PhUSE US Connect 2018

Paper DH11

Practical solutions when analyzing incomplete disposition data

Srinivas Veeragoni, Bayer, Whippany, USA
Dr. Juergen Lembcke, Bayer, Berlin, Germany
David Wei, Bayer, Toronto, Canada

ABSTRACT

Subject disposition in clinical trials provides important information such as protocol mile stones (i.e. informed consent, screening, and randomization) and disposition events (i.e. end of screening, end of treatment), and is sometimes directly related to efficacy events. There are fewer challenges if we have clean data and information about date/time for all disposition events as well as exact information about what happens after the disposition event. But in reality this is not the case for the ongoing trials. So how are we able to present summarized disposition information for interim deliverables? Now, think about also reconciling the subject visits data with the disposition data to further evaluate subject's status, incidence of adverse events and treatment discontinuations. This paper will discuss challenges of working with such data and propose solutions to develop & validate algorithms allowing for clear and meaningful presentation of summarized subject disposition.

INTRODUCTION

Analyses of disposition data with the subjects still ongoing in the study treatment combined with cutoff dates and dirty data pose challenges to the statistical programmers and statisticians. In addition, in some studies the survival status on cutoff date is used for censoring purposes and a so-called survival sweep is performed shortly after the cut-off date. These data, even collected after cutoff, will not be cut, because they are used for efficacy analysis. There might be subjects who withdraw consent or discontinue from a particular epoch but still might appear in following epochs. Conversely the subject might complete the previous epoch but may not appear in subsequent epochs. In order to generate disposition tables with these complex scenarios it is important to have an idea of all such possible cases and develop algorithms accordingly.

The most common tables to display disposition information are: Overall Disposition, Disposition by epoch (i.e. Screening, Treatment, and Long-term follow-up), premature discontinuations by cycle or visit etc. Each of these disposition-by-epoch tables in general constitute the number of subjects entering an epoch, number completed, number discontinued and number ongoing. The specific definitions of these disposition endpoints might vary from study to study based on a variety of factors. A common practice is to describe each disposition endpoint in the study protocol and/or Statistical Analysis plan however often the written definitions can be vague. When creating the summary tables, often we are forced to make assumptions and this paper will highlight such scenarios and identify work-arounds to summarize the disposition data appropriately.

Disposition	Treatment A	Treatment B	Total
Number (%) of subjects			
Enrolled			xx
Screening failures			xx
Assigned to treatment	xx (100.0%)	xx (100.0%)	xx (100.0%)
Study drug never administered	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Started treatment	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Ongoing with study treatment	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Discontinued study treatment	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Ended survival follow-up	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)

Table 1: Sample - Overview of Subject Disposition Sample

TYPES OF DATA CUTS:

There are different scenarios of data cuts depending on the analysis which is planned for the database after cut.

1. SNAPSHOT:

A snapshot cutoff is a complete copy of the database which is stored in a certain folder, e.g. for some DMCs. No further data cleaning is performed after the cut. After study started, on a certain date (cut-off date) all data will be taken without changes. So data are stored as clean as they are at the date of cut-off.

2. DATE BASED CUT-OFF WITH ADDITIONAL TIME FOR CLEANING:

A cut-off is done for pre-defined date (cut-off date). There is data cleaning needed to be performed so that data are as clean as possible until that cut-off date. After this "data cleaning period", a data cut procedure is run to remove all data with a date after the cut-off date as shown in the Figure 1.

3. VISIT BASED CUT-OFF:

Visit based cut-off is related to a planned visit, which is used for analysis, e.g. one year analysis even if the study is running for two years. There is data cleaning needed to be performed so that data are as clean as possible until that cut-off visit.

