

Multiple Applications of ADaM Time-to-Event Datasets

Huei-Ling Chen, Merck Sharp & Dohme Corp., Whitehouse Station, NJ USA¹
Helen Wang, Merck Sharp & Dohme Corp., Whitehouse Station, NJ USA¹

ABSTRACT

The CDISC ADaM team recently created guidelines for Time to Event (TTE) datasets. In the guidelines, a TTE dataset is specifically used to support survival analysis. This paper demonstrates that the TTE dataset can be used to support Exposure-Adjusted Incidence Rate (EAIR) analysis. This exploratory use of the ADaM TTE datasets structure filled a void in the CDISC ADaM analysis dataset structure for EAIR analysis. Detailed procedure and code for generating an EAIR analysis dataset will be presented. In addition, step-by-step instructions for generating a TTE analysis dataset and survival analysis result will also be covered in this paper.

KEYWORDS

AE, EXPOSURE-ADJUSTED INCIDENCE RATE (EAIR), ADAM, SDTM, TIME TO EVENT (TTE), SURVIVAL ANALYSIS

INTRODUCTION

Time to Event (TTE) has been an important analysis endpoint in the clinical trial. For example, time to discharge from hospital is an efficacy measurement for the anesthesia drug injected during the surgery. Time to rescue medication taken is considered a way to measure the therapy drug competence. When there is an event, the duration time from study start to the time the occurred is of study interest. When there is no event, the duration time will be from study start to the last available time information is available for. With an event, the duration is not censored. Without an event, the duration is censored.

The analysis that pertains to a TTE dataset is called survival analysis. The CDISC ADaM team has published a guideline for the TTE dataset. The guideline introduces what variables are required, conditionally required, or permissible for a TTE dataset. In addition, the guideline also has example code for the survival analysis.

The common interest on adverse events (AE) has been the crude rate of AE incidence. This crude rate is a direct way to capture the proportion of subjects who have at least one AE while taking the treatment drug. Compared to the crude rate, the Exposure-Adjusted Incidence Rate (EAIR) is a more appropriate incidence rate as it takes study drug exposure time into account. Instead of calculating the rate for the number of subjects with at least one adverse event over the number of subjects in the population, EAIR is the incidence rate over patient-years.

Since the calculation of EAIR is related to the time to event and also related to whether the event happened or not, the Time to Event dataset is conveniently prepared for EAIR analysis.

CREATE TTE DATASET

The TTE dataset is a special case of Basic Data Structure (BDS) dataset. Sharing the same characteristics of the BDS dataset, the TTE dataset should have two important features:

- Traceability
- Be Analysis-ready

Procedure to Generate a TTE Dataset from SDTM Dataset

Step 1: Convert the SDTM character date variable to numeric date with ISO8601 format

In the TTE dataset, the Date/Time variable is very important. The first step is to identify the relevant Date/Time variables and check if the Date/Time variable is a character variable. If it is a character date variable, it should be converted to a numeric date variable. For example, the RFSTDTC in the SDTM DM domain is a character variable. When the ADaM TTE STARTDT variable is RFSTDTC, the following code is used to convert RFSTDTC to a numeric date variable:

```
STARTDT = INPUT(RFSTDTC, IS8601DA.);  
ADT      = INPUT(AESTDTC, IS8601DA.);
```

Step 2: Pick the first eligible event

Assuming the anemia event is the analysis endpoint, the first anemia event that occurs during the treatment epoch for a given subject will be selected, as the following code illustrates.

```
PROC SORT DATA= AE(KEEP=SUBJID AEDECOD ADT AESEQ EPOCH )
      OUT=AE
      ;
      BY SUBJID ADT;
      WHERE EPOCH = 'TREATMENT' AND AEDECOD = 'ANEMIA' ;
      RUN;

DATA AE1;
      SET AE;
      BY SUBJID ADT;
      IF FIRST.SUBJID;
      RUN;
```

Step 3: Create key variables: AVAL, CNSR, EVNTDESC

In the ADaM guideline, the variables AVAL and CNSR are the variables time to event, and the censor variable, respectively. These two variables are the key analysis variables. When there is an event, AVAL is the duration from study start date to the first event date. The value of CNSR is set to 0 which means the time to event for this subject is not censored as the time to event for this subject is known exactly. When the subject does not have an event in the study, AVAL is the duration from study start date to his/her individual study end date, and usually it is the treatment end date. The value of CNSR is set to 1 which means the time to event for this subject is censored because it is an unknown, and is only known to exceed a certain value (from study start to study end) if any.

The example code below shows the logic to derive the two variables AVAL and CNSR.

```
IF AESTDTC NE '' THEN DO;
      ADT = INPUT(AESTDTC ,IS8601DA.);
      CNSR = 0;
      EVNTDESC = 'AE';
END;
ELSE DO;
      ADT = INPUT(RFENDTC ,IS8601DA.);
      CNSR = 1;
      IF STATUS = 'COMPLETED' THEN EVNTDESC = 'COMPLETED THE STUDY';
      ELSE EVNTDESC = 'DISCONTINUED';
END;

AVAL = ADT - STARTDT + 1;
```

Step 4: Keep essential variables to ensure traceability.

