PharmaSUG 2018 – DS19

Implementation of Data Cut Off in Analysis of Clinical Trials

Mei Dey, AstraZeneca, USA
Ann Croft, ARC Statistical Services Ltd, UK

ABSTRACT

Interim analysis can result in key decisions on early stopping for futility, continuation of the trial or early declaration of trial success. Data from an interim analysis may form the basis for a regulatory submission in the case of early declaration of trial success. Data cut off (DCO) is primarily utilized in interim analysis. In oncology trials, overall survival (OS) and progression free survival (PFS) are key efficacy endpoints. Where patient survival is paramount, the correct application of DCO methodology has a major impact on trial interpretation. Not applying or incorrectly applying DCO methods can negatively affect the trial result and has the potential to turn a positive trial into a negative one. In this paper, we explore the application of DCO principles as illustrated in its specific application in oncology trials. We will look at the operational considerations and how to manage operational and programming challenges when applying DCO for an on-going trial.

INTRODUCTION

For the purpose of this paper, DCO application refers to the process of restricting data up to a specific data cut off date for analysis.

Oncology trials are distinct from other therapeutic area clinical trials in many ways, such as adverse event (AE) toxicity monitoring and endpoint assessment. In oncology trials, analyses are often event driven. Reaching a required number of events for overall survival or progression free survival analysis that are pre-specified in the protocol and statistical analysis plan triggers formal analysis.

As the event goal nears, the sponsor performs event prediction modeling and proposes a data cut-off date (DCO date) upon which all events up to and including the DCO date will be included in the analysis. Conversely, data collected following the DCO date will not be considered as part of the analysis.

DCO application for some data points can be straightforward - that is any record that has a date post-DCO should be removed. For other data points, such as adverse events and overall survival, DCO application may require more complex processing with regards to post-DCO data handling.

DATAFLOW OVERVIEW

The DCO principles are based on the guideline that individual data points after the DCO should not be included in the statistical analysis or submission package. The application of DCO could be carried out at different steps in the process of analyzing data – each has its own advantages and disadvantages. It is possible to apply DCO at these stages:

- RAW data
  - this is the data collected from the CRF or other data collection tool
- SDTM or ADAM
  - these are the CDISC standards for study data tabulation (SDTM) and analysis data (ADaM)

Once the DCO has been applied it is advisable to continue to use that data throughout the rest of the study/timepoint analysis. Therefore, if applied at the RAW dataset level then the SDTM and ADAM datasets would be created using the DCO applied RAW data. If it's decided to apply the DCO at the SDTM level then the RAW and pre-DCO SDTM datasets would also be available but the ADAM datasets would be based on the DCO-applied SDTM datasets.

There are arguments about directly applying the DCO to the SDTM datasets. Some sponsors have the capability of extracting/mapping data points directly from electronic data capture system (EDC) into SDTM via validated tools. SDTM offers standard structures that can make DCO processing simpler if the source data are SDTM. However, if source data are RAW and not SDTM data, discrepancies can exist between RAW data and SDTM data if DCO is applied to SDTM data only. The

key is to apply the DCO on the source data (whether RAW or SDTM) so traceability can be maintained.

When the source data is not SDTM, the preferred option is to apply DCO to RAW data – that is to remove the post DCO data prior to creating SDTM datasets, so that:

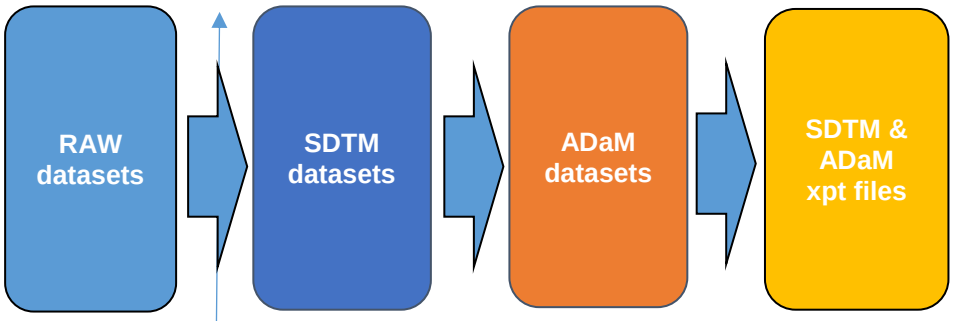
- data points collected after the DCO are not included in the Study Data Tabulation Model (SDTM) and analysis datasets (e.g. ADaM datasets) used for analysis and reporting

AND

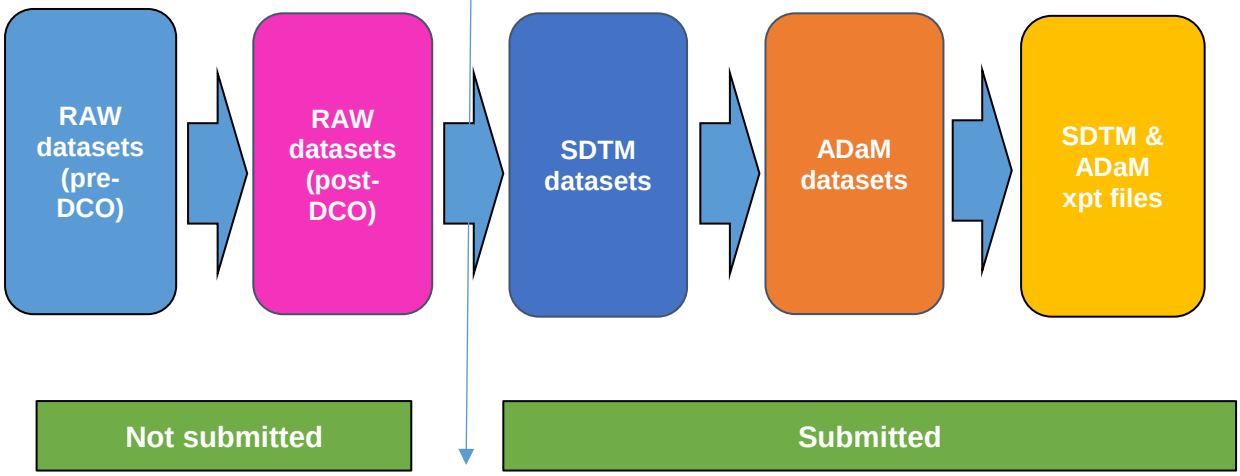
- SDTM datasets, ADaM datasets and their transport files are consistent that they are based on the same version of post-DCO RAW data.