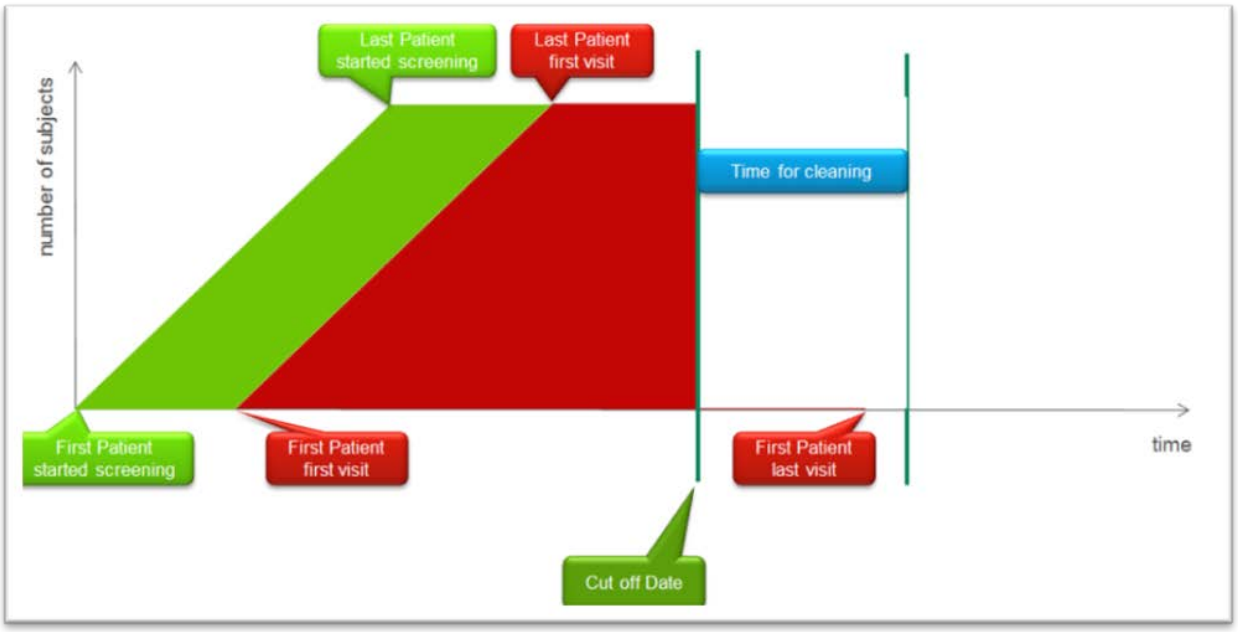


Figure 1: Date based cut-off (after cut-off algorithm is applied)

CHALLENGES IN DATA CUT AND PROGRAMMING DISPOSITION TABLES

CHALLENGE 1: CUTOFF DATA

As discussed in the above section analyses in some of the studies are usually done on cutoff data. The cutoff date is decided based on the purpose of the analyses (like DMC, interim analyses) or for stopping the study (i.e. the date when approximately x number subjects were anticipated to have progressed or died etc.). Cutoff of the disposition data (DS) based on this cutoff date might be simple for some records which have a clear disposition date collected. But there might be some records which have their dates collected in the subject visit domain (SV). In such cases it is necessary to reconcile the dates from SV domain with the DS domain before doing the cutoff. One additional challenge would be with the missing date of the disposition event even after applying imputation rules specified in the study SAP. In such cases a defensive coding approach should be considered before considering worst case scenarios. For example if an informed consent date is after cutoff date at the early stage DMC analyses then all subsequent disposition records for that subject could be cutoff irrespective of their dates as it is clear that all subsequent disposition events would occur only after the informed consent date. In the same way if an End of Treatment (EOT) visit occurs after cutoff date then the safety and survival follow-up records for this subject could be deleted as these visits occur after EOT visit as per the study flow.

Example SAS code to reconcile SV and DS before applying cutoff algorithm:

```
DATA dssv;
  MERGE DS(in=ds) SV(IN=sv KEEP= usubjid epoch visit svstdt);
    BY usubjid epoch visit;
    IF ds;
    IF dsstdt = . AND svstdt ^ = . THEN dssvstdt = svstdt;
     ELSE dssvstdt = dsstdt;
RUN;
```

Cutoff the DS data based on the new reconciled disposition date variable created in above step. Note that records with missing disposition date will be left off in the database using worst case scenario.

```
DATA cutoffds;
 SET dssv;
  WHERE dssvstdt <= cutoffdate;
RUN;
```

