

## Sense and Censorability: Learn censoring techniques with ADTTE for your survival

Shilpakala Vasudevan, Efficacy Lifescience Analytics

### ABSTRACT

In many studies in clinical trials, the time to event is the primary endpoint, with an event being either an adverse event, death or recurrence of a disease. Time to event is calculated as the time from a defined starting point to the time of occurrence of the event of interest. In some cases, the event might not occur within the duration of study. An interesting feature of survival data is that at the end of the observation period, the event of interest may not have occurred for all subjects, or the exact event time may be unknown. In these cases, censoring rules need to be applied to determine the time to event. Depending on the nature of the study, ADTTE can accommodate single events or composite events, with binary or multiple censoring.

This paper will present how the endpoints of these subjects are accounted for, by applying censoring rules.

### INTRODUCTION

An event in clinical trials is any occurrence that is considered as significant for that study. For example, a better response to a treatment or an improvement in a subject's condition, can be treated as an event. While the event is considered as a significant aspect of the study, the time to reach the event becomes also equally important, in many cases. If the subject's condition becomes better in a shorter period of time, it adds to the efficacy of the treatment. This is the reason why time to event is increasingly becoming an important endpoint in efficacy analyses in many studies.

Time to event is straightforward if the event occurs during the observation period or treatment duration of the study planned. In this case, the date of the event and the start date is both known and time to event is easily derived. This subject is what is called an uncensored subject. But not all subjects in a study get to reach the event in the expected time frames. If the event is not reached within the duration of the study for any reason, the subject is censored to calculate the time to event.

### TIME TO EVENT

Time to event (TTE) is the time between start date of observation and first date of documentation of event. In a clinical trial, the start date might be the randomization date, or the first treatment start date, or the enrolment date; and the event of interest might be disease diagnosis, the occurrence of an adverse event, disease progression, or even death.

$$\text{TTE} = \text{event date} - \text{observation start date} + 1$$

Since survival analysis is a key component in oncology studies, time to event is an important feature in oncology and is considered as one of the primary endpoints in many oncological studies.

Below is a list of some common oncological time to event endpoints:

- Overall Survival (OS) – Time from observation start date to the date of death (death due to any cause)
- Progression-Free Survival (PFS) – Time from observation start date to the date of first documented disease progression or death, whichever occurs first
- Time to Progression (TTP) – Time from observation start date to the date of first documented disease progression
- Time to Treatment Failure (TTF) – Time from observation start date to the date of first failure (failure due to any cause like progression, death, discontinuation from treatment due to adverse event, symptomatic

deterioration)

- Event-Free Survival (EFS) or Recurrence-Free Survival (RFS) – Time from observation start date to the date of first documentation of disease recurrence or death, whichever occurs first
- Disease-Free Survival (DFS) -- Time from observation start date to the date of first documentation of disease occurrence or death, whichever occurs first
- Time to Next Treatment (TTNT) – Time from end of primary treatment to start date of next therapy
- Time to Symptom Severity (TSS) – Time from observation start date to the date of increase in symptom severity
- Duration of Response (DOR) – Time from date of first documentation of positive (complete or partial) response to the date of progression or death, whichever occurs first
- Duration of Complete Response (DCR) -- Time from date of first documentation of complete response to the date of progression or death, whichever occurs first

Besides oncology, a lot of other therapeutic areas also are increasingly incorporating time to event endpoints:

- Time to Exacerbation (Respiratory)
- Time to readmission to hospital after discharge of an asthma patient
- Time to Seizure (Neurology)
- Time from infertility treatment to conception
- Time to healing of a wound
- Time to recurrence of a gallstone
- Time to dialysis after a kidney transplant
- Time for better health related quality of life and other wellness factors.

## CENSORING

In all the above endpoints listed in the previous section, time to event is calculated based on the date of the event in consideration. If the event is not reached within the observation period, the time to event is calculated as the observation start date to the censor date.

$$TTE = \text{censor date} - \text{observation start date} + 1$$

Censor date could be the last available date or study completion date, or any other date as required by the study. The statistical analysis plan details the censoring rules for the study. Below is an example:

### Figure 1: Example of censoring rules in a Statistical Analysis Plan (SAP)

#### Censorship for Progression Free Survival:

- 1) Patients last known to be alive and progression-free, and who have a baseline and at least one disease assessment after dosing, are censored at the date of the last objective disease assessment that verified lack of disease progression.
- 2) Patients with inadequate baseline disease assessment are censored at the date of randomization.
- 3) Patients starting new anti-cancer treatment prior to progression are censored at the date of last objective disease assessment documenting no progression prior to the new treatment.

Let us consider we are working on a study, where subjects with lung cancer recently underwent a surgery to remove their tumor. In this study, disease free survival (DFS) is considered as its primary endpoint, and we are observing to see how long it takes for the tumor to recur and for the patients to be disease free. The duration of the study is from, 3 months from the time of surgery to 1 year.

In this case, disease free survival is calculated as:

$$DFS = \text{date of first recurrence of tumor} - \text{start date} + 1$$

If date of first recurrence of tumor was at 9 months, then DFS will be the time from date of surgery to the date of the 9 months review. In this case, there is no censoring required as event has been reached

within the duration of the study.