To accomplish this, DCO rules are applied to the original pre-DCO RAW data to create post-DCO RAW datasets as input into SDTM datasets. The flow charts below demonstrate the process differences between non-DCO applied and DCO applied analysis.

Data flow without DCO processing:



Data flow with DCO processing:



It is the post-DCO RAW datasets that are used to created CDISC compliant SDTM and ADaM datasets. The SDTM and ADaM datasets are submitted to the FDA as SAS transport files (XPT). A copy of the original RAW datasets, post-DCO datasets, SDTM and ADaM datasets must be retained along with DCO programs to ensure data traceability and all submitted datasets are reproducible.

METHODOLOGY

The following sections discuss how to apply a DCO date to a set of study data. Examples are given for a DCO date being applied to RAW data. Data integrity and an ability to maintain traceability is

paramount when applying DCO processing. The methods described below are not in any specific order, as the sponsor should apply the logic to the study in a method that is suitable for their data collection methods keeping in mind data points that influence each other. This paper has covered the more likely types of pages / collected data that one would come across in a study. It is expected that similar methods could be applied to other types of collected data that are not discussed in this paper.

Applicability of all subjects

Informed consent is given by a subject when they allow their data to be used as part of a clinical study. A signed informed consent form is required by all subjects to enter into a clinical study. If a subject has not given their consent by the date of DCO then the subject and any records related to that subject should be removed from all data provided for a data cut. An informed consent date is not always the same as the screening date. A subject may provide their informed consent prior to the DCO and then attend a screening visit after. In relation to applying the DCO logic based on informed consent, only the date of consent is considered.

Missing and partial dates

DCO programming does not impute missing date variables. Date imputation should be a separate process from DCO processing.

If dates are partial, an algorithm per the sponsor's rules should be applied prior to applying DCO. Using worst case scenario principles is generally recommended. It may be that partial start dates are set to the date of first dose or set to the first of the month or year; end dates set to the end of the month or year. See PhUSE Wiki for more information. Sponsors will need to decide on date imputation rules and implementation prior to applying the DCO algorithm.

It is especially important when applying DCO to ensure that a date is not applied to a subject that is beyond their participation in the study or that is beyond their date of death. For patients where a partial date of death is collected and the partial date shows that the death occurred in the same month as the data cut-off then it could be assumed that the death occurred prior to the data cut-off.

Simple data

Simple data is referring to a RAW data module where only one date is collected and not connected to a visit. Date of termination or date of death are two examples of simple data that may be collected on a study. These are not connected to any visit and as such can have a straightforward DCO method applied for these records.

For RAW data the following logic could apply: Any record that has a date post-DCO should be removed.

```
if DEATHDT > DCODT then delete;
```

Listed items

Information may be collected on a CRF in a list format which is not specific to any visit. Concomitant Medications, Medical History, and Adverse Events are just a few examples of listed items. Adverse Events will be discussed separately in a later section. This section relates to other non-visit related, not Adverse Event collected, information. If a date is collected as part of the information then a review of that date against the DCO date is needed.

Listed items are usually collected following an indicator question; for example "Has the subject taken any concomitant medication?". The information entered following a positive response will have a record for each item. The positive response (Yes) may be applied to all of the following records or only to the first record. For each record where a date is collected the date will be validated against the DCO.

The RAW data will have a variable for this indicator question set to 1 (Yes) for at least one item or 0 (No) if the subject does not have any relevant information. If the date of the record is after the DCO then the entire record can be removed from the list, if at least one record for the subject remains. If all of the collected records have a date that is post-DCO then all but one record can be deleted. For the one record that remains or if only one record was collected then the collected information will need to be set to missing (blank) and the indicator question amended. Where there are multiple records collected and only some are post-DCO then only the post-DCO records need be deleted.

The example below demonstrates applying DCO to a list where the indicator response (for example CMANY) is applied to all observations. Data collections systems that apply the indicator response to only the first observation will need some slight amendments to the coding where the datasets are merged.

Pre-DCO subset of RAW CM dataset

SUBJECT	LINE	CMANY	CMSTDAT	DCODT
SAMPLE_111	1	0		2017-10-01
SAMPLE_112	1	1	2016-08-24	2017-10-01
SAMPLE_112	2	1	2017-10-20	2017-10-01
SAMPLE_112	3	1	2016-12-30	2017-10-01
SAMPLE_113	1	1	2017-11-01	2017-10-01

```
*** Remove observations where the line number is greater     ***;
*** than 1 that are after DCO                               ***;
*** For observations on the first line mark all that should  ***;
*** be removed (0)                                          ***;
*** All observations that are pre-DCO will remain marked (1) ***;
data pre_1_cm;
    set input.cm;
    if line>1 and cmdate > DCODT then delete;
    else if line=1 and cmdate > DCODT then do;
      cmdate='';
      cmany ='0';
      *** Amend any other variables that need adjusting ***;
    end;
run;

*** List subjects with more than 1 record available after ***;
*** deletions                                              ***;
proc sort data=pre_1_cm out=subj (keep=subject) nodupkey;
  by subject;
  where cmany='1' and line>1;
run;

proc sort data=pre_1_cm;
  by subject line;
run;

*** Keep all line=1 records if no other line exists.        ***;
*** If subject has at least one positive record that is not ***;
*** line=1 then line=1 can be deleted                       ***;
data output.cm
  merge pre_1_cm (in=a) subj (in=b);
  by subject;
  if b and line=1 and cmany='0' then delete;
  *** If end dates need to be considered then add in code ***;
  *** for adjustments here ***;
run;
```

Once the code above has been applied then it can be seen that subject SAMPLE_111 remains as before as they have no concomitant medications (conmed) to report. For subject SAMPLE_112 then line=2 had a conmed that has been removed as the start date was after the DCODT, this leaves

records where line is 1 and 3. For subject SAMPLE_113 the conmed started after the DCO date but as it was the subjects only record then no records are removed but the collected information is set to missing (only date is shown here); the record indicator CMANY is set to 0 to infer that no records were collected for that subject.

