

## Calculating Restricted Mean Survival Time

Tony Qi, ImmnoGen. Inc., Waltham, MA; Jiuzhou Wang, ImmunoGen. Inc., Waltham, MA

### ABSTRACT

The commonly used descriptive statistics for survival analysis include median survival time, survival rate at given time point(s), all derived from Kaplan-Meier estimate. Mean survival time (MST), however, has received less attention in the field of clinical research, partly because it is often subject to underestimation due to the largest observation being censored. Restricted mean survival time (RMST) can provide additional insight to the survival distribution and is not subject to underestimation by definition. In this paper, we discuss the calculation of RMST at any given time point (up to the largest observation) using PROC LIFETEST. A SAS macro is also developed to streamline the calculation when inference of RMST is required at multiple time points.

### INTRODUCTION

Survival analysis is widely used in oncology study for efficacy measurement. Survival data consist of a response variable that measures the duration of time to specific event happened. Due to nature of clinical trial practice, which is conducted over a finite period of time, not all the event time will be collected. For example, in oncology study, patient may start a new therapy, lost follow-up, etc. Patient will be censored when event of interest has not occurred for patient at a given time. By using survival analysis, investigator can get the distribution of survival for a given population, compare the survival distributions among different groups, or study the relationship between the survival time and some concomitant variables.

Kaplan-Meier method is a powerful tool in survival analysis. It is mainly used to estimate a population survival curve from a sample in the presence of censoring. The following figure (Figure 1) is an example from data of bone marrow transplant patients (sashelp.BMT).

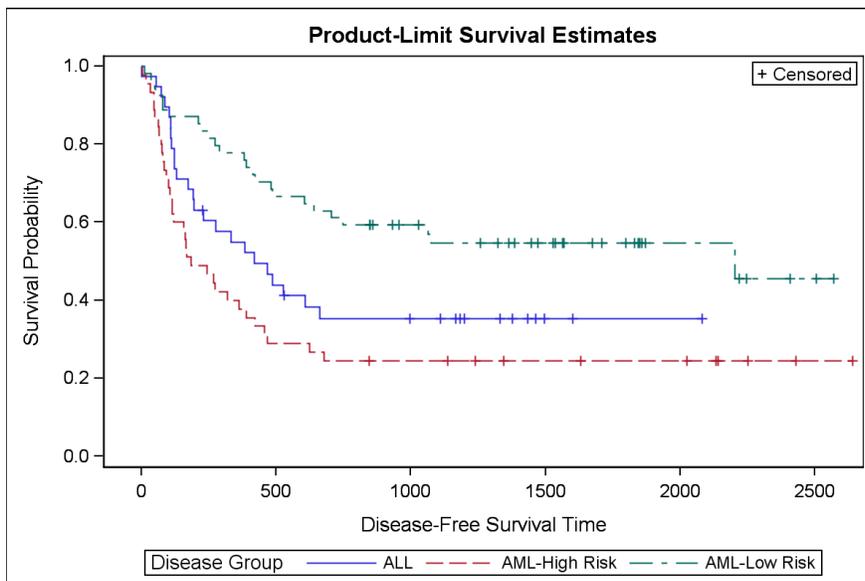


Figure 1: Kaplan-Meier Plot Example

This plot shows the survival curve of three groups: ALL (acute lymphoblastic leukemia), AML (acute myelocytic leukemia)-Low Risk, and AML-High Risk. Variable T means “Disease-Free Survival Time”. Variable Status is “Event Indicator: 1=Event 0=Censored”. Example data as below figure 2:

