PharmaSUG 2021 - Paper AP-067

Imputation for Missing Dosing Time in NONMEM PopPK Datasets

Shuqi Zhao, Merck & Co., Inc., Kenilworth, NJ, USA

ABSTRACT

For late stage clinical trials with oral daily dosing regimen, complete dosing history data is generally not available. Following study design, patients get PK samples drawn at clinic sites, but between scheduled visits, patients are expected to take daily doses at home. Both date and time for PK samples are collected, while accurate time as to when the dose is taken for most of the doses is not available. When building NONMEM®-ready PopPK datasets, however, both date and time for dosing records are required for derivation of actual relative time from first dose and actual relative time from last dose. Due to incomplete dosing history data, time imputation comes into play. Imputation methods can be different depending on study design and data collection design, but the general rule is that the relative time of dose to PK sample doesn't get changed. This paper provides a step-by-step programming guide on how to impute time for dosing records when actual time information is only available for dose prior to PK sample. Topics such as how to unfold dosing records in EX domain into individual daily records and commonly seen data issues in relevant datasets will also be discussed.

INTRODUCTION

NONMEM®-ready PopPK datasets are datasets that PK/PD modelers use for modeling and analysis, in which dosing history data is one of the major components. In late stage studies, subjects might take oral daily doses for a long time, for which most of the doses are taken at home and thus exact time for most of the doses is not collected.

In late stage clinical trials, a common way to collect dosing data is that an interval dosing history is recorded in EX domain with start and end date without actual time. Datetime of the dose that is prior to a PK sample will be captured on PK domain as a time point reference.

In the following sections, this paper will discuss how required information are normally collected in eCRF and datasets, and then provide a step-by-step programming guide on how to impute missing time of some doses in order to create NONMEM®-ready popPK dataset.

BACKGROUND – STUDY DESIGN, HOW DATA ARE COLLECTED AND ETC.

The studies that are in scope of the discussion have the following characteristics. The study has an oral daily dosing regimen. Subjects are asked to take oral doses once a day for a period of time. During the study, subjects pay visits to clinic sites as scheduled, at which a PK sample is collected. At the time when a PK sample is drawn, datetime of the last dose prior to this PK sample is recorded; yet datetime for most of doses remain missing as subjects take doses at home for most of the time. When a subject visits the clinic site as planned, the clinician will first enter the datetime of the last dose prior to this visit into the form, a PK sample is drawn and then the subject takes medication afterwards.

Let's take a look at how datetime related information is collected on eCRF and how it displays in SDTM datasets. Please note the tables and figures below only show information relevant to the topic instead of showing a complete eCRF form or SDTM domain.

STUDY MEDICATION - Form	
…	
Start Date/Time	Month/Date/Year
Stop Date/Time	Month/Date/Year
…	

Table 1. Illustration of Study Medication Form

DOMAIN	SUBJID	EXSTDTC	EXENDTC	EXDOSTOT
EX	1	2021-01-01	2021-01-07	5
EX	1	2021-01-08	2021-01-10	10
EX	1	2021-01-12	2021-01-14	10
EX	1	2021-01-15	2021-01-21	15

Figure 1. Sample EX Domain

For a late stage study with long dosing duration, the dosing records are not collected on a daily basis in the form. Instead, on SM form, the start date (EXSTDTC) and end date (EXENDTC) of a dosing duration are collected. In EX domain, one record represents dosing history for a period instead of a single day. For example, in Figure 1, the first record indicates subject 1 takes 5 unit of medication every day from 2021-01-01 to 2021-01-07.

