 101 | Day 1 | 12 | 12 | BLQ | ng/mL | 0 | | PLASMA | | | | | |
| 101 | Day 1 | 24 | 24 | BLQ | ng/mL | 0 | | PLASMA | | | | | |
| 101 | Day 11 | 0 | 0 | BLQ | ng/mL | 0 | 0 | PLASMA | Extravascular | 6.083 | mg | 0 | 24 |
| 101 | Day 11 | 0.25 | 0.25 | 55.83 | ng/mL | 55.83 | 55.83 | PLASMA | | | | | |
| 101 | Day 11 | 0.5 | 0.5 | 70.39 | ng/mL | 70.39 | 70.39 | PLASMA | | | | | |
| 101 | Day 11 | 1 | 1 | 150.34 | ng/mL | 150.34 | 150.34 | PLASMA | | | | | |
| 101 | Day 11 | 2 | 2 | 180.84 | ng/mL | 180.84 | 180.84 | PLASMA | | | | | |
| 101 | Day 11 | 4 | 4 | 160.24 | ng/mL | 160.24 | 160.24 | PLASMA | | | | | |
| 101 | Day 11 | 6 | 6 | 45.62 | ng/mL | 45.62 | 45.62 | PLASMA | | | | | |
| 101 | Day 11 | 8 | 8 | 23 | ng/mL | 23 | 23 | PLASMA | | | | | |
| 101 | Day 11 | 12 | 12 | BLQ | ng/mL | 0 | | PLASMA | | | | | |
| 101 | Day 11 | 24 | 24 | BLQ | ng/mL | 0 | | PLASMA | | | | | |

**Table 1: Example PK Merge File Structure and Content**

Example of a merge between dosing and PK collection information:

```
data pc_ex;
     merge ex(in=b)  pc (in=a);
     by subject grpid;
     if a;
     pctatnum=(pcdtc-exdtc)/3600;
run;
```

Example of a merge between bioanalytical data and dosing/PK collection information:

```
data final;
     merge conc(in=x)  pc_ex(in=y);
     by subject grpid pctptnum;
     if x ;
run;
```

## WHAT IS THE STUDY DESIGN?

As shown in figure 2, many typical clinical studies include multiple doses, several routes of administration, different treatment arms and/or treatment sequences, metabolite data, and/or urine data. Clinical trials have been conducted in the logical sequence of single ascending dose, multiple ascending dose, food effect, and potential drug-drug interaction studies, along with assessments of bioavailability and bioequivalence, as applicable.