In order to ensure traceability, it is recommended that some SDTM variables be retained in the ADaM dataset. For example, the sequence number variable. In example 1 below, the AE domain is the event-feeding dataset, hence, AESEQ is retained. In example 2 below, the HO domain is the event feeding dataset, hence, HOSEQ is retained.

The two examples below are the analysis-ready ADaM datasets. The data structure is one record per event per subject.

Example 1 is the time to a hypoglycemia event. When no event happened and subjects are discontinued or completed, the variable CNSR is set to 1. When an event happens, the value of variable CNSR is set to 0. With the AE dataset, the AESEQ value is retained to enable traceability.

Example 1

SUBJID	PARAM	PARAMCD	STARTDT	ADT	AVAL	CNSR	EVNTDESC	ASEQ
1234	HYPOGLYCEMIA	HYPO	2008-07-16	2008-07-29	14	1	DISCONTINUED	
1235	HYPOGLYCEMIA	HYPO	2008-01-08	2008-04-06	90	1	COMPLETED	
1236	HYPOGLYCEMIA	HYPO	2007-12-20	2008-02-19	62	0	AE	1
1237	HYPOGLYCEMIA	HYPO	2008-01-09	2008-02-29	52	0	AE	2

Example 2 is the time to discharge from a hospital. The time is calculated with hours as the unit. In this data, the start point (STARTDTM) is the hospitalization start date and time. The event point (ADTM) is the discharge date and time. Both STARTDTM and ADTM are collected from the HO domain. Assume that in this particular study, the study design has presumed the maximum hospital stay is 120 hours. When the value of ADTM is missing, the value of AVAL is set to 120 hours and the value of the variable CNSR is set to 1. When a subject stays in the hospital for more than 120 hours, AVAL is still set to 120 hours and CNSR is set to 1 as well. In this example, the input data is stored in the HO domain so the HOSEQ value is kept for the purpose of traceability.

Example 2

SUBJID	PARAM	STARTDTM	ADTM	AVAL	CNSR	EVNTDESC	HOSEQ
11111	Time to Discharge	2009-05-15T21:27:00	2009-05-17T21:00:00	48	0	Discharge	1
11113	Time to Discharge	2009-07-13T21:27:00		120	1	No Event Info	2
11114	Time to Discharge	2009-06-16T10:20:00	2009-06-21T11:00:00	120	1	Event Over 120 Hours	3

APPLICATION: EXPOSED ADJUSTED INCIDENCE RATE (EAIR)

Patient-years are calculated per the following rule:

- For subjects with at least one adverse event, the patient-year is the duration from first day study medication is taken till the first adverse event. Duration uses year as unit.
- For subjects without an adverse event, the patient-year is the treatment duration with year as unit.

Summarize all the patient-years as the denominator. The EAIR is the number of subjects with at least one AE incidence divided by total patient-years. Subjects who are exposed to study medication for a longer period of time tend to have a higher incidence of AEs. The EAIR adjusts the number of incidences by the length of exposure to study medication. Therefore, the EAIR is a more appropriate measure compared to the crude incidence rate to account for variations in exposure time among subjects.

Procedure to Generate EAIR from TTE Dataset

Step 1: Calculate the total number of subjects, number of subjects with no AEs, and patient-exposure-days

Refer to Example 1; the value of CNSR is 0 when there is an AE and 1 when there is not an AE. The PROC MEANS statement is used to summarize AVAL (Time to AE) as patient-exposure-days (SUMEXPO). The number of subjects in the study is the variable TOTPAT. The total number of censored events is the variable COUNTS. The following code illustrates Step 1:

```
PROC MEANS DATA= ADTTE NOPRINT NWAY;  
  CLASS TRTPN;  
  VAR CNSR AVAL;  
  OUTPUT OUT=EAIR (DROP=_TYPE_ _FREQ_)  
         N(CNSR)=TOTPAT  
         SUM(CNSR AVAL)=COUNTS SUMEXPO;  
RUN;
```

Step 2: Calculate the number of subjects with at least one AE and patient-exposure-years

Note that the variable COUNTS translates into the number of non-AEs. The number of subjects with an AE is TOTPAT minus COUNTS. Patient-exposure-years is obtained from patient-exposure-days divided by 365.25.

```
DATA EAIR1;
  SET EAIR;
  N = TOTPAT - COUNTS;
  TOTEXPO = SUMEXPO / 365.25;
  KEEP TRTPN TOTPAT TOTEXPO SUMEXPO N ;
  LABEL TOTEXPO = "SUM OF EXPOSURE IN YEARS"
        N       = "NUMBER OF SUBJECTS WITH AE"
        TOTPAT  = "NUMBER OF SUBJECTS IN THE POPULATION"
        ;
RUN;
```

Step 3: statistical methods to assess the EAIR among treatment groups

After EAIR for each treatment group is derived, various statistical methods can be used to assess the EAIR among different treatment groups. These statistical methods could be normal approximation, Miettinen and Nurminen (M&N) method, or the Mantel-Haenszel method (Cao et al., 2011).