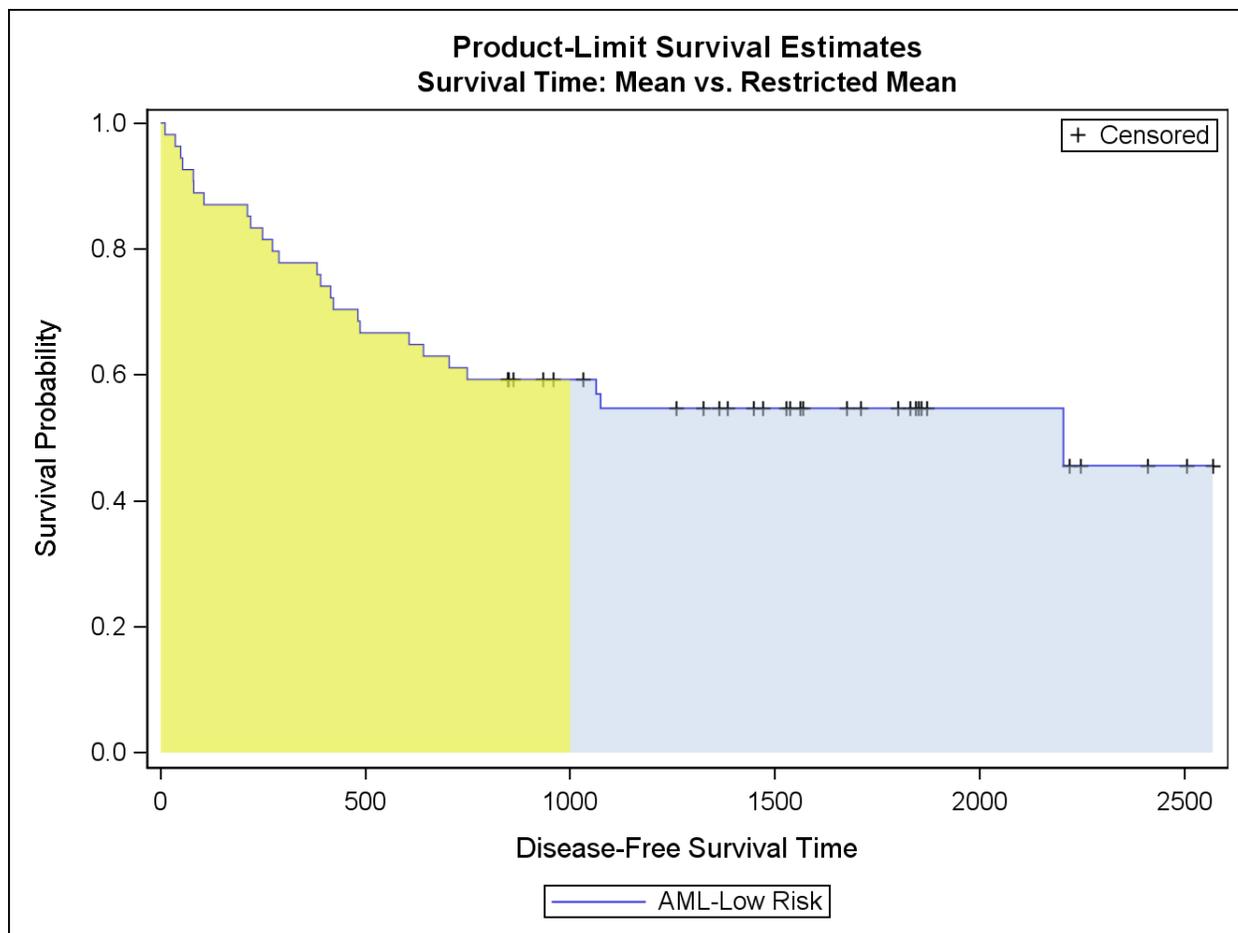
	Group	T	Status
30	ALL	86	1
31	ALL	466	1
32	ALL	192	1
33	ALL	109	1
34	ALL	55	1
35	ALL	1	1
36	ALL	107	1
37	ALL	110	1
38	ALL	332	1
39	AML-Low Risk	2569	0
40	AML-Low Risk	2506	0
41	AML-Low Risk	2409	0
42	AML-Low Risk	2218	0
43	AML-Low Risk	1857	0
44	AML-Low Risk	1829	0
45	AML-Low Risk	1562	0
46	AML-Low Risk	1470	0
47	AML-Low Risk	1363	0
48	AML-Low Risk	1030	0
49	AML-Low Risk	860	0
50	AML-Low Risk	1258	0

**Figure 2: Bone Marrow Transplant Patient Data**

Some key descriptive statistics of survival curve includes median survival time. It refers to half population patients survive after a certain treatment. Mean survival time (MST), which measures the area under the survival curve, however, has received less attention in the field of clinical research, partly because it is often subject to underestimation due to the largest observation being censored. Restricted mean survival time (RMST) can provide additional insight to the survival distribution.

The following figure shows the difference of Mean Survival Time (MST) and Restricted Mean Survival Time (RMST).

