The dataset generation for survival analysis with the ADaM Basic Data Structure for Time-to-Event Analyses (ADTTE) standard

Chengxin Li

Boehringer Ingelheim Pharmaceuticals Inc., Ridgefield, CT, USA

This paper describes the method of survival and time to event analyses dataset generation with ADaM Basic Data Structure for Time-to-Event Analyses (ADTTE) standard with one clinical trial case study. The paper uses the simplest ADTTE model — single event with binary values for censoring variable among the three models. In this clinical trial case, the hierarchy structure is adopted to derive ADTTE analysis dataset transitioning among continuous analysis values and binary analysis values. To develop ADTTE dataset, in addition to following ADaM general rules, the key points are how to define STARTDT (Time to Event Origin Date for Subject) and ADT (Analysis Date) for study events, right censored, and other competing events for each parameter.

Keywords: Dataset, Survival analysis, Time-to-event, ADTTE, ADaM

Introduction

In the clinical trial, to build up time-to-event analysis dataset, the following concepts need to be clear:

- defined events;
- study periods, including the starting point (e.g. the date of randomization or of medical intervention) and the terminal endpoint (e.g. the end date of treatment or end date of observation);
- the time to occurrence of defined events within the study periods;
- the censored event within the study periods, including the events not occurring or discontinuation or occurrence of other specific events. The censored event is referred to right censored.

The ADaM Working Group has specifically defined the analysis data model for time to event — the ADaM Basic Data Structure for Time-to-Event Analyses (ADTTE).[1] ADTTE can be treated as ADaM basic data structure (BDS)[2] plus additional time-to-event variables and requires time-to-event data specified separately from other ADaM BDS datasets. ADTTE defines three models compliant with the ADaM BDS:

1. single event with binary values for censoring variable;
2. single events with multiple values for censoring variable;
3. composite event.

This paper will depict how to develop ADTTE dataset based on the first model of ADTTE standard with the example of a clinical trial. In the case of clinical trial, the primary and secondary endpoints for efficacy analyses are binary events-based.

Correspondence to: CX Li, Boehringer Ingelheim Pharmaceuticals Inc., Ridgefield, CT, USA. Email: chengxin.li@boehringer-ingelheim.com

Dataset Generation

Figure 1 illustrates the general design diagram for ADTTE dataset. The SDTM domains and ADaM datasets, e.g. the Subject-Level Analysis Dataset (ADSL), can be the input data to derive the ADTTE dataset. The ADaM standard (as the general rules), ADTTE standard, and the trial specifications are the documents that the ADTTE dataset should comply with.

Documents

Statistical Analysis Plan, including Table of Contents and Display Templates, is key input to determine what ADaM datasets are required, and what variables are needed in an ADaM dataset. Project level document is another important input, which may result in the dataset containing more variables than needed for one trial, but required for another trial in the same project. For an ADTTE dataset, those documents further determine the parameters and other core or covariate variables in the dataset.

General Rules

Since ADTTE dataset is defined with ADaM BDS structure, to specify ADTTE datasets, the ADaM general rules that the ADTTE dataset should comply with are:

- using ADaM standard variables if already defined in ADaM IG;