PK INVOICE - Form	
…	
Date and Time Sample Obtained	Month/Date/Year Hour: Minute
Last Dose Date and Time Prior to PK Sample Obtained	Month/Date/Year Hour: Minute
…	

Table 2. Illustration of PK INVOICE Form

DOMAIN	SUBJID	PCDTC	PCTPTREF	PCRFTDTC
PC	1	2021-01-01T09:00:00		
PC	1	2021-01-09T09:10:00	Last Dose Prior to PK Sample	2021-01-08T09:30:00
PC	1	2021-01-16T09:20:00	Last Dose Prior to PK Sample	2021-01-15T08:45:00

Figure 2. Sample PK Domain

On PK invoice form, datetime of a PK sample (PCDTC) is collected, as well as datetime of the last dose prior to PK sample (PCRFTDTC). The datetime of the last dose prior to PK sample is crucial as exact elapsed time since last dose is required for PK analysis. In PC domain, datetime of last dose prior to PK sample is shown as variable PCRFTDTC. In Figure 2, PCRFTDTC is missing for the first record because this PK sample is drawn pre-dose.

STEP 1: UNFOLD INTERVAL DOSING RECORDS IN EX DOMAIN INTO INDIVIDUAL DAILY RECORDS

Since dosing data is collected as periods (with start and end date without time) in EX domain, the first thing to do is open up those interval records into individual daily dosing records. This step is to prepare dosing data so that actual time of some individual dosing records will be derived in the following steps.

SAS code below is to open up dosing data from Figure 1 :

```
data AP067.ex02;
    format ADT yymmdd10. ;
    set AP067.ex01;
    by SUBJID EXSTDTC EXENDTC;
    EVID = 1;
    retain ADT ;
    ADT = input(EXSTDTC, yymmdd10.);
    output;
    do while (ADT < input(EXENDTC, yymmdd10.));
          ADT = ADT + 1 ;
          output;
```

```
    end;
run;
```

The code above creates a new variable ADT, which represents date of a daily dose. When executing the code, SAS will output dosing records for each day within a certain duration, which is defined by EXSTDTC (start date) and EXENDTC (end date). The output dataset looks like this:

DOMAIN	SUBJID	ADT	EXSTDTC	EXENDTC	EXDOSTOT	EVID
EX	1	2021-01-01	2021-01-01	2021-01-07	5	1
EX	1	2021-01-02	2021-01-01	2021-01-07	5	1
EX	1	2021-01-03	2021-01-01	2021-01-07	5	1
EX	1	2021-01-04	2021-01-01	2021-01-07	5	1
EX	1	2021-01-05	2021-01-01	2021-01-07	5	1
EX	1	2021-01-06	2021-01-01	2021-01-07	5	1
EX	1	2021-01-07	2021-01-01	2021-01-07	5	1
EX	1	2021-01-08	2021-01-08	2021-01-10	10	1
EX	1	2021-01-09	2021-01-08	2021-01-10	10	1
EX	1	2021-01-10	2021-01-08	2021-01-10	10	1
EX	1	2021-01-12	2021-01-12	2021-01-14	10	1
EX	1	2021-01-13	2021-01-12	2021-01-14	10	1
EX	1	2021-01-14	2021-01-12	2021-01-14	10	1
EX	1	2021-01-15	2021-01-15	2021-01-21	15	1
EX	1	2021-01-16	2021-01-15	2021-01-21	15	1
EX	1	2021-01-17	2021-01-15	2021-01-21	15	1
EX	1	2021-01-18	2021-01-15	2021-01-21	15	1
EX	1	2021-01-19	2021-01-15	2021-01-21	15	1
EX	1	2021-01-20	2021-01-15	2021-01-21	15	1
EX	1	2021-01-21	2021-01-15	2021-01-21	15	1

Figure 3. EX Individual Daily Records

From Figure 3, subject 1 takes 5 unit of medication once daily from 2021-01-01 to 2021-01-07 for 7 days. The subject takes 10 unit of medication once daily from 2021-01-08 to 2021-01-14 but skips medication on 2021-01-11. Then the subject takes 15 unit of medication once daily from 2021-01-15 to 2021-01-21 for 7 days. EVID (event ID) is a variable commonly seen in NONMEM® modeling, where EVID = 0 denotes PK observations while EVID = 1 denotes dosing records.

