Paper 2980-2015

A Set of SAS® Macros for Generating Survival Analysis Reports for Lifetime Data with or without Competing Risks

Zhen-Huan Hu, MPH, Center for International Blood and Marrow Transplant Research, Milwaukee, WI

ABSTRACT

The paper introduces users to how they can use a set of SAS® macros, %LIFETEST and %LIFETESTEXPORT, to generate survival analysis reports for data with or without competing risks. The macros provide a wrapper of the LIFETEST procedure and an enhanced version of the AUTOCALL macro %CIF to give users an easy-to-use interface to report both survival estimates and cumulative incidence estimates in a unified way. The macros also provide a number of parameters to enable users to flexibly adjust how the final reports should look without the need to manually input or format the final reports.

INTRODUCTION

The standard Kaplan-Meier estimator of survival probability is widely used to estimate the survivor function from right censored lifetime data, assuming all censors are independent of the event of interest [1]. For example, in a medical research study, it can be used to measure the survival time of cancer patients with exposures of different risk factors. An observation from such a study can be right censored when a patient is dropped from the study due to lost follow-up, which is independent to the event of interest - here the overall survival. The LIFETEST procedure is implemented in SAS to generate the Kaplan-Meier estimator of the survivor function.

The cumulative incidence function, on the other hand, is used to describe life time data with competing risks [2]. For example, in a study for bone marrow transplantation, a patient can have multiple post-transplantation outcomes, such as relapse (patient's disease returns), or treatment-related mortality (death due to viral infection, secondary cancer, etc.). To properly estimate the rates of relapse across time, treatment-related mortality should be treated as its competing risk - meaning the event of treatment-related mortality precludes the possibility of the same individual to get relapse. Since SAS/STAT 12.1 (SAS 9.3 TS1M2), an AUTOCALL macro %CIF has been added to provide crude cumulative incidence function estimation for such lifetime data [3].

The uses of these two approaches are quite different, while in practice, it is quite often that users would come across lifetime data requiring the implementations of both approaches. The set of macros introduced here, namely %LIFETEST and %LIFETESTEXPORT, provide users a unified and straightforward interface to estimate both the survivor function based on the Kaplan-Meier estimator and the cumulative incidence function, and automatically generate well-formatted publication-ready final reports. The macros can also generate logrank p-values and Gray's test p-values when comparing overall differences for the survival and/or cumulative incidence functions across multiple groups [4]; as well as point-wise p-values when comparing the differences of the survival and/or cumulative incidence estimates at any specific time point. Users can also modify the way how the results being reported in the final report by adjusting additional parameters of the macros.

DATA SOURCE

Hematopoietic stem cell transplantation data from the Center for International Blood and Marrow Transplant Research (CIBMTR) were used in the paper to illustrate the use of the macros. Adult patients (age ≥ 18) diagnosed with acute myeloid leukemia (AML) undergoing allogeneic hematopoietic stem cell transplantations with matched sibling donors between 2001 and 2011 in the US were selected (N = 1,967). Baseline variables included disease stage prior to the transplantation and graft source. Outcomes of interest included overall survival, and relapse, with treatment-related mortality treated as its competing risk.

The data set contained following variables:

- **dead**: overall survival
- **intxsurv**: time from transplantation to death/lost follow-up, by months
- **rel**: relapse after the transplantation
- **trm**: treatment-related mortality
- **intxrel**: time from transplantation to relapse/treatment-related mortality
- **stage**: disease stage prior to the transplantation (early vs. intermediate vs. advanced)
- **graftype**: graft source (bone marrow vs. peripheral blood stem cell)

HOW TO USE THE MACROS

Both the macros and a modified version of the AUTOCALL macro %CIF are included in the supplement files. Before using the macros described in the paper, users need to make sure to use the %INCLUDE statement to include in their programs the references of both the macros and the modified %CIF.

THE KAPLAN-MEIER ESTIMATOR OF THE SURVIVOR FUNCTION

To create reports for the Kaplan-Meier estimates of the survival probability, users can use the following code:

```
%lifetest(indata = data, event = dead,
   intv = intxsurv, timelist = 12 36 60);
%lifetestexport(outdata = report,
   rtftitle = "Survival analysis for overall survival");
```

Among the parameters of the %LIFETEST function, parameter INDATA specifies the name of the data set used in the analysis; EVENT specifies the variable name for the event, which is overall survival here - The label of the EVENT variable, when available, will be used in the final report as the label of the event; INTV defines the time for the events; TIMELIST specifies the time points at which Kaplan-Meier estimates will be measured, with multiple time points separated by spaces.

For the parameters of the %LIFETESTEXPORT function, parameter OUTDATA specifies the file name of the final report (in RTF format); RTFTITLE specifies the title of the report table.

The Kaplan-Meier estimates at the specified time points will be generated and the confidence limits will be calculated via arcsine-square root transformation.