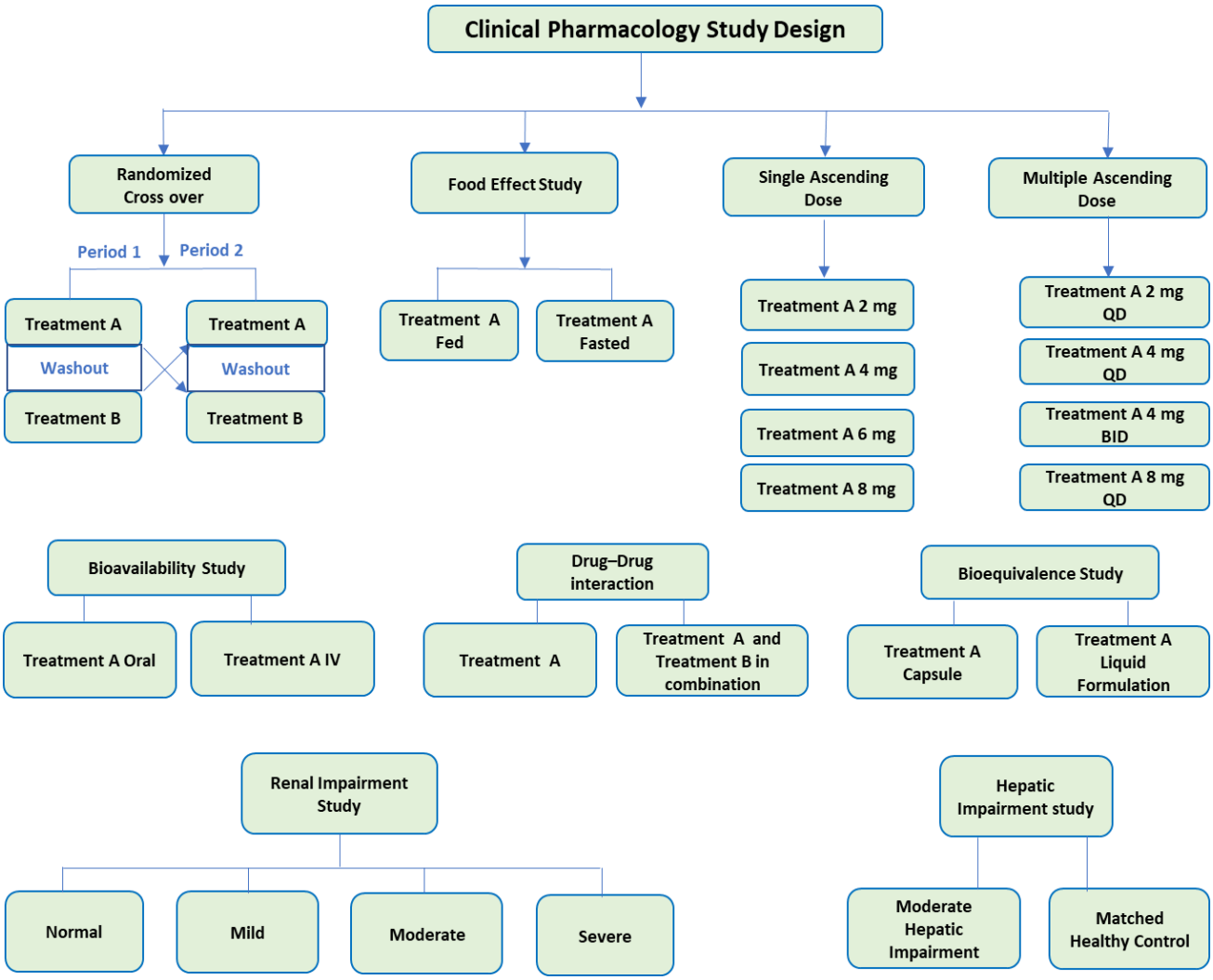



**Figure 2: Standard Clinical Pharmacology Study Designs**

## CROSS-OVER STUDIES

In a crossover design, each participant is randomized to a sequence of treatments that will be sequentially administered in different treatment periods. For instance, in a three-period crossover trial designed to compare Treatments A and B and C, a participant may be randomized into one of six sequences (ABC, ACB, BAC, BCA, CAB, CBA) with a washout period between each period so that the drug can be completely metabolized or excreted from the body.

For crossover studies, dosing and PK collection information entered into the clinical database (from case report forms, CRFs) is merged using Subject and Period variables:

```
data pc_ex;
     merge ex  pkc1(in=a);
       by subject period;
run;
```

Then dosing/PK collection information is merged with bioanalytical data using the Subject, Period, and Nominal time point of PK sample collection.

```
data pkinput1;
     merge pc_ex(in=a) conc(in=b);
       by subject period pctptnum;
       if a and b;
run;
```

## FOOD EFFECT STUDIES

Food effect studies are usually conducted for new drugs to assess the effect of food on the rate and extent of absorption of a drug when administered in a fed condition (shortly after a meal) as compared to administration under fasting conditions. Generally, a randomized two-period, two-sequence crossover design is recommended for studying food effects. The formulation to be tested should be administered on an empty stomach (fasting condition) in one period and fed condition in the other period. The merge for food effect studies is usually the same as in crossover studies, where the merge between dosing and PK collection datasets uses subject and period to merge the data.

## BLINDED STUDIES

In a double-blind study, participants and investigators are unaware of the assigned treatment to facilitate unbiased psychological or physical responses to the treatment. Randomization datasets are used to define whether a subject has been given active treatment or a placebo (note: replacement randomization numbers should be included to ensure all subjects are captured). Before study unblinding, a surrogate randomization dataset is created for programming purposes. Once study unblinding occurs, the actual randomization dataset is used to capture the correct treatment as shown below in table 2.

