

DH06: Can You Cut It? Implementing the Data Cut-off

Lewis Meares, Phastar

Agenda

- What is the data cut-off?
- Why do we need it?
- When should the cut-off be performed?
- Different date types
- Adverse events
- Using the DCO date beyond the DCO process
- Example code

What is the data cut-off?

- To represent the status of the data at a certain date – referred to as the DCO date
- Creates a subset of the data so that your datasets contain only data up to and including the DCO date
- DCO date will be chosen by the study team and can be picked in different ways
- e.g. a certain number of events occurring or a fixed time period
- Most commonly seen in oncology studies

Why is it needed ?

- Need the datasets to contain all observations up to and including the DCO date
- Data needs to be clean
- Data management given X number of days after DCO date to clean the data
- During this time any new pre **or post** DCO data could be added

Why is it needed ? - example

- DCO date is 1st January 2018
- Data up to and including the DCO date is considered to be fully clean on January 8th
- Any observations occurring between 2nd January 2018 and 8th January 2018 will be included in the datasets
- Therefore, the records with dates after 1st January **need to be removed from the final datasets.**

When should the cut-off be performed?

RAW level

- SDTM and ADaM based on same data – increases traceability
- Disadvantage is that you are manipulating your source data
- One Solution is to create a second set of RAW datasets
- Original raw datasets become “PRERAW”
- Perform DCO on PRERAW datasets to create new RAW datasets

When should the cut-off be performed?

SDTM level

- Possible, but more complicated than RAW level
- E.g. death date after DCO, need different approaches for the same date in DM/DS/CE datasets

USUBJID	AGE	RACE	DTHFL	DTHDTC
1	35	ASIAN	Y	2018-02-20

- If the date was removed at the RAW level this would not be an issue
- Hard to follow all the interdependencies between SDTM datasets

When should the cut-off be performed?

ADaM level

- Same complications as performing it at the SDTM level
- Issue of traceability between SDTM and ADaM in any required submissions.

In summary, RAW is the preferred level

Date Types

Assessment dates

Laboratory data example – DCO date = 1st January 2018:

Obs. Number	Subject	Visit number	Visit date	Sample Date	Sodium
1	1	1	2017-08-28	2017-08-30	120
2	1	2	2018-01-15	2018-01-15	130

- Observation 1: visit and sample date before DCO date = **Keep**
- Observation 2: visit and sample date after DCO date = **Remove**

Date Types

Start Dates

- e.g. start date of medication
- > DCO date = **Remove**

End Dates

- e.g. end date of medication
- > DCO date = **Make end date missing**
- Considered as ongoing at DCO (update any corresponding flags)

Date Types – Start and end dates

Obs. Number	Subject	Concomitant Medication	Start Date of Concomitant Medication	End Date of Concomitant Medication
1	1	Insulin	2017-11-30	2017-12-31
2	1	Paracetamol	2017-11-30	2018-02-15
3	1	Vitamin D	2018-01-24	2018-02-22

- Observation 1: Start & End before DCO = keep and no changes
- Observation 2: Start before DCO & End after DCO = keep and make end date missing
- Observation 3: Start & End after DCO = remove

Date Types

Obs. Number	Subject	Concomitant Medication	Start Date of Concomitant Medication	End Date of Concomitant Medication
1	1	Insulin	2017-11-30	2017-12-31
2	1	Paracetamol	2017-11-30	2018-02-15
3	1	Vitamin D	2018-01-24	2018-02-22

Obs. Number	Subject	Concomitant Medication	Start Date of Concomitant Medication	End Date of Concomitant Medication
1	1	Insulin	2017-11-30	2017-12-31
2	1	Paracetamol	2017-11-30	

Date Types

Baseline Dates

Patients recruited before the DCO date:

- If data contains any post DCO dates = Query this with data management.