### LEFT CENSORING

Left censoring occurs when a subject leaves the study before an event occurs, or the study ends before the event has occurred.

For example, if the subject was examined at 3 months at the start of the study and was found to have the tumor recurred. Since 3 months was the first date of review, the patient had not yet been observed as disease-free, even for a day. In this case, the subject had already experienced the event before the study started.

The time to event in this case is undecided. This kind of censoring is not usually common in clinical trials.

### INTERVAL CENSORING

Sometimes, the onset of an event is known to have occurred between a time interval, and hence the exact visit of onset is unknown. The time of occurrence of these events is said to be interval censored.

In the same example, if a subject was found to have no disease recurrence at 3 months, and skipped the 6 months review, and was found to have tumor recurrence at 9 months, the date when the tumor occurred again is not exactly known.

There are 2 ways of dealing with interval censoring. i) One is doing midpoint imputation, where the midpoint of the censoring interval is taken as the censor date (in the above example, 6 months as the midpoint is taken as censor date). ii) The other method is to do right imputation, where the date at which the recurrence is first noted is taken as the censor date (in the above example, 9 months as censor date).

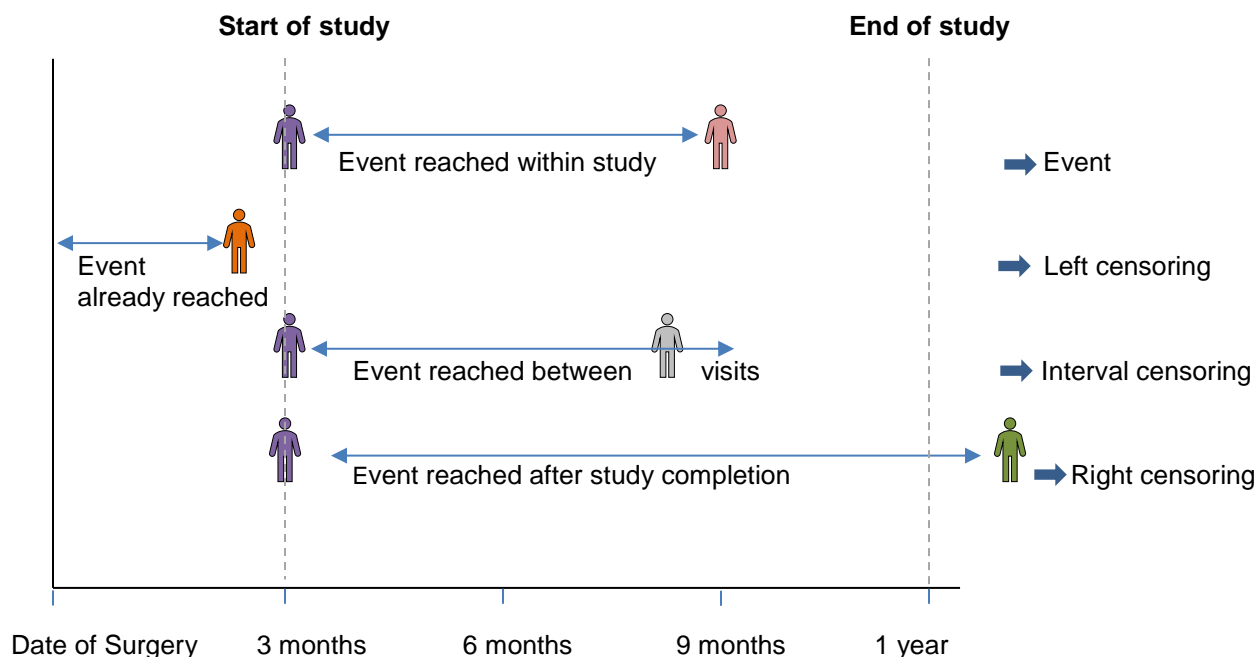
### RIGHT CENSORING

Right censoring is when the event did not occur within the defined study period.

For instance, if the subject did not experience a recurrence of disease throughout the 1 year, and the event did not happen, then the subject will be censored at the last available visit (1 year as the censor date).

Right censoring is the most common kind of censoring observed in clinical trial studies.

**Figure 2: Examples of types of censoring**



Additionally, censoring can be categorized into three types:

- *Type I censoring* occurs when the study starts and ends at a predetermined time, for all subjects. Any subject remaining in the study after are right censored.
- *Type II censoring* occurs when the study starts at the same time for all subjects, and it is terminated when a predetermined portion of subjects have experienced the event. The rest of the subjects are right censored.
- *Random censoring (or non-informative censoring)* is when the censoring is random, and not all subjects start the study at the same time, resulting in unequal follow-up. The censoring time of each subject is statistically independent of their event date. The observed value is the minimum of the censored and event dates; subjects whose event dates are greater than their censoring dates are right-censored.

