PharmaSUG 2024 - Paper 10092

The Imperative Role of Data Cut-Off (DCO) in Clinical Trials: Ensuring Integrity and Precision in Oncology

Nana Yang, Chunpeng Zhao, BeiGene

ABSTRACT

In the realm of clinical trials, particularly in oncology, the process of Data Cut-Off (DCO) is a pivotal juncture that significantly influences the trial's outcome and interpretation. DCO involves restricting data analysis to a predefined date, ensuring that the dataset is static complete, consistent, and ready for submission to regulatory authorities and decision-making. The DCO ensures that the data reflects the patient's status at a specific point in time, which is crucial for the validity and reliability of the trial results, especially the event-driven end points such as overall survival and progression-free survival. The paper describes the general cut off rules based on subject level, visit level and record level, and the difference consideration aspects from survival sweep studies and non-survival sweep studies on three levels. The paper also delineates the procedural framework for implementing DCO process on minus datasets, including all the aCRF variables and MedDRA coding variables and Rave EDC system default variables, which is based on SAS macro.

INTRODUCTION

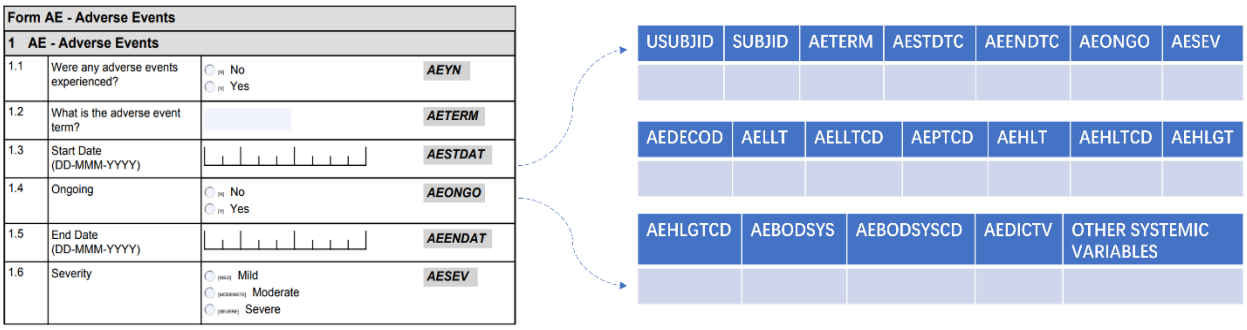
The DCO is the point for analysis, creating a static dataset that is complete and consistent up to a predefined date. The DCO process involves meticulous planning and execution, with rules that apply to the subject, visit, and record levels. These rules ensure that only data relevant to the predefined DCO date is included in the final analysis. The paper presents a procedural framework for implementing the DCO process using minus datasets, which are a streamlined version of the final SDTM datasets.

The paper also addresses the complexities of applying the DCO in different scenarios, such as taking studies with survival sweeps for a special example, highlighting the tailored approach required for various trial designs. Survival sweep study refers to the process of regularly checking or assessing the survival status of patients during a clinical trial. This concept is closely related to the monitoring of Overall Survival, which is a primary endpoint in cancers' clinical trials[1]. By detailing the procedural steps and considerations for implementing DCO, the paper aims to guide researchers in conducting clinical trials with scientific rigor, ultimately contributing to the advancement of medical science and patient care.

WHAT ARE THE MINUS DATASETS?

Minus datasets are the mini-SDTM version compared with final SDTM datasets, which include all the variables from aCRF and the MedDRA coding variables and Rave EDC system default variables. Date variables in minus datasets are in the standard format of yymmdd10. However, other derived SDTM variables do not belong to minus datasets, including EPOCH, LBTOX, VISIT, VISITNUM, XXDY, XXBLFL, XXSEQ, and so on. Common derived variables are all created by different macros outside the minus datasets process.

The Figure 1 is an example of Form AE - Adverse Events from the CDISC website[2]. EDC raw data variables are transformed to SDTM variables, for example AETERM, AESTDTC, AEENDTC, AESEV. Then base on EDC variables to create standard format MedDRA coding variables (AEDECOD, AELLT, AEHLT, AEHLGT, AEBODSYS, AEDICTV) and corresponding class code variables. EDC system default variables are also kept deriving other SDTM variables. After applying cut-off process for minus datasets and then creating derived variables, final SDTM datasets are completed.



Example of AE minus dataset based on Form AE - Adverse Events on the CDISC website.

