

ADINTDT and ADTTE for Survival Sweep in Oncology Studies

Wenyu Hu, Merck Research Laboratories, Merck & Co., Inc., Upper Gwynedd, PA USA

Christine Teng, Merck Research Laboratories, Merck & Co., Inc., Rahway, NJ USA

ABSTRACT

Many oncology studies have either a primary or a secondary endpoint in Overall Survival; therefore, the survival data is critical to study success. The event of interest is death from any cause, and subjects without any event are censored to the latest follow-up date or the last known alive date. When the Interim-Analysis database lock for a certain data cut-off date is approaching, the sites must enter the most up-to-date data for those subjects who are still alive. Of interest are those subjects who are either (1) on study treatment but not yet have the next visit planned, or (2) off treatment but on follow-up phase and have yet to reach the planned visit, or (3) discontinued from the study but agreed to be (though have not yet been) contacted by phone for survival status.

This paper utilizes the oncology ADaM datasets, ADINTDT and ADTTE, to create a list of censored subjects. The list provides information such as disposition date and status from treatment or study, source data of last known alive date, and other key study related dates. The list can assist the sites in following up with the currently censored subjects.

Key Words: CDISC, ADaM, Oncology, OS, ADTTE, ADINTDT, Survival analysis, Time-to-event

INTRODUCTION

Overall Survival (OS) is an objective endpoint and it is the gold standard to demonstrate clinical benefit in oncology trials. An important consideration in trial design and analysis when OS endpoint is involved is to minimize loss to follow-up. Every effort should be made to ensure that patients are followed up for OS and long term safety after terminating treatment. It is common to focus on deaths when the number of deaths may be used to trigger the analysis. Emphasis should also be given to the adequacy of follow-up to patients who are censored. It is important to address the gaps in follow-up and censored patients using the data cut-off date, if event date is beyond cut-off date of an interim analysis.

SURVIVAL SWEEP

Survival sweep collects all patient survival status data (alive or dead) in a short defined period of time (i.e. 2 or 3 weeks). The purpose of a sweep is to assess death event rates, more accurately reflect survival status, and reduce censoring on the overall survival curve. This step ensures complete accounting of patient survival data across all arms of the study.

When preparing interim analysis, programming team can help sites identify gaps by generating a list of censored subjects whose survival status (alive or dead) is not known within a maximum of 3 weeks prior to

the database lock data cut-off date, and these subjects will then be contacted by sites for their survival status. While it might seem simple to just capture the randomization date and compare it to the event date, the source of the data can be in multiple pages of CRFs. There are also additional considerations such as whether the event date is beyond the cutoff date for interim analysis. All of these are taken into consideration for the analysis, and in many cases, require the programmer to produce a spreadsheet showing all possible conditions for the team to review. The spreadsheet provided to clinical and data management is very helpful to identify subjects to be contacted during survival sweep. Figure 1 below shows the 2 example subjects needing follow-up during survival sweep. Every effort is made to contact these subjects or their caregivers during survival sweep. If subjects could no longer be reached, public database is checked against survival status.

Figure 1

USUBJID	STARTDT	ADT	AVAL	EVNTDESC	DCSREAS	PMSREAS
1001-0002	2/28/2016	4/14/2016	47	LAST KNOWN ALIVE AT LB		SURVIVAL FOLLOW-UP
1011-0003	6/22/2016	6/22/2016	1	LAST KNOWN ALIVE AT DS	WITHDRAWAL BY SUBJECT	

USUBJID	DSSTDT	RFXENDT	CUTOFFDT	DAYS_FROM_CUTOFF
1001-0002	4/14/2016	4/7/2016	2/15/2017	-307
1011-0003	6/22/2016		2/15/2017	-238

With ADaM time to event ADTTE and intermediate dataset ADINTDT, this paper shows the structure that supports both validation and statistical review of the results.

ADTTE - TIME TO EVENT

We start by examining what needs to be produced to support the OS analysis. Based on the Statistical Analysis Plan (SAP), “Overall Survival (OS) is defined as the time from randomization to death due to any cause. Subjects without documented death at the time of the final analysis will be censored at the date of the last follow-up.” The following is an example of ADaM TTE Analysis Dataset (ADTTE), where OS is the event of interest.

Figure 2

USUBJID	PARMCD	PARAM	STARTDT	ADT	AVAL	CN SR	EVNTDESC
1001-0001	OS	Overall Survival	6/1/2016	2/1/2017	246	1	LAST KNOWN ALIVE AT FA
1001-0002	OS	Overall Survival	2/28/2016	4/14/2016	47	1	LAST KNOWN ALIVE AT LB
1002-0003	OS	Overall Survival	5/25/2016	8/12/2016	80	0	DEATH
1010-0001	OS	Overall Survival	1/5/2016	2/15/2017	408	1	CENSORED AT DATA CUTOFF DATE
1011-0003	OS	Overall Survival	6/22/2016	6/22/2016	1	1	LAST KNOWN ALIVE AT DS
1012-0005	OS	Overall Survival	3/1/2016	2/15/2017	352	1	CENSORED AT DATA CUTOFF DATE