Censoring is not to be considered the same as truncation. Truncation is where the study is deliberately terminated and truncated. Right truncation occurs when the entire study population has already experienced the event of interest before the study starts. Left truncation happens when subjects have been at risk and have not experienced the event, before entering the study. Most clinical trials studies are right censored and sometimes left truncated.

## ADTTE DATASET

ADTTE is the ADaM Basic Data Structure (BDS) for Time-To-Event (TTE) analyses. The structure of the dataset is very simple, with one record for every subject and parameter (or time to event endpoint). ADTTE caters to the longitudinal nature of survival analysis, where there are repeated observations of the same variables over shorter or longer periods of time. Majority of the time to event analyses in clinical studies are readily well supported by the ADTTE dataset.

Below are the key variables in ADTTE, that contribute to event, censoring and time to event derivations:

**Table 1: ADTTE variables**

Variable Name	Variable Label	Type	Note
USUBJID	Unique Subject Identifier	Char	Unique Subject Identifier
PARAM	Parameter	Char	The description of the analysis parameter (or time to event endpoint). Example: "Time to Progression"
PARAMCD	Parameter Code	Char	The short name of the analysis parameter in PARAM. Example: "TTP"
CNSR	Censor	Num	Indicates if record is censored, and type of censoring done. Example: 0 for event and >0 for censored records
EVNTDESC	Event or Censoring Description	Char	Describes the event of interest or an event that warrants censoring
CNSTDTC	Censor Date Description	Char	Describes the circumstance represented by the censoring date if different from the event date that warrants censoring
STARTDT	Initial Date of Event Risk	Num	The original date of risk for the time-to-event analysis. Example: randomization date or the date of first study therapy exposure
ADT	Analysis Date	Num	Analysis date of event or censoring date
AVAL	Analysis value	Num	The elapsed time to the event of interest from the origin. Example: if aval is measured in days, aval would be  adt – startdt or adt – startdt + 1.

## CASE ILLUSTRATIONS

Now that we know, what time to event is, how censoring is done and the key variables of ADTTE, let us look at a few case scenarios:

1. Single event endpoint with only two values for censoring variable
2. Single event endpoint, with many components, and multiple values for censoring variable
3. Multiple events endpoint or composite endpoint

### CASE 1: OVERALL SURVIVAL (SINGLE EVENT POINT WITH ONLY TWO VALUES FOR CENSORING VARIABLE)

Let us consider a study for pancreatic cancer (pancreatic ductal adenocarcinoma), where equal number of subjects parallelly undergo 2 different treatments for the cancer. The primary endpoint is Overall Survival, and the efficacy is checked by observing the number of months of survival of the patients. Since the endpoint is only survival, only death is taken as the event of interest. Hence this is a single event case, where subjects are termed to have attained the event if there is death, and the remaining are censored if death did not happen within the duration of the study. Censoring value is assigned as 0 if event is reached and no censoring is done, and is otherwise assigned as 1 if it is censored, irrespective of the censoring reason.

Below is the dataset with the dates for deriving overall survival: randomization, death and last contact dates:

**Table 2: Dates for Overall Survival**

usubjid (Unique Subject Identifier)	randdt (Randomization date)	dthdt (Death date)	lastctdt (Date of last contact when alive, before death or end of study)
01-1001	29MAY2012	25MAY2017	23APR2017
01-1002	18MAR2014		16MAR2017
01-1003	06MAY2016	18JUL2017	10FEB2017
01-1004	28JAN2016	18APR2017	25JAN2017
01-1005	01MAY2014		26APR2017
01-1006	20JAN2015		04JUL2017

The start date is randomization date in this study, the event date is the death date, and the censor dates for subjects with no death will be the last contact date when alive. Analysis dates are derived accordingly.

The following code derives the time to event and censoring variables:

```

paramcd = "OS"; *** Parameter Code ***;
param = "Overall Survival (Days)"; *** Parameter ***;
startdt = randdt; *** Time to Event Origin Date for Subject ***;

if dthdt ne . then cnsr = 0; *** Censor ***;
    else cnsr = 1;
if cnsr = 0 then adt = dthdt; *** Analysis Date ***;
    else if cnsr = 1 then adt = lastctdt;
if cnsr = 0 then evtndesc = "Death"; *** Event Description ***;
if cnsr = 1 then do;
    if sdsreas = "OTHER" then evtndesc = "Other";
    else if sdsreas = "LOST TO FOLLOW-UP" then evtndesc = "Lost to Follow-up";
    else if sdsreas = "WITHDRAWAL BY SUBJECT" then evtndesc = "Withdrawal";
end;

if dthdt = . then cnsdtdsc = "Date of last contact when alive";
else cnsdtdsc = ""; *** Censoring Date Description ***;

if nmiss(adt,startdt) = 0 then aval = adt - startdt + 1; *** Analysis Value ***;
if aval ^= . and aval <= 0 then aval = 1;

```

The final adtte dataset is in the below format:

