

Applying Different Data Cutoff Dates for the Analysis of a Study Milestone

Guowei Zhao, Hong Qi, Sapan Patel, Merck & Co., Inc., Rahway, NJ, USA.

ABSTRACT

In clinical trials, applying data cutoff in SDTM (Study Data Tabulation Model) data is crucial for supporting analysis and reporting activities, and subsequent regulatory submissions. Typically, a single data cutoff date is assigned to each study milestone, and the data collected up to that cutoff is utilized for analysis. However, certain trials pose a unique challenge where distinct cohorts have different cutoff dates. It is ideal to apply specific data cutoffs to each corresponding cohort to optimize data analysis while maintaining adherence to industry standards. This paper explores the importance of applying variable data cutoff dates for different cohorts associated with a single milestone in a clinical trial. It addresses the challenges posed by this situation and examines the potential impact on data analysis and interpretation. Furthermore, the paper proposes an effective solution that addresses these challenges while optimizing data analysis and streamlining the submission process.

INTRODUCTION

The application of a predetermined data cutoff (DCO) date is pivotal in ensuring the accuracy for the analysis and reporting of clinical trial data. By excluding incomplete or inconsistent data from ongoing or late-reported patient outcomes, the DCO minimizes the risk of introducing bias, thus safeguarding the integrity of study results. This uniform evaluation of all study participants at the same timepoint ensures consistency in between treatment group comparisons, enhancing the reliability of statistical outcomes based on well-defined datasets for a milestone. In addition, applying a DCO date is often required by regulatory agencies for a successful submission and approval. The requirement ensures that the analysis and reporting are based on clearly specified datasets, thereby enhancing reproducibility.

Throughout the course of a clinical trial, it is common to have multiple cutoff dates that correspond to different milestones including safety monitoring, interim analysis, final analysis, etc. While typically only one DCO date is associated with a specific milestone, unique clinical trial scenarios may necessitate different DCO dates within a milestone. For instance, in trials involving multiple cohorts, distinct DCO dates for each cohort ensure that the data analysis and reporting accurately reflect the specific requirements and objectives of each cohort, facilitating a comprehensive understanding of the trial outcomes.

In this paper, we aim to highlight a unique scenario demonstrating how the application of different DCO dates for distinct cohorts plays a pivotal role in ensuring accurate data analysis and reporting within a milestone. We will present a method detailing how these distinct DCO dates can be precisely applied and discuss its potential applicability in clinical trials facing similar circumstances. For example, in trials with changing interventions over time, the use of different DCO dates in one milestone can enhance the evaluation of varied intervention periods, enabling a more comprehensive understanding of treatment efficacy and safety at specific time points.

KEY CONSIDERATIONS

JUSTIFICATION OF APPLYING ONE-CUTOFF VS. TWO-CUTOFF

In our study population, there are two distinct cohorts characterized by a different mismatch repair (MMR) status, i.e., proficient MMR (pMMR) or deficient MMR (dMMR) as determined through a central or an institutional laboratory test. To ensure unbiased analysis, these two cohorts were randomized in a 1:1 ratio to two treatment groups and both were included in efficacy analyses based on the intention-to-treat (ITT) principle. A different DCO date was assigned to each cohort, with the pMMR cohort having a DCO date that was 10 days earlier.

The one-cutoff strategy is to apply the later DCO date to both cohorts. This approach is straightforward and easy to implement. By using the later DCO date, it ensures that all relevant events necessary for analysis and reporting are included. However, it is important to note that there were nine additional

participants randomized into the pMMR cohort between the two DCO dates, but their data were not included in the marketing application. The exclusion of these participants beyond the distinct DCO date becomes necessary to maintain the accuracy of the statistical analyses and prevent any potential compromises.

On the other hand, the two-cutoff strategy involves applying a specific DCO date for each corresponding cohort. This approach ensures that only data up to and including the specific DCO date are included in the analysis for each cohort, which leads to reliable results and facilitates informed decision-making by regulatory agencies.