The total shaded area (yellow and blue) is the mean survival time, which underestimates the mean survival time of the underlying distribution. The yellow shaded area, where the time interval is restricted to [0, 1000 days], is the restricted mean survival time at 1000 days. Comparison as below figure (Figure 3)



**Figure 3: Mean Survival Time and Restricted Mean Survival Time**

In this paper, we apply two methods to calculate the RMST. The automation method, a SAS macro, was also generated to streamline this process.

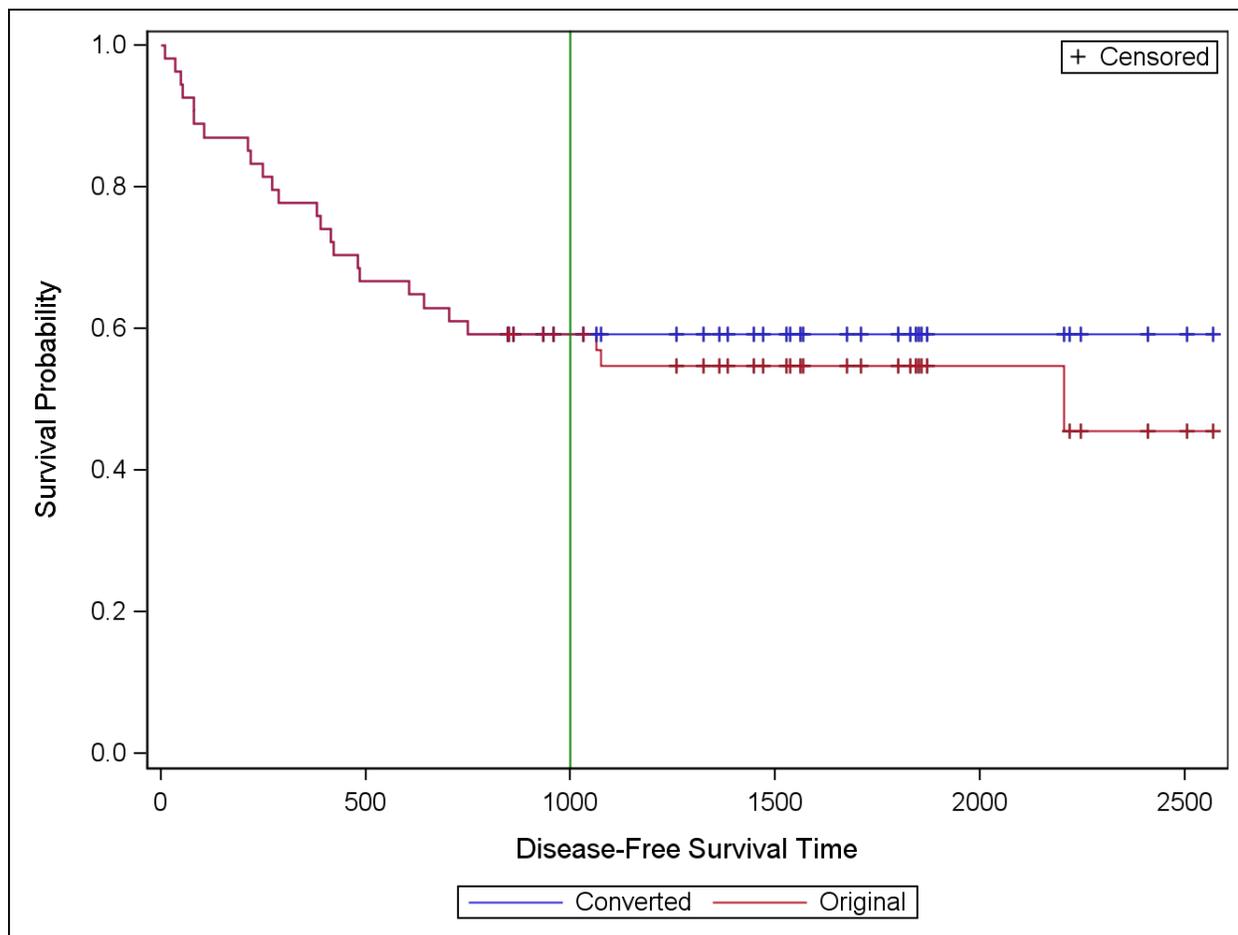
## CALCULATION

We will use the BMT (Bone Marrow Transplant Patients) dataset, AML Low Risk group as example to demonstrate two methods calculating RMST. The interval of Disease-Free Survival Time is set as [0, 1000 days].