Example SAS code using defensive coding to cutoff the disposition events that occur after an earlier event/milestone which is after cutoff date. Note that from the code below, the later events will be cutoff even if they have missing dates because the “CUTFLAG” variable retains its value from the previous event which has a date before cutoff.

```
DATA ds_cutflag;
 SET ds;

  BY usubjid epochn
  dsstdt; RETAIN cutflag;

     IF first.usubjid THEN cutflag=0;

  **to cutoff all subsequent records on or after informed consent for the
     subject**;

     IF infconsdate > cutoffdate THEN cutflag=1;

  **to cutoff safety and survival follow-up records which occur after an
   end of treatment (EOT) visit which is after the cutoff date**;

  **dsstdt corresponds to EOT visit**;

     IF dsstdt > cutoffdate THEN cutflag=1;

RUN;

DATA cutoffds;
 SET
 ds_cutflag;
  IF cutflag =1 THEN DELETE;
RUN;
```

CHALLENGE 2: SURVIVAL SWEEP

For the primary analyses of some studies a survival sweep is sometimes performed to have a complete survival information data about the subjects in the study up to cutoff. This data is usually used in efficacy analyses as an indicator to find out whether the subject was alive until the cutoff date to be censored on cutoff date. So we need to make sure that the survival sweep data is available in the DS domain even after cutoff but should not be used for the disposition summaries.

Example SAS code to avoid cutting off survival sweep data from the DS database:

```
DATA cutoffds;
 SET ds;
  IF (dsstdt > cutoffdate)
    AND (dsdecod ^= “SURVIVAL ASSESSMENT” or dsdecod ^= “DEATH”) THEN DELETE;
RUN;
```

The resulting cutoff DS domain can be further cutoff in the disposition table programs using the algorithms mentioned in the CUTOFF data challenge 1 above with some minor tweaks.

CHALLENGE 3: HOW TO POPULATE TREATMENT END DATE AFTER CUTOFF

If treatment (cycle) start date is before cut-off and end-date is at or after cut-off, EXENDTC could be set to cut-off date and add flag variable (CUTFL) to identify cut records. If the treatment (cycle) start date is before cut-off and end-date is unknown or if the subject is ongoing in treatment then it should be clearly defined in the SAP whether or not to consider the cutoff date as treatment end date. The treatment end date could be populated as +1 day after the start date of the treatment cycle or the cutoff date could be used as the treatment end based on the clinical team's input.

PRESENTING THE SUBJECT DISPOSITIONS AND VALIDATING THE RESULTS

The formal way of presenting the subject dispositions is by disposition summaries with frequency counts and corresponding percentages under each treatment arm. Each of these disposition-by-epoch tables in general constitute the number of subjects entering an epoch, number completed, number discontinued and number ongoing. Though these summaries give detailed information about the disposition events in that epoch they do not show a clear connection between different epochs and sometimes the numbers look confusing if we look across the summaries in different epochs. So the better way to present these disposition summaries is by Disposition Flow Diagrams. There are several approaches of creating these flow diagrams using SAS and other tools. The following two cases illustrate how the flow charts will make it easy to track the subjects through different phases of the study.

CASE 1: COMPLETED STUDY:

In this example a sample completed study is presented. As per the study design the subjects are screened and assigned to treatment. Subjects that completed treatment or discontinued prematurely should be followed up for few months after end of treatment. However the flow chart below shows some possible challenges and probable solutions to handle those challenges.

X Numbers (X) in the circle indicates the time points in the flow chart (Display 1) at which the described challenges could occur. Repeated numbers indicate same or similar scenarios at different time points.

CHALLENGE 1 : OVERLAPPING EPOCHS

Since there are, by definition, no gaps between Elements, the value of SEENDTC for one Element will always be the same as the value of SESTDTC for the next Element. This means an overlap if we do not have a real time. For example the treatment Epoch ends with the end of dosing, but if in the trial only the date, and not the time, of the dosing has been collected then some challenges can occur. If an event has occurred on treatment end day which could also be the Follow-up epoch's start day as shown in the example below, then it could be challenging to assign an Element or an Epoch on the basis of dates alone. When sponsors choose to collect only dates, there could be ambiguity in the algorithms they use to assign the events to Elements or Epochs. These algorithms should be clearly defined by taking clinical team's input and documented in the define file and reviewers guide. While displaying these events in the disposition tables by epoch it would appropriate to show the event only in one epoch to keep the patient flow clear from one epoch to another.