DETERMINING THE DISTINCT COHORTS BEFORE APPLYING THE DCO

To implement the specific DCO date for each distinct cohort, an accurate method to distinguish these cohorts need to be established as the first step. This step becomes particularly critical when direct information is unavailable in the source SDTM datasets. Due to the unique study design and statistical analysis plan, the MMR status in our study is determined through either a central laboratory test or an institutional laboratory test with the consideration of the randomization date. Based on this information, an intermediate dataset is created from the SDTM datasets, containing the final MMR status for each participant. This intermediate dataset is then used to subset the SDTM datasets into two distinct sets based on the MMR status.

COMBINING COHORTS AFTER APPLYING THE DCO

After applying the MMR-status-specific DCO to each set of SDTM datasets, it is important to properly combine the resulting datasets for the pMMR and dMMR cohorts, and perform a subsequent validation check to confirm the effectiveness of the DCO in the final SDTM datasets. This step ensures compliance with regulatory agency requirements and facilitates the submission of complete and accurate SDTM datasets for review and approval.

It is worth noting that trial domains do not go through the DCO process due to their special data structures. Instead, they are copied directly to the final SDTM library.

PROGRAMMING APPROACH

This section provides a description of how the key considerations have been implemented using the programming approach.

STEP 1 DERIVE DATE VARIABLES WITH NULL VALUE

This step is only necessary when the SDTM date variables used for the DCO process contain null values. Such null values in SDTM date variables typically occur in studies with data collection or SDTM mapping that does not complaint with CDISC standards. In our case, we encountered this challenge as our study is an external collaboration trial with unconventional data collection and SDTM mapping.

This issue affected two SDTM date variables: DSDTC (Date/Time of Collection) in the SDTM DS (Disposition) dataset and SSDTC (Date/Time of Assessment) in the SDTM SS (Subject Status) dataset. To accurately apply the DCO to the two datasets using an existing macro, we derived DS.DSDTC and assigned SS.SSDTC based on the available date information in the SDTM datasets. Following the completion of the DCO process, we reassigned these derived date values back to their original null values.

DS.DSDTC was derived as the latest date among the values of DS.DSSTDTC (Start Date/Time of Disposition Event), and SUPPDS.QVAL for which SUPPDS.QNAM is LCNTDTC (Date of Last Contact) or LSTRDTC (Date of Last Treatment). On the other hand, SS.SSDTC was assigned as the value of SUPPSS.QVAL when SUPPSS.QNAM is SSLCDTC (Date of Last Contact). The macro utilized to derive DSDTC is illustrated in Display 1.

```

%macro dsdtc;
**** Merck global standard macro to combine SDTM domain with suppqual ****;
%build0sdtmpplus(domain=ds, sdtmlib=lptsse);
**** create macro variable for all variable names in DS domain ****;
proc sql noprint;
  select name into: vars separated by " "
    from dictionary.columns
   where upcase(libname)="LPTSSE" and upcase(strip(memname))="DS"
  ;
quit;
**** rederive DSDTC as the latest date from DSDTDC, LCNTDTC, LSTRDTC ****;
data ds_;
  retain &vars;
  length dsdtc $100.;
  set dsplus(drop=dsdtc);
  call missing(dsdtc);
  if missing(dsdtc) then do;
    if ~missing(lstrdtc) and missing(lcntdtc) then dsdtc=strip(lstrdtc);
    else if missing(lstrdtc) and ~missing(lcntdtc) then dsdtc=strip(lcntdtc);
  end;
run;
data lptsse.ds (keep=&vars label="Disposition");
  retain &vars;
  set ds_;
  label dsdtc = "Date/Time of Collection";
  dsstdtc1=input(dsstdtc, ?? is8601da.);
  dsdtc1=input(dsdtc, ?? is8601da.);
  if n(dsdtc1,dsstdtc1)=2 then do;
    if dsdtc1 ne dsstdtc1 then do;
      if max(dsdtc1,dsstdtc1)=dsstdtc1 then do;
        dsdtc=dsstdtc;
      end;
    else if max(dsdtc1,dsstdtc1)=dsdtc1 then do;
      dsdtc=dsdtc;
    end;
  end;
  if dsdtc1 eq dsstdtc1 then dsdtc=dsdtc;
  if missing(dsdtc1) and ~missing(dsstdtc) then dsdtc=dsstdtc;
  if ~missing(dsdtc1) and missing(dsstdtc) then dsdtc=dsdtc;
run;
%mend dsdtc;

```

Display 1 Derive DS.DSDTC before the DCO Process

For compliance and traceability, it is crucial to preserve the SDTM variable name, label, and sequence from the original datasets, as well as the SDTM dataset name and label.