Prior to Unblinding: Surrogate Randomization

| RANDNUM | REPLNUM | TMT_TX |
|---|---|---|
| 101 | 111 | Placebo |
| 102 | 112 | Treatment A |
| 103 | 113 | Treatment A |
| 104 | 114 | Placebo |
| 105 | 115 | Treatment A |
| 106 | 116 | Treatment A |

After Unblinding: Actual randomization

| RANDNUM | REPLNUM | TMT_TX |
|---|---|---|
| 101 | 111 | Treatment A |
| 102 | 112 | Treatment A |
| 103 | 113 | Placebo |
| 104 | 114 | Treatment A |
| 105 | 115 | Treatment A |
| 106 | 116 | Placebo |

**Table 2: Randomization in Blinded studies**

```
data rand;
     set lib.Surrogate_1   lib. Surrogate_2 (in=b);
       randno=strip(randnum);
       trtp=strip(tmt_txt);
       if b then randno=strip(replnum);
       keep randno trtp;
run;
```

```
data rand;
     set lib. actrand_1   lib. actrand_2 (in=b);
       randno=strip(randnum);
       trtp=strip(tmt_txt);
       if b then randno=strip(replnum);
       keep randno trtp;
run;

data pkinput1;
     merge conc_ex (in=k) rand;
       by randno;
       If k;
run;
```

## SINGLE DOSE/MULTIPLE ASCENDING DOSE STUDIES

Single Ascending Dose (SAD) and Multiple Ascending Dose (MAD) studies are first-in-human studies conducted to elucidate the safety, tolerability, PK, and PD characteristics of a compound. SAD studies involve a small group of subjects that receive a single dose of the compound and dose escalation usually continues until the maximum dose has been attained per the protocol. In a MAD study, each subject receives multiple doses, and starting dose for a MAD study is usually based on results from the SAD study.

In MAD studies, a profile day (PDAY) variable is created, which is the day of each dosing where PK profile samples are taken. In the following example, full PK profiles are collected on Days 1 and 29 with doses given on Days 1, 15, and 29. The PK Sample on Day 15 is a trough sample, used to monitor the accumulation of concentrations before the next dose. For PK parameter estimation, the actual times are estimated relative to the previous dose. As PK collections may span multiple days after each dose, it is not always possible to use study or visit day to categorize PK profiles; instead, a PK profile day variable (PDAY) is created to define each separate PK profile relative to the previous dose.

Profile day being created in dosing dataset:

```
data ex;
     set ex;
       if folder = 'Day 1' then pday = 1;
       else if folder = 'Day 15' then pday = 15;
       else if folder ='Day 29' then pday = 29;
run;
```

Profile day being created in PK collection dataset:

```
data conc;
     set pk;
       day = input(scan(folder, 2, ''),best.);
       if day<15 then pday = 1;
       else if day=15 then pday = 15;
       else if day>15 then pday = 29;
run;
```

Merging dosing and PK collection datasets using on subject and profile day:

```
data conc_ex;
     merge conc (in=a) ex (in=b);
       by subject pday;
       if a and b;
 run;
```

## BID STUDIES – DOSING TWICE A DAY

Following BID dosing, the second dose of study treatment is typically administered approximately 12 hours after the morning dose. In this case, dosetpt (AM/PM dose) will be derived and used for merging PC and EX datasets.

```
data pc_ex ;
     merge ex (in=d) pc(in=a);
       by subject pday dosetpt;
run;
```

## DRUG-DRUG INTERACTION STUDIES

A common drug-drug interaction study involves co-administration of a compound with a strong inhibitor or a strong inducer of the metabolizing enzyme cytochrome P450, or a substrate. These inducers/inhibitors/substrates may be administered alone or in combination with the compound of interest. For example, Losartan (CYP2C9), omeprazole (CYP2C19), dextromethorphan (CYP2D6), caffeine (CYP1A2), and midazolam (CYP3A4) are used as probe substrates and their concentrations may increase or decrease if there is a drug-drug interaction with the compound of interest. Summaries of the PK parameters of the compound as well as metabolite are determined. In this case, both the probe substrate and its metabolite have the same collection date and time and dosing date and time and therefore we create an additional variable (analyte) as shown below to merge the bioanalytical data and ex dataset.

```
data conc;
     set lib.pkdata;
       analyte=scan(cat);
run;

data pc;
     set lib.pc;
       if paramcd  in ('Caffeine' 'Paraxanthine') then
       analyte='Caffeine ';
       day=scan(FolderName,2,' ');
       grpid=strip(day);
run;


data conc_pc;
     merge conc(in=a) pc;
       by Subject analyte grpid pctptnum;
       if a;
run;
```

