PharmaSUG 2023 - Paper DS-41

Leverage and Enhance CDISC TAUGs to Build More Traceability for and Streamline Development of Efficacy ADaM in Oncology Studies

Xiangchen (Bob) Cui, CRISPR Therapeutics AG, Boston, MA

ABSTRACT

CDISC Breast Cancer Therapeutic Area User Guide (TAUG-BrCa) [1] and Prostate Cancer Therapeutic Area User Guide (TAUG-PrCa) [2] presented '**ADEVENT**' and '**ADDATES**', independently in 2016, and 2017. One of primary reasons for the creation of the intermediate datasets is to support traceability by building into event dataset and/or date dataset through the triplet of SRCDOM, SRCVAR, SRCSEQ variables, and all potential dates from them are used to generate ADTTE (Data for the Time to Event Analyses). ADEVENT can also support another analysis dataset 'ADRESP' for best overall response, etc.

FDA's guideline [3] provided examples for primary and supportive analysis of progression-free survival (PFS). The derivation of dates from tumor assessments is not straight forward and much more complex, especially when Response Evaluation Criteria in Solid Tumors (RECIST 1.1) [4] is applied in the derivation, where the confirmation of a complete response (CR) and partial response (PR) is required. Hence the traceability of these derivations is also very critical to build the confidence of the analysis. The triplet from **ADEVENT** is not sufficient for the traceability of the events derived from tumor assessments due to the complexity of the derivation. The triplet from **ADDATES** only provides the traceability of the derivation of the dates of independent of tumor assessments.

This paper explains the pros and cons of them and introduces a new approach to enhance them so that they can be broadly used to other areas of oncology studies to build more traceability, further streamline the development of efficacy ADaM datasets: **ADEVENT**, **ADRESP**, **ADDATES**, and **ADTTE** for both categorical analysis of tumor response and a TTE analyses and follow the best programming practice.

INTRODUCTION

CDISC Breast Cancer Therapeutic Area User Guide (TAUG-BrCa) [1] presented a 'BDS' dataset '**ADEVENT**' and CDISC Prostate Cancer Therapeutic Area User Guide (TAUG-PrCa) [2] presented ADaM class of 'Other' to create an event dates analysis dataset '**ADDATES**' independently, in 2016, and 2017, respectively. Both intermediate datasets support traceability by building into event dataset through the triplet of SRCDOM, SRCVAR, SRCSEQ variables. '**ADEVENT**' can also support the derivation of **ADRESP** (Analysis of Best Overall Tumor Response) for categorical analysis of tumor response, in addition to **ADTTE** for Time-to-Event (TTE) analyses, while '**ADDATES**' can only support ADTTE. TAUG-BrCa recommends that all possible events should be included in '**ADEVENT**', and TAUG-PrCa recommends that the best practice is to include all dates used for both efficacy events and censoring events. However, the triplet from **ADEVENT** is not sufficient for the traceability of the events derived from tumor assessments, while one from ADDATES does not provide any traceability of the derivation.

We will explain these two standards only about their metadata through their examples, pinpoint their pros and cons. These cons limit their wide use of other areas of oncology studies. We will present the corresponding enhancement and solution to their cons in this paper. The flowchart in Figure 1 below depicts the overall logic and data flow of the new approach.

FDA's guideline [3] provided examples of censoring scheme for primary and supportive PFS analysis. Only TAUG-BrCa [1], TAUG-PrCa [2], and FDA guideline [3] are used in this paper. The metadata, derivations, and examples presented in this paper are hypothetical and for illustrative purpose only, and they are not meant to imply a universally accepted definitions or derivation of the variables. They depend on specific study statistical analysis plan (SAP) of each study. We will use RECIST 1.1 in the examples for the illustration throughout this paper, which is the most complex and challenging situation for efficacy ADaM programming in Oncology Studies.

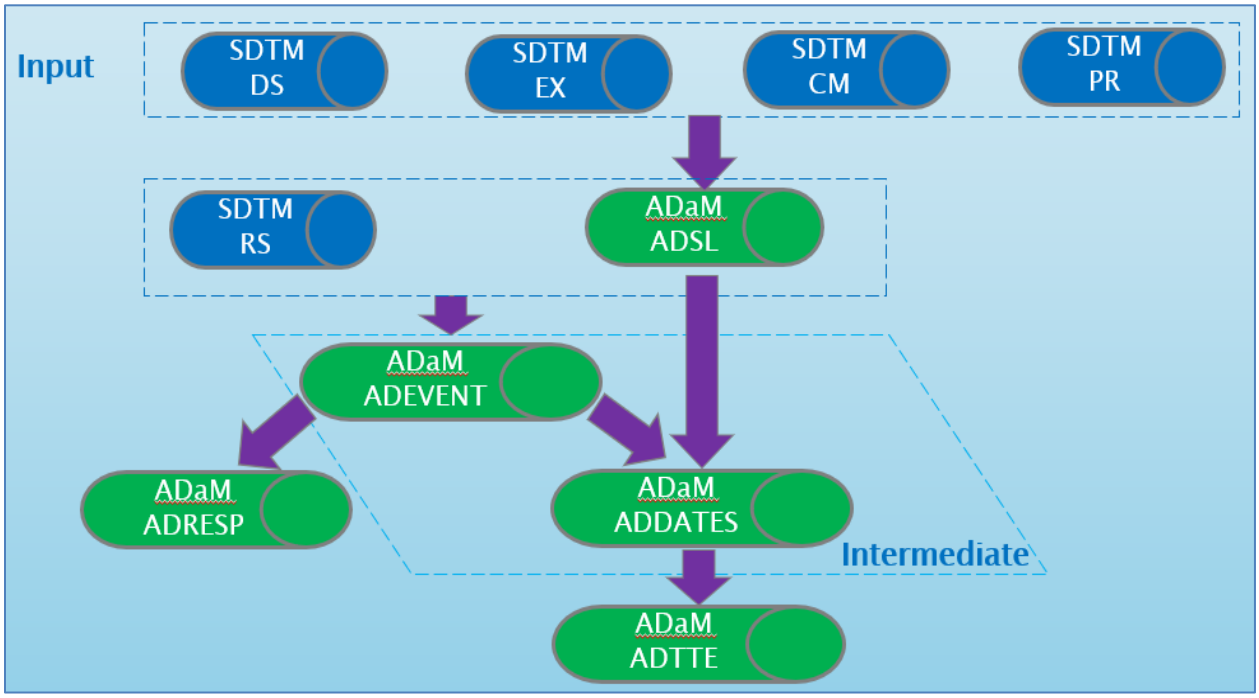


Figure 1. Leveraging CDSIC Standards from ADEVNET and ADDATES for ADRERSP and ADTTE

The new approach separates the derivation of ***dates related to tumor assessments*** and ***dates of independent of tumor assessments*** into two independent programming: ADEVENT.sas and ADDATES.sas. The analysis flags (ANLxxFL) and CRIT1/CRIT1FL (when RECIST1.1 [4] and/or iRECIST [5] are used) are derived and added to the original records of tumor assessments to build more traceability into ADEVENT. ADRESP will be "easily" derived from ADEVENT with the aid of "new" structure and newly created traceability variables. The programming for ADDATES.sas starts by reading ADEVENT data and selecting the records with the events related to tumor assessments for ADTTE. The next step is to add one block SAS® codes and specification for the derivation of each date of independent of tumor assessments, which is very similar to ADSL programming for demographic variables. Firstly, ADTTE.sas simply converts ADDATES from the vertical structure into the horizontal one by SAS transpose procedure, and then follows the study Statistical Analysis Plan (SAP) and CDISC ADaM guideline to build ADTTE, along with traceability variables.

Hence the new approach streamlines the programming for the generation of ADEVENT, ADRESP, ADDATES, and ADTTE. The enhancement of these two CDSIC standards make them to become a more powerful tool to build more traceability and streamline the development of efficacy ADaM datasets in other areas of oncology studies, besides Breast Cancer Therapeutic Area and Prostate Cancer Therapeutic Area. Accordingly, the best programming practice is followed.

Of note, the date of first new anti-cancer therapy from ADCM, and the date of first new stem cell transplant, cancer related surgery, and/or radiotherapy from ADPR could be built into ADSL, which is our ADSL design. For the simplicity and ease of explanation, ADCM and ADPR were dropped from Figure 1. Of course, ADCM and ADPR would be easily built to further support the traceability of ADSL.NEWCTDT (First Date of New Anticancer Therapy). In appendix, the metadata of these two TUGs, and ADRESP are provided for ease of reference.

IMPORTANCE OF TRACEABILITY

The Analysis Data Model Implementation Guide (ADaMIG) v1.2 [6] defines the traceability as:
"Traceability – The property that enables the understanding of the data's lineage and/or the relationship between an element and its predecessor(s). Traceability facilitates transparency, which is an essential component in building confidence in a result or conclusion. Ultimately, traceability permits the understanding of the relationship between the analysis results, the ADaM datasets, the SDTM datasets, and the data collection instrument. Traceability is built by clearly establishing the path between an element and its immediate predecessor. The full path is traced by going from one element to its predecessors, then on to their predecessors, and so on, back to the SDTM datasets, and ultimately to the data collection instrument."

The FDA Study Data Technical Conformance Guide (SDTCG) [7] states *"An important component of a regulatory review is an understanding of the provenance of the data (e.g., traceability of the sponsor's results back to the CRF data). Traceability permits an understanding of the relationships between the analysis results (tables, listings and figures in the study report), analysis datasets, tabulation datasets, and source data."*

RULES TO BE CONSIDERED FOR BOR DERIVATION PER RECIST 1.1

The best overall response (BOR) is one of endpoints in oncology studies per FDA guideline [3]. Paper [8] presents a new approach to simplify the derivation of BOR. For more details, please refer to Appendix 1 and/or the corresponding section in [8]. We will further illustrate how to build more traceability for BOR and explain why our method to build traceability has the advantage to the triplet of SRCDOM, SRCVAR, and SRCSEQ proposed by CDISC TAUG-BrCa.

FDA GUIDELINE FOR PROGRESSION-FREE SURVIVAL (PFS) ANALYSIS

Progression-free survival (PFS) is commonly used as a primary/co-primary endpoint in Phase III of oncology studies, or secondary endpoint in other phases of oncology studies. There are two key elements to calculate PFS duration: how to define the progression (event) date and censoring date, and how to define event versus censoring. FDA guidelines [3] provides the examples of prespecified censoring schemes for primary and secondary PFS analysis. To illustrate the overall logical flow and its setup of the programming approach in this paper, we choose primary PFS analysis from Table C1 [3] only as an example, shown Table 1 below.

Situation	Date of Progression or Censoring	Outcome
Incomplete or no baseline tumor assessments	Randomization	Censored
Progression documented between scheduled visits	Earliest of: Date of progression assessment showing new lesion (if progression is based on new lesion); or Date of last progression assessment	Progressed
No progression	Date of last progression assessment with no documented progression	Censored
Treatment discontinuation for undocumented progression	Date of last progression assessment with no documented progression	Censored
Treatment discontinuation for toxicity or other reason	Date of last progression assessment with no documented progression	Censored
New anticancer treatment started	Date of last progression assessment with documented nonprogression before start of new treatment	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed
Death or progression after more than one missed visit	Date of last progression assessment with documented nonprogression	Censored

Table 1. An Example for Censoring Scheme for Primary PFS Analysis from Table C1 [3]

DATES OF EVENTS ARE NEEDED FOR TIME-TO-EVENT ANALYSE AND BOR/DOR

ADTTE supports Time-to Event analyse, including PFS analysis and overall survival analysis as examples. Of note, the rules to be followed in ADTTE derivation should be clearly specified by the study SAP, which should consider FDA Guideline [3]. Hence the following dates should be ready to be used before the start of the programming per the last section.

1. The first date of progressive disease
2. Date of last progression assessment with no documented progression
3. Date of last progression assessment with documented nonprogression before start of new treatment
4. Date of last progression assessment with documented nonprogression before the missed visits
5. Date of the first anticancer therapy
6. Date of death if the patient died
7. Date of Randomization or the first treatment date
8. Date Last Known Alive
9. Date of Last Exposure to Treatment
10. Date of Analysis Cut-off
11. Date Lost to Follow-up
12. End of Study Date

The following three dates are also in the interest of efficacy analysis when time to response (TTR), when duration of response (DOR) are endpoints.

13. Date of First Occurrence of BOR with CR or PR
14. Date of DOR Start
15. Date of DOR End

The dates listed above can be categorized as either dates related to tumor assessments (Item 1 - Item 4 and Item 13 - Item 15 from SDTM.RS domain) or dates independent of tumor assessments (Item 5 - Item 12) in green, which would be built in ADSL for simplicity.

RATIONALE OF SEPARATING THE DERIVATION OF DATES RELATED TO TUMOR ASSESSMENTS AND DATES INDEPENDENT OF TUMOR ASSESSMENTS

The derivation for dates related to tumor assessments (Item 1 - Item 4 and Item 13 - Item 15 from SDTM.RS domain) is much more complex than ones independent of tumor assessments (Item 5 - Item 12) regarding the SAS programming (SAS program and metadata), which is very similar to ADSL programming for demographic variables, where each variable has its own block of SAS codes and derivation rule in its metadata/specification. There is less interdependence among them. However, the programming for the derivation of dates from tumor assessments depends on the ***hierarchy orders*** described above.

Hence, if these two different programming tasks can be segregated, the developing SAS codes and writing specification would be simpler and easier, which enhances efficiency and high quality. It also makes QC process, the maintenance and updates of SAS programming, and the specification writing much easier, regarding the understanding of the programming logic and the possible assignment to different programmers to do the validation as modules. Accordingly, the best programming practice is fulfilled.