STEP 2 APPLY DCO

2.1 DATA PREPROCESSING

To prepare the intermediate dataset containing the final MMR status information, the following procedures were undertaken.

1. Two intermediate cohort variables were derived using the data from SDTM MI (Microscopic Findings) domain. One variable was created to store the central laboratory results (CMMR), while the other stored the institutional laboratory results (CRFMMR).
2. The randomization date variable (RANDDT) was derived for each randomized participant using the information from the SDTM DS domain.
3. Lastly, a dataset was generated with cohort variable (COHORTI) for determining the distinct cohorts after comparing the randomization date with a protocol-specified date. If a participant was randomized after the specified date, COHORTI was assigned as CRFMMR. Otherwise, COHORTI was assigned as CMMR (refer to Display 2).

```

data subjcoh(keep=usubjid cohorti);
  merge _randdtds2(in=a) cmmr(in=b) crfmmr(in=c);
  by usubjid;
  if a;
  if randdt<='ddmmmyyyy'd then do; cohorti=cmmr; cohortin=cmmrn; end;
  else if randdt>'ddmmmyyyy'd then do; cohorti=crfmmr; cohortin=crfmmrn; end;
run;

```

Display 2 Derive COHORTI with Central or Institutional Laboratory Result

The above intermediate dataset was instrumental in dividing the SDTM datasets into two subsets for pMMR and dMMR cohorts. To avoid inadvertent loss of information, participants with COHORTI value of "INDETERMINATE" or missing were included in dMMR cohort, which had a later DCO date. These participants were confirmed not being randomized or included in the statistical analysis. The resulting

datasets were exported to two separate data libraries for dMMR and pMMR respectively. The macro utilized to carry out this process is exhibited in Display 3.

```
%macro mmr(crit=);
*** select subjects from subjcch dataset ***;
data _subj(keep=usubjid);
set subjcch;
  if &crit eq pmmr then %do;
    where strip(upcase(cohorti)) in ("PMMR");
  %end;
  if &crit eq dmmr then %do;
    where strip(upcase(cohorti)) in ("DMMR", "INDETERMINATE", " ");
  %end;
run;
*** create macro variables for SDTM datasets count, names, and labels ***;
proc contents data=lptsse._all_ out=__ds(where=(upcase(memname) not = 'Z') keep=memname memlabel);
run;
data __ds;
set __ds;
  where memname not in ('TA','TE','TI','TV','TS','LB_ERR');
run;
proc sort; by memname memlabel; run;
proc sql noprint;
  select count(distinct memname) into :nds
  from __ds;
  select distinct memname, memlabel into :inds1-:inds%left(&nds) ,:inlbl1-:inlbl%left(&nds)
  from __ds
  order by memname
  ;
quit;
*** subset SDTM datasets and output to different libraries based on MMR status ***;
%macro subj;
%do i=1 %to %left(&nds);
proc sql;
  if &crit eq pmmr then %do;
    create table lptsspr.&inds%i(Label=&inlbl%i) as
  %end;
  if &crit eq dmmr then %do;
    create table lptssdr.&inds%i(Label=&inlbl%i) as
  %end;
  select a.* from lptsse.&inds%i a,
  _subj b
  where a.usubjid=b.usubjid;
quit;
%end;
%end subj;
%end;
%mmr(crit=pmmr);
%mmr(crit=dmmr);
```