**Table 3: ADTTE dataset for Overall Survival**

usubjid	trtp	param	paramcd	Startdt	adt	aval	cnsr	evntdesc	cnstdtc
01-1001	Treatment A	Overall Survival (Days)	OS	29MAY2012	25MAY2017	1813	0	Death	
01-1002	Treatment A	Overall Survival (Days)	OS	18MAR2014	16MAR2017	1095	1	Lost to Follow-up	Date of last contact when alive
01-1003	Treatment B	Overall Survival (Days)	OS	06MAY2016	18JUL2017	1168	1	Withdrawal	Date of last contact when alive
01-1004	Treatment A	Overall Survival (Days)	OS	28JAN2016	18APR2017	447	0	Death	
01-1005	Treatment B	Overall Survival (Days)	OS	01MAY2014	26APR2017	1092	0	Death	
01-1006	Treatment B	Overall Survival (Days)	OS	20JAN2015	04JUL2017	897	1	Other	Date of last contact when alive

The most common types of analyses done on survival analyses are: Kaplan-Meier estimates, Cox regression and Log-rank test.

Below is a table displaying the statistics for Overall Survival:

**Table 4: Summary table for Overall Survival**

Summary of Overall Survival (All Patients), (Stratified), As Randomized

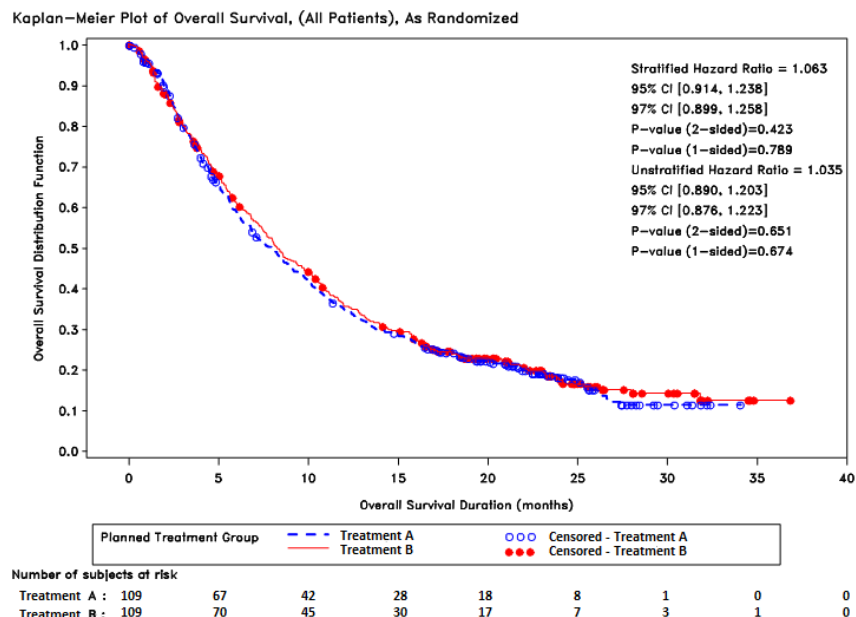
	Treatment A n (%)	Treatment B n (%)
Number of Subjects	109 ( 100)	109 ( 100)
Number of deaths	85 ( 77.9)	84 ( 77.1)
Number censored	25 ( 22.9)	26 ( 23.9)
Reason for censorship		
Alive	17 ( 15.6)	18 ( 16.5)
Withdrawal	6 ( 5.5)	5 ( 4.6)
Lost to Follow-up	1 ( 0.9)	2 ( 1.8)
Other	1 ( 0.9)	1 ( 0.9)
Kaplan-Meier estimates of Time to Event (months) [1]		
Quartiles (95% CI)		
25%	3.7[ 3.2, 4.2]	3.8[ 3.2, 4.5]
50%	7.9[ 6.8, 9.0]	8.3[ 7.4, 9.6]
75%	17.3[ 14.3, 21.6]	17.1[ 15.6, 21.9]
Versus Treatment B		
Hazard Ratio (95% CI) [2]	1.063 [0.914, 1.238]	
(97% CI) [2]	[0.899, 1.258]	
P-value (2-sided) [3]	0.423	
(1-sided) [3]	0.789	

[1] Estimated from the Kaplan-Meier Method;

[2],[3] HR and its CIs are estimated from the stratified Cox Regression and p-values are based on the stratified log-rank test with EGFR status(Wild Type vs. Mutant), KRAS status(Wild Type vs. Mutant) and ECOG at Randomization (0-1 vs. 2) as the stratification factors.

The frequencies for the number of patients reaching the event and the number censored with the reasons is listed. Besides these, Kaplan-Meier analyses is done which determine the estimates of survival over time. The frequencies and estimates indicate the results to be very similar between both treatments. Cox regression analyses is done to derive Hazard Ratio and Log-rank tests are used to generate p-values. A Hazard ratio of 1.063 (value very close to 1) indicates that the treatments have very similar effects, and are experiencing an equal number of events at any given point in time. The confidence intervals (values are crossing null hypothesis of 1) and p-values indicate insufficient substantiation to conclude that the two treatment groups are statistically significantly different from each other.

**Figure 3: Kaplan-Meier Plot of Overall Survival**



Proc Lifetest is used to generate Kaplan-Meier estimates:

```
PROC LIFETEST data=adtte
method=km outsurv=dsplot;
time aval*cnsr(1);
by trtpn trtp;
run;
```

With the probabilities generated from the procedure, we can use proc gplot to plot the data.