Post-DCO subset of RAW CM dataset

SUBJECT	LINE	CMANY	CMSTDAT	DCODT
SAMPLE_111	1	0		2017-10-01
SAMPLE_112	1	1	2016-08-24	2017-10-01
SAMPLE_112	3	1	2016-12-30	2017-10-01
SAMPLE_113	1	0		2017-10-01

Assessment based information dates

Visits are part of a clinical trial and can occur on a scheduled basis or unscheduled basis. The date of visit is not always the same date that an assessment is carried out or that a sample is taken from a subject for that specific visit. Laboratory samples, ECG, vital signs, etc. are examples of assessment-based information. For this collected data it is likely there will be a visit date and a sample or collection date as well per record. Per the example below the date of sample could be before or after the date of visit.

Pre-DCO subset of RAW Lab dataset

ROW	VISITDT	LBDT	DCODT
1	2017-09-01	2017-09-01	2017-10-01
2	2017-10-28	2017-10-30	2017-10-01
3	2017-09-29	2019-10-02	2017-10-01
4	2017-10-02	2017-09-30	2017-10-01

It is advisable to remove the record if at least one of the dates is post DCO. This then removes the likelihood of data accidentally existing in the dataset when the visit had not yet happened or the result was not available to the subject.

```
if visitdt > DCODT or lbdt > DCODT then delete;
```

The resulting table from the example will have 1 row. Row 2 will be removed as both dates (VISITDT and LBDT are after the DCODT. Row 3 and row 4 will be removed as only one of the dates is after the DCO; row 3 will be removed due to LBDT and row 4 will be removed due to VISITDT.

Post-DCO subset of RAW Lab dataset

ROW	VISITDT	LBDT	DCODT
1	2017-09-01	2017-09-01	2017-10-01

Adverse events

Adverse events have many connected fields that affect each other. Date of adverse event, outcome of adverse event, action taken with medication, seriousness, relatedness, etc. all have connected qualities. It is not advisable to individually check the adverse event data fields separately. The information should be joined with records for exposure, death, discontinuation, overdose, etc. to effectively apply a DCO to adverse event information.

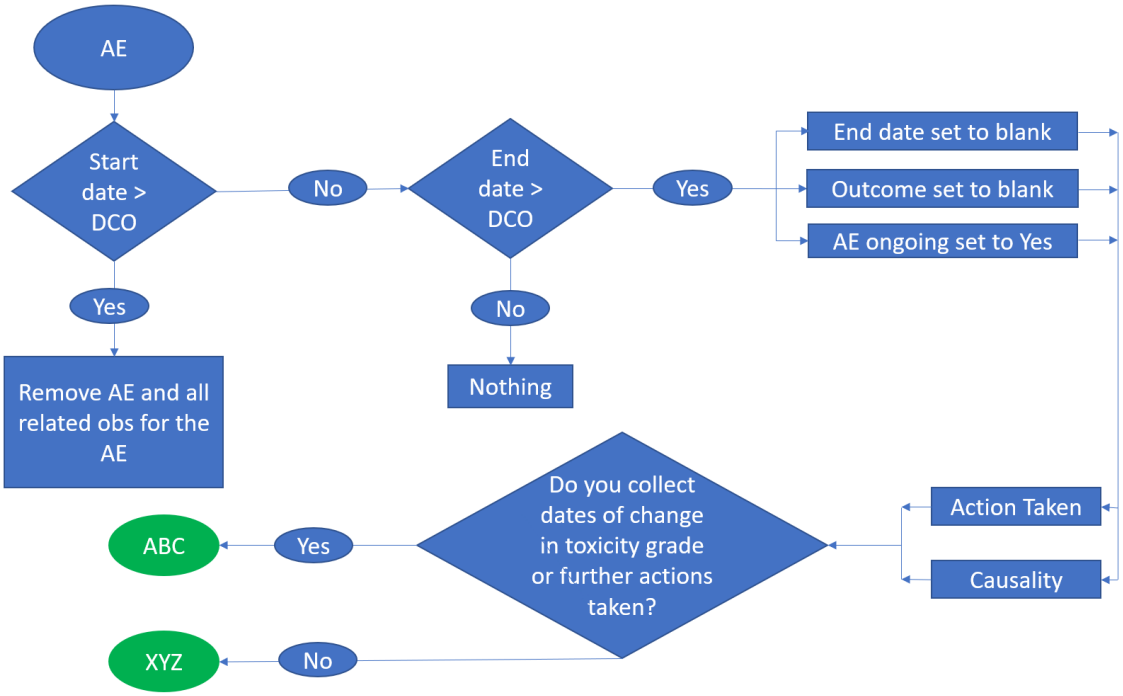
There are a number of steps to follow in order to apply a DCO to adverse events. A diagram showing the flow of steps to the method is shown below. The diagram is used as an illustrative purpose, each study should review the steps required to apply DCO independently.