Patients recruited after DCO date:

- If data contains any post DCO dates = Remove

Other Dates

- e.g. date of death; date of withdrawal = Remove

Adverse Events

Serious Adverse Events

Adverse event	AE Start Date	Date AE Became Serious	Serious AE	Requires or Prolongs Hospitalisation
Nausea	2017-12-15	2018-01-17	Y	Y

- AE Start date before DCO = **Keep**
- AE becomes an SAE after DCO = **Need to make the AE a non-serious AE**

Adverse Events

Outcome of AE

Adverse event	AE start date	AE end date	Outcome
Nausea	2017-12-15	2018-01-15	RECOVERED/RESOLVED

- AE Start date before DCO = **Keep**
- End date after DCO = **Make end date missing and change outcome**
- Outcome will be changed to **NOT RECOVERED/NOT RESOLVED**

Adverse Events

Action Taken

AE data:

Adverse event	AE Start Date	AE End Date	Action Taken
Headache	2017-12-15	2018-02-30	DOSE REDUCED

- End date after DCO = Make end date missing
- DOSE REDUCED is most severe action taken
- Need to check patients dose data

Adverse Events

Dose data:

Dose	Frequency	Dose Start Date	Dose End Date	Action Taken	Reason for Action Taken
80	BID	2017-11-30	2017-12-15	DOSE NOT CHANGED	
0		2017-12-16	2017-12-16	DRUG INTERRUPTED	Subject forgot to take dose
80	BID	2017-12-17	2018-01-16	DOSE NOT CHANGED	
40	BID	2018-01-17	2018-02-30	DOSE REDUCED	Adverse Event

DOSE REDUCED is after DCO so doesn't count

DRUG INTERRUPTED isn't due to adverse event so also doesn't count

Adverse Events

Pre DCO

Adverse event	AE Start Date	AE End Date	Action Taken
Headache	2017-12-15	2018-02-30	DOSE REDUCED

Post DCO

Adverse event	AE Start Date	AE End Date	Action Taken
Headache	2017-12-15		DOSE NOT CHANGED

Using the DCO date beyond the DCO process

After cut-off has been performed, will still need to use the DCO date

ADAE:

AEDECOD	ASTDT	TRTSDT	TRTEDT	TRTEMFL
Nausea	28MAY2017	30APR2017		

If $ADSL.TRTSDT \leq ASTDT \leq ADSL.TRTEEDT + x \text{ days}$ then $TRTEMFL = 'Y'$

AEDECOD	ASTDT	TRTSDT	TRTEDT	TRTEMFL
Nausea	28MAY2017	30APR2017	01JAN2018	Y

Using the date beyond the DCO process

- Will have to use the DCO date in multiple programs
- Consider having the macro variable in a separate SAS macro / autoexec

```
%macro cutoff;  
  %global dcodate;  
  %let dcodate =2018-01-01;  
%mend cutoff;
```

Example code

```
%macro del  
(ds= /*input dataset*/  
,dat= /*date variables from ds dataset to compare to DCO date*/);  
  data raw.&ds.;  
    set preraw.&ds.;  
    array orig_dat &dat.;  
    do over orig_dat;  
      if substr(orig_dat,1,10)>"&dcodate." then delete;  
    end;  
  
  run;  
  
  proc compare base=preraw.&ds. compare=raw.&ds.;  
  run;  
  
%mend del;  
  
%del(ds=labraw, dat=labdat visdat);
```

Example code

```
%macro startend
(ds=,sdat=none,edat=none);
  data raw.&ds.;
  set preraw.&ds.;
  %if &sdat. ne none %then %do;
    array a_sd &sdat.;
    do over a_sd;
      if substr(a_sd,1,10)>"&dcodate." then delete;
    end;
  %end;
  %if &edat. ne none %then %do;
    array a_ed &edat.;
    do over a_ed;
      if substr(a_ed,1,10)>"&dcodate." then a_ed="";
    end;
  %end;
run;
proc compare base=preraw.&ds. compare=raw.&ds.;
run;
%mend startend;
%startend(ds=ae,sdat=aest,edat=aeen);
```

Conclusion

Many things to consider when performing the data cut-off:

- When to do the DCO – RAW level is recommended
- Dataset specific issues – mainly in adverse events
- DCO date needed after DCO process – store DCO date in a separate macro variable

Thank You!

Lewis Meares

Principal Programmer at PHASTAR



Address Unit 2, 2A Bollo Lane, London W4 5LE

Tel +44 (0)20 7183 7062

Email lewis.meares@phastar.com **Website** www.phastar.com