Censored subjects are indicated as dots for both treatments in the graph. The curves indicate similar results between the two treatments.

**CASE 2: TIME TO DETERIORATION OF QUALITY OF LIFE (SINGLE EVENT ENDPOINT, WITH MANY COMPONENTS, AND MULTIPLE VALUES FOR CENSORING VARIABLE)**

Many studies integrate health-related quality of life as one of the major endpoints in determining the effectiveness of the therapies on the subjects. Quality of Life (QoL) includes many factors that indicate the general well-being of patients in life.

In this example study, while overall survival is the primary endpoint, time to deterioration (TTD) of quality of life is taken as the secondary endpoint. Time to deterioration can be taken as the time it takes for the quality of life scores to deteriorate with a minimal clinically important score change, from the baseline score. Time to deterioration is a single event endpoint that considers only deterioration in quality of life as an event. Here even death is considered as a censoring reason and not an event.

**Table 5: Dates for Time To Deterioration of Quality of Life (QoL)**

usubjid (Unique Subject Identifier)	Randdt (Randomization date)	Deterdt (Deterioration date)	lastctdt (Date of last contact when alive, before end of study)
02-1001	10FEB2016		07NOV2017
02-1002	25JUN2013		28APR2017
02-1003	17SEP2012	25NOV2016	12JUN2016
02-1004	09MAY2014		01MAY2017
02-1005	29JAN2015	10APR2017	28FEB2017
02-1006	30OCT2014		14JAN2017

Here, deterioration alone is considered an event and so, if the event is reached, we have cnsr as 0. For the subjects which do not have the event occurred within the time span, censoring values are different for each censoring reason. Censoring values in this case can be more than one (as 1,2,3,4 and so on):

```
if dthdt ne . then cnsr = 0; *** Censor ***;
else if sdreas="End of Study" then cnsr = 1;
else if sdreas="Withdrawal by Subject" then cnsr = 2;
else if sdreas="Subject Discontinuation" then cnsr = 3;
else if sdreas="Lost to Follow-up" then cnsr = 4;
```

Below is the final addtte dataset with different cnsr values:

**Table 6: ADTTE dataset for Time To Deterioration of Quality of Life (QoL)**

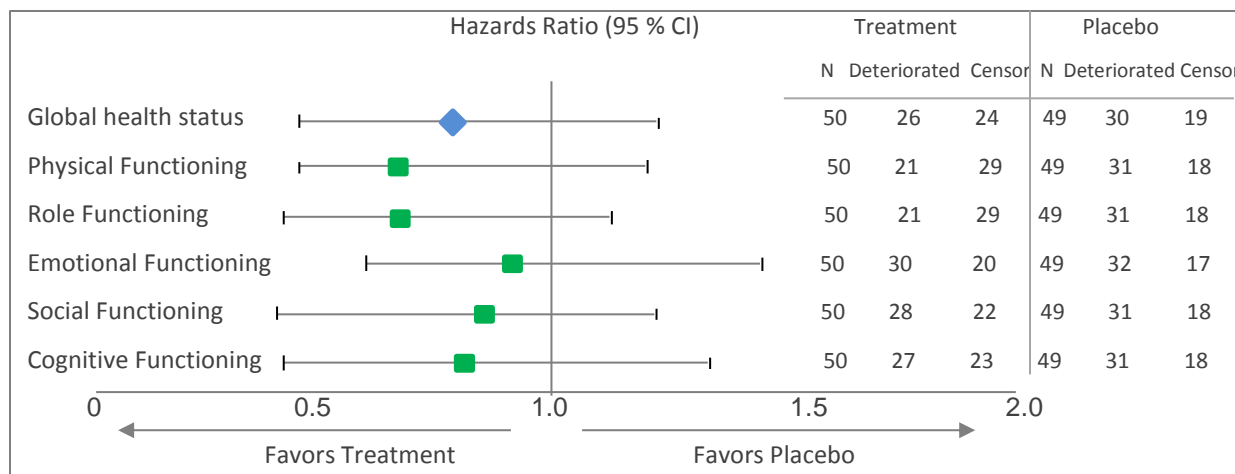
usubjid	param	paramcd	randdt	adt	aval	cnsr	evntdesc	cnsdtc
02-1011	Time to deterioration of QoL (Days)	TTD	10FEB2016	07NOV2017	637	2	Withdrawal by Subject	Date of last contact
02-1012	Time to deterioration of QoL (Days)	TTD	25JUN2013	28APR2017	1404	1	End of study	Date of last contact
02-1013	Time to deterioration of QoL (Days)	TTD	17SEP2012	12JUN2016	1365	0	Deterioration	
02-1014	Time to deterioration of QoL (Days)	TTD	09MAY2014	01MAY2017	1089	3	Subject Discontinuation	Date of last contact
02-1015	Time to deterioration of QoL (Days)	TTD	29JAN2015	28FEB2017	762	0	Deterioration	
02-1016	Time to deterioration of QoL (Days)	TTD	30OCT2016	14JAN2017	77	4	Lost to Follow-up	Date of last contact