The final report will look like follows (Table 1):

	Study population (N = 1969)	
Outcomes	N Eval	Prob (95% CI)
Overall survival	1967	
1-year		58 (56-60)%
3-year		43 (40-45)%
5-year		37 (34-39)%

Table 1. Survival analysis for overall survival

THE CUMULATIVE INCIDENCE FUNCTION

Similarly, to create reports for cumulative incidence estimates, users can use the following code:

```
%lifetest(indata = data, event = rel, competerisk = trm,
   intv = intxrel, timelist = 12 36 60);
%lifetestexport(outdata = report,
   rtftitle = "Survival analysis for relapse");
```

The additional parameter added here is COMPETERISK, used to specify the competing risk variable to the event of interest, as defined by EVENT. Once the COMPETERISK parameter is specified, the macro %LIFETEST will automatically call the enhanced %CIF macro internally to estimate the cumulative incidence function instead of the Kaplan Meier estimates.

The estimates of the cumulative incidence function at the specified time points will be generated, as shown in Table 2. The confidence limits of each estimate will be calculated via arcsine-square root transformation.

	Study population (N = 1969)	
Outcomes	N Eval	Prob (95% CI)
Relapse	1936	
1-year		36 (34-38)%
3-year		43 (41-45)%
5-year		46 (43-48)%

Table 2. Survival analysis for relapse

COMBINING ESTIMATES OF MULTIPLE OUTCOMES

Users can easily combine the estimates of multiple events in a single report by executing multiple %LIFETEST statements before calling the %LIFETESTEXPORT function:

```
%lifetest(indata = data, event = dead,
   intv = intxsurv, timelist = 12 36 60);
%lifetest(indata = data, event = rel, competerisk = trm,
   intv = intxrel, timelist = 12 36 60);
%lifetestexport(outdata = report,
   rtftitle = "Survival analyses for overall survival and relapse");
```

The report will look like follows (Table 3):

	Study population (N = 1969)	
Outcomes	N Eval	Prob (95% CI)
Overall survival	1934	
1-year		51 (49-54)%
3-year		40 (38-42)%
5-year		36 (33-38)%
Relapse	1936	
1-year		36 (34-38)%
3-year		43 (41-45)%
5-year		46 (43-48)%

Table 3. Survival analyses for overall survival and relapse

COMPARING DIFFERENT GROUPS

To compare the Kaplan-Meier estimates of the survivor function or the estimates of the cumulative incidence function across different groups, users can use the following code:

```
%lifetest(indata = data, strata = stage, event = dead,
   intv = intxsurv, timepoint = 12 36 60);
%lifetest(indata = data, strata = stage, event = rel, competingrisk = trm,
   intv = intxrel, timepoint = 12 36 60);
%lifetestexport(outdata = report,
   rtftitle = "Overall survival and relapse, grouped by disease stage");
```

The additional parameter STRATA of the %LIFETEST function can be used to specify the group variable - or the main effect variable in most cases. If available, the format of the group variable will be used to label the header of the report table. By default, logrank test will be used to estimate the differences across multiple survivor functions, while Gray's test will be used when comparing cumulative incidence functions. Their p-values will be provided in the report table, as shown in Table 4.

	Early (N = 1091)		Intermediate (N = 344)		Advanced (N = 534)		
Outcomes	N Eval	Prob (95% CI)	N Eval	Prob (95% CI)	N Eval	Prob (95% CI)	p-value
Overall survival	1077		343		514		<0.001
1-year		60 (57-63)%		58 (52-63)%		29 (25-33)%	
3-year		49 (46-52)%		45 (40-50)%		19 (16-23)%	
5-year		44 (41-47)%		40 (35-46)%		15 (12-19)%	
Relapse	1079		343		514		<0.001
1-year		28 (25-31)%		31 (26-36)%		55 (51-60)%	
3-year		35 (32-38)%		40 (35-45)%		62 (57-66)%	
5-year		38 (35-41)%		43 (37-48)%		64 (59-68)%	

Table 4. Overall survival and relapse, grouped by disease stage

THE POINT-WISE P-VALUE

Apart from generating logrank p-values and Gray's test p-values to evaluate the differences across the entire survivor/cumulative incidence function, the macros also support to generate point-wise p-values at any specified time point. The point-wise p-values are calculated based on the estimates of the survivor/cumulative incidence function and their standard errors.

Point-wise p-values are hidden by default. Users can add point-wise p-values in the report by setting the parameter PWPVALUE = 1 of the %LIFETEST function:

```
%lifetest(indata = data, strata = stage, event = dead,
   intv = intxsurv, timelist = 12 36 60, pwpvalue = 1);
%lifetest(indata = data, strata = stage, event = rel, competingrisk = trm,
   intv = intxrel, timepoint = 12 36 60, pwpvalue = 1);
%lifetestexport(outdata = report,
   rtftitle = "Overall survival and relapse, grouped by disease stage");
```

The final report will look like follows (Table 5):