Published by Maney
DOI 10.1179/1757092112Z.0000000001

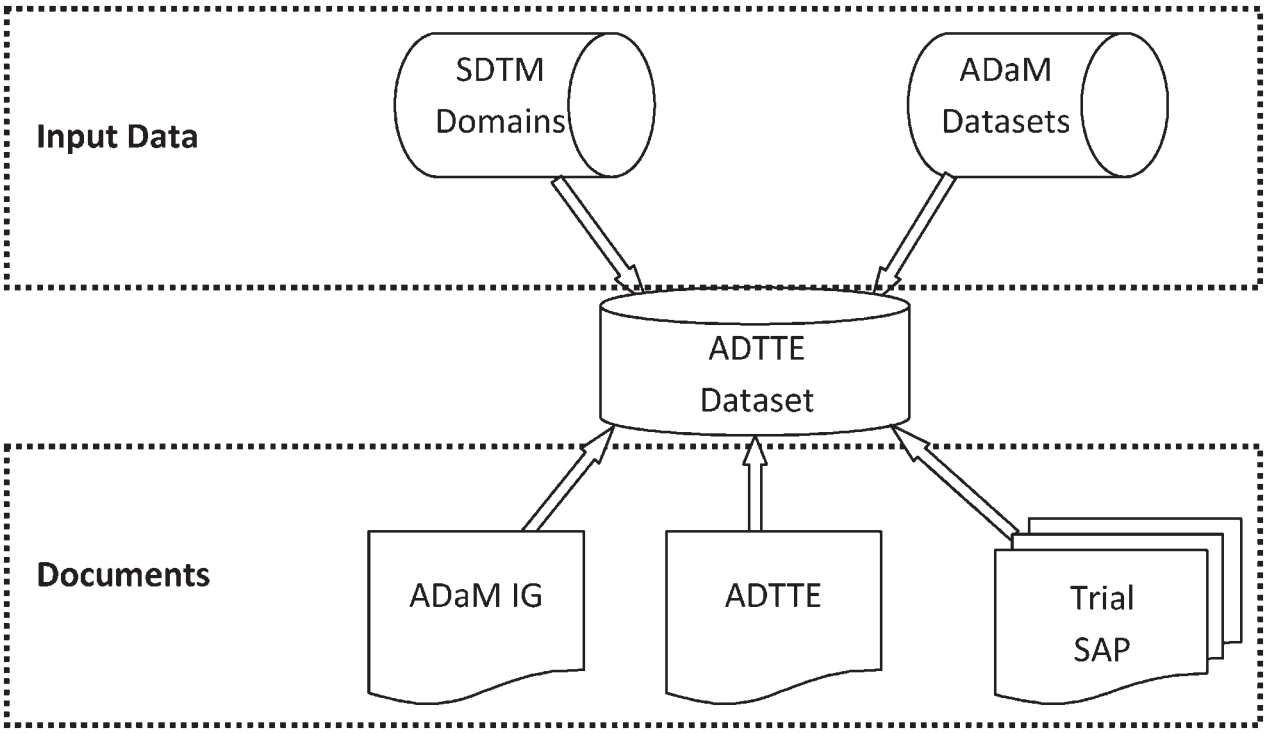

Figure 1 Design diagram of ADTTE dataset

- following naming conventions, e.g. harmonization principle, SAS V5 transport file naming, and labeling conventions, etc.;
- analysis ready, containing all the variables needed for the specific analysis without further data manipulation;
- traceability to facilitate reviews (including as much supportive SDTM variables from SDTM as needed, including SRCDOM, SRCVAR, and SRCSEQ to support data point traceability whenever practical).

Input Data

ADSL, generated directly from SDTM domains, contains subject demographics, derived population flags, treatment information (planned and actual), trial dates, etc. The core variables in ADSL shall be included in the ADTTE dataset as needed. ADTTE dataset can also carry over other required information from ADaM BDS dataset and SDTM domains.