Cox proportional hazards model is a regression method frequently used for survival data and provides an estimate of the hazard ratio and confidence interval. Hazard ratios are very widely used in survival analysis, as they help estimate the probability that the event occurs at any given point of time. The following code generates the hazards ratios and confidence intervals, with the aval and cnsr:

```
PROC PHREG data=addtte;
  model aval * cnsr (0) = trtpn / alpha=0.05 rl ties=efron;
run;
```

Forest plots are very useful in depicting estimates across different sub-groups and also in representing the overall estimated effect of the exposure of interest. Time To Deterioration contains many components, and hence the statistics are displayed for each component so a comparison can be made by group. As can be seen in the below plot, the hazard ratios represented by the square markers in all the components are favoring the treatment when compared to the placebo. The overall effect represented by the diamond marker in the global health status also indicates effectiveness for the treatment.

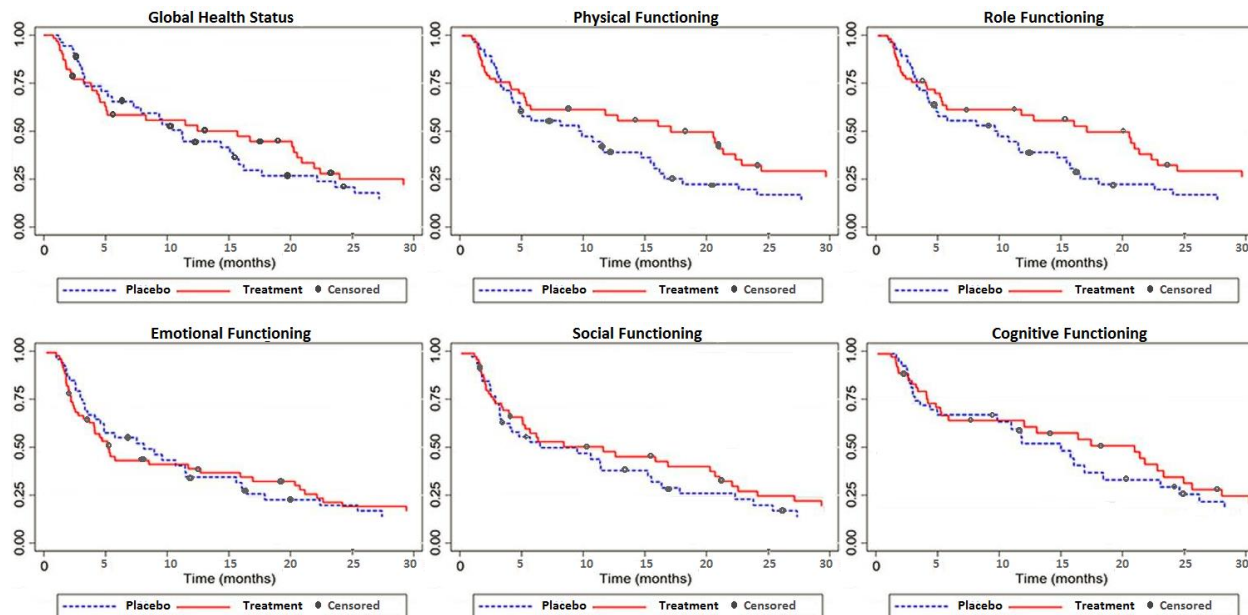
**Figure 4: Forest plot with Hazard Ratios for Time To Deterioration of Quality of Life (QoL)**





Since Quality of Life has different components (Global Health Status, Physical Functioning, Role Functioning, Emotional Functioning, Social Functioning and Cognitive Functioning), Kaplan meier plots can be generated for each component to get the estimates:

**Figure 5: Kaplan-Meier plots for Time To Deterioration of Quality of Life (QoL), Individually by Component**



### CASE 3: HEART FAILURE-FREE SURVIVAL (MULTIPLE EVENTS ENDPOINT OR COMPOSITE ENDPOINT)

A composite endpoint is where two or more events are considered together for one common endpoint. For example, when we look for Heart failure-Free Survival, we consider all kinds of heart failure and death as events of interest. Let us consider a study where 20 subjects are randomized to Treatment A and 20 remaining subjects to Treatment B, with heart conditions. All 40 subjects are at a risk of heart failure at the time of randomization.

Apart from heart failure, death is also considered as an event in this endpoint. In both cases, since there is no censoring, censoring value will be 0. For all other subjects, where there is no failure or death, censoring values will be assigned according to the reason for censoring.