Display 3 Subset SDTM Datasets Based on MMR Status

2.2 APPLYING DCO

In this step, each distinct DCO date was applied to the corresponding subset of SDTM datasets utilizing a Merck global standard macro %apply0date0cutoff. The sample codes illustrating this process are provided in Display 4.

```
%apply0date0cutoff(cutoffdt =%gmvar_cutoff_dmmr
,input_libname =lptsspr
,domains_use_startdtc =%str(AE, BE, CM, DV, EX, EC, FE, HO, PR, PX, SE, SV)
,domains_use_dtc =%str(BS, DA, EG, FA, FF, IE, LB, MB, MH, MI, PC, PE, PF, PY, QS, RS,
SC, TR, TU, VS, CF, HM, LS, PH, ZE, CE, CC, DD, RP, SS, DS)
,output_libname = lptsspr
,debug=N);
```

Display 4 Applying DCO Utilizing Macro %apply0date0cutoff

For SDTM domains listed for the macro parameters “domains_use_startdtc” and “domains_use_dtc”, the macro processes the data to keep records up to the cutoff date. However, some date variables in the SDTM DM (Demographics) and EX (Exposure) domains are processed differently by this macro. If DM.RFSTDTC (Subject Reference Start Date/Time) and DM.RFXSTDTC (Date/Time of First Study Treatment) are greater than the DCO date, DM.RFSTDTC, DM.RFXSTDTC, DM.RFENDTC (Subject Reference End Date/Time), and DM.RFXENDTC (Date/Time of Last Study Treatment) are set to missing. If only DM.RFENDTC and DM.RFXENDTC are greater than the DCO date, they are replaced with the value of DCO date. If DM.RFPENDTC (Date/Time of End of Participation) is greater than the DCO date, DM.RFPENDTC is set to missing. Also, for the record with DM.RFSTDTC greater than the DCO date, both DM.ACTARM and DM.ACTARMCD are set to missing. Similarly, if EX.EXENDTC (End Date/Time of Treatment) is greater than the DCO date, it is replaced with the value of DCO date, and EX.EXENDY (Study Day of End of Treatment) will be re-derived accordingly.

2.3 COMBINING DATASETS

To ensure consistency in data analysis and reporting for all study participants in both cohorts, the two sets of SDTM datasets were combined after completing each DCO step. This is a reverse process of separating datasets in the previous steps following the same programming strategies. The combination of datasets is essential to achieve accuracy, efficiency, and compliance with regulatory submission requirements. However, if only a specific cohort is required for submission, there is no need to undergo the additional step of combining the subset data after the DCO step.

STEP 3 VERIFY ACCURACY

The global macro, %apply0date0cutoff, has undergone extensive validation to ensure accurate application of a single DCO date to the SDTM datasets for a given study milestone. Consequently, the need for a validation step is usually eliminated for studies falling within this category. However, the DCO process for this study has become more complex with the addition of steps such as Step 2.1, as well as additional tasks within Step 2.2 and Step 2.3. Therefore, it is necessary to conduct a validation process to confirm the successful application of the DCO in the final SDTM datasets.

To accomplish this, a dedicated validation program has been developed, encompassing the following functions:

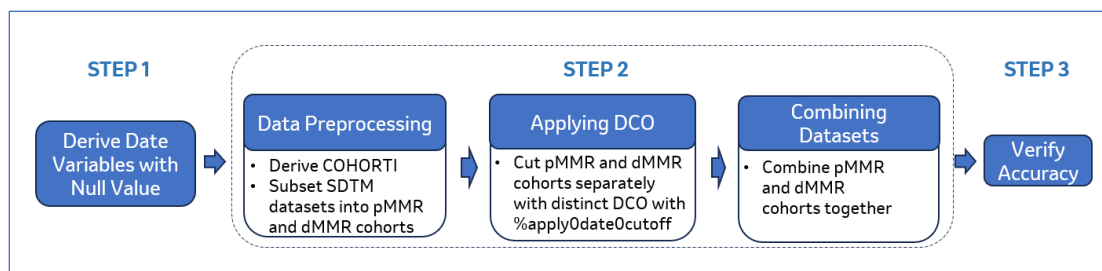
1. Identification of the maximum value for each date variable listed in the %apply0date0cutoff parameters, specific for each cohort.
2. Comparison of the aforementioned maximum dates with the DCO date, considering the participant's MMR status.
3. Printing any date that exceeds the specified DCO date for further investigation.