INTRODUCTION OF ADEVENT FROM CDISC BREAST CANCER THERAPEUTIC AREA USER GUIDE

TAUG-BrCa [1] introduced an intermediate data set ADEVENT for Breast Cancer Therapeutic in 2016. Figure 2 below describes data flow from using an intermediate dataset ADEVENT [9].

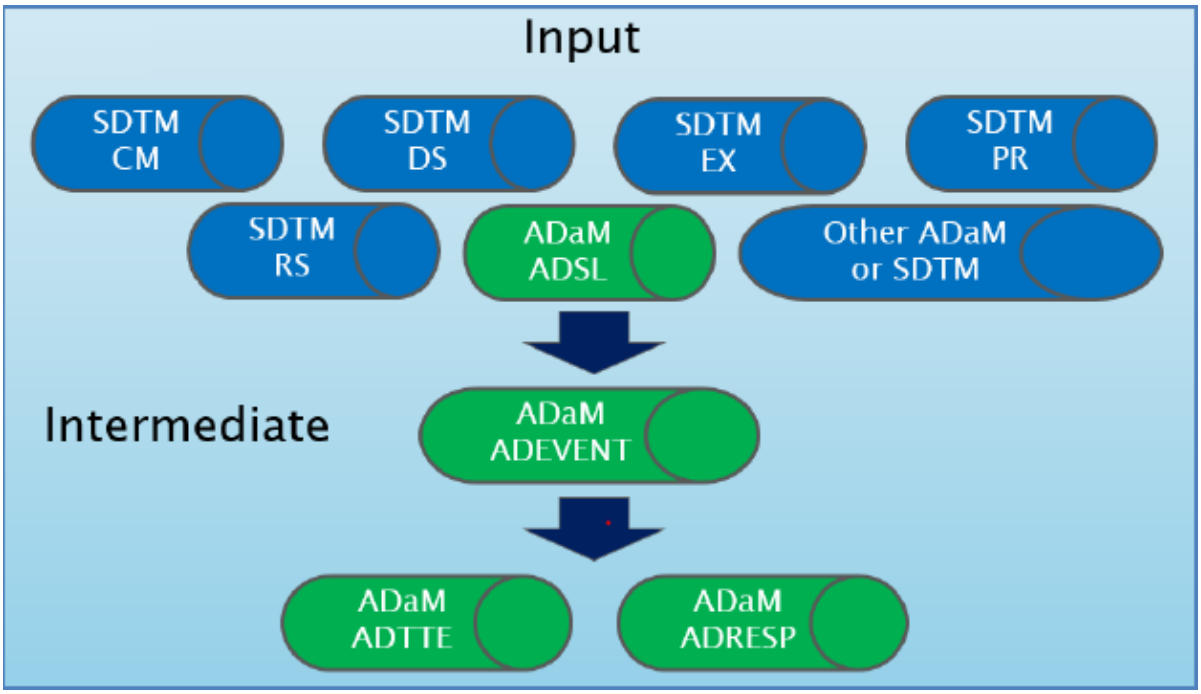


Figure 2. The Example of Using an Intermediate Dataset ADEVENT for Traceability Data Flow [9]

Please refer to Appendix 2 for ADEVENT metadata. 'ADEVENT' was defined with parameters of 'Disposition', 'Assessment', 'Event', with ASTDT (Analysis Start Date) storing the date that event occurred, AVALC descripting the type of event and an optional variable SRCDESC providing additional information. Table 2 shows an example of ADEVENT from TAUG-BrCa [1].

Row	USUBJID	ASEQ	ASTDT	ASTDY	PARQUAL	PARAMCD	AVALC	ANL01FL
1	ABC-123-001	1	2013DEC29	-4	PROTOCOL	DISPOSIT	RANDOMIZED	
2	ABC-123-001	2	2013DEC30	-2	INVESTIGATOR	ASSESS	PD	Y
3	ABC-123-001	3	2013DEC31	-1	CENTRAL	ASSESS	SD	Y
4	ABC-123-001	4	2014JAN01	1	PROTOCOL	DISPOSIT	TREATMENT	Y
5	ABC-123-001	5	2014JAN21	20	INVESTIGATOR	ASSESS	SD	Y
6	ABC-123-001	6	2014JAN22	22	CENTRAL	ASSESS	SD	Y
7	ABC-123-001	7	2014FEB13	44	INVESTIGATOR	ASSESS	PR	Y
8	ABC-123-001	8	2014FEB14	45	CENTRAL	ASSESS	PR	Y
9	ABC-123-001	9	2014MAR06	65	INVESTIGATOR	ASSESS	PR	Y
10	ABC-123-001	10	2014MAR07	66	CENTRAL	ASSESS	PR	Y
11	ABC-123-001	11	2014MAR28	87	INVESTIGATOR	ASSESS	PD	Y
12	ABC-123-001	12	2014MAR29	88	CENTRAL	ASSESS	PD	Y
13	ABC-123-001	13	2014MAR30	89	PROTOCOL	DISPOSIT	TREATMENT	Y
14	ABC-123-001	14	2014MAR31	90	PROTOCOL	EVENT	PROHIB MED	
15	ABC-123-002	1	2013NOV10	-3	PROTOCOL	DISPOSIT	RANDOMIZED	

Table 2. An Example ADEVENT from CDISC Breast Cancer Therapeutic Area User Guide [1]

Per ADEVENT metadata and the example above, it seems that ADEVENT.sas simply "stacks" all tumor assessments from SDTM.RS, and "***other dates***" listed in the last section: Item 5 (Date of the first anticancer therapy) to Item 8 (Date Last Known Alive) in a "proper" order. Including the derivation of

"***other dates***" in ADEVENT programming does not follow the best programming practice, explained above.

ADEVENT does not have the sorting keys, even though BDS.PARAMN is added. We will add it in the following section.

TAUG-BrCa [1] also provides the metadata and examples for ADTTE with PFS, OS, and DOR as the values of PARAMCD, and ADRESP with BOR as the single value of PARAMCD. ADTTE and ADRESP are constructed from ADEVENT. It further shows the benefit of the intermediate dataset: ADEVENT. How to derive dates of progression or censoring and outcome for the primary PFS analysis in ADTTE.sas and how to derive BOR in ADRESP.sas are not clearly explained in the guideline. It seems that it is up to the user for the implementation. The specification for AVALC of ADRESP (BOR derivation) is not applicable to the derivation where the confirmation of CR and PR is required, when RECIST1.1 and iRECIST are used for non-randomized trials. In fact, the programming for this derivation is NOT "simple" or straight forward! Readers can refer to [8] for the details.

The variable: AVALC stores the "codes" of dates, and ASTDT stores "dates", when PARAMCD is set to "EVENT". The value of AVALC is assigned to be as intuitive as possible. For example, when the censoring date: "***Date of last progression assessment with documented nonprogression before the missed visits***" is to be output inside ADEVENT.sas, what value is assigned to AVALC to facilitate the programming in ADTTE? "**LBFMISDT**" can be one of the choices. However, "**LBFMISDT**" is still very difficult for the users to understand its meaning during the downstream programming. It needs the carry-over of the meaning/description of the event (ADEVENT.AVALC) in the implementation of ADEVENT. In our approach, AVALC is set to **"3:LBFMISDT (Date Last Tumor Asses.Bef. Missed Visits)**" so that "**LBFMISDT"** will be the variable name and "**Date Last Tumor Asses.Bef. Missed Visits"** will be its label in the downstream dataset. We will elaborate it later. Please refer to Table 7.

TAUG-BrCa [1] also introduces the triplet of SRCDOM, SRCVAR, and SRCSEQ variables in ADEVENT to trace each event back to the original source dataset. It does provide clear traceability of events independent to tumor assessments. However, the triplet is not sufficient for the traceability of the events derived from tumor assessments due to the complexity of the derivation. For example, in order to derive the date of the last adequate tumor assessment (CR, PR, SD) for no PD subjects, the tumor assessments are required to satisfy the "eligibility" for the derivation: prior to the first new anti-cancer therapy, and the first PD date. For another example, the derivation of BOR also requires the confirmation of CR and PR, in addition to the "eligibility" above, when RECIST1.1 and/or iRECIST are used for non-randomized trials. How to trace back if the confirmation is met, and what value of time-point overall response will be used for BOR derivation if the confirmation is not met. Hence, we need more traceability in ADEVENT programming. Readers can refer to [8] for the details re the derivation of BOR.

In summary, ADEVENT builds some traceability by the triplet of SRCDOM, SRCVAR, and SRCSEQ, but it is not "enough"; ***Secondly***, the derivation of "**other dates**" complicates the programming development of both SAS program and metadata and it is the root of cause of the difficulty in maintaining and updating of both SAS program and its specification (metadata traceability); It does not follow the best programming practice; ***Thirdly***, there are no sorting keys to sort the data; ***Fourthly***, ADEVENT should have the capacity of the carry-over of the meaning/description of the event to the downstream dataset to facilitate the programming in ADTTE. Hence***, there is an urgent need for an improvement of ADEVENT from TAUG-BrCa [1] so that ADEVENT can build more traceability, facilitate the programming in ADTTE, and follow the best programming practice.***

INTRODUCION OF ADDATES FROM CDISC PROSTATE CANCER THERAPEUTIC AREA USER GUIDE

TAUG-PrCa [2] introduced an intermediate data set ADDTAES for Prostate Cancer Therapeutic in 2017. Similarly to ADEVENT, it supports the development of ADTTE with the same triplet to build traceability. Figure 3 below describes a possible approach for the order of creating the analysis datasets [2].

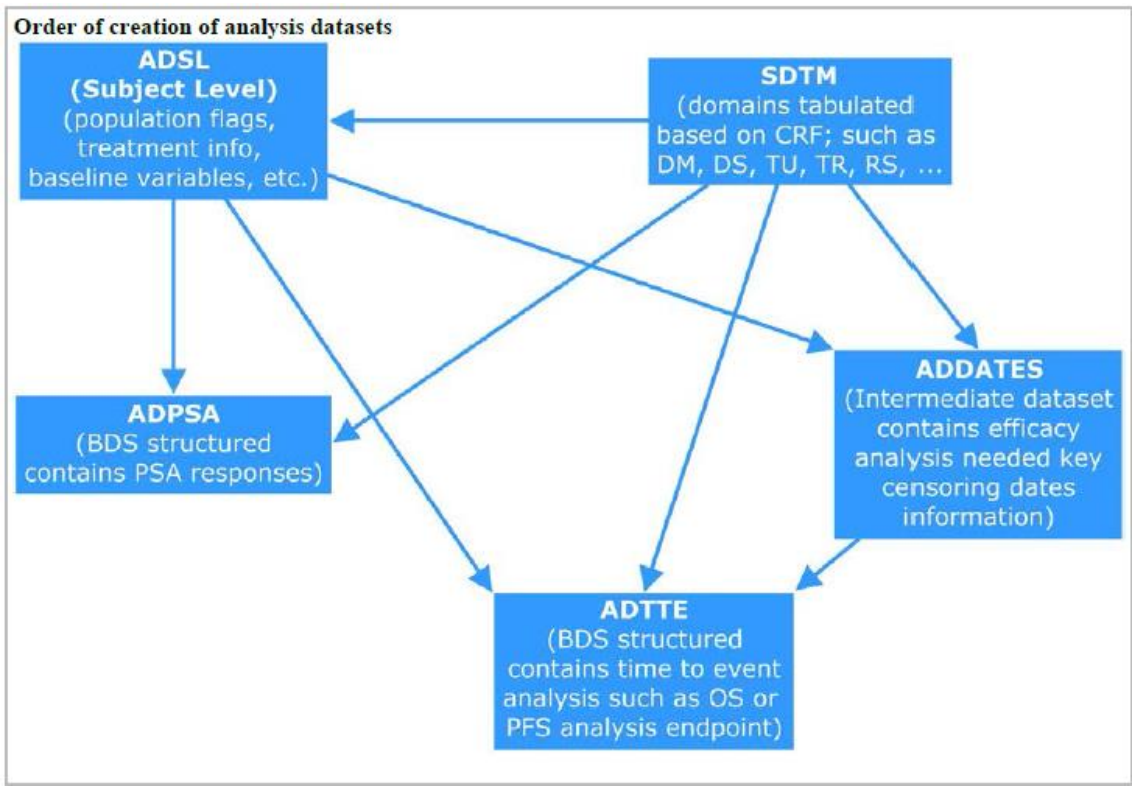


Figure 3. An Intermediate Dataset ADDATES for Traceability Data Flow

Please refer to Appendix 3 for ADDATES metadata. 'ADDATES' is defined by the variables: **ADTDESC** (Description of Analysis Date), **ADTDESCD** (Description of Analysis Date Code), and **ADT** (Analysis Date) storing the date that event occurred. Table 3 shows an example of ADDATES from TAU-PrCa [2].

Row	USUBJID	ASEQ	ADT	ADTDESC	ADTDESCD	ADY
1	ABC-123-001	1	03MAR2014	Date of Randomization	RANDDT	1
2	ABC-123-001	2	15OCT2014	Change in Anti-Cancer Therapy	RXCHGDT	227
3	ABC-123-001	3	15SEP2014	Date of Last Tumor Assessment with No PD	LNOPDDT	197
4	ABC-123-001	4	03DEC2014	Date Last Known Alive	LSTALVDT	276
5	ABC-123-001	5	01NOV2014	Date of Analysis Cut-off	CUTOFFDT	244

Table 3. An Example ADDATES from CDISC Prostate Cancer Therapeutic Area User Guide [2]

The pair of variables: ADTDESCD and ADTDESC are very similar to the one of PARAMCD and PARAM in BDS dataset. To sort the dataset, we need an extra variable to be added, similar to BDS.PARAMN beside ADDATES standard metadata, which will be introduced in the following section.