The first step is to check the start date of the adverse event. If this is after the date of DCO then the record from Adverse Events and any connected records will be removed. Connected records are from CRF pages that collect additional information about an AE; the questions do not exist on the main AE CRF page as they are often optional pages based on a specific response from the main AE CRF page. Additional information could be the date of change of toxicity, findings about the AE, or actions taken as part of an SAE, etc. If the start date of the AE is prior to or starts on the date of DCO then it should remain as part of the DCO directive.

For the AEs that remain, other dates need to be checked against the date of DCO. These can include but not limited to:

- AE end date
- Hospital admission dates
- ICU admission dates
- Toxicology change dates

These additional dates have further implications for other AE CRF collected information and also for any derived fields.



Pre-DCO subset of RAW AE dataset

ROW	SUBJECT	AESTDT	AEENDT	AEOUT	ONGOING
1	SAMPLE_121	2017-10-01	2017-10-01	Recovered/Resolved	NO
2	SAMPLE_121	2017-10-02	2017-10-15	Recovered/Resolved	NO
3	SAMPLE_121	2017-10-07		Recovering	YES
4	SAMPLE_121	2017-10-15			YES

ROW	DCODT
1	2017-10-13
2	2017-10-13
3	2017-10-13
4	2017-10-13

For the dataset above row 1 will remain unchanged as the entire AE started and completed prior to the DCO; row 3 will also remain unchanged as there is nothing to suggest that any information after the DCO is being displayed. Row 4 will be removed for subject SAMPLE_121 as the start date (AESTDT) is after the DCO date. For row 2 the fields will need to be amended as the AE ended after the DCO therefore at the time of the DCO the AE was ongoing and no outcome was available. The end date of the adverse event (AEENDT) is removed and set to missing as is the outcome (AEOUT); the collection variable for an ongoing adverse event (ONGOING) is amended and set to YES.

Post-DCO subset of RAW AE dataset

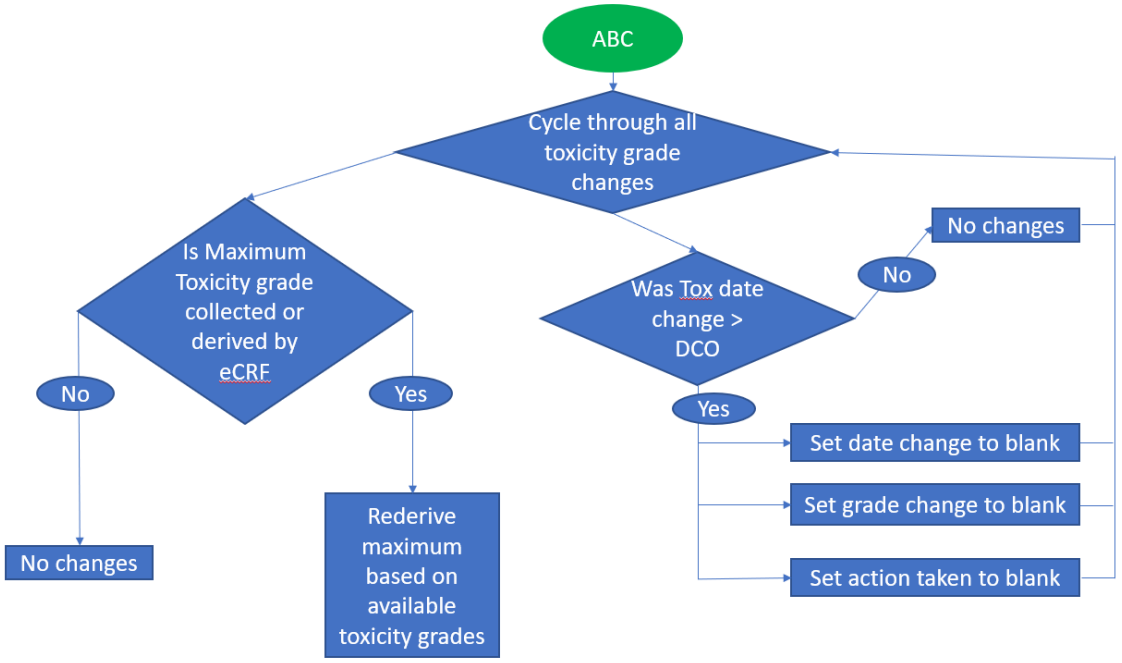
ROW	SUBJECT	AESTDT	AEENDT	AEOUT	ONGOING
1	SAMPLE_121	2017-10-01	2017-10-01	Recovered/Resolved	NO
2	SAMPLE_121	2017-10-02			YES
3	SAMPLE_121	2017-10-07		Recovering	YES

Toxicity grade

In oncology trials, National Cancer Institute (NCI) grading scale - Common Terminology Criteria for Adverse Events (CTCAE) is used to indicate the severity of the adverse events. A grade of 1 corresponds to mild, 2 to moderate, 3 to severe, 4 is a life-threatening and 5 is a death related to the adverse event. In non-oncology trials AEs use the severity scoring of mild, moderate and severe.

Toxicity grade changes in an oncology study may be collected as individual fields within a record or as list items. For list items see section on lists. For where the grade changes are collected as individual fields (within the same observation) then for each toxicity grade in turn – review the date of change to the DCO. If the date of change was prior to DCO then no change needs applied to the fields. If any of the toxicity grade change dates are post-DCO then that date should be set to blank. Ensure that any other fields collected at the same time as that specific toxicity grade change are also set to blank.

If the maximum toxicity grade is collected or derived as part of the eCRF database and passed in the RAW data then once all the toxicity grade changes have been reviewed relative to DCO the maximum toxicity grade should be derived again based on all remaining grades.