	Early (N = 1091)		Intermediate (N = 344)		Advanced (N = 534)		
Outcomes	N Eval	Prob (95% CI)	N Eval	Prob (95% CI)	N Eval	Prob (95% CI)	p-value
Overall survival	1089		344		534		<0.001
1-year		67 (64-70)%		65 (60-70)%		36 (32-40)%	<0.001
3-year		51 (48-54)%		48 (43-53)%		22 (19-26)%	<0.001
5-year		45 (42-48)%		42 (36-47)%		17 (14-20)%	<0.001
Relapse	1079		343		514		<0.001
1-year		28 (25-31)%		31 (26-36)%		55 (51-60)%	<0.001
3-year		35 (32-38)%		40 (35-45)%		62 (57-66)%	<0.001
5-year		38 (35-41)%		43 (37-48)%		64 (59-68)%	<0.001

Table 5. Overall survival and relapse, grouped by disease stage

SECONDARY COMPARING GROUP

A secondary group variable can be added to the report by specifying the parameter SUBGROUP of the %LIFETEST function. If available, the format of the secondary group variable will be used to label the first column of the report table. Depending on whether the competing risk is defined, logrank p-values and/or Gray's test p-values will be provided for each subgroup. This could be quite helpful when doing cross comparisons:

```
%lifetest(indata = data, strata = stage, subgroup = graftypecat,
   event = dead, intv = intxsurv, timelist = 12 36 60);
%lifetestexport(outdata = report,
   rtftitle = "Overall survival, grouped by disease stage and graft
source");
```

The final report will look like follows (Table 6):

	Early (N = 1091)		Intermediate (N = 344)		Advanced (N = 534)		
Outcomes	N	Prob (95% CI)	N	Prob (95% CI)	N	Prob (95% CI)	p-value
Overall survival							
Bone marrow	95		24		43		<0.001
1-year		65 (55-75)%		67 (47-84)%		40 (26-54)%	
3-year		54 (43-64)%		NE		NE	
5-year		49 (38-60)%		NE		NE	
PBSC	994		320		491		<0.001
1-year		67 (64-70)%		65 (59-70)%		36 (32-40)%	
3-year		51 (47-54)%		48 (43-54)%		22 (18-26)%	
5-year		45 (41-48)%		42 (37-48)%		17 (14-21)%	

Table 6. Overall survival, grouped by disease stage and graft source

MISCELLANEOUS CONFIGURATIONS

Less than 15 Observations Left

Some entries in the table will show “NE” (not estimable) by default (as shown in Table 6) if there are less than 15 observations left at the specified time point. In such cases, the number of observations is considered too small to have enough statistical power to do any valid survival analysis. Users can nevertheless force the macro to show the estimates for these entries by specifying the parameter LEFTCHECK = 0 of the %LIFETEST function. Since doing so would often results in producing very wide confidence intervals, a warning message will be left in the log file to indicate that there are less than 15 observations left.

Specified Time Point beyond 6 Months from the Actual Last Event

A footnote will be appended to the final report if the last event happening before a specified time point is more than certain threshold earlier than the specified time point. By default, the threshold is set to be 6 months. For example, a user may want to know the Kaplan-Meier estimates at the 5-year time point while the actual time of the last event is 52 months. In such cases, the actual time - 52 months - will be denoted in the footnote.

Users can overwrite the default value for the threshold by specifying the parameter FOOTNOTE of the %LIFETEST function to any value ≥ 0. If FOOTNOTE is set to be 0 however, no footnote will be appended to the report.

Hiding p-values

User can hide all p-values in the final report by setting %LIFETESTEXPORT parameter PVALUE = 0.

CONCLUSION

By using the %LIFETEST and %LIFETESTEXPORT macros, users can conduct survival analyses in a unified way for lifetime data with or without competing risks. The macros provide users an easy-to-use yet highly flexible way to create publication-ready survival analysis reports without the hassle of manually inputting and formatting results each time when the analysis is redone. By doing so, the macros can greatly improve the workflow of analyzing lifetime data.

REFERENCES

1. Kaplan, E.L. and P. Meier, *Nonparametric Estimation from Incomplete Observations.* J Am Stat Assoc, 1958. **53**: p. 457-481.
2. Klein, J.P. and M.L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data*. 2003, New York: Springer-Verlag.
3. Lin, G., Y. So, and G. Johnston, *Analyzing Survival Data with Competing Risks Using SAS® Software*, in *SAS Global Forum 2012*. 2012: Orlando, Florida.
4. Gray, R., *A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk.* The Annals of Statistics, 1988. **16**: p. 1141-1154.

CONTACT INFORMATION

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

Zhen-Huan Hu, MPH
Center for International Blood and Marrow Transplant Research
Medical College of Wisconsin, Milwaukee, WI
Email: zhu@mcw.edu

The data presented here are preliminary and were obtained from the Statistical Center of the Center for International Blood and Marrow Transplant Research. The CIBMTR is a research collaboration between the National Marrow Donor Program/Be The Match and the Medical College of Wisconsin. The analysis has not been reviewed or approved by the Advisory or Scientific Committees of the CIBMTR.