STEP 2: JOIN INDIVIDUAL DOSING RECORDS WITH PK RECORDS BY DATE OF DAILY DOSE FROM EX DOMAIN AND DATE OF LAST DOSE FROM PK DOMAIN

Now that daily dosing records are available, the next step is to get time information of some dosing records from PC domain, as datetime of dose prior to PK sample is collected in PC domain.

Before joining EX and PC domain, some preprocessing work on PC domain is necessary:

```
data AP067.pc02;
    format ADT LASTDOSEDATE yymmdd10. ATM LASTDOSETIME time5. ;
    set AP067.pc01;
    EVID = 0 ;
    ADT = datepart(input(PCDTC, is8601dt.));
    ATM = timepart(input(PCDTC, is8601dt.));
    if PCRFTDTC ^= "" then do;
          LASTDOSEDATE = datepart(input(PCRFTDTC, is8601dt.));
          LASTDOSETIME = timepart(input(PCRFTDTC, is8601dt.));
    end;
run;
```

The preprocessed PC domain looks like this:

DOMAIN	SUBJID	PCDTC	ADT	ATM	PCTPTREF	PCRFTDTC	LASTDOSEDATE	LASTDOSETIME	EVID
PC	1	2021-01-01T09:00:00	2021-01-01	9:00			.	.	0
PC	1	2021-01-09T09:10:00	2021-01-09	9:10	Last Dose Prior to PK Sample	2021-01-08T09:30:00	2021-01-08	9:30	0
PC	1	2021-01-16T09:20:00	2021-01-16	9:20	Last Dose Prior to PK Sample	2021-01-15T08:45:00	2021-01-15	8:45	0

Figure 4. Preprocessed PC Domain

Datetime of PK sample (PCDTC) is split up into date (ADT) and time (ATM) for the purpose of sorting in the following steps. Datetime of last dose prior to PK sample (PCRFTDTC) is split up into date (LASTDOSEDATE) and time (LASTDOSETIME) so that date of last dose is used as a condition when EX domain joins PC domain.

Next, PROC SQL is used to join EX domain and PC domain together on SUBJID and date of daily dosing records (ADT) from EX domain and date of last dose (LASTDOSEDATE) from PC domain so that time of last dose prior to PK can be passed to the corresponding dosing record in EX domain:

```
proc sql;
    create table AP067.ex_pc as
    select a.*, b.LASTDOSETIME as ATM
    from AP067.ex02 as a
    left join AP067.pc02 as b
    on a.SUBJID = b.SUBJID and a.ADT = b.LASTDOSEDATE
    order by SUBJID, ADT, ATM;
quit;
```

The joined dataset looks like this:

DOMAIN	SUBJID	EXSTDTC	EXENDTC	EXDOSTOT	ADT	ATM	EVID
EX	1	2021-01-01	2021-01-07	5	2021-01-01	.	1
EX	1	2021-01-01	2021-01-07	5	2021-01-02	.	1
EX	1	2021-01-01	2021-01-07	5	2021-01-03	.	1
EX	1	2021-01-01	2021-01-07	5	2021-01-04	.	1
EX	1	2021-01-01	2021-01-07	5	2021-01-05	.	1
EX	1	2021-01-01	2021-01-07	5	2021-01-06	.	1
EX	1	2021-01-01	2021-01-07	5	2021-01-07	.	1
EX	1	2021-01-08	2021-01-10	10	2021-01-08	9:30	1
EX	1	2021-01-08	2021-01-10	10	2021-01-09	.	1
EX	1	2021-01-08	2021-01-10	10	2021-01-10	.	1
EX	1	2021-01-12	2021-01-14	10	2021-01-12	.	1
EX	1	2021-01-12	2021-01-14	10	2021-01-13	.	1
EX	1	2021-01-12	2021-01-14	10	2021-01-14	.	1
EX	1	2021-01-15	2021-01-21	15	2021-01-15	8:45	1
EX	1	2021-01-15	2021-01-21	15	2021-01-16	.	1
EX	1	2021-01-15	2021-01-21	15	2021-01-17	.	1
EX	1	2021-01-15	2021-01-21	15	2021-01-18	.	1
EX	1	2021-01-15	2021-01-21	15	2021-01-19	.	1
EX	1	2021-01-15	2021-01-21	15	2021-01-20	.	1
EX	1	2021-01-15	2021-01-21	15	2021-01-21	.	1

Figure 5. EX Domain with Time

By joining EX domain and PC domain on (EX.SUBJID and PC.SUBJID), and (EX.ADT and PC. LASTDOSEDATE), exact time of some dosing records are obtained from last dose in PC. For example, for the dosing record on 2021-01-08, the exact time 9:30 is obtained from PK domain.