Pre-DCO subset of RAW AE dataset

ROW	SUBJECT	AESTDT	AEENDT	AEOUT	DCODT
1	SAMPLE_122	2017-09-27	2017-10-12	Recovered/Resolved	2017-10-13
2	SAMPLE_122	2017-10-01	2017-10-07	Recovered/Resolved	2017-10-13
3	SAMPLE_122	2017-10-07	2017-10-16	Not Recovered/Not Resolved	2017-10-13
4	SAMPLE_122	2017-10-15	2017-10-16	Fatal	2017-10-13

ROW	TOXGRDST	TOXGR1CH	TOXGR1DT	TOXGR2CH	TOXGR2DT
1	2				
2	2	3	2017-10-02		
3	2	3	2017-10-09	4	2017-10-14
4	4	5	2017-10-16		

ROW	TOXGR3CH	TOXGR3DT	ONGOING
1			NO
2			NO
3	5	2017-10-16	NO
4			NO

In the above example maximum toxicity grade was derived for SDTM, not collected. For the dataset, rows 1 and 2 will remain unchanged as the entire AE started and completed prior to the DCO; the toxicity grade change in row 2 happened prior to the AE end date and as a result does not need any changes. Row 4 will be removed completely as the AE did not start until after the DCO. Row 3 toxicity grade changes will need to be amended relevant to the DCO. As the AE started prior to DCO then no change is needed to the starting toxicity grade (TOXGRDST). See example code below:

```
*** Amend each toxicity grade in turn up to the ***;
*** maximum expected changes ***;
%do i=1 %to &maxgrd;
  if toxgr&i.dt>DCODT then do;
    toxgr&i.dt='';
    toxgr&i.ch='';
  end;
%end;
```

Toxicity grade change 1 (TOXGR1CH) needs no change as TOXGR1DT < DCODT, however toxicity grade changes 2 (TOXGR2CH) and 3 (TOXGR3CH) will need to be set to missing as the grade change happened after DCO.

Post-DCO subset of RAW AE dataset

ROW	SUBJECT	AESTDT	AEENDT	AEOUT	DCODT
1	SAMPLE_122	2017-09-27	2017-10-12	Recovered /Resolved	2017-10-13
2	SAMPLE_122	2017-10-01	2017-10-07	Recovered /Resolved	2017-10-13
3	SAMPLE_122	2017-10-07			2017-10-13

ROW	TOXGRDST	TOXGR1CH	TOXGR1DT	TOXGR2CH	TOXGR2DT
1	2				
2	2	3	2017-10-02		
3	2	3	2017-10-09		

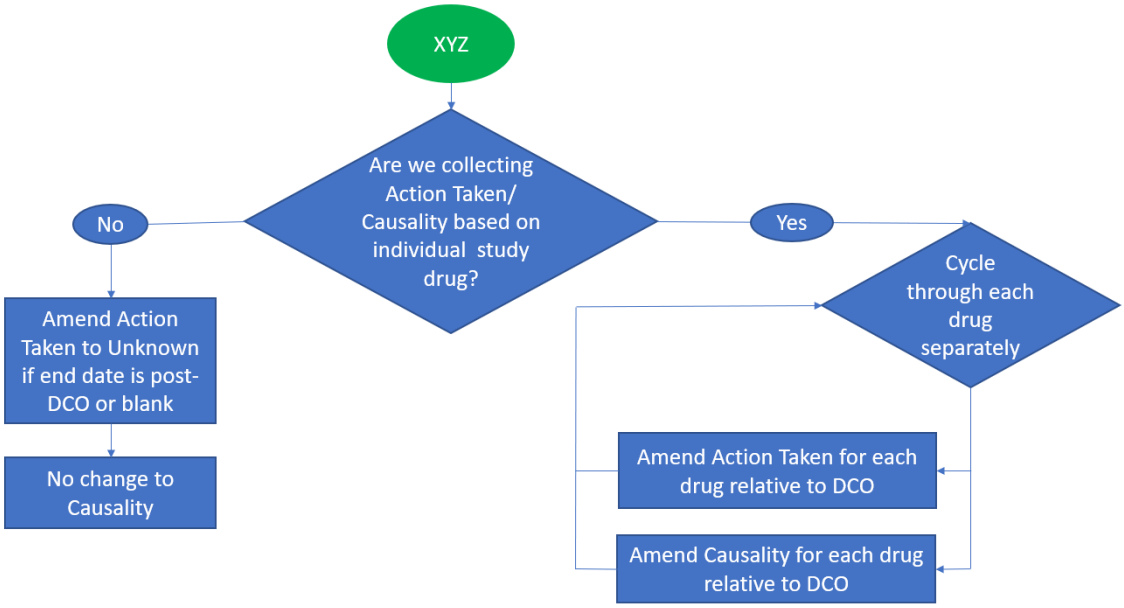
ROW	TOXGR3CH	TOXGR3DT	ONGOING
1			NO
2			NO
3			YES

Action taken/causality

Action Taken with study drug may be collected overall for the AE or individually for each drug on combination or device studies (COMBO). If collected overall then the Action Taken field should be

amended if there is a dose change / interruption / discontinuation as a result of the AE where the change happens post-DCO. Alternatively, for a COMBO study then the AE may have collected information on the Action Taken for each drug/device. It is recommended that the field is set to "Unknown" or "Not recorded" if amendment is required. Alternatively, if either of these code/terms is unacceptable within the study parameters then the Action Taken should be set to blank.

Causality is another field that may be collected for each drug/device in a COMBO study. As with Action Taken the field should be amended if the Causality result happens post-DCO. The result of an amendment to Action Taken or to Causality should be set to "Unknown" or "Not recorded" if these are an acceptable value for the study otherwise the field should be set to blank.



The example below demonstrates some of the issues is dealing with Open-Label study data when applying DCO when corresponding exposure information has been collected.