Example:

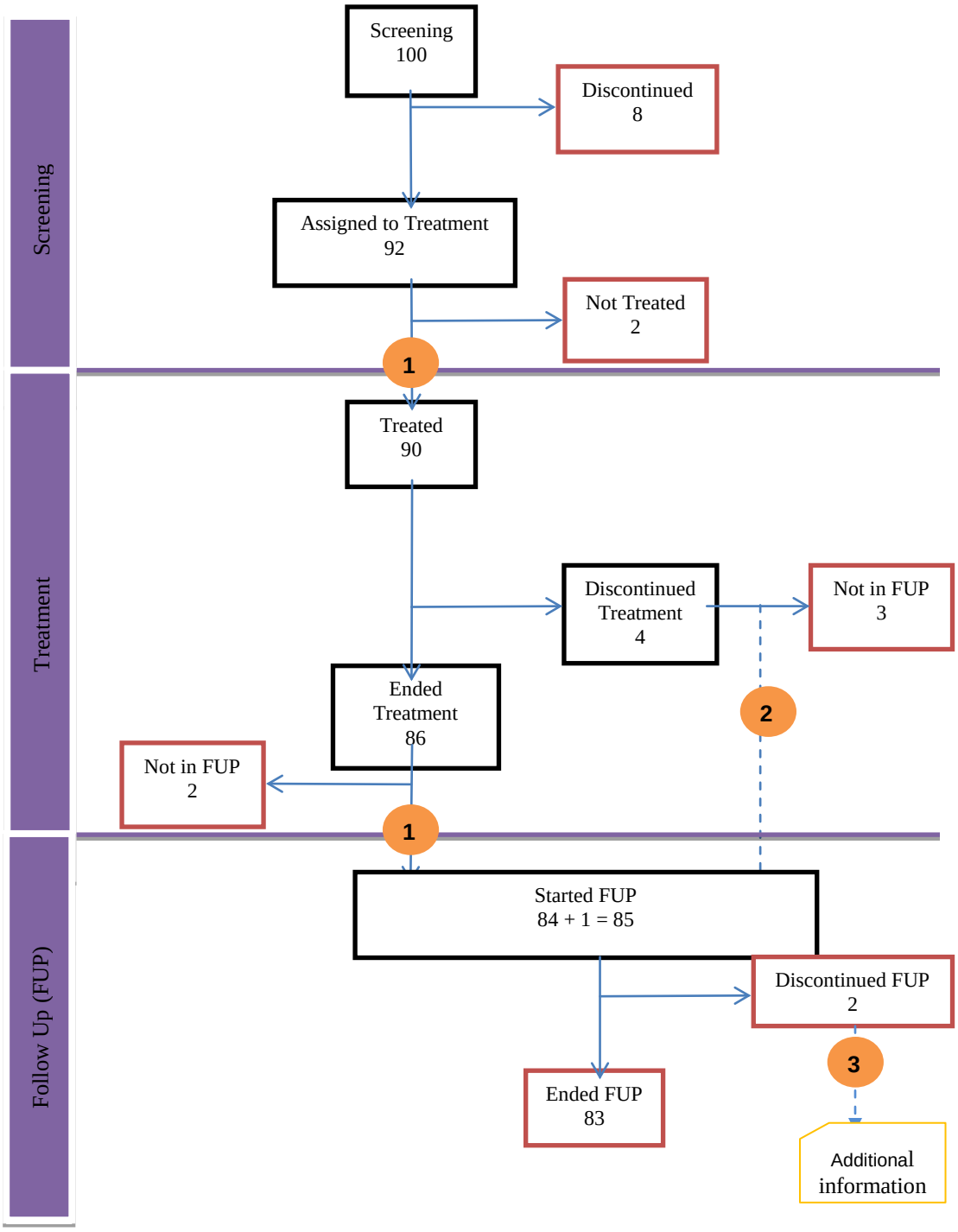
Row	DOMAIN	USUBJID	SESEQ	ETCD	SESTDTC	SEENDTC	EPOCH
1	SE	101	1	TRT	2018-01-01	2018-01-15	TREATMENT
2	SE	101	2	FU	2018-01-15	2018-03-30	FOLLOW-UP