- AVAL is defined as the elapsed time to death if a subject died on or prior to the cutoff date. If a subject's live/death status is known between cutoff date and the actual interim analysis date, then the subject is censored on the cutoff date. If a subject is known to be alive prior to the cutoff date, then the subject is censored at the last known alive date.
- STARTDT is the time to event origin date for subject. STARTDT can be the randomization date, treatment start date, etc.
- ADT is the analysis date for the event or censoring associated with AVAL.
- CNSR indicates an event or censoring. If the subject died on or before cutoff date of 2/15/2017, then CNSR=0. If the subject does not have documented death by cutoff date, then CNSR=1.
- EVNTDESC indicates the death event or the domain which has last known alive date for censoring. It contains texts clarifying the event of interest or reasons for censoring subjects who did not have the event, which will help the team to summarize why the subject was censored.

From the above figure 2, we can see that subject (1002-0003) has an event, and the other four subjects are still alive till the cutoff date and are censored. Two subjects have last follow-up visits before date cutoff, and two subjects have last follow-up visits beyond the cutoff date so ADT are set to cutoff date. To derive ADT above, many dates of interest are used, such as Date of Last Known Alive Prior to Cutoff at domain XX, Date of Last Known Alive at domain XX, Data Cutoff Date, Date of Death, etc. To get the last follow-up date, many domains need to be checked and only the latest date among those domains will be selected.

In ADTTE dataset, we are deriving OS dates as follows:

1. If DTHDT is present then it will be censored as "0"
2. If UNCUTDT (Dates from multiple domains) is after cutoff, it will be censored as "1" (We are using data without applying cutoff date to derive uncut date)

The list of survival subjects will be followed before Interim Analysis are those with CNSR="1".

ADINTDT - INTERMEDIATE DATASET TO SUPPORT TRACEABILITY

A fundamental principle of ADaM is traceability. To assist review, analysis datasets and metadata must clearly communicate how the analysis datasets were created. Traceability is built by clearly establishing the relationship between SDTM and ADaM. The data point traceability is very helpful when a reviewer tries to trace a complex data manipulation path. This traceability is established by providing SRCSEQ, SRCVAR and SRCDOM to the specific data values that are used as input for an analysis value.

The ADT in the ADTTE above involves multiple domains and has very complex derivation rules. An intermediate analysis dataset like ADINTDT which stores dates of interests can provide traceability. It can serve as a link between SDTM and ADTTE. The intermediate dataset clarifies which source dataset is used in ADTTE analysis and why. It also has huge benefits for later review and QC process. The following is an example of ADINTDT dataset which shows the dates of interest that are used to derive overall survival.

Figure 3

USUBJID	PARAM	PARAMCD	ADT	ADTF	SRC DOM	SRC VAR	SRC SEQ
1001-0001	Date of Last Known Alive Prior to Cutoff	LKALDT	2/1/2017	D	FA	FADTC	58
1001-0001	Data Cutoff Date	CUTOFFDT	2/15/2017				
1001-0001	Date of Last Known Alive	LKAL2DT	2/1/2017	D	FA	FADTC	58
1001-0002	Date of Last Known Alive Prior to Cutoff	LKALDT	4/14/2016		LB	LBDMTC	84
1001-0002	Data Cutoff Date	CUTOFFDT	2/15/2017				
1001-0002	Date of Last Known Alive	LKAL2DT	4/14/2016		LB	LBDMTC	84
1002-0003	Date of Death	DTHDT	8/12/2016		DD	DDDMTC	1
1002-0003	Date of Last Known Alive Prior to Cutoff	LKALDT	7/20/2016		LB	LBDMTC	64
1002-0003	Data Cutoff Date	CUTOFFDT	2/15/2017				
1002-0003	Date of Last Known Alive	LKAL2DT	7/20/2016		LB	LBDMTC	64
1010-0001	Date of Last Known Alive Prior to Cutoff	LKALDT	2/9/2017		EX	EXDMTC	90
1010-0001	Data Cutoff Date	CUTOFFDT	2/15/2017				
1010-0001	Date Beyond Cutoff	UNCUTDT	3/6/2017		RS	RSDMTC	21
1010-0001	Date of Last Known Alive	LKAL2DT	2/15/2017				
1011-0003	Date of Last Known Alive Prior to Cutoff	LKALDT	6/22/2016		DS	DSDMTC	2
1011-0003	Data Cutoff Date	CUTOFFDT	2/15/2017				
1011-0003	Date of Last Known Alive	LKAL2DT	6/22/2016		DS	DSDMTC	2
1012-0005	Date of Last Known Alive Prior to Cutoff	LKALDT	2/10/2017		FA	FADTC	4
1012-0005	Data Cutoff Date	CUTOFFDT	2/15/2017				
1012-0005	Date Beyond Cutoff	UNCUTDT	2/23/2017		DD	DDDMTC	2
1012-0005	Date of Last Known Alive	LKAL2DT	2/15/2017				