Pre-DCO subset of RAW AE dataset

ROW	SUBJECT	AESPID	AESTDT	AEENDT	AEOUT	DCODT
1	SAMPLE_123	1	2017-09-27	2017-10-10	Recovered /Resolved	2017-10-13
2	SAMPLE_123	2	2017-10-12	2017-10-20	Recovered /Resolved	2017-10-13

ROW	ONGOING	AEACNA	AERELA	AEACNB	AERELB
1	NO	No Action	Not Related	No Action	Not Related
2	NO	Dose Reduced	Probably Related	No Action	Not Related

ROW	AEACNC	AERELC
1	No Action	Not Related
2	No Action	Possibly Related

Pre-DCO subset of RAW EX data

ROW	SUBJECT	EXTRT	EXSTDT	EXENDT	EXDOSE	DCODT
1	SAMPLE_123	Dummy_A	2017-08-27	2017-10-14	200mg	2017-10-13
2	SAMPLE_123	Dummy_B	2017-08-27	2017-10-25	200mg	2017-10-13
3	SAMPLE_123	Dummy_C	2017-08-27	2017-10-25	200mg	2017-10-13
4	SAMPLE_123	Dummy_A	2017-10-14	2017-10-25	200mg	2017-10-13

ROW	EXFREQ	EXACN	ACNAE1	ACNAE2	ACNAE3	ACNAE4
1	2 tablets / BID					
2	1 tablets / BID					
3	1 tablets / QD					
4	1 tablets / BID	Dose Reduced	2			

The example above the subject SAMPLE_123 will have no change to their first AE (AESPID=1). During their 2nd AE (AESPID=2) there was action taken with the study drug Dummy_A (AEACNA) however that change was not until after the DCO. The Action Taken variable for the AE needs to be amended to "No action" for that which applies to the first treatment as that would have been the action taken with the study drug at the time of DCO. A change to causality with the first treatment (AERELA) will also need to be done. Causality to another treatment, Dummy_C, (variable AERELC) will not be amended as it had no effect on a change in the treatment. Action taken with Dummy_B (AEACNB) and causality by Dummy_B (AERELB) will also remain unchanged.

Post-DCO subset of RAW AE dataset

ROW	SUBJECT	AESPID	AESTDT	AEENDT	AEOUT	DCODT
1	SAMPLE_123	1	2017-09-27	2017-10-10	Recovered /Resolved	2017-10-13
2	SAMPLE_123	2	2017-10-12			2017-10-13

ROW	ONGOING	AEACNA	AERELA	AEACNB	AERELB
1	NO	No Action	Not Related	No Action	Not Related
2	YES	No Action	Not Related	No Action	Not Related

ROW	AEACNC	AERELC
1	No Action	Not Related
2	No Action	Possibly Related

A further example of DCO applied to subject SAMPLE_124 in the RAW AE dataset where Action Taken with treatment and Causality per treatment is taken on a combo study.

Pre-DCO subset of RAW AE dataset

ROW	SUBJECT	AESPID	AESTDT	AEENDT	AEOUT	DCODT
3	SAMPLE_124	1	2017-09-07	2017-10-27	Recovered /Resolved	2017-10-13
4	SAMPLE_124	2	2017-09-27	2017-10-30	Recovered /Resolved	2017-10-13

ROW	ONGOING	AEACNA	AERELA	AEACNB	AERELB
3	NO	Dose Reduced	Probably Related	Dose Reduced	Probably Related
4	NO	No Action	Not Related	Dose Reduced	Probably Related

ROW	AEACNC	AERELC
3	No Action	Not Related
4	No Action	Not Related

Pre-DCO subset of RAW EX data

ROW	SUBJECT	EXTRT	EXSTDT	EXENDT	EXDOSE	DCODT
5	SAMPLE_124	Dummy_A	2017-09-03	2017-10-01	200mg	2017-10-13
6	SAMPLE_124	Dummy_B	2017-09-03	2017-10-27	200mg	2017-10-13
7	SAMPLE_124	Dummy_C	2017-09-03	2017-11-13	200mg	2017-10-13
8	SAMPLE_124	Dummy_A	2017-10-02	2017-11-13	200mg	2017-10-13
9	SAMPLE_124	Dummy_B	2017-10-28	2017-11-13	200mg	2017-10-13

ROW	EXFREQ	EXACN	ACNAE1	ACNAE2	ACNAE3	ACNAE4
5	2 tablets / BID					
6	1 tablets / BID					
7	1 tablets / QD					
8	1 tablets / BID	Dose Reduced	1			
9	1 tablets / QD	Dose Reduced	1	2		

The first AE for subject SAMPLE_124 will amend the Action Taken and Causality fields related to treatment Dummy_B as the change to them happened after DCO but there will be no change to fields related to Dummy_A as that change happened prior to DCO. For the 2nd AE the Action Taken and Causality are amended for fields related to treatment Dummy_B.

Post-DCO subset of RAW AE dataset

ROW	SUBJECT	AESPID	AESTDT	AEENDT	AEOUT	DCODT
3	SAMPLE_124	1	2017-09-07			2017-10-13
4	SAMPLE_124	2	2017-09-27			2017-10-13

ROW	ONGOING	AEACNA	AERELA	AEACNB	AERELB
3	YES	Dose Reduced	Probably Related	No Action	Not Related
4	YES	No Action	Not Related	No Action	Not Related

ROW	AEACNC	AERELC
3	No Action	Not Related
4	No Action	Not Related

Overall survival

Overall survival is an efficacy endpoint in most oncology trials. Knowing the status of a subject (alive, dead, or lost to follow-up) at the time of DCO is part of the collected information. A survival sweep may be carried out more than once within a trial to determine survival for multiple endpoints. This survival sweep takes information regarding the last known survival information by the investigator. Death information may be collected from sources other than investigator notes (i.e. death registers). Merging survival with information about the date of death is required when applying DCO to survival. Amendments are generally made to roll the information back to what was known to be true at the time of DCO.