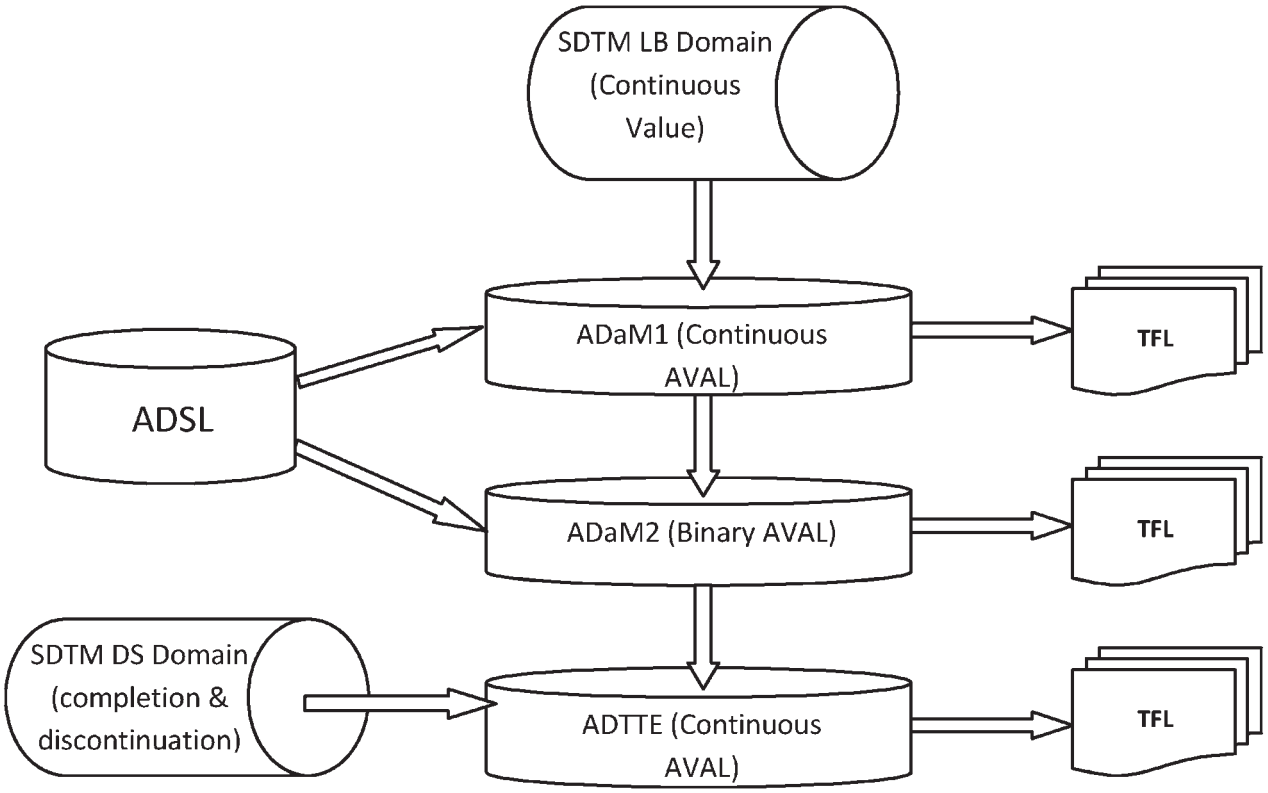

Figure 2 Process flow of ADTTE dataset

Process Flow

In the case study, the analysis data are originated from LB domain LBSTRESN (Numeric Result/Finding in Standard Units), which is a continuous variable. However, the endpoint analyses are based on binary events. Therefore, for simplicity, the hierarchy design structure is adopted to derive ADaM analysis datasets transitioning from continuous analysis values to binary analysis values, and then to continuous analysis values (Fig. 2).

Figure 2 depicts the process flow to derive ADTTE dataset, and also reflects the architecture of dataset generations for efficacy analysis in the case study. The first analysis dataset ADaM1, derived from LB SDTM domain and ADSL and one record per visit per parameter per subject, is a BDS dataset with continuous type AVAL. Time windowing, imputation for missing data, and treatment phase derivation are applied when deriving this dataset. SRCDOM, SRCVAR, and SRCSEQ variables are added for traceability.

The second analysis dataset, based on the first dataset and ADSL, is also a BDS dataset, but containing binary AVAL events, one record per parameter and per subject.

The third analysis dataset is an ADTTE dataset, smoothly derived from the second ADaM dataset for the interested events and the associated date. The disposition information from SDTM DS domain (e.g. the subject's completion status and reason for discontinuation) can be merged for censored events and censoring descriptions of ADTTE dataset. The ADTTE dataset contains continuous type AVAL recording elapsed days, one record per parameter and per subject.

The three datasets have the corresponding tables, figures, and listings to support respectively.

Design

The ADTTE keeps ADaM BDS structure along with ADaM TTE-related variables: STARTDT (Time to Event Origin Date for Subject), CNSR (Censor), and EVNTDESC (Event or Censoring Description). The ADTTE.AVAL is defined explicitly as the elapsed time to the event of interest from the origin date.

Under the process flow described in Fig. 2, it becomes simple to develop ADTTE dataset:

- study identifier variables, subject demographic variables, population indicator variables, and treatment variables can be directly carried over from the second ADaM dataset or from ADSL;
- the EVNTDESC (Event or Censoring Description) is derived from the corresponding disposition information from DS;
- the logic design for analysis parameter variables, timing variables, and censoring variables are defined in Table 1;
- the generated physical dataset from the above logical designs is listed in Table 2 (the data listed here are simulated data).