Table 4. summarizes the dates of interest listed in ADDATES metadata. Two dates from Number 9 to Number 10 are among the five dates required for PFS analysis introduced in the section prior to the last two sections. Similar to the derivation of "other dates" in ADEVENT, the dates from Number 1 to Number 8 are independent of tumor assessments.

Since the tumor assessments used for the derivation of the dates of progression or censoring for PFS analysis are not included in ADDATES, all the traceability of the derivation is "**totally lost"** from the both the dataset and meta data, even though the triplet of SRCDOM, SRCVAR, SRCSEQ pinpoints the records with the dates of interest. In fact, the triplet only provides the traceability of the derivation of the dates independent of tumor assessments: Number 1 to Number 8.

No.	Dates of Interest
1	Change in Anti-Cancer Therapy
2	Date of Analysis Cut-off
3	Date of Death
4	Date of Randomization or the First Treatment
5	Date of Toxicity Leading to Discontinuation
6	Date Last Known Alive
7	Date Lost to Follow-Up
8	End of Study Date
9	**Date of Last Tumor Assessment with No PD**
10	**Date of Tumor Assessment with PD**

Table 4. Dates of Interest Listed in Metadata from ADDATES from CDISC TAUG-PrCa [2]

It seems that ADDATES guideline was not designed to support the categorical analysis of tumor response: BOR and DOR. The date of first occurrence of BOR, the start date of DOR, and the end date of DOR can be added to the dates of interest above so that BOR and DOR can be "indirectly" derived from them. However, there is no traceability of the derivation if we followed ADDATES data flow!

In summary, ADDATES **ONLY** builds traceability by the triplet of SRCDOM, SRCVAR, and SRCSEQ for the derivation of the dates of independent of tumor assessments, no traceability for the dates related to tumor assessments is built; ***Secondly***, similarly to ADEVENT, ADDATES programming does not separate the derivation of ***dates related to tumor assessments*** and ***dates of independent of tumor assessments*** into two independent programming sections. Therefore, it does not follow the best programming practice; ***Thirdly,*** ADDATES does not have right variable to sort the data; ***Fourthly,*** the capacity of supporting the categorical analysis of tumor response: BOR and DOR from ADDATES should be added with the traceability of the complicated derivation from tumor assessment. Hence, there is ***an urgent need*** for an improvement of ADDATES [2] so that it can be broadly used to other areas of oncology studies, build more traceability, facilitate the programming in ADTTE about the ordering of ADTDESCD, and follow the best programming practice.

PROS AND CONS OF ADEVENT AND ADDATES FROM TWO CDISC STANDARDS

Table 5. summarize the pros and cons of these two CDISC standards, which shows the rationale to enhance them for an extensive use in oncology studies.

Order	Feature/Benefit	ADEVENT	ADDATES
1	Support ADTTE	Yes	Yes
2	Support BOR and DOR when no confirmation of CR and PR is required	Yes	No/Indirectly
3	Support BOR and DOR when the confirmation of CR and PR is required	No	No
4	Traceability of the Derivation of the Dates of Independent of Tumor Assessments	Yes	Yes
5	Traceability of the Derivation of the Dates Related Tumor Assessments	Not Sufficient (RECIST 1.1/ iRECIST)	**"Totally Lost"** from both the dataset and metadata

6	The meaning of the event (ADEVENT.AVALC)/date (ADDATES. ADTDESCD) can be carried over to the subsequent programming.	AVALC does not store the description of event.	ADTDESC provides the description of ADTDESCD.
7	Sorting Keys to Sort the Data	No	No
8	Best Programming Practice	Does not follow	Does not follow

Table 5. The Summary of PROS and CONS of ADEVENT and ADDATES from CDSIC Standards

INTRODUCTION OF OUR NEW DATA FLOW

Figure 1 in Introduction Section depicts the overall logic and data flow of the new approach.

We start the programming of 'ADEVENT' [1] first, and only select the records of tumor assessments with overall assessment from SDTM.RS domain, which are used for the derivation of the BOR/DOR for both categorical analysis of tumor response and a Time-to-Event (TTE) analysis. Extra records are created as listed in Table 6 and Table 7 below. Please refer to Appendix 4 for the flowchart depicting the overall logic flow of the programming approach inside ADEVENT.sas [8].

Table 6 shows the values of PARAMCD, PARAM, and PARAMN with one-to-one mapping from ADEVENT. Table 7 shows all possible values of PARAMN, PARAMCD, PARAM, and AVALC for PARAMCD='EVENT'. Each subject can have up to seven (7) records, for "PDDT (Date of Documented Progression (PD))" and "LANOPDDT (Date of Last Adequate T. Asses. of No PD)" are exclusive. The first column in Table 7 shows one-to-one relationship among ASEQ, PARAMN, PARAMCD, PARAM, and AVALC for PARAMCD='EVENT'.

PARAMCD	PARAM	Comments
OVERALLR	Overall Evaluation	Original RS Records with PARAMN=1 to maintain the traceability of the derivation of TP_BOR and EVENT
TP_BOR	Derived Overall Evaluation	Derived One Per RECIST 1.1 or iRECIST with PARAMN=2 Of note, please refer to [8] for the details
EVENT	Event Date	Derived Eight Dates of Progression or Censoring for PFS Analysis (the first five in Table 7), as well as BOR/DOR (the last three in Table 7), with PARAMN=3. **Of Note,** SRCDOM='RS', SRCVAR='RSDTC', and SRCSEQ is set to RSSEQ where the respective ANLxxFL has the value of 'Y'.

Table 6. Tabulation of the Values of PARAMCD, PARAM, and PARAMN from ADEVENT Dataset

ASEQ	PARAMN	PARAMCD	PARAM	AVALC
100	3	EVENT	Event Date	1:PDDT (Date of Documented Progression (PD))
200	3	EVENT	Event Date	2:LANOPDDT (Date of Last Adequate T. Asses. of No PD)
300	3	EVENT	Event Date	3:LBFMISDT (Date Last Tumor Asses.Bef. Missed Visits)
400	3	EVENT	Event Date	4:LAPNCTDT (Date Last Tumor Asses. Pri. to New CT)
500	3	EVENT	Event Date	5:BORDT (Date of First Occurrence of BOR)
600	3	EVENT	Event Date	6:DORSTDT (Date of DOR Start)
700	3	EVENT	Event Date	7:DORENDT (Date of DOR End)
800	3	EVENT	Event Date	8:NEWCTDT (First Date of New Anticancer Therapy)

Table 7. All Possible Values of AVALC from ADEVENT with PARAMCD='EVENT'

Of note, the last row in Table 7 for 'First Date of New Anticancer Therapy' is added due to the fact it is needed to derive other variables, even though it is independent of tumor assessment. Definitely, it could be placed in ADDATES programming. ***Secondly***, ADEVENT metadata from TAUG-BrCa specifies AVALC as "***Reported Assessment associated with the ASTDT***". As explained above, the meaning of the event dates would be lost in both ADEVENT dataset and the downstream programming. The meaning/description of the event is attached to the "original" AVALC, and the numbering (1-8) is prefixed to it for the sorting key, shown above. It improves the readability of the dataset so that it further facilitates

the downstream programming, for example, ADDATES and ADRESP. Secondly, number (1-8) and symbol ':' are prefixed to AVALC so that the developer can define the sorting order for the event dates later. Table 8 shows the metadata for ASEQ, which will be the sorting key for ADEVENT. In fact, ***ASEQ*** enhances the readability of the data, and provides further traceability of each record, and it will used to track back from ADTTE to ADEVENT, which will be shown in the later section

Variable Name	**Variable Label**	**Type**	**Source/Derivation/Comment**
RSSEQ	Sequence Number	integer	RS.RSSEQ
ASEQ	Analysis Sequence Number	integer	Derived: For the records with PARAMCD = 'OVERALLR', Sort by STUDYID, USUBJID, ASTDT, and PARAMCD then assign value. Start at 1 for each subject. No duplicates allowed within a subject; For the records with PARAMCD = 'TP_BOR', increment ASEQ by 0.5; For the records with PARAMCD ='EVENT', ASEQ=100*input(scan(AVALC,1,':'), 1.). Note: For the records with PARAMCD = 'OVERALLR', ASEQ ranges from 1 to the largest number of tumor assessments; For the records with PARAMCD = 'TP_BOR', ASEQ ranges from 1.5 to the largest number of tumor assessments+0.5; For the records with PARAMCD ='EVENT', ASEQ ranges from 100 to 800.

Table 8. The Metadata for ASEQ as the Sorting Key for ADEVENT

In general, analysis flags (ANLxxFL) enhance the readability of a data. If these flags have certain level of ***interdependency***, they also provide the traceability of the derivation and its logic flow.

Furthermore, eleven (11) analysis flags (ANL01FL - ANL11FL) and CRIT1/ CRIT1FL are derived and added to the records of tumor assessments to build ***more traceability*** of the derivation of the dates of progression or censoring related variables for Time-to-Event analysis, for example PFS, as well as BOR/DOR. These flags indicate whether the record meets the respective conditions specified in Table 9 below.

Of note, ANL01FL, ANL02FL, ANL04FL, and CRIT1/ CRIT1FL are specific to the analysis for the studies when RECIST 1.1 is required.

Variable Name	Derivation/Comments
CRIT1FL	CRIT1FL= 'Y' if RSDTC-TRTSDT +1 >= 49
CRIT1	For adequate assessment of SD, CRIT1='Meets Minimum Duration (49 Days)' if CRIT1FL= 'Y'
ANL01FL	ANL01FL='Y', if CR/PR was confirmed or SD met minimum duration (CRIT1FL= 'Y') Note: For the records with PD or NE, ANL01FL is set to 'Y' to facilitate the derivation and its programming. Of note, the derivation does not consider the anti-cancer therapy.
ANL02FL	ANL02FL='Y', if the record was the first progression disease (PD) assessment, on or prior to the first anti-cancer therapy for the derivation of BOR and DOR per RECIST 1.1
ANL03FL	ANL03FL='Y' if the record posted the first PD record
ANL04FL	ANL04FL='Y', if tumor assessments "eligible" for the derivation of BOR and DOR, which satisfied the conditions: on or prior to the first new anti-cancer therapy and the first PD date for RECIST 1.1
ANL05FL	ANL05FL='Y', if the record with ANL04FL='Y' was the last adequate tumor assessment (CR, PR, SD) for no PD subjects, and it was used for the derivation of DOR and PFS

Variable Name	Derivation/Comments
ANL06FL	ANL06FL='Y', if the record with ANL04FL='Y' was the last adequate tumor assessment (CR, PR, SD) prior to the first record of more than one missed visit prior to the first PD
ANL07FL	ANL07FL='Y', if the record with ANL04FL='Y' was last adequate tumor assessment (CR, PR, SD) prior to the first new anti-cancer therapy, and prior to the first PD. It is for primary analysis of PFS.
ANL08FL	ANL08FL='Y', if the record with ANL04FL='Y' was the first tumor assessment with Best Overall Response
ANL09FL	ANL09FL='Y', if the record with ANL04FL='Y' was the start date of response of the subjects whose Best Overall Response was either CR or PR
ANL10FL	ANL10FL='Y', if the record with ANL04FL='Y' was the end date of response of the subjects whose Best Overall Response was either CR or PR.

Table 9. Analysis Flags Built in ADEVENT for the Traceability for Both Time-to-Event Analysis and BOR/DOR

The following provides a little more details for each flag.

ANL01FL indicates whether the confirmation of CR/PR and SD is met with the aid from CRIT1FL per RECIST 1.1;

ANL02FL flags the first record with PD, which is on or prior to the first anti-cancer therapy; for the tumor assessments post the first date of progressive disease and the date of the first anticancer therapy would be excluded in the derivation of BOR and DOR when RECIST 1.1 is applied;

ANL03FL indicates whether the records are after the first PD. It is used for the derivation of ALN04FL, in addition to the indication of the records past the first PD;

ANL04FL is critical for the derivation of BOR/DOR and the event (date), and it is set to "Y' if both ANL03FL^='Y' and RSDTC<=the date of the first anti-cancer therapy. It indicates whether tumor assessment is "eligible" for being used for the derivation.

ANL05FL flags the last adequate tumor assessment (CR, PR, SD) for no PD subjects, and it was used for the derivation of DOR end date and the event (date) of "***LANOPDDT (Date of Last Adequate T. Asses. of No PD)***" for primary PFS analysis;

ANL06FL indicates whether the record was last adequate tumor assessment (CR, PR, SD) prior to the first record of more than one missed visit prior to the first PD, which is used to derive the event (date) of "***LBFMISDT (Date Last Tumor Asses.Bef. Missed Visits)***" for primary PFS analysis;

ANL07FL indicates whether the record was last adequate tumor assessment (CR, PR, SD) prior to the first anti-cancer therapy. It is used to derive the event (date) of "***LAPNCTDT (Date Last Tumor Asses. Pri. to New CT***" for primary PFS analysis;

ANL08FL indicates whether the record was the first tumor assessment with Best Overall Response (BOR);

ANL09FL indicates whether the record was the start date of response of the subjects whose Best Overall Response was either CR or PR;

ANL10FL indicates whether the record was the end date of response of the subjects whose Best Overall Response was either CR or PR.