Pre-DCO subset of RAW SURVIVE dataset

SUBJECT	STATUS	SURDAT	CONDAT	DCODT
SAMPLE_131	ALIVE	2017-10-30	2017-10-30	2017-10-29
SAMPLE_132	LTFU	2017-10-10	2017-11-01	2017-10-29
SAMPLE_133	ALIVE	2017-07-07	2017-07-20	2017-10-29
SAMPLE_133	ALIVE	2017-10-28	2017-10-28	2017-10-29
SAMPLE_134	ALIVE	2017-07-07	2017-07-08	2017-10-29
SAMPLE_134	DEAD	2017-10-31	2017-11-02	2017-10-29

Pre-DCO subset of RAW DEATH dataset

SUBJECT	DTHDAT	DCODT
SAMPLE_132	2017-10-12	2017-10-29
SAMPLE_133	2017-10-30	2017-10-29
SAMPLE_134	2017-10-31	2017-10-29

From the example above, it can be seen that for subject SAMPLE_131 the survival date (SURDAT) and the date of last contact (CONDAT) are both after DCO. Therefore, at the time of DCO the subject was alive hence the date of survival and date of contact are both amended to be set to the same date as the DCO. For subject SAMPLE_132 the date of death is before the DCO but the subject was lost to follow-up prior to date of death. Therefore, the only amendment needed is to the date of contact which is set to the DCO date. Subject SAMPLE_133 has been involved in 2 survival sweeps for this study, only the latest survival sweep will be of interest, no change is needed to earlier records. As they were alive at time of contact and were not known to have died by the time of contact we do not amend the information. For subject SAMPLE_134 their latest survival sweep was carried out after the DCO when they were recorded as DEAD. According to the DEATH information their death was not until after the DCO therefore the record needs to be amended to change the subject to being alive at the time of the DCO.

Post-DCO subset of RAW SURVIVE dataset

SUBJECT	STATUS	SURDAT	CONDAT	DCODT
SAMPLE_131	ALIVE	2017-10-29	2017-10-29	2017-10-29
SAMPLE_132	LTFU	2017-10-12	2017-10-29	2017-10-29
SAMPLE_133	ALIVE	2017-07-07	2017-07-20	2017-10-29
SAMPLE_133	ALIVE	2017-10-28	2017-10-28	2017-10-29
SAMPLE_134	ALIVE	2017-07-07	2017-07-08	2017-10-29
SAMPLE_134	ALIVE	2017-10-29	2017-10-29	2017-10-29

RECIST

An efficacy measure used in evaluating certain oncology study endpoints is RECIST. RECIST (Response Evaluation Criteria in Solid Tumors) is a set of published rules that are used as a basis to evaluate tumor response in oncology solid tumor trials. We refer to data collected for RECIST evaluation (from both investigator assessment and independent assessor review) as RECIST data in this paper. There are special considerations for RECIST data for DCO processing.

The definition of RECIST per NCI is -

"A standard way to measure how well a cancer patient responds to treatment. It is based on whether tumors shrink, stay the same, or get bigger. To use RECIST, there must be at least one tumor that can be measured on x-rays, CT scans, or MRI scans. The types of response a patient can have are a complete response (CR), a partial response (PR), progressive disease (PD), and stable disease (SD). Also called Response Evaluation Criteria In Solid Tumors."

As with other clinical tests in the research industry these scans may then be reviewed by one or more Independent Assessor(s). Due to the independence in their nature, the independent reviews may be carried out months after the subject's actual scan date. The Investigator or Independent Assessor results can be amended or changed which may happen based on later scans of subject tumors. To ensure traceability and maintain integrity of the data DCO needs to be applied carefully and sensitively.

For RECIST the most important date is that of the subject scan date, not the actual review date of the scan. Any scans that happen prior to DCO will be included in the data cut. As Independent Assessor reviews, as stated above, can take place days or even months after scans these reviews may happen after the DCO. A scan that happens prior to DCO will have all reviews of that scan included in the data cut even if the review takes place after the DCO. Scans and their results which happen post-DCO will, as with other data, be removed from the data cut.

OPERATIONAL CONSIDERATIONS

Data base lock/clean file considerations

Analysis can only be as accurate as trial data collected. It is important that the study team be aware of expected DCO date for the analysis, and discuss the approach (timing, data to include, etc.) at the **earliest possible stage** to ensure all team members are on the same page. Some of the considerations are:

- What data needs to be included/cleaned (e.g. DCO patient population, critical data points)
- Specific plans around conducting the survival sweep, including the time window to perform the sweep
- Plans to chase missed RECIST scans and plans for complete central review of RECIST scans (if applicable)
- Plan for completing central review of RECIST scans (if applicable)

Once the DCO date has been established, the agreed upon DCO date should be clearly communicated to cross-functional study teams, which generally includes, but is not limited to, clinical operations and data management functions. The study teams would plan, agree, and document timing and activities leading up to the data cut-off point as well as subsequent clean file and data base lock activities. This is to ensure all data prior and up to the DCO date are entered, cleaned and verified for inclusion in data extract. A period of several weeks are commonly required for data cleaning and query resolution from the time of DCO to clean file. This is not a simple undertaking and requires careful planning and coordination of cross-functional teams. Any risks associated with the proposed DCO approach or timing should be evaluated as well.

For a live clinical database, data entry across all sites may occur at any given moment either before or after DCO date. This is especially true if a survival sweep is required. The survival sweep, where survival follow-up is conducted, usually does not complete until after the DCO date. An up-to-date query on whether a subject, usually in follow-up, is required to be carried out in a survival sweep by

either contacting the subject, family or using a death registry in order determine a subject's survival status. Since overall survival is a key oncology efficacy endpoint, any data points related to survival need to be collected and entered prior to locking the database.

If an independent assessor review of scan data is required for analysis, the study team will need to work closely with the central review vendor to resolve missing scans and ensure all scans are read completely, entered and data transferred in time for database lock.