Table 1 The logic design of ADTTE dataset

PARAM	PARAMCD	PARAMN	STARTDT	ADT	AVAL	CNSR	EVNTDESC
Time to Event_1 (days)	Event_1	10	RANDDT (Date of Randomization)	Event date or censored date ADSL.TRTEDT	ADT-STARTDT + 1	0 if event; 1 if censored	DSDECOD
Time to Event_2 (days)	Event_2	20	(Event_1 at EOT) TRTEDT (Date of Last Exposure to Treatment)	Event date or censored date as ADSL.RFPENDT (Date of End of Participation)	ADT-STARTDT + 1	0 if event; 1 if censored	DSDECOD
			(not Event_1 at EOT) Missing	Missing	Missing	missing	Not Applied
Time to Event_3 (days)	Event_3	30	RANDDT (Date of Randomization)	Event date or censored date as ADSL.TRTEDT	ADT-STARTDT + 1	0 if event; 1 if censored	DSDECOD

Table 2 The physical design of ADTTE dataset

ROW	PARAM	PARAMCD	PARAMN	AVAL	STARTDT	ADT	CNSR	EVNTDESC
1	Time to Event_1 (days)	Event_1	10	57	11/2/2009	12/28/2009	0	Event_1
2	Time to Event_1 (days)	Event_1	10	192	11/7/2009	5/17/2010	1	Completed
3	Time to Event_1 (days)	Event_1	10	29	11/20/2009	12/18/2009	0	Event_1
4	Time to Event_1 (days)	Event_1	10	92	10/18/2009	1/17/2010	1	Lost to follow-up
5	Time to Event_2 (days)	Event_2	20	186	4/23/2010	10/25/2010	1	Completed
6	Time to Event_2 (days)	Event_2	20					Not applied
7	Time to Event_2 (days)	Event_2	20	24	5/2/2010	5/25/2010	0	Event 2
8	Time to Event_2 (days)	Event_2	20	167	5/1/2010	10/14/2010	1	Completed
9	Time to Event_3 (days)	Event_3	30	173	11/2/2009	4/23/2010	1	Completed
10	Time to Event_3 (days)	Event_3	30	126	11/6/2009	3/11/2010	0	Event_3
11	Time to Event_3 (days)	Event_3	30	174	10/19/2009	4/10/2010	1	Completed
12	Time to Event_3 (days)	Event_3	30	90	10/19/2009	1/16/2010	1	Lost to follow-up

To develop ADTTE dataset, the key points are how to define STARTDT (time to event origin date for subject) and ADT (analysis date) for study events, right censored, and other competing events for each parameter. For parameters of Time to Event_1 and Time to Event_3 (rows 1–4 and rows 9–12 in Table 2), the Time to Event Origin Date is subject randomization date, and the Analysis Date is event date or censored date. For parameters of Time to Event_2 (rows 5–8 in Table 2), the Time to Event Origin Date is Date of Last Exposure to Treatment (EOT) if Event_1 at EOT, and the Analysis Date is Event_2 date or censored date. If not Event_1 at EOT, then Time to Event Origin Date and the Analysis date are all set missing. The permissible variable PARAMN, one-to-one mapping with PARAM, is for ordering and programmatic manipulation.

Conclusion

The ADTTE structure is simple and standardized. Based on ADTTE dataset, one general macro can be implemented for survival analyses. To design ADTTE dataset, besides complying with ADaM general rules and further with ADTTE requirements, we need to be clear of the time-to-event endpoints to determine the parameters. We also need to know the censored events, the corresponding event origin date, and the analysis date.

Acknowledgements

The author would like to thank Luyan Dai for the valuable inputs to the dataset design. The author would also like to thank Jingwei Gao and Nancy Bauer for the paper reviews.

References

1 CDISC Analysis Data Model (ADaM) Team. The ADaM Basic Data Structure for Time-to-Event Analyses Version 1.0. Round Rock (TX): CDISC; 8 May 2012.
2 CDISC Analysis Data Model (ADaM) Team. The Analysis Data Model (ADaM) Implementation Guide Version 1.0. Round Rock (TX): CDISC; 17 December 2009.