After this step, exact time is only available for some dosing records, which are the last doses prior to PK samples collected in PK domain. For most of the dosing records shown in Figure 5, time is not available. This also aligns with the study design as subjects are expected to take most of the medications at home.

STEP 3: SET DOSING RECORDS AND PK RECORDS TOGETHER AND SORT

As shown in the Figure 5, time of most dosing records is missing so it's necessary to impute them. Before imputing missing time for dosing records, dosing records and PK records need to be set together and sorted for the ease of imputation.

One thing to note here is that the imputed time won't be directly used in a PK analysis, as PK analysis only cares about the exact elapsed time from the last dose for a PK sample. The exact elapsed time from every previous dosing record for a PK sample is not of interest. However, the time imputation is still necessary as it's beneficial for sorting the order of dosing history and PK observations.

To set and sort dosing records and PK observations together, run SAS code as below:

```
data AP067.expc00;
    set AP067.ex_pc AP067.pc02;
proc sort;
    by SUBJID ADT EVID ;
run;
```

The output dataset looks like this:

DOMAIN	SUBJID	EVID	ADT	ATM	EXDOSTOT	EXSTDTC	EXENDTC	PCDTC	PCRFTDTC
PC	1	0	2021-01-01	9:00	.			2021-01-01T09:00:00	
EX	1	1	2021-01-01	.	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-02	.	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-03	.	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-04	.	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-05	.	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-06	.	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-07	.	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-08	9:30	10	2021-01-08	2021-01-10		
PC	1	0	2021-01-09	9:10	.			2021-01-09T09:10:00	2021-01-08T09:30:00
EX	1	1	2021-01-09	.	10	2021-01-08	2021-01-10		
EX	1	1	2021-01-10	.	10	2021-01-08	2021-01-10		
EX	1	1	2021-01-12	.	10	2021-01-12	2021-01-14		
EX	1	1	2021-01-13	.	10	2021-01-12	2021-01-14		
EX	1	1	2021-01-14	.	10	2021-01-12	2021-01-14		
EX	1	1	2021-01-15	8:45	15	2021-01-15	2021-01-21		
PC	1	0	2021-01-16	9:20	.			2021-01-16T09:20:00	2021-01-15T08:45:00
EX	1	1	2021-01-16	.	15	2021-01-15	2021-01-21		
EX	1	1	2021-01-17	.	15	2021-01-15	2021-01-21		
EX	1	1	2021-01-18	.	15	2021-01-15	2021-01-21		
EX	1	1	2021-01-19	.	15	2021-01-15	2021-01-21		
EX	1	1	2021-01-20	.	15	2021-01-15	2021-01-21		
EX	1	1	2021-01-21	.	15	2021-01-15	2021-01-21		

Figure 6. Set of PK and Dosing Records

After dosing records and PK records are set together, the records are sorted first by ADT (date of daily dosing records or date of PK sample) and then by EVID. Sorting by EVID implies the subject takes a PK sample first when visiting the clinic and then takes the medication, which aligns with the study design.