Row	DOMAIN	USUBJID	DSTERM	DSDECOD	DSCAT	EPOCH	DSDTC	DSSTDTC
1	DS	101	Heart Failure	PD	DISPOSITION EVENT	TREATMENT? Or FOLLOW-UP?	2018-01-15	2018-01-15

In the example above it could be clear that the subject stopped the treatment before PD if the time was collected or the end of treatment disposition page is completed with such information and the FOLLOW-UP epoch could be assigned. However, if such information is not available taking the worst case approach the death could be assigned to TREATMENT Epoch and consider PD as reason for treatment termination and document the approach in corresponding documents.

.

Display 1: Disposition Overview – Study completed



Legend:

The boxes in black indicate the subjects still in the study at that time point

The boxes in red indicate the subject disposition from the study

X Numbers (X) in the circle indicates the time points at which the described challenges could occur as marked on the flow chart. Repeated numbers indicate same or similar scenarios at different time points.

CHALLENGE 2: PROTOCOL ALLOWED DEVIATIONS IN PATIENT FLOW

Some study designs allow subjects to skip some epochs based on some criteria like disease status. For example, if subjects have disease progression during treatment phase, the study design could allow the subjects to skip Active Follow-Up period and directly go to Long-term Follow-Up. In this case it is important to have a thorough understanding of the study design to know the different paths the subjects could take during the course of the study. Cross checks should be built in based on multiple data sources like DS, EX, AE, RS and SV to account the subject in appropriate disposition tables.

CHALLENGE 3: DOCUMENTATION OF DATA AFTER PATIENT WITHDRAWN FROM STUDY

There could be scenarios where the subjects could withdraw consent from further follow-up and the database captures some information from external sources about the patient status after the withdrawal. This could be an additional challenge whether or not to rely on that information for our analysis. These cases should be discussed with the DM team to verify the source documents and build a standard process to perform analysis based on this information and following standard SOPs.

CASE 2: ONGOING STUDY

In the Display 2 a sample ongoing study is presented. As per the study design the subjects are screened and assigned to treatment. Subjects that completed treatment or discontinued prematurely should be followed up for few months after end of treatment. Since the study is still ongoing, data cutoff is applied for an interim analysis. The flow chart shows some possible areas of challenges and some solutions to handle those challenges are discussed below.

CHALLENGE 1: ACCOUNTING FOR SUBJECTS WITH MISSING INFORMATION

As shown in the display, cutoff could leave many data points open ended with no further information. In this example, 100 subjects enrolled for screening, 6 subjects discontinued during screening period and 92 were assigned to treatment (randomized). However there are 2 subjects who are still unaccounted for as there is no further information for those subjects after they have signed the informed consent and it is expected that those 2 subjects would complete all screening procedures in the next few days. From the cutoff database we would not have any information if they completed the screening period and were assigned to treatment after the cutoff. So in these cases an informed decision has to be made by the study teams based on the study design and protocol to account those subjects in the appropriate epoch disposition tables. Here the flow chart could be very helpful to figure out those unaccounted subjects.

Similarly subjects in the boxes “Treatment not started (1)”, “Ongoing Treatment (10)”, “Follow-up not started (4)”, and “Ongoing follow-up (10)” might not have enough information to classify them under this groups. In this case we would have to take the approach of elimination to classify them to appropriate stage. For example 76 subjects ended treatment and 4 subjects have been put under “Follow-up not started”. We should verify all the 76 subjects who ended the treatment to see if there is any follow-up record and if not available check the reason for treatment termination. If the reason is “Death” or “Lost to follow-up” or “Subject withdrawal” then we should consider those subjects under “Not in Follow-up” and put the remaining subjects under “not started FUP” – in this case 4 subjects.