### METHOD 1: TIMELIM OPTION IN PROC LIFETEST

The first method we use is the TIMELIM option in PROC LIFETEST. TIMELIM option “specifies the time limit used in the estimation of the mean survival time and its standard error. The mean survival time can be shown to be the area under the Kaplan-Meier survival curve.” The limitation of this method is that if the largest observed time in the data is censored, the area under the survival curve is not a closed area. The estimation was restricted to the largest event time. To overcome this limitation, we need an intermediate dataset so that the K-M estimate from that dataset is identical to the K-M estimate from the original dataset up to  $T=1000$  days and the remaining K-M curve is flat. This can be simply achieved by censoring all values in variable T in the original dataset that are greater than 1000 (Figure 4).

The following figure shows the difference. The blue line is from the converted data and the red line is from the original data, Any event after interval time 1000 was set as censored.



**Figure 4: Converted vs Original**

The following code was used to apply TIMELIM option to calculate RMST

```
DATA NEWBMT;
  SET SASHELP.BMT;
  IF GROUP='AML-Low Risk';
  IF T>=1000 THEN STATUS=0;
RUN;

PROC LIFETEST DATA=NEWBMT PLOTS=SURVIVAL TIMELIM=1000;
  time T * STATUS(0);
RUN;
```

The result is 720.74 days;

**METHOD 2: DIFFERENTIATION AND INTEGRATION**

The 2nd method we use is called differentiation and integration. The area under the survival curve is the sum of D small rectangulars, where D is the number of events up to T=1000 days (Figure 5). We calculate each part then sum them together to get the RMST value. The tricky part is the last section

between the last observation prior to T=1000 and 1000 days. We need to add extra code to handle this last rectangular.

