

A SAS® Macro to Provide Survival Functions along with Cox Regression Fast and Easy

Chia-Ling Ally Wu, Seattle Genetics

ABSTRACT

Depending on study design and analysis needs, multiple statistics generated from different SAS® procedures may be needed for time-to-event analyses. This paper describes a SAS® macro that combines those SAS® procedures to generate survival functions and the Cox proportional hazard ratio in one shot, which can help users save time in coding and generate a customized output quickly and easily. Details of the execution and application of the macro will be demonstrated through examples. The macro call returns two different output layouts, i.e., optional estimates in rows and independent variables in columns, or reversely, the independent variables in rows and estimates in columns. Macro users can choose the optional estimates, such as the number of subjects at risk, the number of events and censored observations, the quartiles of the survival function, the coefficient of predictors, the Cox proportional hazard ratio, and the corresponding p-value.

This macro applies the SAS® PROC LIFETEST procedure to compute the survival function by the log-rank test and the Wilcoxon test, as well as running the PROC PHREG procedure based on the Cox proportional regression model to estimate the effect of predictors on hazard ratio.

INTRODUCTION

Time to event is one of the most often used analyses in the oncology field. In time-to-event data, investigators follow subjects until they reach a pre-specified endpoint (for example, death or progression of disease). Subjects sometimes withdraw from a study, or the study is completed before the endpoint is reached for some subjects. In these cases, the survival times are censored. The uncensored survival times are referred to as event times. Unlike with ordinary statistical tests and modeling methods, survival analysis considers event status as well as the time of event. The survival and hazard functions are key concepts in survival analysis for describing the distribution of event times. The survival function is the probability of surviving up to that time. The hazard function gives the potential that the event will occur, per time unit, given that an individual has survived up to the specified time (Despa, 2010). The Cox Proportional model is the most commonly used multivariable approach for analyzing survival data in medical research. It is essentially a time-to-event regression model, which describes the relation between the event incidence, as expressed by the hazard function, and a set of covariates.

In SAS®, the PROC LIFETEST procedure creates a Kaplan-Meier survival plot, computes the survival function from lifetime data, and compares the survivor function between groups by the log-rank test and the Wilcoxon test. The PROC PHREG procedure is used in the Cox proportional regression model to estimate the effect of predictors on hazard rates. Using SAS® ODS SELECT in PROC LIFETEST and PROC PHREG provides statistical information in data set format for further analysis use.

PROC LIFETEST and PROC PHREG procedures provide many optional functions to fulfill analysts' needs, but it may take time for users to understand and use functions appropriately. Besides, the estimates come from so many ODS SELECT datasets in the PROC LIFETEST and PROC PHREG procedures that users may need to take time to combine all the information. Hence, this paper would like to present a macro, *mCox*, which creates a customized Cox regression table and provides estimates in one step.

All the examples in this macro use the Bone Marrow Transplant Patients data <*Sashelp.bmt*> from the SAS library. At the time of transplant, each patient is classified into one of three risk categories: ALL (acute lymphoblastic leukemia), AML-Low Risk (acute myelocytic leukemia, low risk), and AML-High Risk, which are defined in the GROUP variable. The endpoint of interest is the disease-free survival time, which is the time in days to death, relapse, or the end of the study, which is defined in the STATUS variable (1 is an event and 0 is censored). In order to demonstrate, the author created dummy AGE (from ages 5 to 15) and SEX variables.

OVERVIEW OF THE MACRO

The SAS® macro in this paper, *mCox*, aims to create a dataset using Cox regression modeling and survival analysis, which provide descriptive statistical estimates including number of total subjects, event and censored, and the quantiles (25%, median, and 75%), minimum and maximum of the survival function, as well as the Cox proportional hazard ratio with 95% CI and the corresponding p-value. Those estimates are combined in a dataset, and can be assigned to present in horizontal or vertical style. The macro enables requesting an optional dataset with the Cox proportional hazard ratio at the customized level of the variable, i.e. interaction term. For example, if we consider that male and female may have different hazard ratios among disease types, we would like to see the hazard ratio among disease types when sex is male and female separately. In addition, a filter for the data is included to allow subsetting analysis data.

Figure 1, Figure 2, and Figure 3 present the output datasets generated by the macro. This is the example modeling disease-free survival among ALL, AML-Low risk and AML-High risk, taking age, sex and interaction of age with disease group into account. The event of interest is disease relapse including death. Figure 1 presents independent variables in the parameter column, the labels captured from the coded value in parameter value, with the survival and Cox hazard ratio statistical estimates depicted horizontally. Figure 2 reversely, presents main effect - disease groups in columns, and presents statistical estimates vertically. The user can choose either Figure 1 or 2 by using the macro parameter OutHorizontal: Y is presented in Figure 1, N as in Figure 2, and the default is Y. Figure 3 offers the Cox hazard ratio between groups at different levels of interaction age. Figure 3 is optional and requires an input value in the macro parameter CovLevel. In this example, the author used aged 5, 10, 15 years old. The code for these examples is provided in the example section of this paper.