Of note, ANL05FL - ANL10FL require that the record meets ANL04FL='Y'. Secondly, ANL09FL and ANL10FL can be used together to select the records for the simple derivation of DOR. These flags have interdependency on the hierarchy orders.

COMPARISON OF TRACEABILITY BETWEEN ANLXXFL AND THE TRIPLET OF SRCDOM, SRCVAR, AND SRCSEQ

Table 10 shows the one-to-one relationship between each flag among ANL02FL, ANL05FL - ANL10FL and its respective event (date).

ADEVENT Flags	ADEVENT.AVALC
ANL02FL	PDDT (Date of Documented Progression (PD))
ANL05FL	LANOPDDT (Date of Last Adequate T. Asses. of No PD)
ANL06FL	LBFMISDT (Date Last Tumor Asses.Bef. Missed Visits)
ANL07FL	LAPNCTDT (Date Last Tumor Asses. Pri. to New CT)
ANL08FL	BORDT (Date of First Occurrence of BOR)
ANL09FL	DORSTDT (Date of DOR Start)
ANL10FL	DORENDT (Date of DOR End)

Table 10. One-to-one Relationship between ANLxxFL and AVALC for PARAMCD='EVENT'

For each event of AVALC with PARAMCD='EVENT', one can easily use its respective ANLxxFL to locate the record, where the event occurred. It seems that it plays the same role as the triplet of SRCDOM, SRCVAR, and SRCSEQ does. Per the note for PARAMCD='EVENT' in Table 6: "***SRCDOM='RS', SRCVAR='RSDTC', and SRCSEQ is set to RSSEQ where the respective ANLxxFL has the value of 'Y'***", the triplet's role is really from RSSEQ, for 'RS' and RSDTC are only values for SRCDOM and SRCVAR in our setup. The triplet or RSSEQ can tell one the location of each event in RS records by searching each record starting from the first record until the target sequentially, the worst-case scenario: going through all the values of RSSEQ until the last record, to locate SRCSEQ from the triplet. However, ***it does not provide the details re the conditions or logics of the derivation in a logical order***. It has the analogy to a mailman to deliver the mail per the mailing address. The triplet was a new mailman who needs more time to deliver the mail to the address by door-by-door, which contrasted with respective ANLxxFL to locate the record, where the house was sitting on the top of the mountain, which was easily found and located for the mail delivery. For each event in the data block with PARAMCD='EVENT' which is listed in Table 7, the user of ADEVENT can directly and quickly find the unique 'Y' of its corresponding ANLxxFL among ANL02FL, ANL05FL - ANL10FL by simply looking at the data. This unique 'Y' is used as a "***pointer***" to pinpoint the location (record), instead of searching the target value of RESEQ to find the "mailing address". Secondly, the viewer of the data can "look around" the nearby other time-point values of tumor assessments as well as the current one, and other flags and "figure out" why the event occurs at this record, for these flags have interdependency on the hierarchy orders described above. Hence it has the analogy to the blueprint of the house (how the house was built!), in addition to a mailing address.

Hence, these flags provide more traceability of the derivation of records for events (dates) regarding the complicated rules discussed above, for examples, RECIST 1.1/iRECIST, and censoring scheme for primary PFS analysis from FDA guideline, compared to the triplet of SRCDOM, SRCVAR, and SRCSEQ. In fact, the triplet is still kept in ADEVENT just for the compliance with CDSIC TAUG-BrCa standard. We will provide further explanation in the following section from simulated examples.

THREE EXAMPLES OF ADEVENT FROM SIMULATED SUBJECTS

Table 11 shows an example of ADEVENT from a simulated subject. One block of data for 'EVENT' highlighted in blue are the derived events (dates) from tumor assessments. The other block of data are tumor assessments from RS domain for the derivation, including the derived record highlighted in green per RECIST 1.1 for BOR. Please refer to [8] for the details.

From this example, ANL04FL indicates all records from RS satisfied the conditions***: "prior to the first new anti-cancer therapy, and the first PD date for RECIST 1.1"***, and no new anti-cancer therapy received. Hence, there was no event (date) for "***LAPNCTDT (Date Last Tumor Asses. Pri. to New CT)***" and all values for ANL07FL were blank.

The subject had 'PD' at Cycle 10, and ANL05FL was 'blank' for all records. Hence there was no record for "***LANOPDDT (Date of Last Adequate T. Asses. of No PD)***" for the event (date).

ANL06FL indicates that the record from Cycle 4 was" the ***last progression assessment with documented nonprogression before the missed visits***". Hence the record from Cycle 4 was the source for the "***LBFMISDT (Date Last Tumor Asses.Bef. Missed Visits)***" indicated by ANL06FL.

ANL01FL indicates that the record from Cycle 4 with 'PR' was not confirmed per RECIST 1.1, and 'PR' was 'downgraded' to 'SD', which was stored by the newly created record highlighted in green for BOR derivation. ANL08FL and ANL09FL indicate that the record from Cycle 2 was the source for *BOR date* and *DOR start date, respectively*, and ANL10FL indicates that the record from Cycle 10 was the source for *DOR end date*. Please note that there is no censoring rules applicable to the derivation of the duration of response in this paper. Otherwise, Cycle 4 would be the source for DOR end date.

USUBJID	ASEQ	PARAMN	PARAMCD	PARAM	VISIT	ASTDT	ASTDY	AVALC	CRIT1FL	RSSEQ
simu_097	1	1	OVERALLR	Overall Evaluation	Cycle 2	2018-06-23	69	PR	Y	33
simu_097	2	1	OVERALLR	Overall Evaluation	Cycle 4	2018-08-02	109	PR	Y	34
simu_097	2.5	2	TP_BOR	Derived Overall Evaluation	Cycle 4	2018-08-02	109	SD	Y	34
simu_097	3	1	OVERALLR	Overall Evaluation	Cycle 6	2018-09-11	149	NE	Y	35
simu_097	4	1	OVERALLR	Overall Evaluation	Cycle 8	2018-10-21	189	NE	Y	36
simu_097	5	1	OVERALLR	Overall Evaluation	Cycle 10	2018-11-30	229	PD	Y	37
simu_097	100	3	EVENT	Event Date	Cycle 10	2018-11-30	229	1:PDDT (Date of Documented Progression (PD))		
simu_097	300	3	EVENT	Event Date	Cycle 4	2018-08-02	109	3:LBFMISDT (Date Last Tumor Asses.Bef. Missed Visits)		
simu_097	500	3	EVENT	Event Date	Cycle 2	2018-06-23	69	5:BORDT (Date of First Occurrence of BOR)		
simu_097	600	3	EVENT	Event Date	Cycle 2	2018-06-23	69	6:DORSTDT (Date of DOR Start)		
simu_097	700	3	EVENT	Event Date	Cycle 10	2018-11-30	229	7:DORENDT (Date of DOR End)		

ASEQ	CRIT1	ANL01 FL	ANL02 FL	ANL03 FL	ANL04 FL	ANL05 FL	ANL06 FL	ANL07 FL	ANL08 FL	ANL09 FL	ANL10 FL	SRCDOM	SRCVAR	SRCSEQ
1	Meets Minimum Duration (49 Days)	Y			Y				Y	Y				
2	Meets Minimum Duration (49 Days)				Y		Y							
2.5	Meets Minimum Duration (49 Days)				Y									
3	Meets Minimum Duration (49 Days)	Y			Y									
4	Meets Minimum Duration (49 Days)	Y			Y									
5	Meets Minimum Duration (49 Days)	Y	Y		Y						Y			
100												RS	RSDTC	37
300												RS	RSDTC	34
500												RS	RSDTC	33
600												RS	RSDTC	33
700												RS	RSDTC	37

Table 11. An Example of ADEVENT Data from A Simulated Subject

Note: The relationship between RSSEQ and SRCSEQ could be easily understood by adding the column "RSSEQ".

Table 12 shows the second example of ADEVENT from a simulated subject. The subject had PD at Cycle 2 indicated by ANL02FL. ANL03FL indicates that all records were past the PD, except for Cycle 2. ANL04FL shows that only the record at Cycle 2 is eligible for the derivation of events and BOR/DOR. Hence BOR was PD, which is indicated by ANL08FL. Only ANL02FL and ANL08FL had a record with 'Y'! Hence there were only two records with events from tumor assessments highlighted in blue, in addition to the record from ADCM for ADEVENT.AVALC='**8: NEWCTDT (First Date of New Anticancer Therapy)**'.

ANL05FL shows that no record exists for the date of last adequate tumor assessments for no PD subjects, for the subject had PD at Cycle 2 indicated by ANL02FL.

USUBJID	ASEQ	PARAMN	PARAMCD	PARAM	VISIT	ASTDT	ASTDY	AVALC	CRIT1FL	RSSEQ
simu_094	1	1	OVERALLR	Overall Evaluation	Cycle 2	2018-06-23	69	PD	Y	33
simu_094	2	1	OVERALLR	Overall Evaluation	Cycle 4	2018-08-02	109	PR	Y	34
simu_094	3	1	OVERALLR	Overall Evaluation	Cycle 6	2018-09-11	149	SD	Y	35
simu_094	4	1	OVERALLR	Overall Evaluation	Cycle 8	2018-10-21	189	PR	Y	36
simu_094	5	1	OVERALLR	Overall Evaluation	Cycle 10	2018-11-30	229	CR	Y	37
simu_094	5.5	2	TP_BOR	Derived Overall Evaluation	Cycle 10	2018-11-30	229	SD	Y	37
simu_094	100	3	EVENT	Event Date	Cycle 2	2018-06-23	69	1:PDDT (Date of Documented Progression (PD))		33
simu_094	500	3	EVENT	Event Date	Cycle 2	2018-06-23	69	5:BORDT (Date of First Occurrence of BOR)		33
simu_094	800	3	EVENT	Event Date	Cycle 6	2018-08-22	129	8:NEWCTDT (First Date of New Anticancer Therapy)		

ASEQ	CRIT1	ANL01 FL	ANL02 FL	ANL03 FL	ANL04 FL	ANL05 FL	ANL06 FL	ANL07 FL	ANL08 FL	ANL09 FL	ANL10 FL	SRCDOM	SRCVAR	SRCSEQ
1	Meets Minimum Duration (49 Days)	Y	Y		Y				Y					
2	Meets Minimum Duration (49 Days)	Y		Y										
3	Meets Minimum Duration (49 Days)	Y		Y										
4	Meets Minimum Duration (49 Days)	Y		Y										
5	Meets Minimum Duration (49 Days)			Y										
5.5	Meets Minimum Duration (49 Days)			Y										
100												RS	RSDTC	33
500												RS	RSDTC	33
800												ADSL	NEWCTDT	1

Table 12. The Second Example of ADEVENT Data from A Simulated Subject

Table 13 shows the third example of ADEVENT from a simulated subject. The subject had a new anticancer therapy on 2018-08-03, which can be seen at the last record, where AVALC='**8: NEWCTDT (First Date of New Anticancer Therapy)**'. It was the root of cause of the "ineligibility" of the tumor assessments after Cycle 4 to be included in the derivation. ANL01FL was set to 'Y' for the first four records, for only RECIST 1.1 is applied to the derivation of the confirmation flag without consideration of anti-cancer therapy. ANL04FL can help the readers to quickly identify that the records only from Cycle 2 and 4 can be used for the derivation. The BOR is 'PR', and ANL08FL indicates that the first occurrence is from Cycle 2, which is also the DOR start date as shown by ANL09FL. ANL10FL shows that the DOR end date was at Cycle 4. If ANL04FL did not exist, it would be very difficult for the readers to understand why BOR were not chosen at Cycle 8 with the value: 'CR'.

USUBJID	ASEQ	PARAMN	PARAMCD	PARAM	VISIT	ASTDT	ASTDY	AVALC	CRIT 1FL	RSSEQ
simu_091	1	1	OVERALLR	Overall Evaluation	Cycle 2	2018-06-23	69	PR	Y	33
simu_091	2	1	OVERALLR	Overall Evaluation	Cycle 4	2018-08-02	109	PR	Y	34
simu_091	3	1	OVERALLR	Overall Evaluation	Cycle 6	2018-09-11	149	SD	Y	35
simu_091	4	1	OVERALLR	Overall Evaluation	Cycle 8	2018-10-21	189	CR	Y	36
simu_091	5	1	OVERALLR	Overall Evaluation	Cycle 10	2018-11-30	229	CR	Y	37
simu_091	5.5	2	TP_BOR	Derived Overall Evaluation	Cycle 10	2018-11-30	229	SD	Y	37
simu_091	200	3	EVENT	Event Date	Cycle 4	2018-08-02	109	2:LANOPDDT (Date of Last Adequate T. Asses. of No PD)		
simu_091	400	3	EVENT	Event Date	Cycle 4	2018-08-02	109	4:LAPNCTDT (Date Last Tumor Asses. Pri. to New CT)		
simu_091	500	3	EVENT	Event Date	Cycle 2	2018-06-23	69	5:BORDT (Date of First Occurrence of BOR)		
simu_091	600	3	EVENT	Event Date	Cycle 2	2018-06-23	69	6:DORSTDT (Date of DOR Start)		

simu_091	700	3	EVENT	Event Date	Cycle 4	2018-08-02	109	7:DORENDT (Date of DOR End)		
simu_091	800	3	EVENT	Event Date	Cycle 4	2018-08-03	110	8:NEWCTDT (First Date of New Anticancer Therapy)		

ASEQ	CRIT1	ANL01FL	ANL02FL	ANL03FL	ANL04FL	ANL05FL	ANL06FL	ANL07FL	ANL08FL	ANL09FL	ANL10FL	SRCDOM	SRCVAR	SRCSEQ
1	Meets Minimum Duration (49 Days)	Y			Y				Y	Y				
2	Meets Minimum Duration (49 Days)	Y			Y	Y		Y			Y			
3	Meets Minimum Duration (49 Days)	Y												
4	Meets Minimum Duration (49 Days)	Y												
5	Meets Minimum Duration (49 Days)													
5.5	Meets Minimum Duration (49 Days)													
200												RS	RSDTC	34
400												RS	RSDTC	34
500												RS	RSDTC	33
600												RS	RSDTC	33
700												RS	RSDTC	34
800												ADSL	NEWCTDT	1

Table 13. The Third Example of A Simulated Subject's Assessments to Demonstrate the Role of ANL09FL and ANL10FL for DOR Start and End

The three examples above further demonstrate the ***superior of ANLxxFL to the triplet of SRCDOM, SRCVAR, and SRCSEQ*** regarding the traceability of the events (dates) in ADEVENT.