WHAT IS THE SUJECT LEVEL INCLUSION AND EXCLUSION CRITERIA?

If there are multiple inform consent forms (ICFs) or different kinds of ICF contents and corresponding date on the same page, including optional biopsy, blood collection, biomarker and biological samples and so on, the first informed consent is identified by the earliest informed consent date. If a patient signs the first informed consent after cut-off date, the patient should be excluded in SDTM.

A patient is identified by SUBJID rather than USUBJID. The variable SUBJID is used to uniquely identify each subject involved in a study. In case a single subject is screened and/or enrolled more than once in a study, the subject's SUBJID should be distinct for each separate screening or enrollment[3].

WHAT IS THE VISIT LEVEL INCLUSION AND EXCLUSION CRITERIA?

If a visit of a patient starts after cut-off date, this visit of the patient should be excluded in SDTM. A visit is identified by folder name in Rave EDC system. Visit is assigned as the values of folder name in upper case from study EDC Architect Loader Specification (ALS) document, a visit for survival follow-up visits is assigned as combination of folder name with record position (Rave EDC system variable) for distinction. In Rave EDC system, all the control relationships are one-to-one, within record, form name, instance, subject, and site. It means that record is the data point, form name is the form where the record is located, and instance is the visit information where the form is located. For example, folder name with value such as 'Screening', 'Cycle 1 Day 1' represents data point visit information. But for the unscheduled visit, folder name equals 'Unscheduled Visit', and instance name equals to 'Unscheduled Visit 07 MAY 2024', so use folder name instead of instance name to identify a visit of a patient[4].

Visit start date is identified by the earliest date out of date of visit variable (SVSTDTC) in Subject Visits (SV) minus domain, date of sample collection (--STDTC) in all corresponding minus SDTM datasets, or assessment date (--DTC, -- is domain name) within a Folder name.

WHAT IS THE RECORD LEVEL INCLUSION AND EXCLUSION CRITERIA?

A record in record level is identified by Rave EDC Record ID. If an event, an intervention, or an assessment activity starts after cut-off date, such records should be excluded. Start date of an event, an intervention, or an assessment is identified by the earliest date out of event start dates (--STDTC, -- is event domain name), intervention start dates (--STDTC, -- is intervention domain name), or assessment dates (--DTC, -- is domain name). If there are --STDTC and --DTC in the same minus domain, the earlier one out of two will be as the earliest date.

For without Survival Sweep, if an Adverse Event starts on or before cut-off date, but ends after cut-off date, not set the end date of adverse event, its toxicity grade, related outcome and result in death to null, and the end reference time point is set to cut-off date, and its status of reference point referred to the end reference time point don't be set to ongoing [5].

SPECIAL CASE FOR STUDIES WITH SURVIVAL SWEEP

The Figure 2 is an example of AE and death information in study with survival sweep, how to deal with it?

SUBJID	Cut-off Date	DTHDTC	AESTDTC	AETOXGR	RFICDTC	AEOUT
001-001	2023-07-15	2023-07-16	2023-07-01	1	2023-05-01	FATAL
001-001	2023-07-15	2023-07-16	2023-07-16	5	2023-05-01	FATAL

Figure 2 Example of AE and death information in study with survival sweep.

If a patient signs the 1st informed consent before cut-off date, the patient's data collected in EDC data CRF Form Survival Follow-Up, Form Death Details and Form Adverse Events with fatal outcome where the adverse events happened on or before cut-off date should be kept. That means all records in Subject Status (SS) domain and Death Details (DD) domain will not be dropped for study in order to keep consistency between SS domain, Demographics (DM) domain (death date and death flag in Form Death Details or Form Adverse Events), DD domain (special death details information in Form Death Details), AE domain (death flag in Form Adverse Events with fatal outcome), and Disposition (DS) domain (DSDECOD equals 'Death' in Form End of Study).

1. If not, it can cause the related Pinnacle 21 Issues as following[6].

   Case 1: DTHFL must equal 'Y' when there is a record in the Subject Status (SS) domain with SSSTRESC = 'DEAD'.

   Case 2: DTHFL must equal 'Y' when there is a record in the Death Details (DD) domain.

   Case 3: DTHFL must equal 'Y' when there is a record in the Adverse Events (AE) domain with AEOUT = 'FATAL'.

   Case 4: DTHFL must equal 'Y' when there is a record in the Adverse Events (AE) domain with AESDTH = 'Y'.