The censoring code would be as below:

```

if dthdt ne . then cnsr = 0; *** Censor ***;
else if sdreas="Heart failure" then cnsr = 0;
else if sdreas="Lost to Follow-up" then cnsr=1;
else if sdreas="Given new therapy" then cnsr=2;
else if sdreas="Discontinued" then cnsr = 3;
else if sdreas="Completed Study" then cnsr=4;
    
```

**Table 7: ADTTE dataset for Heart Failure-Free Survival**

usubjid	param	param cd	randdt	adt	aval	cnsr	evntdesc	cnsdtc
03-1111	Heart failure-Free Survival (Days)	HFS	25SEP2013	26APR2017	1310	0	Heart failure	

03-1112	Heart failure-Free Survival (Days)	HFS	01NOV2014	18OCT2017	1083	1	Lost to Follow-up	Date of last contact when alive
03-1113	Heart failure-Free Survival (Days)	HFS	15MAY2015	22DEC2016	588	2	Given new therapy	Last assessment before therapy
03-1114	Heart failure-Free Survival (Days)	HFS	04AUG2014	14AUG2017	1107	0	Death	
03-1115	Heart failure-Free Survival (Days)	HFS	18DEC2016	19JAN2017	33	3	Discontinued	Last assessment showing no failure or death
03-1116	Heart failure-Free Survival (Days)	HFS	22MAR2017	23JUL2017	124	4	Completed Study	Last assessment showing no failure or death

Since composite endpoints involve more than one event and more than one values of censoring, it is likely that there would be a multitude of dates involved for deriving the event date and censoring date. For this purpose, it would be easier to create an intermediate adtte dataset that first retrieves all the dates from the source datasets, and then sets all dates for each subject in a horizontal format, with every date for each subject in a separate record.

For example, the following is an interim dataset with all dates for one subject, set in separate rows:

**Table 8: Interim ADTTE dataset (ADTTE\_INT) for Heart Failure-Free Survival**

usubjid	Param	Paramcd	Aval	srcdom	srcvar	srcseq
03-1113	Date of Randomization	RANDDT	15MAY2015	ADSL	RANDDT	
03-1113	Date of last adequate assessment before new therapy	LSTADQDT	22DEC2016	RS	RSDTC	07
03-1113	Start date of new therapy	NEWSTDT	17JAN2017	CM	CMSTDTC	22
03-1113	Date of Heart Failure	HRFAILDT	09FEB2017	RS	RSDTC	10
03-1113	Date of last adequate assessment	LSTADDT	22FEB2017	RS	RSDTC	03
03-1113	Date of death	DTHDT	20MAR2017	DM	DTHDTC	15

From this intermediate dataset, the dates are again transposed into the below format, so that all required dates are present in one record for every subject:

**Table 9: Transposed ADTTE dataset from the interim dataset**

usubjid	RANDDT	LSTADQDT	NEWSTDT	HRFAILDT	LSTADDT	DTHDT
03-1113	15MAY2015	22DEC2016	17JAN2017	09FEB2017	22FEB2017	20MAR2017

Censoring can then be easily done with the dates to derive the final dataset. Having all the dates in a vertical structure makes it straightforward to assign event and censor dates and derive the adt values. Also, where there is a need to find the first available or last available dates, it is easier to do a comparison of the dates with this data structure format.

Since the final dataset is not directly derived from the source datasets, it is preferable to create source data variables in the interim dataset that could help trace from where the dates were retrieved.

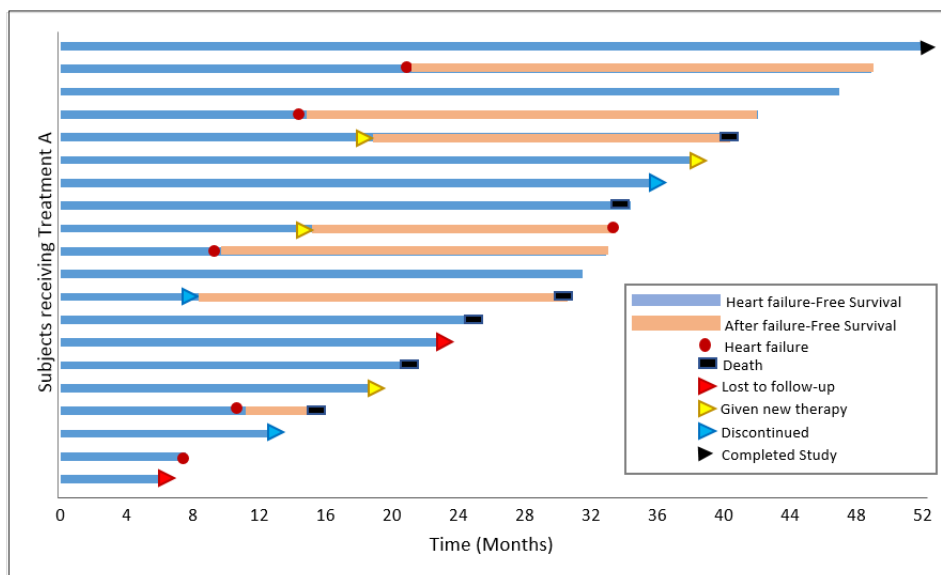
Following are the variables used in ADTTE for traceability:

**Table 10: Traceability variables in ADTTE**

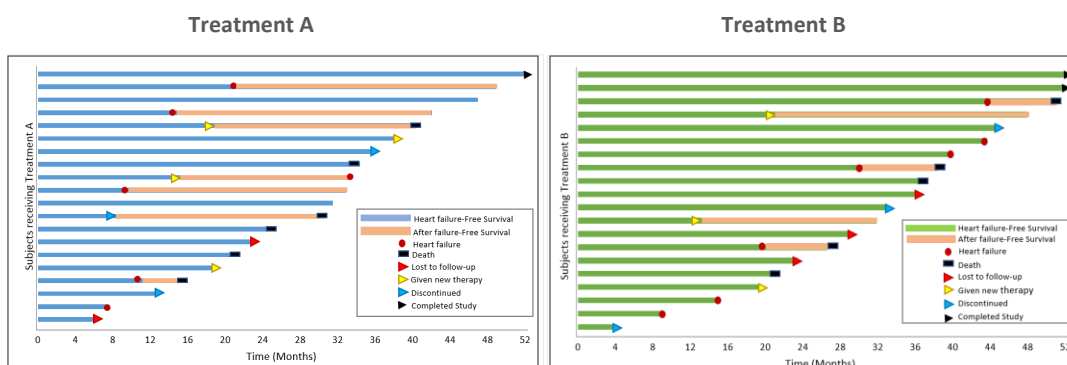
Variable Name	Variable Label	Type	Note
SRCDOM	Source Data	Char	The name of the SDTM domain (or dataset) or analysis dataset that relates to the analysis variable (e.g., AVAL)
SRCVAR	Source Variable	Char	The name of the column (in the domain or dataset identified by SRCDOM) that relates to the analysis variable.
SRCSEQ	Source Sequence Number	Num	The sequence number --SEQ or ASEQ of the row (in the domain or dataset identified by SRCDOM) that relates to the analysis variable.

Since this study has only 20 subjects as sample size for each treatment, swimmer plots can be generated to compare the effects of the two treatments. Below is a swimmer plot generated for the first treatment. In a swimmer plot, each subject is depicted as a bar lane in the graph, and the events and the censored events are marked along the bars, over time. From the plot, we can identify how long each subject had a Heart failure-Free Survival, and what was the reason for their censoring/event.

**Figure 6: Swimmer Plots for Heart Failure-Free Survival**



Below figure is a side by side comparison of the swimmer plots for both treatments:



It can be noticed that Treatment B has longer and more failure-free bar lines than Treatment A, indicating Treatment B to be more effective than Treatment A. Swimmer plots are ideal only when there are lesser number of subjects involved in the analysis. The more the subjects or the variables in consideration, the more the plots would become cluttered and difficult to read.

## CONCLUSION

In this paper, different censoring techniques were explained with examples. Survival analysis is an integral part of many studies now, and time to event endpoints are vastly included in different therapeutic areas, apart from oncology. The ADTTE dataset is designed to handle most of the survival analysis data. However, some time to event endpoints might involve dealing with multiple dates from many different sources. Also, in some studies where interval censoring is done, and the exact event date is unknown, there could be other dates into consideration. In these cases, it is always recommended to go step by step, and create interim datasets with the dates, so that censoring can be done in an easier way. A well written, easy to read data specification and a simple code will help to keep the entire process comprehensible. Depending on the study, even multiple ADTTE datasets may be created, based on the number of time to event endpoints involved in the study, and as is required by the sponsor. It is always better to keep the dataset simple in order to maintain clarity. In line with this, it is also recommended to not keep the non-time events along with the time events in the same dataset, even if they fit into the same basic data structure.

## REFERENCES

Website CDISC Analysis Data Model (ADaM) Team May 2012. "The ADaM Basic Data Structure for Time-to-Event Analyses – CDISC".

[https://www.cdisc.org/system/files/all/standard\\_category/application/pdf/adam\\_tte\\_final\\_v1.pdf](https://www.cdisc.org/system/files/all/standard_category/application/pdf/adam_tte_final_v1.pdf)

Book Paul D. Allison. Survival Analysis Using the SAS System. A Practical Guide, Second Edition. SAS Institute Inc. 2010

Journal article Amélie Anota, Zeinab Hamidou, Sophie Paget-Bailly, Benoist Chibaudel, Caroline Bascoul-Mollevi, Pascal Auquier, Virginie Westeel, Frederic Fiteni, Christophe Borg, and Franck Bonnetain 2013 "Time to health-related quality of life score deterioration as a modality of longitudinal analysis for health-related quality of life studies in oncology: do we need RECIST for quality of life to achieve standardization?". Springer

Journal article Shankar Prinja, Nidhi Gupta, and Ramesh Verma 2010 "Censoring in Clinical Trials: Review of Survival Analysis Techniques" Indian Journal of community medicine

Journal article Faraz S. Ahmad, Hongyan Ning, Jonathan D. Rich, Clyde W. Yancy, Donald M. Lloyd-Jones and John T. Wilkins 2016 "Hypertension, Obesity, Diabetes, and Heart Failure-Free Survival The Cardiovascular Disease Lifetime Risk Pooling Project" JACC: Heart Failure

## CONTACT INFORMATION

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

Shilpakala Vasudevan  
Ephicity Lifescience Analytics  
[shilpakala.vasudevan@ephicity.in](mailto:shilpakala.vasudevan@ephicity.in)

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.