Parameter	Parameter Value	HR Coefficient (SE)	HR Estimates (95%CI)	HR p-value	Total Subjects	Censored N (%)	Event N(%)	Survival Q1 (95%CI)	Survival Q2 (95%CI)	Survival Q3 (95%CI)	Survival Min	Survival Max
Group	ALL				76	28 /76 (36.8)	48 /76 (63.2)	4.1 (3.6, 6.5)	13.9 (7.7, 20.3)	-(22.1, -)	0.03	69.37
Group	AML-Low Risk	-1.8 (0.7)	- (-,-)	0.01	108	58 /108 (53.7)	50 /108 (46.3)	13.0 (7.3, 16.2)	73.5 (24.9, -)	- (-,-)	0.33	85.63
Group	AML-High Risk	0.1 (0.6)	- (-,-)	0.84	90	22 /90 (24.4)	68 /90 (75.6)	2.8 (2.1, 3.8)	6.1 (4.0, 10.6)	22.6 (14.1, -)	0.07	88.00
Age	AGE	-0.1 (0.0)	- (-,-)	0.14								
Sex	M	-0.3 (0.2)	0.7 (0.5,1.0)	0.05								
Age*group	Age*AML-Low Risk	0.1 (0.1)	- (-,-)	0.05								
Age*group	Age*AML-High Risk	0.0 (0.1)	- (-,-)	0.58								

Figure 1. Disease-free survival function and Cox regression statistical estimates presented horizontally

Parameter Value	ALL	AML-Low Risk	AML-High Risk
Total Subjects	76	108	90
Censored N(%)	28 /76 (36.8)	58 /108 (53.7)	22 /90 (24.4)
Event N(%)	48 /76 (63.2)	50 /108 (46.3)	68 /90 (75.6)
Survival Q1 (95%CI)	4.1 (3.6, 6.5)	13.0 (7.3, 16.2)	2.8 (2.1, 3.8)
Survival Q2 (95%CI)	13.9 (7.7, 20.3)	73.5 (24.9, -)	6.1 (4.0, 10.6)
Survival Q3 (95%CI)	-(22.1, -)	- (-,-)	22.6 (14.1, -)
Survival Min	0.03	0.33	0.07
Survival Max	69.37	85.63	88.00
HR Coefficient (SE)		-0.7 (0.7)	0.9 (0.6)
HR Estimates (95%CI)		- (-,-)	- (-,-)
HR p-value		0.32	0.15
HR Coefficient (SE): AGE	0.0 (0.0)		
HR Estimates (95%CI): AGE	- (-,-)		
HR p-value: AGE	0.92		
HR Coefficient (SE): AGE*GROUP		0.0 (0.1)	-0.1 (0.1)
HR Estimates (95%CI): AGE*GROUP		- (-,-)	- (-,-)
HR p-value: AGE*GROUP		0.97	0.35
HR Coefficient (SE): M	-0.3 (0.2)		
HR Estimates (95%CI): M	0.7 (0.5,1.0)		
HR p-value: M	0.05		

Figure 2. Disease-free survival function and Cox regression statistical estimates presented vertically

Description	HR Estimates (95%CI)
GROUP 0 vs 1 At AGE=5	1.2 (0.6, 2.4)
GROUP 0 vs 2 At AGE=5	1.0 (0.5, 2.0)
GROUP 1 vs 2 At AGE=5	0.8 (0.4, 1.6)
GROUP 0 vs 1 At AGE=10	1.9 (1.2, 2.8)
GROUP 0 vs 2 At AGE=10	0.7 (0.5, 1.0)
GROUP 1 vs 2 At AGE=10	0.4 (0.3, 0.6)
GROUP 0 vs 1 At AGE=15	2.9 (1.2, 7.1)
GROUP 0 vs 2 At AGE=15	0.5 (0.2, 1.0)
GROUP 1 vs 2 At AGE=15	0.2 (0.1, 0.4)

Figure 3 Cox proportional hazard ratio at different levels of interaction term

MACRO PARAMETERS

Table 1 describes the parameters required or optional parameters in the macro. To generate the basic descriptive survival function and hazard function in univariate Cox regression, only the input dataset name, time, censoring variable and the value that identifies censored observations, and analysis group identifier (parameter InData, TimeVar, CensorVar, and StrataVar respectively) are required. When the user wants to take more independent variables into the model, then covariates (CovVarN, CovVarC) or the interaction with the main effect (IntVar) may be needed to specify user needs.