   Case 5: DTHFL must equal 'Y' when there is a record in the Disposition (DS) domain with DSDECOD = 'DEATH'.

   Case 6: Fatal Adverse Event records (AEOUT = 'FATAL') in the Adverse Events (AE) domain should have End Date/Time of Adverse Event (AEENDTC) equal to subject' s Date/Time of Death (DTHDTC) populated in the Demographics (DM) domain. Subject' s Date/Time of Death info should be consistent across all domains.

2. If yes, it can save related communication cost due to the consistency and transparency from several aspects:

   Case 1: If a patient died in Survival follow-up, whether the death happened before or after cut-off date can show in death details form.

   Case 2: Adverse Event Form can also show the cause of death is Adverse Event indeed rather than PD due to Disease under study. This information may not be found in SDTM data, if fatal AE started after cut-off date.

   Case 3: If there is a fatal AE, related death details can also be available in cut data.

HOW TO APPLY CUT-OFF BASED ON MINUS DATASETS?

Based on all minus datasets, cut-off process will be divided into two steps:

Step 1:

The first step is to collect all related dates by Form with flagging how to use each date in data cut in Data Cut Specification. Besides YN, the value 'SW' will be used instead of 'Y', if the date usages/handling are different between studies with/without survival sweep. For example, AEENDTC of Fatal AE. Then review and update three columns (subject level cut flag, visit level cut flag, record level cut flag) in the Data Cut Specification. Subject level cut flag means whether the date or datetime variable will contribute to the cut-off of subject level cut. For example, DSSTDTC in the inform content form is contributed to the subject

level cut, so DSSTDTC is marked as ‘Y’. Visit level cut flag means whether variable will be used to the cut-off of visit level. For example, Form Medical History is not related to the visit level cut, so MHSTDTC is marked as ‘N’. Record level cut flag means whether date will play a role in the cut-off of record level. For example, all --ENDTC variables will be not involved in record level cut, so they are marked as ‘N’. The Figure 3 is an example of Data Cut Specification[7].

UForm	SourceDataset	FieldName	FieldLabel	FieldSEQ	Page	Domain	DomainLabel	Dataset	VariableName	VariableLabel	ForSubjectCutFlag	ForVisitCutFlag	ForRecordCutFlag
Date of Visit	SV	SVSTDAT	Date of visit	2	3	SV	Subject Visits	SV	SVSTDTC	Start Date/Time of Visit	N	Y	Y
Informed Consent	DS_IC	DSSTDAT	Date informed consent signed	1	24	DC	Demographics as Collected	DC	RFICDTC	Date/Time of Informed Consent	N	N	N
Informed Consent	DS_IC	DSSTDAT	Date informed consent signed	1	24	DS	Disposition	DS	DSSTDTC	Start Date/Time of Disposition Event	Y	N	Y
Informed Consent	DS_IC	ICOBBSDAT	Date of consent to storage and future research with biomarker and biological samples?	4	24	DS	Disposition	DS	DSSTDTC	Start Date/Time of Disposition Event	Y	N	Y
Demographics	DM	BRTHYY	Year of birth	1	26	DC	Demographics as Collected	DC	BRTHDTC	Date/Time of Birth	N	N	N
Enrollment	DS_ENR	RAND_DSSTDAT	Date of randomization	2	30	DS	Disposition	DS	DSSTDTC	Start Date/Time of Disposition Event	N	N	Y
Medical History	MH	MHSTDAT	Date started	2	36	MH	Medical History	MH	MHSTDTC	Start Date/Time of Medical History Event	N	N	Y
Medical History	MH	MHENDAT	Date stopped	4	36	MH	Medical History	MH	MHENDTC	End Date/Time of Medical History Event	N	N	N
Prior Systemic Therapy	CM_PST	CMSTDAT	Start date	4	39	CM	Concomitant/Prior Medications	CM	CMSTDTC	Start Date/Time of Medication	N	N	Y
Prior Systemic Therapy	CM_PST	CMENDAT	Stop date	5	39	CM	Concomitant/Prior Medications	CM	CMENDTC	End Date/Time of Medication	N	N	N
Prior Systemic Therapy	CM_PST	PDDAT	Date of disease progression	9	40	RS	Disease Response and Clin Classification	RS	RSDTC	Date/Time of Assessment	N	Y	Y

Figure 3 Example of Data Cut Specification

Step 2:

Run macro to create all final SDTM minus datasets. There is a parameter to control cut whether or not in macro. ‘Y’ means that will call the nested cut-off macro. For study with survival sweep, cut-off macro will convert the visit level cut flag and record level cut flag of SSDTC and DTHDTC to ‘N’, otherwise, the default study is without survival sweep. The cut-off macro will follow the rules mentioned above and carry out cut-off process from subject level, visit level and record level step by step respectively as shown in Figure 4, then merge three levels to form final SDTM datasets.