STEP 4: IMPUTE TIME USING RETAIN STATEMENT

Once dosing records and PK sample records are nicely sorted, use the last observation carried forward (LOCF) method to impute the missing time, meaning the missing time of dosing records can be imputed using retain statement. Run SAS code as below:

```
data AP067.expc01;
    length _ATM 8 ATMF $1 ;
    set AP067.expc00;
    by SUBJID ADT EVID ;
    retain _ATM ;
    if first.SUBJID then _ATM = . ;
    if EVID = 0 then _ATM = ATM+60;
    else if EVID = 1 and ATM = . then do;
```

```
            ATM = _ATM;
            ATMF = "H" ;
        end;
    run;
```

The output dataset looks like this:

DOMAIN	SUBJID	EVID	ADT	ATM	ATMF	EXDOSTOT	EXSTDTC	EXENDTC	PCDTC	PCRFTDTC
PC	1	0	2021-01-01	9:00		.			2021-01-01T09:00:00	
EX	1	1	2021-01-01	9:01	H	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-02	9:01	H	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-03	9:01	H	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-04	9:01	H	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-05	9:01	H	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-06	9:01	H	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-07	9:01	H	5	2021-01-01	2021-01-07		
EX	1	1	2021-01-08	9:30		10	2021-01-08	2021-01-10		
PC	1	0	2021-01-09	9:10		.			2021-01-09T09:10:00	2021-01-08T09:30:00
EX	1	1	2021-01-09	9:11	H	10	2021-01-08	2021-01-10		
EX	1	1	2021-01-10	9:11	H	10	2021-01-08	2021-01-10		
EX	1	1	2021-01-12	9:11	H	10	2021-01-12	2021-01-14		
EX	1	1	2021-01-13	9:11	H	10	2021-01-12	2021-01-14		
EX	1	1	2021-01-14	9:11	H	10	2021-01-12	2021-01-14		
EX	1	1	2021-01-15	8:45		15	2021-01-15	2021-01-21		
PC	1	0	2021-01-16	9:20		.			2021-01-16T09:20:00	2021-01-15T08:45:00
EX	1	1	2021-01-16	9:21	H	15	2021-01-15	2021-01-21		
EX	1	1	2021-01-17	9:21	H	15	2021-01-15	2021-01-21		
EX	1	1	2021-01-18	9:21	H	15	2021-01-15	2021-01-21		
EX	1	1	2021-01-19	9:21	H	15	2021-01-15	2021-01-21		
EX	1	1	2021-01-20	9:21	H	15	2021-01-15	2021-01-21		
EX	1	1	2021-01-21	9:21	H	15	2021-01-15	2021-01-21		

Figure 7. PK and Dosing Records with Time Imputation

At this step, as a convention, when imputing missing time an offset of 1 minute or 5 minutes is implemented to specify the order of the PK sample and dosing record on the same day – PK sample is drawn first, followed by medication administration. An imputation flag ATMF is also included to show the level of imputation of missing time, which is hour.

After this step, the imputation of missing time of dosing records are completed. Again, the imputed time of dosing records won't be directly used in PK analysis, but time imputation is necessary to ensure that the order of dosing history and PK observations is correct.

COMMONLY SEEN DATA ISSUES REGARDING DOSING DATE AND TIME

When working on PK and dosing data, one should be cautious whether EX and PC data are "communicating" well, meaning that information from EX data and PC data are not supposed to contradict with each other. In this section, two common data issue scenarios are discussed.

SCENARIO 1

The study design implies that it's oral daily dosing, but last dose datetime in PC domain may point to different times on the same day.

DOMAIN	SUBJID	PCDTC	PCTPTREF	PCRFTDTC
PC	2	2021-01-01T09:00:00		
PC	2	2021-01-09T09:15:00	Last Dose Prior to PK Sample	2021-01-08T09:30:00
PC	2	2021-01-10T09:20:00	Last Dose Prior to PK Sample	2021-01-08T10:30:00

Figure 8. Data Issue 01 - PC Domain

This subject has PK tests performed on 2021-01-09 and 2021-01-10, respectively, both of which points to the same date 2021-01-08 as its last dose date. However, one is at 09:30 and the other is at 10:30, implying that the subject has taken two doses on 2021-01-08, which contradicts with the study design. A query should be sent to the site for this kind of scenario.

SCENARIO 2

Scenarios 2 is that last dose datetime of a PK sample in PC domain disagrees with what is collected in EX domain.