- PARAM describes the dates of interest used to derive OS in ADTTE.
 - Date of Last Known Alive Prior to Cutoff captures the latest known alive date from exposure, subject visits, tumor results, disease response, laboratory assessment, vital sign assessments, survival follow-up alive status, disposition, etc.
 - Date of Death contains dates from death form or survival follow-up form with death status.
 - Data Cutoff Date is the cutoff date for the interim analysis.
 - Date Beyond Cutoff only appears if subjects have data available after cutoff date, besides the domains Last Known Alive Prior to Cutoff checks, it also adds death form or survival follow-up form with death status to see whether any dates beyond cutoff date
 - Date of Last known Alive contains latest alive date from all sources. If the latest date is greater than the cutoff date, then it is set to the cutoff date.
- PARAMCD is the short name of the analysis parameter in PARAM, max of 8 characters and should be as intuitive as possible to represent the one-to-one mapping with PARAM.
- ADT contains numeric dates of interest for the OS analysis.

- ADTF is used to indicate the level of imputation for the analysis date. ‘D’ indicates that “day” is imputed. ‘M’ indicates that month/day is imputed. ‘Y’ indicates that year/month/day is imputed. If no imputations were done, then ADTF is null.
- SRCDOM indicates the SDTM domain that relates to ADT. SRCVAR indicates the name of the column that relates to ADT. SRCSEQ indicates the sequence number of the variable that relates to ADT. These traceability variables greatly improve the review process.

PROCESS FOR CREATING ADINTDT AND ADTTE

Step 1: Create ADINTDT

- Create two separate folders; a new one to keep SDTM data with cutoff applied, and the original without cutoff during interim analysis.
- Pre-process all source datasets and perform imputation if there is any partial date.
- Create the five parameters described above.
 - Date of Last Known Alive Prior to Cutoff is obtained from the folder with cutoff applied. Date of interest which indicate subject still alive are set together and the latest date is set to PARAMCD=' LKALDT' with corresponding SRCDOM, SRCVAR and SRCSEQ. All randomized subjects have this parameter in the intermediate dataset.
 - Even though Data Cutoff Date is the same for all subjects, still keep it in the intermediate dataset so that the cutoff date applied is clear to the reviewer.
 - Date of Death is obtained from the folder with cutoff applied and only kept for subjects who have events.
 - Date Beyond Cutoff is obtained from the folder without cutoff applied. Date of interest with both alive and death date are set together. And if the latest date is later than cutoff date, it is set to PARAMCD=' UNCUTDT'. Only subjects having date beyond cutoff shall have this parameter created.
 - Date of Last known Alive is derived based on Date Beyond Cutoff and Date of Last Known Alive Prior to Cutoff. If subject has non-missing Date Beyond Cutoff, then Date of Last known Alive is set to Cutoff Date, otherwise, it is set to Date of Last Known Alive Prior to Cutoff. All randomized subjects have this parameter in the intermediate dataset.
- Create the permanent intermediate dataset that sets the five parameters together.

Step 2: Create ADTTE

- Transpose ADINTDT
- Derive OS using the 5 parameters created in above ADTTE and apply censoring rule. If Date of Death is not missing, then the subject has event and ADT=DTHDT (event) as subject 1002-0003 shows. If Date Beyond Cutoff is not missing (subject is known alive or dead between cutoff and actual analysis, then the subject is censored on the cutoff date ADT= CUTOFFDT as subject 1010-0001 and 1012-0005 show above. All other subjects are censored at their last know alive date ADT= LKAL2DT.

CONCLUSION

Creating time-to-event datasets may seem quite daunting at first, especially when so many dates are involved to derive the OS endpoint. The intermediate dataset ADINTDT can provide a clear picture of which dates are being used for analysis and how ADTTE is derived. It can help statisticians and reviewers to see how the dates fall chronologically and relate to each another. This paper lays out the framework for basic TTE analyses for OS. It can be further expanded to more complex analyses, e.g, single event with multiple values for the censoring variable, composite event which use other ADaM datasets as input when building ADINTDT.

REFERENCES

- CDISC ADaM Basic Data Structure for Time-to-Event Analysis Version 1.0
<http://www.cdisc.org/adam>
- Analysis Data Model Implementation Guide 1.0
<http://www.cdisc.org/adam>
- Tom Santopoli, Kim Minkalis, Sandra Minjoe (2014) An ADaM Interim Dataset for Time-to-Event Analysis Needs
<https://www.lexjansen.com/pharmasug/2014/DS/PharmaSUG-2014-DS09.pdf>

ACKNOWLEDGEMENTS

The authors would like to thank their management for their support and encouragement. We also like to thank Merck oncology standard team for their ideas of apply the intermediate dataset for ADTTE analysis, and we have enjoyed many benefit of this great concept.

CONTACT INFORMATION

Wenyu Hu Merck Research Laboratories, Merck & Co., Inc., Upper Gwynedd, PA 19454 wenyu_hu@merck.com	
Christine Teng Merck Research Laboratories, Merck & Co., Inc. Rahway, NJ 07065 christine_teng@merck.com	