Every final post-cut SDTM minus domain includes combined removed flag of three levels (visit level remove flag = Y or subject level remove flag = Y or record level remove flag = Y), and describe removed reason, and XXDTC of three levels, cut-off date and other all minus SDTM variables as shown in Figure 5. Then in every domain code, filter records by adding combined removed flag = Y before deriving other SDTM variables based on minus variables to access final SDTM datasets.

The Figure 6 is the removed reason from subject, visit and record level and scope domains of application. For example, subject level applies to DC DM SS domains, if a patient signs the 1st informed consent after cut-off date, this patient will be excluded from cut-off data and give removed reason = ‘1st Informed Consent was signed off after Cutoff Date’ in DC DM SS domains for such patient.

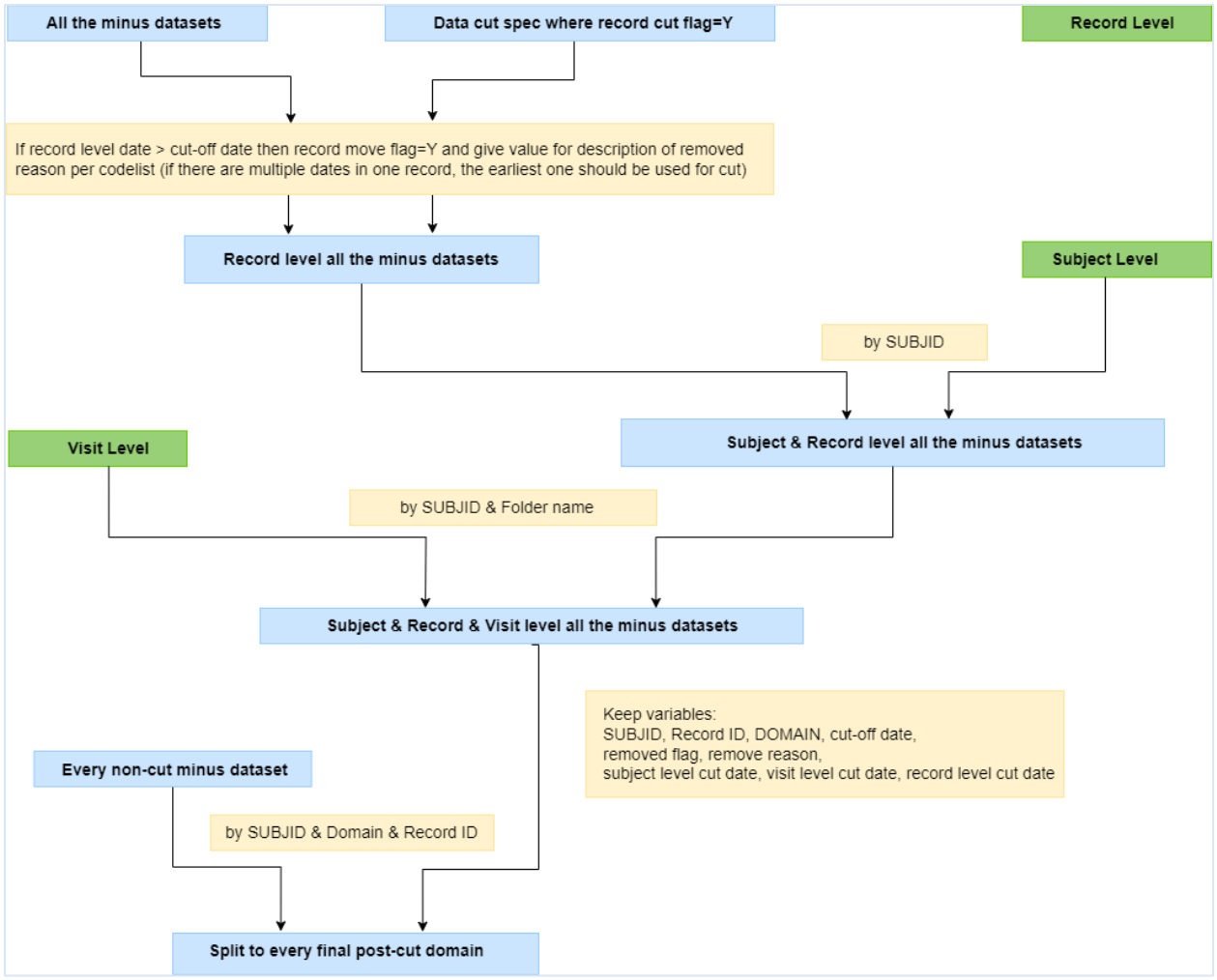


Figure 4 Flow chart of application of cut-off process.

Subjid	Remove_FL	Remove_Description	Subject_Level_DTC	Visit_Level_DTC	Record_Level_DTC	CUTOFFDATE	Other all minus variables
001-001	Subject_Remove_FL = Y or Visit_Remove_FL = Y or Record_Remove_FL = Y	Remove_Description_Subject_level---Remove_Description_Visit_level---Remove_Description_Record_level	2023-09-01	2023-09-30	2023-09-01	2023-10-01	.....

Figure 5 Example of Final post-cut minus SDTM domain columns.

Level	Remove_Description	Domain
Record	Event happened after Cutoff Date	AE BE CE DS HO MH
Record	Intervention happened after Cutoff Date	AG CM PR SU ML EC
Record	Assessment happened after Cutoff Date	SV CV EG LB MB MI QS RS TU TR VS BS CP GF IS MO MK NV OE RP UR PE FT SC FA SR
Subject	1st Informed Consent was signed off after Cutoff Date	DC DM SS
Visit	Visit happened after Cutoff Date	Visit level

Figure 6 Cut-off reason from subject, visit and record level.

CONCLUSION

The meticulous implementation of DCO process stands as a testament to the scientific rigor required in clinical trials, particularly within the demanding field of oncology. The comprehensive approach to DCO, as elucidated, not only guarantees the integrity of clinical trial outcomes but also upholds the trust of patients, healthcare providers, and regulatory authorities. The paper has navigated through the complexities of DCO, from the conceptualization of minus datasets