### RADIOLABELED MASS BALANCE STUDIES

Mass balance studies are designed to elucidate the clearance pathways of new drugs. For mass balance studies, radiolabeled compounds are administered and various biological fluids such as blood, plasma, feces, saliva, and emesis are collected and analyzed. Since concentrations for multiple specimens are collected, specimen type (PCSPEC) is added when merging raw database extracts with bioanalytical data.

```
data final;
     merge conc(in=a) pc_ex(in=b);
       by subject pcspec pctptnum;
       if  a;
run;
```

## EXCLUSION FLAGS

It is often necessary to exclude individual PK concentrations or entire PK profiles from PK parameters and/or statistical analyses. Rather than simply deleting such records, the exclusions are always handled using analysis flags. Some examples of exclusion flags are listed below in the following five scenarios:

### SCENARIO 1 (Full Profile of BLQs):

In this case, the subject has BLQ values recorded at all the nominal timepoints. This profile can be excluded from the PK analysis as it is not possible to generate noncompartmental parameters in this instance and therefore an exclusion flag is added to the merge file to exclude relevant records.

```
proc univariate data = conc (where=(pcstresn=0)) noprint ;
     by  subject cat pday;
     var pctptnum;
     output out = sum  n = count ;
run;
data concs;
     merge conc (in=x) sum;
       by subject  cat pday;
       If x;
       if count = 6  then do;
       exclfl='Yes';
       exreas= 'Incomplete Profile';
 end;
run;
```

### SCENARIO 2 (Emesis):

Subjects who experience emesis or vomiting within 2 times the median $T_{max}$ after study drug dosing are often excluded from the PK analysis, particularly in BE or food effect studies. In such cases, vomiting or emesis time is calculated from the emesis start date (AE dataset) and dosing date (EX dataset) and compared with median $T_{max}$. If the emesis time is less than 2 times of median $T_{max}$ for the affected treatment group, then the subject will be excluded from PK analysis.

```
emtime=round((aestdate-exdate)/3600,0.0001);
if aeterm="EMESIS" and emtime < 2 *median Tmax then do;
     exclfl="YES";
     exreas="EMESIS";
 end;
```

### SCENARIO 3 (Predose > 5% of $C_{max}$):

If the predose concentration is greater than 5% of the $C_{max}$ then PK parameters of the corresponding subjects are dropped from the PK analysis.

```
if pctptnum =0  then do;
     if pcstresn > 0.5*Cmax then do;
     exclfl="YES";
     exreas="Predose > 5% of Cmax";
 end;
```

### SCENARIO 4 (Sparse PK Sampling):

In this case, the subject has sparse samples recorded at certain timepoints that are not needed for PK parameter analysis. In the below example, timepoints collected during the visit (Cycle 3 Day 1) are excluded from PK analysis by adding an exclusion flag.

```
if index(visit, 'Cycle 3 Day 1')>0 then do;
       exclfl="YES";
       exreas="Sparse Sample";
 end;
```

### SCENARIO 5 (%AUC extrapolation > 20%):

In this case, if the extrapolated AUC is greater than 20%, then parameters VZFO, CLFO, and AUCIFO are excluded from the PK analysis.

```
if paramcd='AUCPEO' and aval>20 then do;
     If paramcd in ('VZFO','CLFO','AUCIFO') then do;
     exclfl="YES";
     exreas="%AUC extrapolation > 20% ";
  end;
end;
```

## IS THE MERGE FILE CORRECT?

- The merge file should include all the records present in the bioanalytical data with all the analytes and all the time points for a subject profile.
- Ensure that the merge file always uses the most recent data; check to include the latest extracted datasets/bioanalytical data /randomization datasets.
- Check for negative actual times or if actual times are visibly different from nominal timepoint. Set up warnings/quality checks where the deviation is more than x minutes (or x percent) from the nominal timepoint.
- Set up warnings/alerts for missing lab records in the merge file.
- Check the appropriate dosing records of the analyte and profile day from the dosing dataset.
- Check if the nominal timepoints are in hours or days.
- Ensure actual times account for daylight savings at affected sites.
- Ensure concentration units are correct for all analytes and matrices.
- Check that all data handling rules (including BLQ imputation) have been applied correctly, as per the SAP.