Who should perform DCO

DCO processing should be performed by the programming team responsible for analysis and reporting. The use of programming code to cut the data at the DCO date may consist of many simple and/or complex algorithms applied to a snapshot of the data.

Data management plays a critical role in providing a data snapshot which has been checked for completion, integration and validation. New data can be entered up until the data extract (snapshot) for an on-going study. This snapshot will contain a mixture of cleaned data (data prior to the DCO date) and uncleaned data (data post the DCO date) in the database. This snapshot is sometimes referred to as RAW data as it is the raw, collected data from a sponsor's EDC system.

This snapshot is then made available to the programming team for DCO and further processing. It should be clear to all functions that the programing team for analysis and reporting is ultimately accountable for the accuracy of DCO processing according to DCO specifications.

DCO review in outsourced studies

For outsourced studies, a sponsor's in-house programming team performs risk-based reviews to ensure the quality of the vendor's deliverable. However, DCO review may sometimes be over-looked. The common expected deliverables from vendors are raw data, SDTM, ADaM along with their accompanying documentations.

Because of the potentially high impact on datasets and key outputs, the DCO application should be considered a high-risk review item and warrants a thorough review from the sponsor. Independent DCO programming from the sponsor is highly recommended to check the vendor has correctly applied the DCO directives.

The DCO review is very important as differences at source data level will carry through into analysis for both safety and efficacy endpoints. To perform the DCO check, data from both pre- and post-DCO application should be requested from the vendor. A sponsor can then apply its own DCO programs to pre-DCO RAW data in order to confirm the post-DCO RAW data from the vendor.

DCO specification development

The DCO specification is a joint effort by the statistical and programming teams. A DCO specification should be written at an individual dataset level. Rules about how to apply DCO should be specified for each RAW dataset. Study statistician and lead programmer need to examine together the rules for each RAW dataset to ensure DCO rules are being developed correctly according to DCO guidelines.

Below is a sample DCO specification which states the RAW dataset date variable used for applying the DCO method and detailed rules for the DCO application.

Module	Description	Date variable	Exclude or Reset (if date variable > DCO)	DCO Method	Rules
DOSE	Administration of Study Drug	SD_SDAT	E	SIMPLE	Any records that correspond to an administration after DCO (SD_SDAT > DCO) should be removed
ECG	Electrocardiogram	VIS_DAT	E	ASSESSMENT	Any records that correspond to a visit after DCO (VIS_DAT > DCO) should be removed
ECG	Electrocardiogram	ECG_DAT	E	ASSESSMENT	Any records that correspond to an assessment after DCO (ECG_DAT > DCO) should be removed
TERM	Completion/Withdrawal from Study	TERM_DAT	E	SIMPLE	Any records that correspond to a termination after DCO (TERM_DAT > DCO) should be removed

For an outsourced study, the vendor can develop its own DCO specification according to the DCO guidance document provided by the sponsor. The specification will then need to be reviewed and agreed by the sponsor. It is recommended the sponsor review of the vendor's DCO specification happens early in the process. Another approach to consider would be that an in-house DCO specification (if already developed) can be shared so the vendor can use it to develop code directly for DCO application.

Develop DCO programming code in-house

Once the DCO specification is agreed upon, the programming team can proceed with producing and validating the DCO program using the raw data described above. Since DCO principles are the same, there could be strong efficiency gains from standardized DCO programs. A standardized CRF setup can also be valuable when setting up standardized programs. Macros can be developed for each module or similar types of DCO method. A well-developed DCO program can be reused for subsequent DCO processing or easily adapted by another study. Not all studies are based on the same set of RAW data modules. Therefore, there should be a check at the study level that all applicable raw data have been accounted for and what DCO rules should apply to each dataset.

In-house DCO programs can be developed to compare the resulting datasets against the vendor's post-DCO raw datasets. In addition, having validated DCO programs can help support any in-house deliverable independent of vendor support. Sometimes the vendor is unwilling to take on additional requests from the sponsor due to resource or timeline constraints; to have a readily available and validated DCO program can greatly facilitate programming effort should any task become an in-house deliverable.

CONCLUSION

DCO implementation can potentially have a major impact on the trial results. It's important to understand DCO principles, develop specifications and programs/macros to correctly and accurately apply the DCO algorithms. Data traceability and integrity are major considerations in DCO applications. Defining and implementing a standard process that includes developing standard programs/methods could increase efficiency and reduce variability across studies. In addition, the operational challenges should not be overlooked. A coordinated effort across functional teams to ensure data quality is a MUST before applying DCO analysis. Although oncology DCO principles are used to illustrate DCO applications in this paper, it can be expanded to any other therapeutic areas or studies where DCO is applicable.

REFERENCES

PhUSE Wiki "Imputing Partial Dates" available at
https://www.cancer.gov/publications/dictionaries/cancer-terms/def/recist

NCI Dictionaries "NCI Dictionary of Cancer Terms" available at
https://www.cancer.gov/publications/dictionaries/cancer-terms/def/recist

Bin Yao, Yining Ye and Hua Yu "Data and Analysis Considerations in Oncology Clinical Trials" available at
http://medcraveonline.com/BBIJ/BBIJ-04-00101.php

Byrichetti, Kiranmai. "Macro to get data from specific cutoff period". 2016. Proceedings of PharmaSUG 2016 Conference.
https://www.pharmasug.org/proceedings/2016/BB/PharmaSUG-2016-BB14.pdf

ACKNOWLEDGEMENTS

The authors would like to thank George Hage, Tracy Turschman, Lynne Poole and Yanping Zhou for reviewing the paper and providing valuable feedback.

CONTACT INFORMATION

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

Mei Dey
AstraZeneca
One MedImmune Way
Gaithersburg, MD 20878 USA
E-mail: Mei.Dey@astrazeneca.com

Ann Croft
ARC Statistical Services Limited
Berkshire, UK
E-mail: annRcroft@gmail.com