APPLICATION: SURVIVAL ANALYSIS

In clinical trials, the study interest is the incidence rate of an event. To present the idea of this study interest, the survival rate, which is the reverse of the event rate is analyzed. At time 0 (the beginning of the study), the survival rate is 1. As time elapses, events happen, and the survival rate decreases gradually. Notice that the survival rate is not simply one minus the event rate. Some subjects might be lost to follow up thus unknown whether he/she has an event or not. The survival rate should take this unknown situation into account. In the clinical trial, whether a patient completed the study or was lost to follow up, the time to the last data available date is stored in the variable AVAL. When no event happens and the time to last data available date recorded, the record is considered "CENSORED." The variable CNSR in the ADaM TTE dataset captures the information of "CENSORED."

The commonly used method to estimate the cumulative incidence (survival function) is the Kaplan Meier Estimator (K-M). K-M is based on an idea that the time interval is a very small unit, and hence the survival curve is more close to the actual value.

PROC LIFETEST

The following code creates the Kaplan Meier Estimation for the survival curve. The clinical trial focuses on the performance of different treatment groups. The STRATA statement makes the estimation by TRTPN which represents different treatment groups.

```
ODS OUTPUT PRODUCTLIMITESTIMATES = KMESTIMATION;

PROC LIFETEST DATA = ADTTE METHOD = KM OUTSURV = OUTSURV;
  TIME AVAL*CNSR(1);
  STRATA TRTPN;
RUN;

ODS OUTPUT CLOSE;
```

The LOG-RANK TEST is the most commonly used test statistic for comparing different survival curves. The following code ODS OUTPUT HOMTESTS extracts the LOG-RANK test result to data set OUT2:

```
ODS OUTPUT HOMTESTS=OUT2;
```

```
PROC LIFETEST DATA = ADTTE METHOD = KM;
    TIME AVAL*CNSR(1);
    STRATA TRTPN;
QUIT;

ODS OUTPUT CLOSE;

DATA OUT2;
    SET OUT2;
    IF TEST='LOG-RANK';
RUN;
```

PROC PHREG

Factors may occur that are related to survival and there is a need to control these factors or covariates while estimating the difference of the survival curve. A commonly used model is the Cox proportional hazard regression model. The PROC PHREG statement presented below performs the estimation:

```
ODS OUTPUT PARAMETERESTIMATES = ESTIMATE ;

PROC PHREG data = ADTTE;
    MODEL AVAL*CNSR(1) = TRTPN/RISKLIMITS;
    RUN;

ODS OUTPUT CLOSE;
```

CONCLUSION

The standardized data structure designed by the CDISC ADaM team has helped the regulatory agency and sponsor to facilitate the data review process and prepare analysis datasets for submission. The TTE dataset guideline provides clear instruction, data examples, and sample code for survival analysis. In addition to survival analysis, the code demonstrates that the TTE dataset is also a great analysis-ready dataset when conducting an AE Exposure-Adjusted Incidence Rate (EAIR) analysis.

REFERENCES

- Belle, G., Fisher, L., Heagerty, P., and Lumley, T. (2004) Biostatistics: A Methodology for the Health Sciences, 2nd Edition, Wiley, New York.
- Cao, D. and He, X. (2011) Statistical Analysis of Adverse Events in Randomized Clinical Trials Using SAS, PharmaSUG 2011, SP07.
- Brucken, N., Minjoe, S., and Widel, M. (2011) Using the ADaM Basic Data Structure for Survival Analysis, PharmaSUG 2011, TU09.
- The ADaM Basic Data Structure for Time-to-Event Analyses, Version 1.0, January 5, 2011
- Analysis Data Model Version 2.1 Final, Dec 17, 2009 (CDISC website: <http://www.cdisc.org/adam>)
- ADaM Implementation Guide, Version 1.0 (ADaMIG v1.0), Dec 17, 2009 (CDISC website: <http://www.cdisc.org/adam>)

FOOTNOTE

¹ working under a contract with Accenture, LLP, Chicago, IL

ACKNOWLEDGMENTS

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

The authors would like to thank their management team for their advice on this paper/presentation.

Contact Information

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

Huei-Ling Chen
c/o Merck Sharp & Dohme Corp.
126 Lincoln Avenue
P.O. Box 2000
Rahway, NJ 07065
Phone: 732-594-2249
e-mail: Huei-Ling_Chen@merck.com

Helen Wang
c/o Merck Sharp & Dohme Corp.
126 Lincoln Avenue
P.O. Box 2000
Rahway, NJ 07065
Phone: 732-594-1296
e-mail: Helen_Wang@merck.com