The macro expects that the dataset contains both the code and the decode variables for macro variable StrataVar. In the example below, the dataset should have variable GROUP and GROUP_C. The numeric variable, GROUP, would have values of 1, 2 and 3, while the character variable, GROUP_C, would have values ALL, AML-Low Risk and AML-High Risk, respectively. Same rule applies to CovVarC, where SEX and SEX_C are code 0/1 and decode Male and Female. This information would pass to PROC FORMAT to create labels in the table.

Macro Variable	Required	Description of Function
_InData	Y	Input dataset assumed to be one record per subject.
_TimeVar	Y	Time at risk variable.
_CensorVar	Y	Censoring variable should include the value to identify a censored observation. For instance, STATUS (0) identifies 0 as censored data and 1 as the event of interest.
_StrataVar	Y	Analysis effect for KM analysis and the Cox regression. For example, disease group (ALL, AML high and AML low) in this paper, requires numeric format.
_Ref		Study reference group. For example, disease group 0 (ALL) in this paper. Default is the last sorted value in the dataset.
_CovVarN		Numeric covariates in the modeling. In the paper, the example is age at onset of disease.
_CovVarC		Characteristics covariates in the modeling. In the paper, the example is sex.
_CovLevel		Shows the hazard ratio in the model in customized settings, which assigns to HAZARDRATIO statement in the PROC PHREG. This paper presents the hazard ratio between groups aged 5, 10, and 15 years old.
_IntVar		Specifies the interaction terms in the modeling. This paper takes age with disease group as an interaction example.
_DataFL		Selects partial data to do survival analysis. Acts in the WHERE statement in the PROC LIFETEST and PROC PHREG procedure, should be provided in SAS syntax, for example, AGE>45.
_OutHorizontal		Selects whether the output dataset is presented horizontally or vertically. Default is Y.

Table 1. Definition of the parameters and optional functions.

This macro uses the PROC LIFETEST procedure to have number of total subjects, events, and censored subjects, the 25%, median, 75% of survival function and its 95% CI, as well as min and max of the survival function.

```
ods output survivalplot=zz_plotdata CensoredSummary=zz_esum Quartiles=zz_qtp;
proc lifetest data=&_INDATA
  %if %bquote(&_DATAFL.) ne %then %do; %str((where=(&_DATAFL.))) %end;
method=pl;
ods select SurvivalPlot CensoredSummary Quartiles;
time &_TIMEVAR. * &_CENSORVAR.;
strata &_STRATAVAR.;
run;
```

This macro applies the PROC PHREG procedure to derive the Cox proportional hazard ratio, 95% CI and p-value for each independent variable. The user can specify the main effect, covariates and interactions in the model. The user can also assign to present the hazard ratio at any level of the interaction term.

In this macro, Cox proportional regression estimates are produced using the EXACT method, which is the most accurate method for the continuous time-scale model, but also takes more computational resource. The hazard ratio is calculated based on the reference group specified in the macro, or the default value is like the SAS default using the last sorted group if not specified.

```
ods output ParameterEstimates=zz_hazard HazardRatios=zz_hr;
proc phreg data=&_INDATA. %if %bquote(&_DATAFL.) ne %then
  %str((where=(&_DATAFL.)));;
ods select ParameterEstimates
  %if (%bquote(&_COVVARN.) ne and %bquote(&_COVLEVEL.) ne) %then
HazardRatios;;
class &_STRATAVAR. (ref="&_REF.") &_COVVARC.;
model &_TIMEVAR.* &_CENSORVAR.= &_STRATAVAR. &_COVVARN. &_COVVARC.
  %if %bquote(&_INTVAR.) ne %then &_INTVAR.;
  /ties=EXACT risklimits alpha=0.05;
hazardratio &_STRATAVAR.
  %if %bquote(&_COVVARN.) ne and %bquote(&_COVLEVEL.) ne %then /
  diff=all at(&_COVVARN.=&_COVLEVEL.);;
run;
```