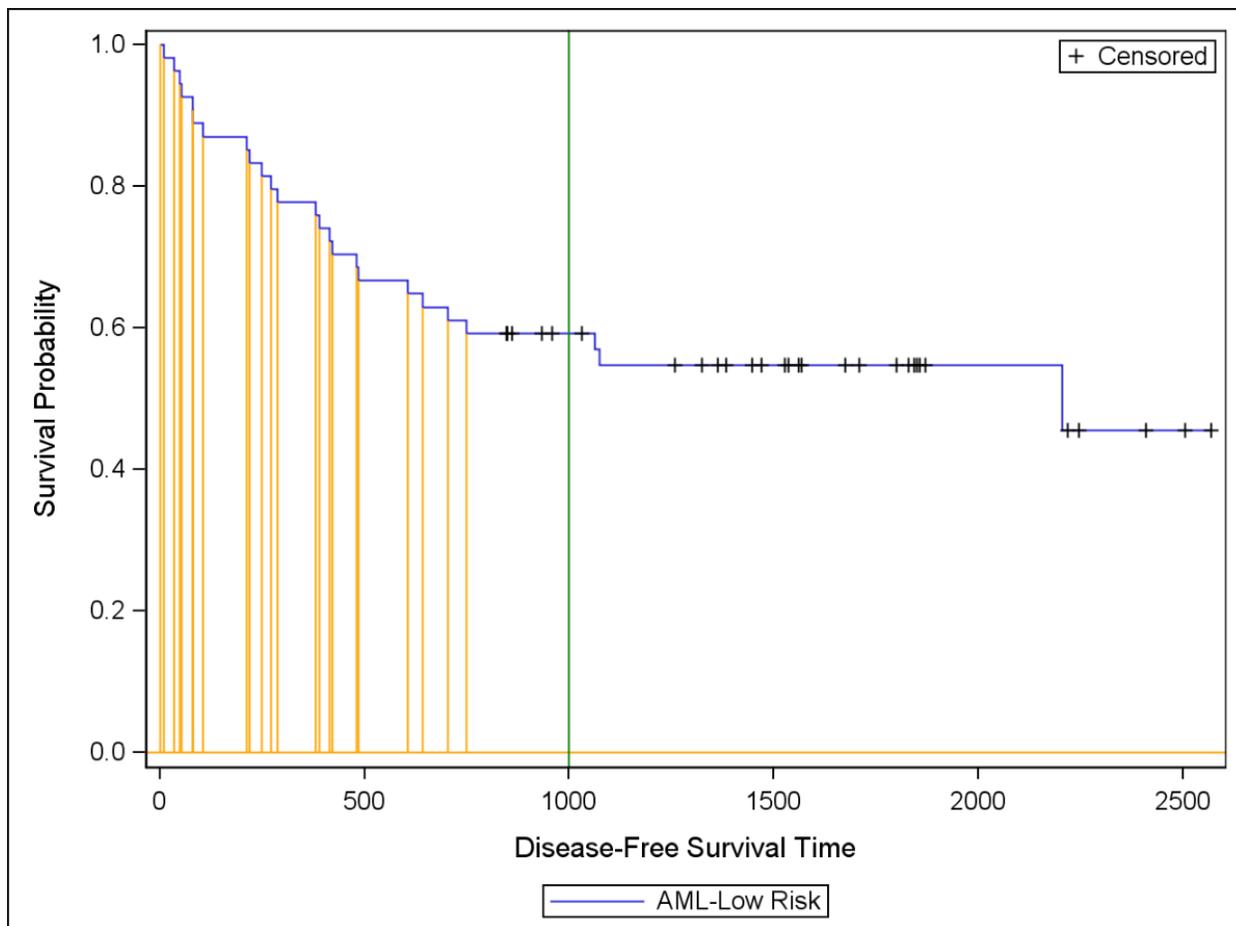


Figure 5: Sum up all small blocks

We use the following code to calculate the RMST

```
DATA ADDUP;
  SET SURVIVALPLOTDATA;
  IF TIME<=1000;
RUN;

DATA ADDUP;
  SET ADDUP END=EOF;
  OUTPUT;
  IF EOF THEN DO;
    TIME=1000;
    OUTPUT;
  END;
RUN;

DATA ADDUP;
  SET ADDUP;
```

```
LAGT=LAG (TIME) ;
LAGS=LAG (MAX (SURVIVAL, CENSORED) ) ;
IF LAGT=. THEN LAGT=0;
WIDTH=TIME-LAGT;
AREA=WIDTH*LAGS;
IF AREA=. THEN AREA=0;
RUN;

DATA ADDUP2;
  SET ADDUP;
  RETAIN TOTAREA 0;
  TOTAREA=TOTAREA+AREA;
RUN;

PROC PRINT DATA=ADDUP2;
  WHERE TIME=1000;
  VAR TOTAREA;
RUN;
```

We also developed the following macro to streamline this process

```
%RMST(dsin=newbmt, time=T, status=status, interval=1000);
```

## DISCUSSION

Method 2 (Differentiation and Integration) is more complicated than Method 1 in terms of coding. The calculation of standard error for RMST (not covered in this paper) is more complicated than the calculation of RMST itself using Method 2. Method 1 takes advantage of the TIMELIM option in PROC LIFETEST, but requires censoring all observations after the given time point T which can be easily achieved. It not only provides an estimate of RMST, but also gives its standard error for use in statistical inference.

## CONCLUSION

In this paper, we talk about the mean survival time and restrict mean survival time in survival analysis. We present two methods to calculate the restricted mean survival time. First one is using timelim option in proc lifetest, with data technique to convert open space shaded area to closed shaded area. Second method is using differentiation and integration to sum up all areas before the set interval time. A macro was also created to streamline and standardize this process. We recommend Method 1 due to its ease for implementation.

## REFERENCES

Website "Why Use a Kaplan-Meier Analysis?" Accessed March 9, 2018  
<http://biostat.mc.vanderbilt.edu/wiki/pub/Main/ClinStat/km.lam.pdf>

Website "Kaplan-Meier Survival Estimates" Accessed March 9, 2018  
[https://www.statsdirect.com/help/survival\\_analysis/kaplan\\_meier.htm](https://www.statsdirect.com/help/survival_analysis/kaplan_meier.htm)

## ACKNOWLEDGMENTS

We would like to thank ImmunoGen Inc. for providing the opportunity to work on this paper. We also appreciate the following manager/colleague for thoughtful reviews and comments.

Bates, Richard (Associate Director of Medical Affairs, ImmunoGen)

## RECOMMENDED READING

- *SAS®/STAT 14.1 User's Guide*

## CONTACT INFORMATION

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

Tony Qi  
ImmunoGen, Inc.  
830 Winter St, Waltham MA 02451  
Tony.Qi@immunogen.com  
781-895-0148

Joe Wang  
ImmunoGen, Inc.  
830 Winter St, Waltham MA 02451

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.