## LESSONS LEARNED

- It is important to derive correct actual times from PK collection and dosing Datetime. Sometimes, CRF PK/EX records can have missing or incorrect datetime. Missing data can create biased estimates of parameters. DM and Clinic need to confirm if the sample was collected, but the collection date/time was not available or if the sample was recorded incorrectly. Additional comments populated in the CRF/ Bioanalytical data can help identify the reason for missing datetimes. Nominal timepoints can be used as actual time in these cases to generate PK parameters.

- Bioanalytical data also needs to be thoroughly reviewed for the presence of duplicate records and the outliers with significant deviations as per SAP.

- CRF data is usually extracted early and so identifying potential reconciliation issues early by the programmers makes it easy to identify potential data issues before receiving Quality Controlled (QC)/Quality Audited (QA) Bioanalytical data.

- Use of appropriate dosing records and AM/PM dose and include a merge variable with AM/PM information if the samples are collected both in AM/PM on the same day. It is also important to identify if any time points in the raw data serve as both a post-dose time point for the first dose, as well as a pre-dose time point for the next profile. This usually happens for MAD studies where the end of 1$^{st}$ profile sample is also treated as the beginning of the next profile sample.

- Adding quality checks for actual times helps in identifying any deviations from nominal time. We can tell from the quality checks, that the differences between actual and nominal time are consistent in most of the cases and are close to nominal timepoints. Daylight savings need to be applied depending on whether the parameters are generated by days/hours and the location of the clinic where samples were collected needs to be considered.

- Concentrations below the range of bioanalytical assay sensitivity are termed as Below the limit of quantification (BLQ) records. Correct imputation rules as per SAP need to be applied for the BLQ values present in between two measurable concentrations and in the terminal phase of a profile. Profiles with a greater number of BLQ records can cause biased estimation of PK parameters and therefore, exclusion flags need to be added.

- For blinded studies after study unblinding, actual randomization datasets need to be replaced with surrogate randomization datasets to derive the correct treatment and we need to make sure to include the replacement random numbers.

## CONCLUSION

A high-quality PK merge dataset supports the successful generation of SDTM PP, ADPP, and TLF outputs. It is vital to understand the study design and identify missing or incorrect data while building the merge dataset. This paper discusses various examples of some of the important study designs that can help readers better understand the merging of PK collection and dosing data. This paper also summarizes some of the lessons learned that can help programmers avoid common errors and identify potential complications. ADNCA ADaM standard dataset for PK merge file creation is currently in development and is expected to provide better guidance in the future to support merge dataset and PK parameter analysis.

## REFERENCES


1. Gieser, Gerlie. "Clinical pharmacology 1: phase 1 studies and early drug development." Food and Drug Administration (FDA) website. https://www. fda. gov/media/84920/download.
2. Food and Drug Administration (FDA). 2002. “Guidance for Industry Food-Effect Bioavailability and Fed Bioequivalance Studies.” (December):1–12.
3. Food and Drug Administration (FDA). 2012. “Guidance for Industry. Drug Interaction Studies Study Design, Data Analysis, Implications for Dosing, and Labeling Recommendations.” (February):79.
4. Scott R. Evans Ph.D., Fundamentals of clinical trial design. J Exp Stroke Transl Med. 2010; 3(1);19-27.
5. Shen J, Swift B, Mamelok R, Pine S, Sinclair J, Attar M. Design and Conduct Considerations for First-in-Human Trials. Clin Transl Sci. 2019;12(1):6-19.


## ACKNOWLEDGEMENTS


The author would like to thank Shallabh Mehta, Mark Beard, Mamta Agarwal, and Edward Elam for their support and for providing their valuable input for this paper.


## CONTACT INFORMATION

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


Durga Kalyani Paturi, PhD
Pharmaceutical Product Development
Email: Durga.paturi@ppdi.com


## DISCLAIMER

The contents of this paper are the work of the author and do not necessarily represent the opinions, recommendations, or practices of PPD.