DOMAIN	SUBJID	EXSTDTC	EXENDTC	EXDOSTOT
EX	3	2021-01-01	2021-01-07	5
EX	3	2021-01-08	2021-01-14	10
EX	3	2021-01-15	2021-01-21	15

Figure 9. Data Issue 02 - EX Domain

DOMAIN	SUBJID	PCDTC	PCTPTREF	PCRFTDTC
PC	3	2021-01-01T09:00:00		
PC	3	2021-01-09T09:05:00	Last Dose Prior to PK Sample	2021-01-08T09:40:00
PC	3	2021-01-16T09:25:00	Last Dose Prior to PK Sample	2021-01-10T08:30:00

Figure 10. Data Issue 02 - PC Domain

This subject has a PK test performed on 2021-01-16, and the last dose prior to this PK sample is on 2021-01-10. However, the collected EX data disagrees with this. From EX domain, it shows that this subject has been taking doses consecutively, implying that the last dose prior to the PK test on 2021-01-16 should be 2021-01-15, not 2021-01-10. A query should be sent to the site for this kind of scenario.

Besides the scenarios discussed above, other commonly seen data issues should also be monitored, such as overlapped dosing records in EX domain, or a collected last dose datetime is after PK sample datetime.

CONCLUSION

This paper presents programming steps on how to impute missing time of dosing records for creation of NONMEM®-ready popPK dataset. The discussion in the paper includes the studies that are in scope, how data is collected in eCRF and organized in datasets, detailed steps on how to impute missing time of dosing records and commonly seen data issues.

This paper gives one method to impute the missing time of dosing records. Based on study design and data collection design, however, the imputation methods on missing time of dosing records can be different. The general rule is that with imputation on missing time, the sequence of dosing history data and PK observation data stay correct. The programming steps discussed in the paper are suitable for studies with oral daily dosing regimen, but the method can also be adjusted for use in studies with other dosing regimens.

LIST OF ABBREVIATIONS

NONMEM®	Non-Linear Mixed Effects Modeling
PK/PD	Pharmacokinetic/Pharmacodynamic
popPK	Population Pharmacokinetics

REFERENCES

Koukuntla, Raghu Kishor. 2014. “NONMEM® – A Programmer point of view.” *Proceedings of PhUSE 2014*. London, UK: *PhUSE*. Available at https://lexjansen.com/phuse/2014/pp/PP12.pdf.

Reza Sharmeen. 2015. “Growing Needs in Drug Industry for NONMEM Programmers Using SAS®.” *Proceedings of PharmaSUG 2015*. Orlando, FL: PharmaSUG. Available at https://www.pharmasug.org/proceedings/2015/SP/PharmaSUG-2015-SP07.pdf.

Reddy, Sree Harsha Sreerama and Subramoney, Vishak. 2018. “Dosing In NONMEM® Data Sets an Enigma.” *Proceedings of PharmaSUG 2018*. Seattle, WA: PharmaSUG. Available at https://www.pharmasug.org/proceedings/2018/BB/PharmaSUG-2018-BB02.pdf.

Su, Jing and Kang, Jiannan (Jane). 2018. “Challenges and Strategies in PKPD Programming.” *Proceedings of PharmaSUG 2018*. Seattle, WA: PharmaSUG. Available at https://www.pharmasug.org/proceedings/2018/AA/PharmaSUG-2018-AA13.pdf.

SAS Help Center: RETAIN Statement. Accessed March 27, 2021. Available at https://documentation.sas.com/?cdcId=pgmsascdc&cdcVersion=9.4_3.5&docsetId=lestmtsref&docsetTarget=p0t2ac0tfzcgbjn112mu96hkgg9o.htm&locale=en.

ACKNOWLEDGMENTS

The author would like to thank all Merck & Co., Inc., Kenilworth, NJ, USA colleagues who have provided their valuable inputs on this paper.

CONTACT INFORMATION

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

Shuqi Zhao
Senior Scientist, Statistical Programming
Merck & Co., Inc., Kenilworth, NJ, USA
E-mail: Shuqi.zhao@merck.com