to the intricate application of SAS macros for data aggregation and analysis. It has highlighted the critical importance of subject-level, visit-level, and record-level considerations, which collectively form the bulwark of data integrity. The distinction between survival sweep and non-survival sweep studies further emphasizes the tailored approach necessary for different trial designs.

In conclusion, the DCO is a linchpin in the clinical trial machinery, pivotal in determining the success and credibility of a study. By adhering to the principles and procedures outlined in this paper, researchers can ensure that their trials are conducted with the utmost precision and integrity. The adoption of standardized DCO processes, such as the SAS macro presented, can significantly enhance efficiency and consistency across trials, ultimately contributing to the advancement of medical science and the improvement of patient care.

REFERENCES

[1] Royle KL, Meads D, Visser-Rogers JK, White IR, Cairns DA. How is overall survival assessed in randomised clinical trials in cancer and are subsequent treatment lines considered? A systematic review. Trials. 2023 Nov 6;24(1):708. doi: 10.1186/s13063-023-07730-1. PMID: 37926806; PMCID: PMC10626781.

[2] eCRF Portal. https://www.cdisc.org/kb/ecrf/ecg-test-results-local-reading.

[3] Study Data Technical Conformance Guide - Technical Specifications Document. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/study-data-technical-conformance-guide-technical-specifications-document.

[4] Medidata Rave system custom function examples and template library management. http://www.gcp-clinplus.com/index/detail/id/765/c_id/194.html.

[5] Shang Shi. “Implementation of Data Cut-Off in Analysis of Oncology Clinical Trials”. 2021. Proceedings of PharmaSUG 2021 Conference. https://www.lexjansen.com/pharmasug-cn/2021/DM/Pharmasug-China-2021-DM023.pdf

[6] https://www.pinnacle21.com/validation-rules/sdtm.

[7] Amelia Pang. “Implementation of Cut-Off Date in Analysis of Solid tumor Clinical Trials”. 2023. Proceedings of PharmaSUG 2023 Conference. https://www.lexjansen.com/pharmasug-cn/2023/CC/Pharmasug-China-2023-CC137.pdf

CONTACT INFORMATION

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

Nana Yang
BeiGene
Nan.yang@beigene.com

Chunpeng Zhao
BeiGene
Chunpeng.zhao@beigene.com