This validation program serves as a crucial component in ensuring the accuracy and integrity of the DCO process within the final SDTM datasets for this study. Display 5 presents the macro responsible for executing this step.

```
%macro checkcut(lib=, dom=, datevar=, date1=, date2=);
*** get COHORTI, COHORTIN from ADSL ***;
proc sql;
  create table &dom.comb as
  select a.*, b.cohorti, b.cohortin
  from &lib.&dom. as a left join lptda.adsl as b
  on a.usubjid = b.usubjid;
quit;
*** pick the maximum date value ***;
data &dom.;
  set &dom.comb;
  if &datevar. ne "" then do;
    if length(compress(&datevar.)) ge 10 then qval = &datevar.;
    else if length(compress(&datevar.)) = 7 then qval = compress(&datevar.||"-01");
    else if length(compress(&datevar.)) = 4 then qval = compress(&datevar.||"-01-01");
    date=input(substr(qval,1,10),yymmdd10.);
  end;
  format date yymmdd10.;
run;
proc sql;
  create table max as
  select cohortin, cohorti, max(date) as maxvalue format =yymmdd10.
  from &dom.
  group by cohortin, cohorti;
quit;
*** print date if not cut correctly ***;
data print;
  set max;
  length comment $100;
  if cohortin = 1 then do;
    if maxvalue gt "&date2."d then comment = "Greater than cutoff value";
  end;
  if cohortin = 2 then do;
    if maxvalue gt "&date1."d then comment = "Greater than cutoff value";
  end;
run;
title "&lib.&dom. - &datevar.";
proc print data=print;
run;
title;
%mend checkcut;
%checkcut(lib=lptss, dom=ae, datevar=aestdtc, date1=ddmmmyyyy, date2=(dd+10)mmmyyyy);
```

Display 5 Verify the Accuracy of DCO

Display 6 illustrates the flow of programming approach that has been implemented to apply two DCO dates to the corresponding cohorts in one study milestone.



Display 6 Programming Approach for Applying Distinct DCO Dates for Each Cohort

CONCLUSION

This paper provides a comprehensive approach, offering detailed steps on the application of two distinct DCO dates to two cohorts within a single study milestone. It addresses the specific challenges encountered in our study and provides corresponding solutions, with potential applicability to other clinical trials. Our focus on verifying the accuracy of the applied DCO dates is of utmost importance in ensuring the reliability and effectiveness of the analysis.

Additionally, the inclusion of SAS sample codes serves to better illustrate the process. By applying distinct DCO dates for each cohort within a clinical trial milestone, the approach ensures accurate and meaningful analysis, taking into account the unique characteristics and timelines of each cohort. This fosters consistency and comparability in interpreting the trial results.

Overall, the application of distinct DCO dates for different cohorts is essential for enhancing accuracy, facilitating meaningful analysis, and ensuring consistency and comparability across cohort-specific interpretations of trial outcomes.

REFERENCES

Merck BARDS Standard Macro Library User Manual

Mei Dey, Ann Croft. "Implementation of Data Cut Off in Analysis of Clinical Trials". PharmaSUG 2018 – DS19. [PharmaSUG-2018-DS19.pdf](#)

Kiranmai Byrichetti, SCRI, Nashville, TN Jeffrey Johnson, SCRI, Nashville, TN. "Macro to get data from specific cutoff period". PharmaSUG 2016- Paper BB14 [PharmaSUG-2016-BB14.pdf](#)

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to their colleagues at Merck BARDS SP who helped in addressing the unique challenges of multiple DCO within a study milestone and made contributions in developing the Merck global standard macro. Additionally, the authors would like to thank Mary Varughese, the Executive Director, and Amy Gillespie, the head, for reviewing the paper and providing valuable feedback.

CONTACT INFORMATION

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

Guowei Zhao
Merck & Co., Inc., Rahway, NJ, USA
+1 (267)305-1325
guowei.zhao@merck.com

Hong Qi
Merck & Co., Inc., Rahway, NJ, USA
+1 (267)305-7589
hong_qi@merck.com

Sapan Patel
Merck & Co., Inc., Rahway, NJ, USA
+1 (267) 305-6064
sapan.patel@merck.com