INTRODUCTION OF ADRESP PROGRAMMING SUBSEQUENT TO ADEVENT PROGRAMMING

TAUG-BrCa [1] introduced an efficacy data, named as ***ADRESP*** (Analysis of Best Overall Tumor Response) - one record per subject per analysis, which supports Categorical Analysis of Tumor Response. Please refer to the Appendix 5 for its metadata,

The following SAS codes shows that ONLY one data step can create ADRESP from ADEVENT dataset.

```
data resp;
     length avalc $4. param $40.;
     set adevent(rename=(avalc=oavalc param=oparam) where=(anl08fl='Y'));
     srcdom='ADEVENT';srcvar='AVALC';srcseq=aseq;
*** (1) BOR=Best Overall Response;
     paramcd='BOR';param='Best Overall Response';paramn=1;
     avalc=strip(oavalc);
     output;
*** (2) BORORR=Objective Response Rate;
     paramcd='BORORR';param='Objective Response Rate';paramn=2;
     if oavalc in ('CR','PR') then do;avalc='Y';aval=1;end;
     else do;avalc='N';aval=2;end;
     output;
*** (3) BORDCR=Disease Control Rate;
     paramcd='BORDCR';param='Disease Control Rate';paramn=3;
     if oavalc in ('CR','PR','SD') then do;avalc='Y';aval=1;end;
     else do;avalc='N';aval=2;end;
     output;
run;
```

Table 14 shows Examples of ADRESP data from three simulated subjects.

USUBJID	ASEQ	PARAMCD	PARAM	PARAMN	AVALC	AVAL	SRCDOM	SRCVAR	SRCSEQ
simu_091	1	BOR	Best Overall Response	1	PR	2	ADEVENT	AVALC	1
simu_091	2	BORORR	Objective Response Rate	2	Y	1	ADEVENT	AVALC	1
simu_091	3	BORDCR	Disease Control Rate	3	Y	1	ADEVENT	AVALC	1
simu_094	1	BOR	Best Overall Response	1	PD	4	ADEVENT	AVALC	1
simu_094	2	BORORR	Objective Response Rate	2	N	2	ADEVENT	AVALC	1
simu_094	3	BORDCR	Disease Control Rate	3	N	2	ADEVENT	AVALC	1
simu_097	1	BOR	Best Overall Response	1	PR	3	ADEVENT	AVALC	1
simu_097	2	BORORR	Objective Response Rate	2	Y	2	ADEVENT	AVALC	1
simu_097	3	BORDCR	Disease Control Rate	3	Y	1	ADEVENT	AVALC	1

Table 14. Examples of ADRESP Data from Three Simulated Subjects

From the example above, one can see that ADRESP can be easily used to generate objective response rate (ORR) table.

INTRODUCTION OF ADDATES PROGRAMMING SUBSEQUENT TO ADEVENT PROGRAMMING

The programming for ADDATES starts to read ADEVENT data to select the records with PARAMCD='EVENT'. The following SAS codes convert ADEVENT.AVALC into ADDATES.ADTDESCN, ADDATES.ADTDESCD, and ADDATES.ADTDESC. Table 15 shows the metadata of ADTDESCN being added to ADDATES derived from ADEVENT.AVALC of the records with PARAMCD='EVENT'. This explains why ADEVENT.AVALC is 'built' by concatenating 'each code' and its meaning, along with the prespecified order so that ADTDESC is directly retrieved from it, in addition to ADTDESCN and ADTDESCD. Table 16 shows the one-to-one relationship among the triplet of ADTDESCN, ADTDESCD and ADTDESC.

Of note, ADTDESCN is a ***new*** variable to be added to the metadata of ADDATES in TAUG-PrCa [2]. It helps the sorting of ADDATES dataset. It plays the same role as BDS.PARAMN for PARAMCD and PARAM. It was prespecified in ADEVENT.AVALC in Table 7. It also serves as the link between the ADDATES and ADTTE, which plays the key role in creating the traceability. We will explain it later section.

Variable Name	Variable Label	Type	Codelist/Controlled Terms	Source/Derivation/Comment
ADTDESCN	Description of Analysis Date (N)	integer	ADTDESCN (ADTDESC): (1) 1=Date of Documented Progression (PD) (2) 2=Date of Last Adequate T. Asses. of No PD (3) 3= Date Last Tumor Asses.Bef. Missed Visits (4) 4=Date Last Tumor Asses. Pri. to New CT (5) 5=Date of First Occurrence of BOR (6) 6=Date of DOR Start (7) 7=Date of DOR End (8) 8=New Anticancer Therapy Start Date (9) 9=Date of Randomization (10) 10=Date of First Exposure to Treatment (11) 11= Date of Last Exposure to Treatment (12) 12=Date of Analysis Cut-off (13) 13=Date of Death (14) 14=Date Last Known Alive (15) 15=Date Lost to Follow-up (16) 16=End of Study Date	Assigned: 1, if ADTDESC='Date of Documented Progression (PD)'; 2, if ADTDESC='Date of Last Adequate T. Asses. of No PD'; 3, if ADTDESC='Date Last Tumor Asses.Bef. Missed Visits'; 4, ADTDESC='Date Last Tumor Asses. Pri. to New CT'; 5, if ADTDESC='Date of First Occurrence of BOR'; 6, if ADTDESC='Date of DOR Start'; 7, if ADTDESC='Date of DOR End'; 8, if ADTDESC='New Anticancer Therapy Start Date'; 9, if ADTDESC='Date of Randomization'; 10, if ADTDESC='Date of First Exposure to Treatment'; 11, if ADTDESC='Date of Last Exposure to Treatment'; 12, if ADTDESC='Date of Analysis Cut-off'; 13, if ADTDESC='Date of Death'; 14, if ADTDESC='Date Last Known Alive'; 15, if ADTDESC='Date Lost to Follow-up'; 16, if ADTDESC= End of Study Date'

Table 15. Metadata of ADTDESCN Being Added to ADDATES

The snip of SAS codes in ADDATES.sas to retrieve ADTDESCN, ADTDESCD and ADTDESC from ADEVENT.AVALC and derive ADT directly from ADEVENT.ASTDT are shown below.

```
** Dates Related Tumor Assessment from ADEVENT where paramcd='EVENT';
data tu_events(rename=(astdt=adt astdy=ady));
     length adtdesc tmp $70. adtdescd $8. srcvar $12.;
     set adevent(drop=srcvar where=(paramcd='EVENT'));
     adtdescn=input(scan(avalc,1,':'),best.);
     tmp=translate(avalc,'',':');
     adtdescd=scan(tmp,2,'');
     len=length(strip(avalc));len1=length(strip(adtdescd))+5;len2=len-len1;
     adtdesc=substr(avalc,len1,len2);
     srcdom='ADEVENT';srcvar='AVALC$ASTDT';srcseq=aseq;
     keep usubjid avalc adtdescd adtdesc astdt astdy trtsdt adtdescn srcdom
          srcvar srcseq aseq;
run;
```

ADTDESCN (Description of Analysis Date (N))	ADTDESCD (Description of Analysis Date Code)	ADTDESC (Description of Analysis Date)
1	PDDT	Date of Documented Progression (PD)
2	LANOPDDT	Date of Last Adequate T. Asses. of No PD
3	LBFMISDT	Date Last Tumor Asses.Bef. Missed Visits
4	LAPNCTDT	Date Last Tumor Asses. Pri. to. New CT
5	BORDT	Date of First Occurrence of BOR
6	DORSTDT	Date of DOR Start
7	DORENDT	Date of DOR End
8	NEWCTDT	First Date of New Anticancer Therapy

Table 16. Tabulation of ADTDESCN, ADTDESCD and ADTDESC

The next step is to add other dates of independent of tumor assessments to ADDATES dataset. The SAS codes are developed by independently adding each programming block to generate one record for each date listed in Table 17 below. This programming is very similar to one for ADSL. The derivation rules for these dates are very simple and straightforward, compared to ones for the dates of progression or censoring for PFS analysis and BOR/DOR.

ADTDESCN (Description of Analysis Date (N))	ADTDESCD (Description of Analysis Date Code)	ADTDESC (Description of Analysis Date)
9	RANDDT	Date of Randomization
10	TRTSDT	Date of First Exposure to Treatment
11	TRTEDT	Date of Last Exposure to Treatment
12	CUTOFFDT	Date of Analysis Cut-off
13	DTHDT	Date of Death
14	LSTALVDT	Date Last Known Alive
15	LOSTFUDT	Date Lost to Follow-up
16	EOSDT	End of Study Date

Table 17. Tabulation of ADTDESCN, ADTDESCD and ADTDESC for Other Dates of Independent of Tumor Assessments

Table 18 shows an example of ADDATES from one of three simulated subjects above.

USUBJID	ASEQ	ADTDESCN	ADT	ADTDESC	ADTDESCD	ADT	ADY	SRCDOM	SRCVAR	SRCSEQ
simu_097	1	1	2018-11-30	Date of Documented Progression (PD)	PDDT	2018-11-30	229	ADEVENT	AVALC$ASTDT	100
simu_097	2	3	2018-08-02	Date Last Tumor Asses.Bef. Missed Visits	LBFMISDT	2018-08-02	109	ADEVENT	AVALC$ASTDT	300
simu_097	3	5	2018-06-23	Date of First Occurrence of BOR	BORDT	2018-06-23	69	ADEVENT	AVALC$ASTDT	500
simu_097	4	6	2018-06-23	Date of DOR Start	DORSTDT	2018-06-23	69	ADEVENT	AVALC$ASTDT	600
simu_097	5	7	2018-11-30	Date of DOR End	DORENDT	2018-11-30	229	ADEVENT	AVALC$ASTDT	700

USUBJID	ASEQ	ADTDESCN	ADT	ADTDESC	ADTDESCD	ADT	ADY	SRCDOM	SRCVAR	SRCSEQ
simu_097	6	9	2018-04-16	Date of Randomization	RANDDT	2018-04-16	1	ADSL	RANDDT	1
simu_097	7	10	2018-04-16	Date of First Exposure to Treatment	TRTSDT	2018-04-16	1	ADSL	TRTSDT	1
simu_097	8	11	2018-12-18	Date of Last Exposure to Treatment	TRTEDT	2018-12-18	247	ADSL	TRTEDT	1
simu_097	9	12	2020-03-31	Date of Analysis Cut-off	CUTOFFDT	2020-03-31	716	ADSL	CUTOFFDT	1
simu_097	10	14	2020-02-07	Date Last Known Alive	LSTALVDT	2020-02-07	663	ADSL	LSTALVDT	1
simu_097	11	15	2019-03-04	Date Lost to Follow-up	LOSTFUDT	2019-03-04	323	ADSL	EOSDT	1
simu_097	12	16	2019-03-04	End of Study Date	EOSDT	2019-03-04	323	ADSL	EOSDT	1

Table 18. An Example of ADDATES from One of Three Simulated Subjects

The first five records highlighted in green are “directly” derived from ADEVENT, and their traceability are built by the triplet of SRCDOM, SRCVAR, and SRCSEQ. It is worthwhile to point out that the values of SRCVAR is assigned to ‘AVALC$ASTDT’, for both AVALC and ASTDT collectively are used to derive ADTDESCN, ADTDESCD, ADTDESC, and ADT. SRCSEQ will be used to trace back from ADDATES to ADEVENT for the source of the event, for ADDATES.SRCSEQ=ADEVENT.ASEQ. It builds the traceability in ADDATES from ADEVENT for the events from tumor assessments.

INTRODUCTION OF ADTTE PROGRAMMING SUBSEQUENT TO ADDATES PROGRAMMING

We will not spend any time here to introduce ADTTE (Data for the Time to Event Analyses), for the readers may already have the knowledge from CDISC guideline, and other resources. However, we will introduce how ADTTE can be “easily” built from ADDATES.

ADDATES has the vertical structure of a BDS dataset. ADTTE needs different dates (ADDATES.ADT) for the derivation, and these dates should be horizontal to support the derivation. The following SAS codes shows that SAS transpose procedure can be used to directly convert ADDATES dataset into one, where one subject has one record per parameter with all value of ADT listed in columns, and each value of ADTDESCD as its variable name and value of ADTDESC as its label.

```
*** get all dates for PFS and OS, ect.;
proc sort data=addates_recist out=addates1;by usubjid adtdescn;run;
proc transpose data=addates1 out=addates2(drop=_name_ _label_);
     by usubjid;
     id adtdescd;
     idlabel adtdesc;
     var adt;
run;
```

Table 19 shows an example of transposed ADDATES with the values of ADTDESCD as variable name from three simulated subjects above. Table 20 shows an example of transposed ADDATES with ADTDESC as labels from same three simulated subjects above. Each subject has a single observation with all the dates of interest shown by columns as variables.