EXAMPLES

Figure 1 and Figure 3 offer a multivariate model of disease-free survival function and Cox regression estimates, presented horizontally. We present the code here. It shows disease-free survival among ALL, AML low risk and high risk subjects, after controlling for age, sex and the interaction between age and disease type. The censor variable STATUS will be 1 if the subject relapsed or died during the study, and 0 if the subject remained disease free or is lost to follow up. The user also has to specify the time variable to be used in the PROC LIFETEST and PROC PHREG model. The survival estimates are based on disease stratification and ALL (uncoded value 0) as reference group.

```
%mCox ( _InData= BMT
  ,_TimeVar= T_M
  ,_CensorVar= STATUS(0)
  ,_StrataVar= GROUP
  ,_Ref= 0
  ,_CovVarn = AGE
  ,_CovVarc = SEX
  ,_CovLevel = %str(5, 10, 15)
  ,_IntVar = AGE*GROUP
  ,_DataFL=
  ,_OutHorizontal = Y);
```

Figure 4 and Figure 5 present univariate modeling. Figure 4 does not assign a reference group, so the last group (AML high risk, uncoded value 2) been applied as SAS® default, and it presents horizontally as the macro default. In contrast, Figure 5 assigns ALL (uncoded value 0) as the reference group, and we can see that the hazard ratio trend also reverses.

Parameter	Parameter Value	HR Coefficient (SE)	HR Estimates (95%CI)	HR p-value	Total Subjects	Censored N (%)	Event N(%)	Survival Q1 (95%CI)	Survival Q2 (95%CI)	Survival Q3 (95%CI)	Survival Min	Survival Max
Group	AML-High Risk				90	22/90 (24.4)	68/90 (75.6)	2.8 (2.1, 3.8)	6.1 (4.0, 10.6)	22.6 (14.1, -)	0.07	88.00
Group	ALL	-0.4 (0.2)	0.7 (0.5, 1.0)	0.04	76	28/76 (36.8)	48/76 (63.2)	4.1 (3.6, 6.5)	13.9 (7.7, 20.3)	-(22.1, -)	0.03	69.37
Group	AML-Low Risk	-1.0 (0.2)	0.4 (0.3, 0.6)	0.00	108	58/108 (53.7)	50/108 (46.3)	13.0 (7.3, 16.2)	73.5 (24.9, -)	-(, -)	0.33	85.63

Figure 4. Univariate model of disease-free survival function and Cox regression presented horizontally

```
%mCox ( _InData= BMT
, _TimeVar= T_M
, _CensorVar= STATUS(0)
, _StrataVar= GROUP);
```

Parameter Value	ALL	AML-Low Risk	AML-High Risk
Total Subjects	76	108	90
Censored N(%)	28/76 (36.8)	58/108 (53.7)	22/90 (24.4)
Event N(%)	48/76 (63.2)	50/108 (46.3)	68/90 (75.6)
Survival Q1 (95%CI)	4.1 (3.6, 6.5)	13.0 (7.3, 16.2)	2.8 (2.1, 3.8)
Survival Q2 (95%CI)	13.9 (7.7, 20.3)	73.5 (24.9, -)	6.1 (4.0, 10.6)
Survival Q3 (95%CI)	-(22.1, -)	-(, -)	22.6 (14.1, -)
Survival Min	0.03	0.33	0.07
Survival Max	69.37	85.63	88.00
HR Coefficient (SE)		-0.6 (0.2)	0.4 (0.2)
HR Estimates (95%CI)		0.6 (0.4, 0.8)	1.5 (1.0, 2.1)
HR p-value		0.00	0.04

Figure 5. Univariate model of disease-free survival function and Cox regression presented vertically

```
%mCox ( _InData= BMT
, _TimeVar= T_M
, _CensorVar= STATUS(0)
, _StrataVar= GROUP
, _Ref= 0
, _OutHorizontal = N);
```

CONCLUSION

This paper presents a SAS® macro to generate customized datasets combining survival and Cox regression time-to-event analyses.

REFERENCES

- Despa, S. (2010). What is Survival Analysis? Stats New #78, Cornell University. (<https://www.cscu.cornell.edu/news/statnews/stnews78.pdf>)
- Hosmer D.W., Lemeshow S., and May S., Applied Survival Analysis: Regression Modeling of Time-to-Event Data, Wiley, 2008.
- Wu, C. A (2018). A SAS® Macro to Generate Information Rich Kaplan-Meier Plots. PharmaSUG 2018. Paper presented at PharmaSUG 2018, Seattle, WA, US, 29 April - 2 May (AD25).

ACKNOWLEDGMENTS

I would like to sincerely thank to my manager Karen Chen, and Statistical Programming Senior Director Michiel Hagendoorn for reviewing our paper and providing valuable feedback. I also would like to thank to Linda MacKeen, Gordon Fancher and Sreeram Kundoor for their inputs to refine this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. For a copy of the macro contact the author at:

Chia-Ling Ally Wu
Seattle Genetics, Inc.
21823 30th Drive S.E. Bothell, WA 98021
425-527-2321
awu@seagen.com

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.