CHALLENGE 2: FINDING MISSING LINKS BETWEEN EPOCHS

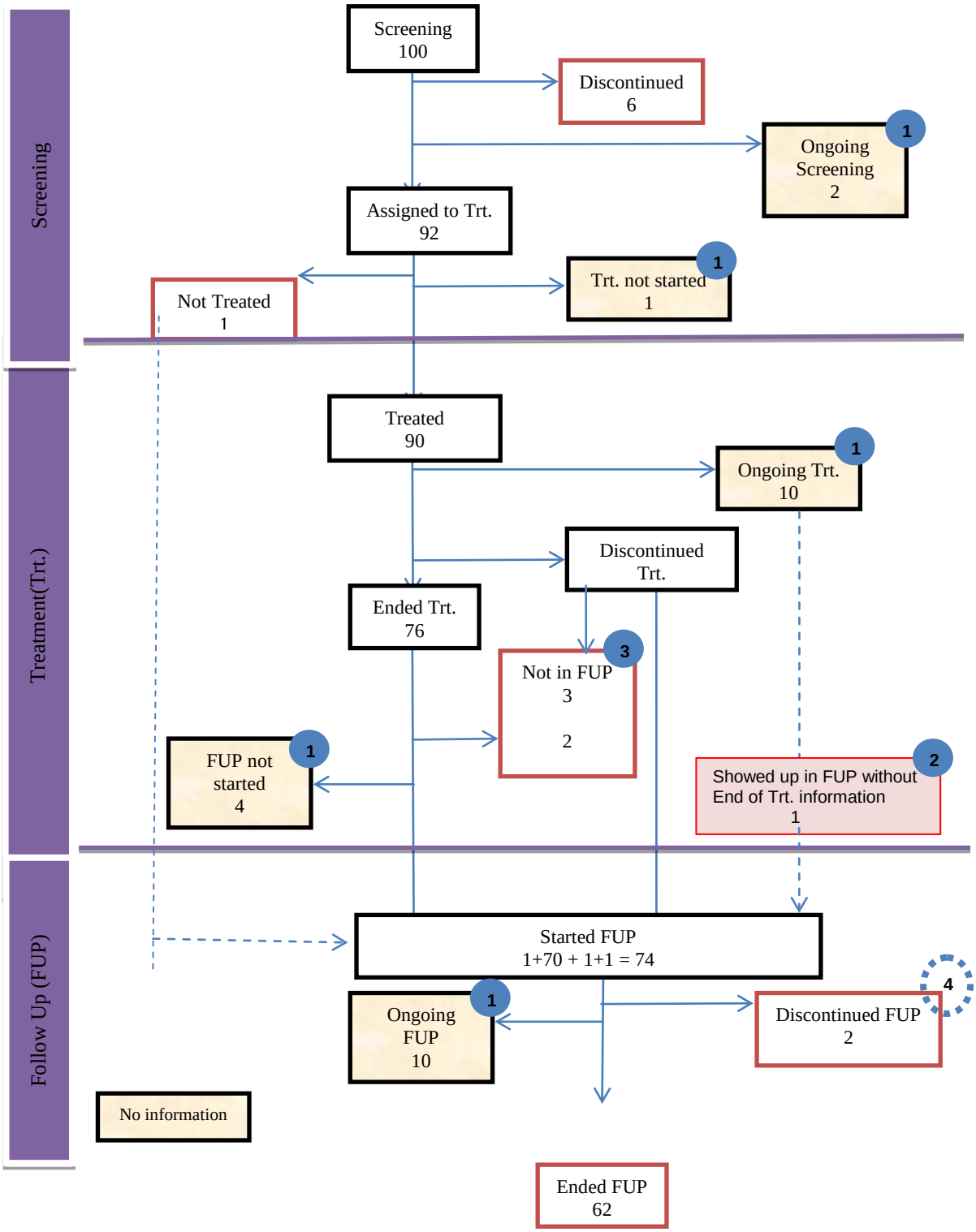
There might be subjects with missing termination information in an epoch but could show up in subsequent epochs. For example, there could be information about a subject’s follow-up epoch but the corresponding end of treatment information could be missing. This could be because of unclean or missing data. For analysis such subjects could still be considered as ended treatment as they could only enter the subsequent epochs if they have completed the previous epochs.

In the chart below there are 10 subjects who are considered as ongoing in treatment as they do not have “end of treatment” page filled yet at the time of cutoff. However one of these subjects had a follow-up record populated in the database which is still before the cutoff date. So it could be assumed that the subject’s end of treatment information is missing and the subjects should be counted under End of treatment epoch disposition table and also counted under “Started Follow-up”.