USUBJID	TRTSDT	PDDT	DTHDT	LANOPDDT	LBFMISDT	LAPNCTDT	NEWCTDT	EOSDT	LSTALVDT	CUTOFFDT
simu_091	2018-04-16			2018-08-02		2018-08-02	2018-08-03	2019-04-25	2020-03-30	2020-03-31
simu_094	2018-04-16	2018-06-23					2018-08-22	2019-02-20	2020-01-26	2020-03-31
simu_097	2018-04-16	2018-11-30			2018-08-02			2019-03-04	2020-02-07	2020-03-31

Table 19. Example Transposed ADDATES with ADTDESCD as Variable Name from Three Simulated Subjects Above

Unique Subject Identifier	Date of First Exposure to Treatment	Date of Documented Progression (PD)	Date of Death	Date of Last Adequate T. Asses. of No PD	Date Last Tumor Asses.Bef. Missed Visits	Date of Last Tumor Asses. Pri. to New CT	New Anticancer Therapy Start Date	End of Study Date	Date Last Known Alive	Date of Analysis Cut-off
simu_091	2018-04-16			2018-08-02		2018-08-02	2018-08-15	2019-04-25	2020-03-30	2020-03-31
simu_094	2018-04-16	2018-06-23					2018-08-22	2019-02-20	2020-01-26	2020-03-31
simu_097	2018-04-16	2018-11-30			2018-08-02			2019-03-04	2020-02-07	2020-03-31

Table 20. Example Transposed ADDATES with ADTDESC as Labels from Same Three Simulated Subjects Above

Each variable in the transposed ADDATES in Table 19 will be used to derive ADTTE for each TTE endpoint in the downstream programming, and each label of the variable in Table 20 can help the users to better understand each variable for use to facilitate the downstream programming. It shows the benefit of ADDATES.ADTDESC, designed by CDISC TAUG-PrCa [2]. For the dates related to tumor assessments, names and their labels are directly from ADEVENT.AVALC in our ***new*** setup, where the meaning/description of the event is attached to the "original" AVALC. Its benefit from the attachment is demonstrated by the example. The example above also shows the readers why ADDATES adheres to the BDS structure, besides the benefit of the repository to store all dates of interest.

It is worthwhile to point out that ADDATES.SRCSEQ, which is establishing the path between each event date in ADDATES and its immediate predecessor in ADEVENT for tumor assessments, is "lost" when ADDATES is transposed. However, the one-to-one relationship between ADTDESCN and ADTDESCD defined in Table 16 and Table 17 can lead the users to find ADTDESCN, and further SRCSEQ for each variable in the transposed ADDATES dataset, exemplified by Table 18. The traceability from each date of interest in the transposed ADDATES dataset to ADDATES, especially dates of interest related to tumor assessments, is fulfilled by the aid of ***ADTDESCN***.

INTRODUCTION OF ADTTE PROGRAMMING FOR PFS

FDA's guideline [10] provides recommendations to applicants on endpoints for cancer clinical trials. Its Table 1 "***A Comparison of Important Cancer Approval Endpoints***" provides time-to-event (TTE) endpoints, for example, Overall Survival (OS), Time to Treatment Failure (TTF), Disease-free Survival (DFS), Event-Free Survival (EFS), Progression-Free Survival (PFS), Time to Progression (TTP).

This section will introduce PFS as an example to show the readers the benefit of ADDATES and the traceability of derivation built from ADEVENT to ADDATES to ADTTE, for PFS is the mostly used primary/co-primary endpoint in oncology studies, in addition to OS, and its more complexity of the derivation of ADTTE compared to one for OS. As the explanation in Introduction section, the censoring scheme for PFS supportive analysis in Table 1 is used to derive ADTTE. The readers could apply this new approach to other TTE endpoints based on the specific context of use, which would be specified in study Statistical Analysis Plan (SAP).

CDISC [11] provides the ADaM standards for time-to-event (TTE) endpoints, named as ADTTE. Its key variables are ADT, AVAL, CNSR, EVNTDESC, and CNSDTDSC, along with the triplet of SRCDOM, SRCVAR, and SRCSEQ as the traceability variables. The derivation of these key variables could be derived through a series of IF-THEN/ELSE-DO statements inside ADTTE SAS programming by applying the censoring rules, per the ***availability*** of transposed ADDATES dataset in horizontal structure, which is exemplified by Table 19.

Table 21 shows the partial of the derivation rules for these ADTTE variables per FDA guideline [3] presented in Table 1 for primary analysis of PFS. IF-THEN/ELSE-DO statements by the column 'CONDITION' are 'from top to bottom' for each analysis. The one-to-one relationship between ADTDESCN and ADTDESCD shown in Table 16 and Table 17 is used for the derivation of ADTDESCN in the last column. This intermediate variable: ADTDESCN can be used as a link between ADTTE and ADDATES so that the triplet of SRDDOM, SRCVAR, and SRCSEQ can be built for ADTTE based on the source from ADDATES. We will explain it next.

IF CONDITION	EVNTDESN	EVNTDESC	CNSDTDSC	CNSR	ADT	ADTDESCN
PDDT>.Z and DTHDT>.Z	1	DOCUMENTED PROGRESSION PRIOR TO DEATH		0	PDDT	1
PDDT>.Z and DTHDT=.	2	DOCUMENTED PROGRESSION		0	PDDT	1
PDDT=. and DTHDT>.Z	3	DEATH		0	DTHDT	14
.Z<LAPNCTDT<PDDT	4	PD AFTER NEW ANTICANCER THERAPY'	LAST RADIOLOGIC ASSESSMENT PRIOR TO NEW ANTICANCER THERAPY	2	LAPNCTDT	4
.Z<LBFMISDT <PDDT	5	PD AFTER MISSING ASSESSMENTS	LAST RADIOLOGIC ASSESSMENT PRIOR TO MISSING ASSESSMENTS	3	LBFMISDT	3
.Z<LAPNCTDT< DTHDT	6	DEATH AFTER NEW ANTICANCER THERAPY'	LAST RADIOLOGIC ASSESSMENT PRIOR TO NEW ANTICANCER THERAPY	2	LLAPNCTDT	4
.Z< LBFMISDT < DTHDT	7	DEATH AFTER MISSING ASSESSMENTS	LAST RADIOLOGIC ASSESSMENT PRIOR TO MISSING ASSESSMENTS	3	LBFMISDT	3
LANOPDDT>.Z	8	NO PROGRESSION	LAST RADIOLOGIC ASSESSMENT SHOWING NO PROGRESSION	1	LANOPDDT	2

Table 21. Derive Rules for ADTTE Key Variables per FDA guideline [3] presented in Table 1 for Primary PFS Analysis

We will provide how to design this logic above for the ***full scenario*** in another paper [12], for the emphasis of this paper is to introduce the new approach to enhance these two CDISC standards and another reason is due to the limitation of paper pages.

CONVENIENTLY BUILD TRACEABILITY FOR ADTTE

After the execution of the programming for ADTTE's key variables specified in Table 21 for PFS, the intermediate SAS dataset is named as adtte_tmp. Before the final SAS data is output for ADTTE, the triplet of SRDDOM, SRCVAR, and SRCSEQ for ADTTE is derived as follows.

```
proc sort data=addates out=addates1(keep=usubjid adtdescn srcdom srcvar
                                    srcseq);by usubjid adtdescn;run;
proc sort data=adtte_tmp;by usubjid adtdescn;run;
data adtte_tmp1;
     merge adtte_tmp(in=a)
           addates1;
     by usubjid adtdescn;
     if a;
run;
```

The SAS codes above show that ADDATES.ADTDESCN plays the key role to derive the triplet for ADTTE directly from the one of ADDATES, which has two categories: one with SRCDOM='ADEVENT', another with SRCDOM='ADSL' which is straight forward to trace the source of the derivation in ADTTE back to ADSL kind of variable. The ADDATES.SRCSEQ among the triplet with SRCDOM='ADEVENT' is directly derived from ADEVENT.ASEQ. Hence ADTTE.SRCSEQ can be used directly to identify the record for the 'event' in ADEVENT along with tumor assessments from SDTM.RS domain to better understand the relationship between ADTTE and the SDTM RS domain. Hence the traceability is built by clearly establishing the path between ADTTE and SDTM RS domain. The full path is traced by going from ADTTE to its predecessor: either ADEVENT, back to the SDTM RS domain, for the source from tumor assessment, or ADSL for the source independent of tumor assessments.

Table 23 and the first three columns in Table 24 show an example of ADTTE's key variables from a simulated subject from Table 11 with the 'temporary/intermediate' variable: ADTDESCN (=3), which corresponds ADDATES.SRCSEQ=300 in Table 18. Column 4-Column 6 in Table 23 show the triplet from ADDATES with SRCSEQ=300. Table 11 for ADEVENT data from this subject is copied below for ease of reference. Hence directly go to ADEVENT and look for the record with ASEQ=300! From the row, there are two methods to locate the record for the source of ADTTE, shown by Table 22 below.

Method 1	AVALC="3:LBFMISDT (Date Last Tumor Asses.**Bef. Missed Visits**)" → ANL06FL='Y' → ASEQ=2 →the second record
Method 2	SRCSEQ=34 → RSSEQ=34 → ASEQ=2 or ASEQ=2.5→the second record or third record → ASEQ=2→the second record, for the record with ASEQ=2.5 was derived for BOR derivation.

Table 22. Two Methods to Locate the Record for the Source of ADTTE

It is up to the readers to decide which method to be used to locate the one for the data source of this record in ADTTE. Please keep in mind that the number of tumor assessments in this example had only five (5) records! If it had more records, it would take more time to locate the target record, which is explained in the section: "**Comparison of traceability between ANLxxFL and the triplet of SRCDOM, SRCVAR, and SRCSEQ**".

Looking at the overall tumor response records with PARAMCD='OVERALLR' from the first to the fifth makes one to easily understand why EVNTDESC='**PD AFTER MISSING ASSESSMENTS'**, and ADTTE.ADT=**'2018-08-02**'! [12] will provide more detailed SAS programming to derive ADTTE.ADTDESCN.

USUBJID	PARCAT1	PARAMCD	PARAM	EVNTDESN	EVNTDESC	CNSDTDSC	ADTDESCN
simu_097	Primary PFS Analysis	PFS	Progression Free Survival (Days)	5	PD AFTER MISSING ASSESSMENTS	LAST RADIOLOGIC ASSESSMENT PRIOR TO MISSING ASSESSMENTS	3

Table 23. An Example of ADTTE from a Simulated Subject with the Temporary Variable: ADTDESCN

ADTDESCN	CNSR	ADT	SRCDOM	SRCVAR	SRCSEQ
3	3	2018-08-02	ADEVENT	AVALC$ASTDT	300

Table 24. An Example of ADTTE from a Simulated Subject with Triplet from ADDATES

USUBJID	ASEQ	PARAMN	PARAMCD	PARAM	VISIT	ASTDT	ASTDY	AVALC	AVAL	CRIT1FL	RSSEQ
simu_097	1	1	OVERALLR	Overall Evaluation	Cycle 2	2018-06-23	69	PR	2	Y	33
simu_097	2	1	OVERALLR	Overall Evaluation	Cycle 4	2018-08-02	109	PR	2	Y	34
simu_097	2.5	2	TP_BOR	Derived Overall Evaluation	Cycle 4	2018-08-02	109	SD	3	Y	34
simu_097	3	1	OVERALLR	Overall Evaluation	Cycle 6	2018-09-11	149	NE	5	Y	35
simu_097	4	1	OVERALLR	Overall Evaluation	Cycle 8	2018-10-21	189	NE	5	Y	36
simu_097	5	1	OVERALLR	Overall Evaluation	Cycle 10	2018-11-30	229	PD	4	Y	37
simu_097	100	3	EVENT	Event Date	Cycle 10	2018-11-30	229	1:PDDT (Date of Documented Progression (PD))			
simu_097	300	3	EVENT	Event Date	Cycle 4	2018-08-02	109	3:LBFMISDT (Date Last Tumor Asses.Bef. Missed Visits)			
simu_097	500	3	EVENT	Event Date	Cycle 2	2018-06-23	69	5:BORDT (Date of First Occurrence of BOR)			
simu_097	600	3	EVENT	Event Date	Cycle 2	2018-06-23	69	6:DORSTDT (Date of DOR Start)			
simu_097	700	3	EVENT	Event Date	Cycle 10	2018-11-30	229	7:DORENDT (Date of DOR End)			

ASEQ	CRIT1	ANL01 FL	ANL02 FL	ANL03 FL	ANL04 FL	ANL05 FL	ANL06 FL	ANL07 FL	ANL08 FL	ANL09 FL	ANL10 FL	SRCD OM	SRCVAR	SRCSE Q
1	Meets Minimum Duration (49 Days)	Y			Y				Y	Y				
2	Meets Minimum Duration (49 Days)				Y		Y							
2.5	Meets Minimum Duration (49 Days)				Y									
3	Meets Minimum Duration (49 Days)	Y			Y									
4	Meets Minimum Duration (49 Days)	Y			Y									
5	Meets Minimum Duration (49 Days)	Y	Y		Y						Y			
100												RS	RSDTC	37
300												RS	RSDTC	34
500												RS	RSDTC	33
600												RS	RSDTC	33
700												RS	RSDTC	37

Table 11. An Example of ADEVENT Data from A Simulated Subject

The example above shows how the traceability for ADTTE is built through data flow: ADEVENT, ADDATES, and ADTTE, and analysis flags: ANL02FL, ANL05FL, ANL06, and ANL07FL and their respective event dates, shown in Table 11, for events (dates) related to tumor assessments, and ADDATES directly to ADTTE for events (dates) independent of tumor assessments from ADSL, which can be traced back to respective SDTM domain, for example, EX, DS, CM, and PR. It provides the users and/or FDA reviewers with a much clearer traceability of how ADTTE is assembled.

Figure 4 and Figure 5 below depict two methods above to show the traceability from ADTTE to ADSL or ADEVENT.

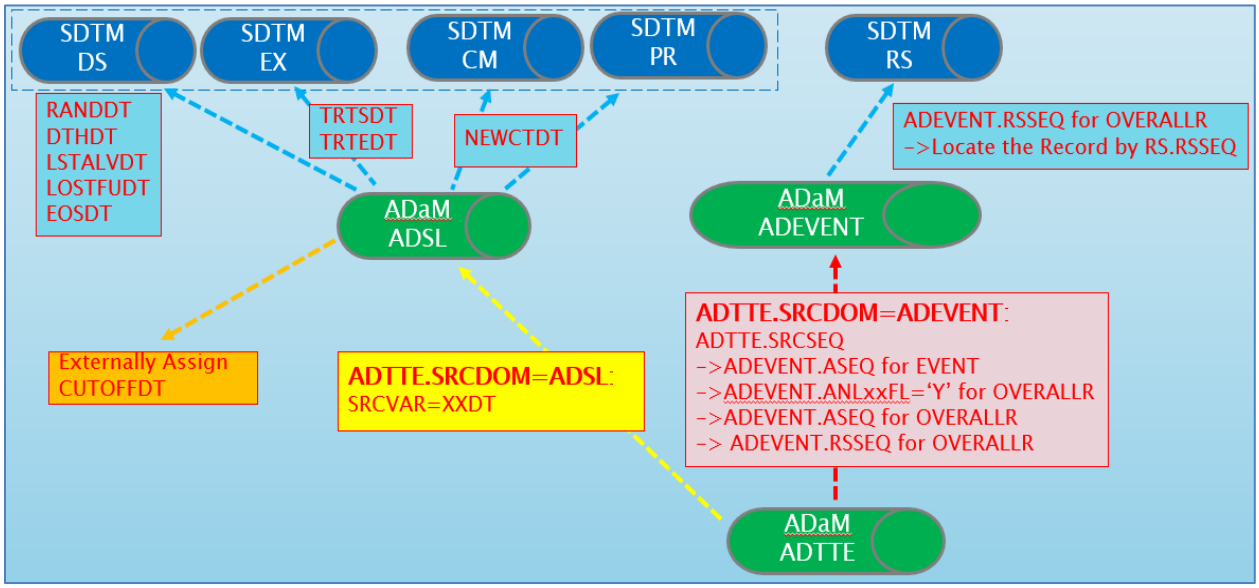


Figure 4. Display of the Traceability from ADTTE to ADSL or ADEVENT by Method 1

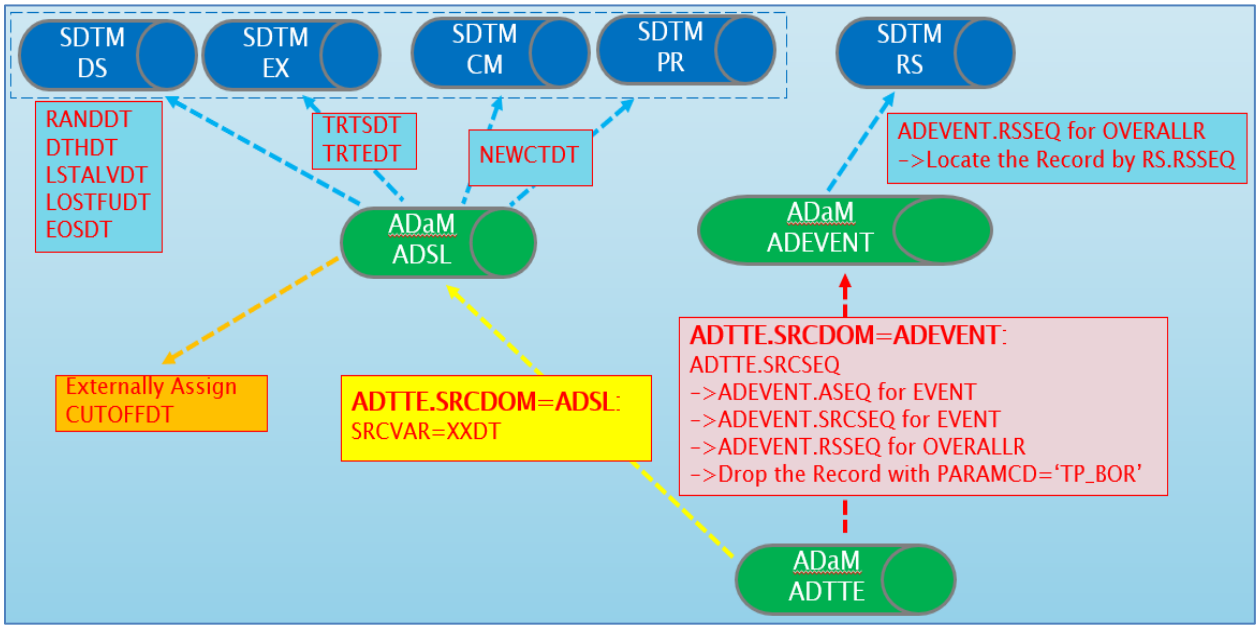


Figure 5. Display of the Traceability from ADTTE to ADSL or ADEVENT by Method 2

IN SUMMARY

Figure 6 below displays how the traceability is built, along with the overall logic and the data flow, shown in Figure 1.

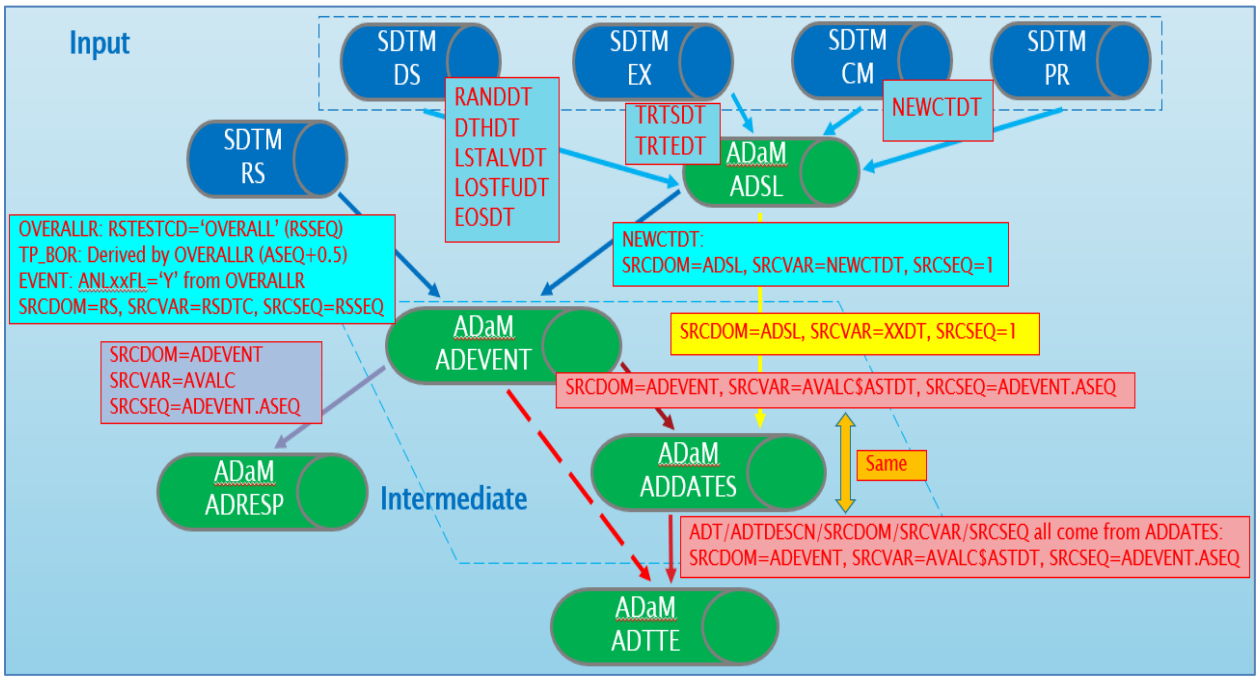


Figure 6. Display of Traceability Assemblies, along with the Overall Logic and Data Flow, Shown in Figure 1

So far, this paper has illustrated the new approach to enhance these two CDISC standards: the ADEVENT of TAUG-BrCa [1] and ADDATES of TAUG-PrCa [2] through their updated metadata and examples. Table 25 and Table 26 summarize the enhancement of the deficiencies of them, which are identified in section "PROS and CONS of ADEVENT and ADDATES from CDSIC Standards".

Order in Table 5	Feature/Benefit	ADEVENT Deficiency	Enhancement from This Paper
3	Support BOR and DOR when the confirmation of CR and PR is required	No	Create a new record with PARAMCD='TP_BOR' PARAM='Derived Overall Evaluation', please refer to [8] for the details
5	Traceability of the Derivation of the Dates Related Tumor Assessments	Not Sufficient (RECIST 1.1/ iRECIST)	Same as above, and add ANLxxFL flags providing more traceability, along with ASEQ
6	The meaning of the event (ADEVENT.AVALC)	AVALC does not store the description of event.	The meaning/description of the event is attached to the "original" AVALC, and the numbering (1-8) is prefixed to it for the sorting key, refer to Table 7
8	Sorting Keys to Sort the Data	No	Same as above, along with ASEQ
7	Best Programming Practice	Does not follow	Separate the derivation of dates related to tumor assessments and dates of independent of tumor assessments into two independent programming

Table 25. The Enhancement of CONS of ADEVENT from CDSIC TAUG-BrCa

Order in Table 5	Feature/Benefit	ADDATES Deficiency	Enhancement from This Paper
2	Support BOR and DOR when no confirmation of CR and PR is required	No/Indirectly	Following ADEVENT, ADRESP supports BOR, and ANL09FL and ANL10FL facilitate the DOR derivation in ADTTE.
3	Support BOR and DOR when the confirmation of CR and PR is required	No	Same as one for ADEVENT enhancement
5	Traceability of the Derivation of the Dates Related Tumor Assessments	"Totally Lost" from both the dataset and metadata	Being Built in ADEVENT through ANLxxFL flags, along with the triplet of SRCDOM, SRCVAR, and SRCSEQ
7	Sorting Keys to Sort the Data	No	Adding a new variable: ADTDESCN (Description of Analysis Date (N)) as the sorting key, refer to Table 16
8	Best Programming Practice	Does not follow	Same as one for ADEVENT enhancement

Table 26. The Enhancement of CONS of ADDATES from CDSIC TAUG-PrCa

It is worth mentioning that these enhancements include new and much clearer traceability for ADTTE through data flow: ADEVENT, ADDATES, and ADTTE, and analysis flags: ANL02FL, ANL05FL, ANL06, and ANL07FL pointing the interested SDTM RS records and their respective event dates for events (dates) related to tumor assessments, and ADDATES directly to ADTTE for events (dates) independent of tumor assessments from ADSL, which can be further traced back to respective SDTM domains, e.g., EX, DS, CM, and PR. Figure 4 and Figure 5 illustrate how the users of ADTTE can use either of them to locate the source data of ADTTE. It further shows that this new process to ensure "***Traceability – The property that enables the understanding of the data's lineage and/or the relationship between an element and its predecessor(s)***" [6]. [12] will further explain the new and much clearer traceability for ADTTE programming, which follows the process presented in this paper.

This new process streamlines the generation of ADaM datasets: **ADRESP**, and **ADTTE** for both categorical analysis of tumor response and a TTE analyses and follows the best programming practice.

CONCLUSION

This paper introduces two intermediate datasets: ADEVENT and ADDATES proposed by two CDISC TAUGs to support traceability. It pinpoints and explains their pros and cons. To leverage their pros and overcome their cons, it presents a new process to enhance them.

Ten (10) analysis flags (ANL01FL-ANL10FL) are proposed to be added to ADEVENT of TAUG-BrCa [1]. A new intermediate variable: ADTDESCN is added to the metadata of ADDATES in TAUG-PrCa [2]. It portrays the critical part of ADTDESCN as a link between ADTTE and ADDATES so that the triplet of SRDDOM, SRCVAR, and SRCSEQ can be built for ADTTE based on the source from ADDATES, and further from ADEVENT and ADSL. It depicts two methods to show how the traceability of ADTTE built in ADTTE can be used to trace back to respective SDTM domains, which provides transparency, and further builds/increases confidence in a result or conclusion for the FDA reviewers.

Three examples in the paper further demonstrate the superior of ten (10) analysis flags (ANL01FL-ANL10FL) to the triplet of SRCDOM, SRCVAR, and SRCSEQ regarding the traceability of the events (dates) in ADEVENT.

The new approach streamlines the programming for the generation of ADEVENT, ADRESP, ADDATES, and ADTTE. The enhancement of CDSIC standards from both ADEVNET and ADDATES make them possible to be applied broadly to other areas of oncology studies, besides Breast Cancer Therapeutic Area and Prostate Cancer Therapeutic Area. The intent of this presentation is to guide readers in developing a CDISC ADaM compliant programming with much clearer traceability that is applicable across multiple projects in oncology studies.