CHALLENGE 3: ACCOUNTING FOR SUBJECTS WHO DISCONTINUE DUE TO SOME SPECIFIC EVENTS

In this case there could be some subjects who discontinued from the study due to specific events which could make it clear that they are not expected to enter the following epoch. For example if a subject discontinues from treatment epoch due to reasons like “withdrawn consent from complete study“ or “death“ then you could assume that the subject might not enter the subsequent epochs. However there could be cases where there could be still some grey areas (events) like “lost to Follow-up” as reason for end of treatment termination but could show-up few months later for the follow-up visit. In this case all the available information available for that subject in the database should be carefully evaluated before summarizing them in the respective epoch tables.

Display 2: Example for Disposition Overview (Ongoing Trial)



Legend:

The boxes in black indicate the subjects still in the study at that time point

The boxes in red indicate the subject disposition from the study

x — Numbers (X) in the circle indicates the time points at which the described challenges could occur as marked on the flow chart. Repeated numbers indicate same or similar scenarios at different time points.

CHALLENGE 4 : "CENSORING" SUBJECTS WITH INFORMATION LOST DUE TO CUTOFF

In Time-to-Event analysis subjects should be censored if they do not have event of interest at the time of analysis. If the event happens after the cutoff and that event's information is lost during cutoff process then it could be tricky to find out at what time point the subject should be censored. For example if there was an assessment short after cut-off, Should we censor time to event on cut-off day or on last assessment day? In other words, do we use the information about assessment after cut-off or not? If we could use the information beyond cutoff then the subject should be censored on cut-off day. If we could not then we should censor the subject on the day of last assessment before cutoff. This needs to be specified in the SAP clearly. As we have seen the Survival sweep section, information is specifically collected about the subject's survival status after cutoff to confidently present the subject's survival information in the overall survival analysis.

CONCLUSION

We have seen the challenges in analyses of disposition data both for a completed study and an ongoing trial with the subjects still ongoing in the study treatment with cutoff dates applied and dirty data. It could be hard to track and visualize the patient flow across different epochs using regular disposition by epoch tables. Disposition flow charts could be very helpful in these cases which could help account for all possible scenarios. In addition care should be taken during study designing, CRF development, data collection and cutoff execution to collect and store required information (example: a flag variable to know if the subject enters to next epoch at the end of each epoch page) to answer some of the critical questions. It is also very important to document all the assumptions in assigning the epochs in the define files and analysis data reviewers guide. In addition submission of analysis programs proactively with supporting Analysis Results Metadata (ARM) will help the reviewers understand the algorithms easily, could reduce some of the information requests and speed up the overall review process.

ACKNOWLEDGMENTS

We would like to thank Rene Wentzeck and Ankur Mathur for their support in writing this paper.

RECOMMENDED READING

Brian Fairfield-Carter and Suzanne Humphreys. 2011. "Alternative Approaches to Creating Disposition Flow Diagrams" Proceedings of 2011 Pharmaceutical SAS Users Group. https://www.pharmasug.org/proceedings/2011/TT/PharmaSUG-2011-TT08.pdf

Mei Dey and Ann Croft. 2018. "Implementation of Data Cut Off in Analysis of Clinical Trials". Proceedings of 2018 Pharmaceutical SAS Users Group. https://www.pharmasug.org/proceedings/2018/DS/PharmaSUG-2018-DS19.pdf

Huijuan (Tracy) He. 2018. "An Exploratory Way to Draw a Flow Diagram for the Subject Disposition of a Two-Arm Randomized Trial Using SAS® 9.4 M3". Proceedings of 2018 Pharmaceutical SAS Users Group. https://www.pharmasug.org/proceedings/2018/AD/PharmaSUG-2018-AD28.pdf

CONTACT INFORMATION

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

Srinivas Veeragoni
Bayer
100 Bayer Boulevard
Whippany, NJ, USA / 07981
Email:
Srinivas.Veeragoni@bayer.com

Dr. Juergen Lembcke
Bayer AG
Building P300, A130
13342 Berlin, Germany
E-mail:
juergen.lembcke@bayer.com

David Wei
Bayer Inc.
2920 Matheson Blvd E
Mississauga, ON L4W 5R6,
Canada
E-mail: david.wei@bayer.com