REFERENCES

[1] CDSIC **"**Breast Cancer Therapeutic Area User Guide v1.0" May 2016 https://www.cdisc.org

[2] CDSIC "Prostate Cancer Therapeutic Area User Guide v1.0" July 2017 https://www.cdisc.org/standards/therapeutic-areas/prostate-cancer/prostate-cancer-therapeutic-area-user-guide-v10

[3] "Clinical-Trial-Endpoints-for-the-Approval-of-Non-Small-Cell-Lung-Cancer-Drugs-and-Biologics" April 2015 https://www.fda.gov/media/116860/download

[4] "New response evaluation criteria in solid tumors: Revised RECIST guideline (version 1.1)"

October 2008 https://ctep.cancer.gov/protocolDevelopment/docs/recist_guideline.pdf

[5] "iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics"

March 2017 https://www.thelancet.com/action/showPdf?pii=S1470-2045%2817%2930074-8

[6] CDISC Analysis Data Model Implementation Guide version v1.2 https://www.cdisc.org

[7] FDA "Study Data Technical Conformance Guide", October 2022 https://www.fda.gov/media/153632/download

[8] Xiangchen (Bob) Cui, and Sri Pavan Vemuri, "Simplifying the Derivation of Best Overall Response per RECIST1.1 and iRECIST in Solid Tumor Clinical Studies", PharmaSUG 2020 in May 2020

[9] Wayne Zhong, Richann Watson, Daphne Ewing, and Jasmine Zhang, "More Traceability: Clarity in ADaM Metadata and Beyond", PharmaSUG 2019 in June 2019

[10] FDA Guidance for Industry, Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics, May, 2007 https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-endpoints-approval-cancer-drugs-and-biologics

[11] CDISC ADaM Basic Data Structure for Time-to-Event Analysis Version 1.0 http://www.cdisc.org

[12] A New Technique to Assemble ADTTE (Data for the Time to Event Analyses) with Ease and Traceability in Oncology Studies - under preparation

ACKNOWLEDGMENTS

Appreciation goes to PK Morrow and Yanqiong Zhang for their invaluable review and comments.

CONTACT INFORMATION

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

Name: Xiangchen (Bob) Cui, Ph.D.
Enterprise: CRISPR Therapeutics AG
Address: 105 West 1st Street
City, State ZIP: Boston, MA 02127
Work Phone: 908-240-4086
E-mail:xiangchen.cui@crisprtx.com

Appendix 1. Rules to Be Considered for BOR Derivation per RECIST 1.1 [4]

RECIST 1.1 [4] states the following rules (Rule 1-Rule 5), and iRECIST [5] states Rule 6. These rules should be followed and/or considered.

1. "In non-randomized trials where response is the primary endpoint, the confirmation of a complete response (CR) and partial response (PR) is required to ensure response identified are not the result of measurement error".
2. "Complete or partial responses may be claimed only if the criteria for each are met at a subsequent time point as specified in the protocol (generally 4 weeks later)".
3. "In the case of SD, measurements must have met the SD criteria at least once after study entry at a minimum interval (in general not less than 6–8 weeks) that is defined in the study protocol."
4. "Protocols must specify how any new therapy (eg, radiotherapy or surgery) introduced before progression will affect best response designation."
5. "In trials where confirmation of response is required, repeated 'NE' time point assessments may complicate best response determination. The analysis plan for the trial must address how missing data/assessments will be addressed in determination of response and progression."
6. "The best overall response is the best response recorded from the start of treatment until disease progression/recurrence (taking as reference for progressive disease the smallest measurements recorded since the treatment started)."

Appendix 2. ADEVENT Metadata from CDISC Breast Cancer Therapeutic Area User Guide (TAUG-BrCa) [1]

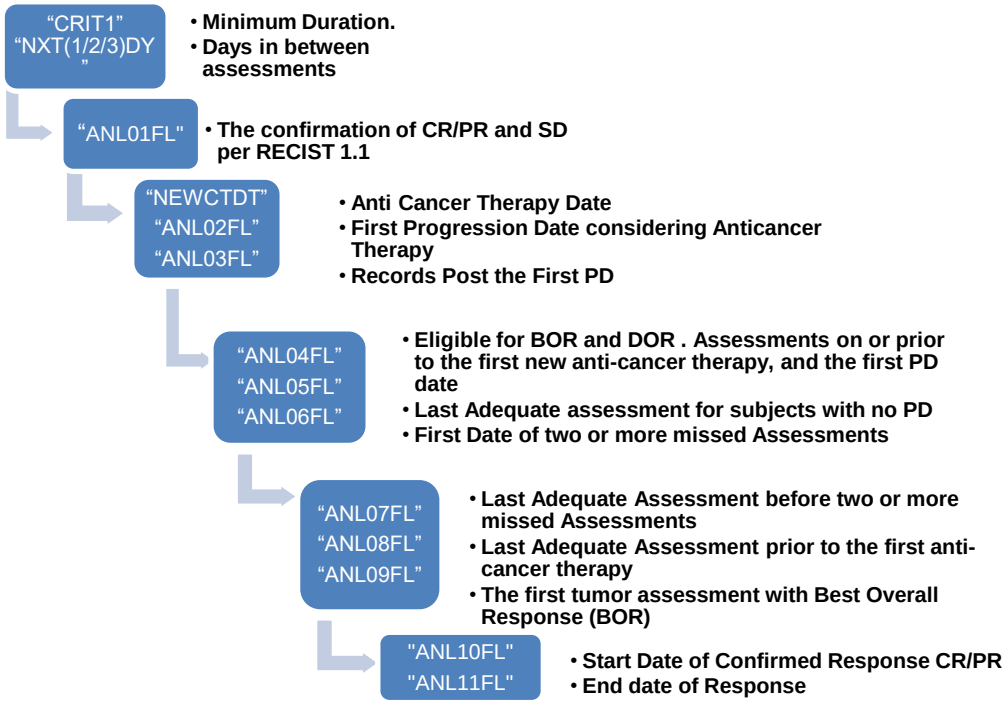
Variable Name	Variable Label	Type	Codelist/Controlled Terms	Source/Derivation/Comment
STUDYID	Study Identifier	Char		DM.STUDYID
USUBJID	Unique Subject Identifier	Char		DM.USUBJID
ASEQ	Analysis Sequence Number	Num		Sequential number for associating a record number in the ADEVENT dataset. A unique number per subject, per parameter, per parameter qualifier, per analysis start date.
ASTDT	Analysis Start Date	Num		The date that the event occurred is the corresponding --DTC variable for each PARAMCD converted to numeric date format. RS.RSDTC when PARAMCD = 'ASSESS' DS.DSSTDTC when PARAMCD = 'DISPOSIT' AE.AESTDTC or MH.MHSTDTC or DV.DVSTDTC or CM.CMSTDTC or PR.PRSTDTC or some other source data for an event which prevents further assessments when PARAMCD = 'EVENT'.
ASTDY	Analysis Start Relative Day			The number of days from randomization to the date of the reported event. ASTDT - ADSL.RANDDT + 1
PARQUAL	Parameter Qualifier	Char	INVESTIGATOR; CENTRAL	INVESTIGATOR for investigator-based tumor response assessments. CENTRAL for central imaging tumor response assessments. Otherwise set to null.
PARAM	Parameter	Char	ASSESSMENT; DISPOSITION; EVENT	These are the different categories of events that can occur during the execution of the study. ASSESSMENT: The RECIST assessments typically collected from the RS domain. DISPOSITION: These are dispositions collected during the study. Typically expected would be the date randomized, date treatment ended, and date withdrew from study. EVENT: These are events that occur during the conduct of a clinical trial. In some cases, they could be protocol violations or events that prevent further assessments from being made.
PARAMCD	Parameter Code	Char	ASSESS; DISPOSIT; EVENT	If RECIST assessment, then PARAMCD = 'ASSESS' If disposition event, then PARAMCD = 'DISPOSIT' If event that is a protocol violation or prevents further assessments, then PARAMCD = 'EVENT'

Variable Name	Variable Label	Type	Codelist/Controlled Terms	Source/Derivation/Comment
AVALC	Analysis Value (C)	Char		Reported Assessment associated with the ASTDT.
SRCDOM	Source Domain	Char		This is the source SDTM domain or ADaM data set to which the record being used for the analysis value can be traced.
SRCVAR	Source Variable	Char		This is the variable in the source SDTM domain or ADaM data set to which the analysis value can be traced.
SRCSEQ	Source Sequence	Num		This is the sequence number -- SEQ or ASEQ of the row in the domain identified in the SRCDOM that relates to the analysis value being derived.
ANL01FL	Analysis Flag 01	Char	Y	Identifies whether the event can be used in time-to-event analysis. If assessment is prior to baseline or after a censoring event, then they are not included.

Appendix 3. ADDATES Metadata from CDISC TAUG-PrCa [2]

Variable Name	Variable Label	Type	Codelist/Controlled Terms	Source/Derivation/Comment
STUDYID	Study Identifier	text		DM.STUDYID
USUBJID	Unique Subject Identifier	text		DM.USUBJID
ASEQ	Analysis Sequence Number	integer		Sequential number for associating a record number in the ADDATES dataset.
ADT	Analysis Date	integer		This is the date that the event occurred.
ADTDESC	Description of Analysis Date	text	Change in Anti-Cancer Therapy; Date of Analysis Cut-off; Date of Death; Date of Randomization; Date of Toxicity Leading to Discontinuation; Date Last Known Alive; Date Lost to Follow-Up; End of Study Date; Date of Last Tumor Assessment with No PD; Date of Missing Tumor Assessment; Date of Tumor Assessment with PD	This is a text description of the event of interest that occurred on ADT and at study day. This variable is restricted to $40 characters so that the value can be used as a label if the data set was transposed.
ADTDESCD	Description of Analysis Date Code	text	RXCHGDT; CUTOFFDT;DTHDT;LNOPDDT;MISEXDT;RANDDT;TOXICDT;PDDT;LSTALVDT;LOSTFUDT;EOSDT	This is an 8-character code for the date. Restricting this variable to 8 characters will allow this value to be used as a variable name if the data set was transposed.
ADY	Analysis Relative Day	integer		This is the analysis day that is the number of days from a specified anchor dates, such as randomization date, to ADT.
SRCDOM	Source Domain	text		This is the source SDTM domain or ADaM data set to which the record being used for the analysis value can be traced.
SRCVAR	Source Variable	text		This is the variable in the source SDTM domain or ADaM data set to which the analysis value can be traced.
SRCSEQ	Source Sequence	integer		This is the sequence number -- SEQ or ASEQ of the row in the domain identified in the SRCDOM that relates to the analysis value being derived.

Appendix 4 Flowchart depicting the overall logic flow of the programming approach [8]



Appendix 5. ADRESP Metadata from CDISC Breast Cancer Therapeutic Area User Guide (TAUG-BrCa) [1]

Variable Name	Variable Label	Type	Length/Display Format	Controlled Terms or Format	Source/Derivation/Comment
STUDYID	Study Identifier	text	20		ADSL.STUDYID
USUBJID	Unique Subject Identifier	text	40		ADSL.USUBJID
ASEQ	Analysis Sequence Number	integer	8		Derived: Sort by STUDYID, USUBJID, and PARAMN then assign value. Start at 1 for each subject. No duplicates allowed within a subject.
PARAMCD	Parameter Code	text	8	PARAMCD (ADTDESC): (1) BOR=Best Overall Response (2) BORORR=Objective Response Rate (3) BORDCR=Disease Control Rate	Assigned: Refer to the comment for AVALC
PARAM	Parameter	text	40	PARAMN (PARAM): (1) 1=Best Overall Response (2) 2=Objective Response Rate (3) 3=Disease Control Rate	Assigned: PARAM='Best Overall Response' if PARAMCD='BOR'; PARAM=' Objective Response Rate' if PARAMCD=' BORORR'; PARAM='Disease Control Rate' if PARAMCD='BORDCR'

Variable Name	Variable Label	Type	Length/Display Format	Controlled Terms or Format	Source/Derivation/Comment
PARAMN	Parameter (N)	integer	8	PARAMN (PARAMCD): (1) 1=BOR (2) 2=BORORR (3) 3=BORDCR	Assigned: 1, if PARAMCD='BOR'; 2, if PARAMCD='BORORR'; 3, if PARAMCD='BORDCR';
AVALC	Analysis value (c)	float	8	8.1	Derived: Step 1: Select records from ADEVENT with the condition: ANL09FL='Y'; Step 2: update PARAMCD by assigning it to 'BOR', keep AVALC and AVAL, and output the record to the final dataset; Step 3: If AVALC in ('CR','PR') , set PARAMCD='BORORR' and AVALC='Y'; otherwise, AVALC='N'; Step 4: If AVALC in ('CR','PR','SD') , set PARAMCD='BORDCR' and AVALC='Y'; otherwise, AVALC='N'
AVAL	Analysis value (N)	float	8	8.1	Derived: For PARAMCD='BOR' AVAL=ADEVENT.AVAL; Otherwise, AVAL=1 if AVALC='Y'; AVAL=2 if AVALC='N'.
SRCDOM	Source Domain	text	8		Derived: Equal to 'ADEVENT'
SRCVAR	Source Variable	text	12		Derived: Equal to 'AVALC'
SRCSEQ	Source Sequence	integer	8		Derived: Equal to ADEVENT.ASEQ