



# SAS® Macros on Performing Competing Risks Survival Data: CIF Plot, Backward Elimination Fine & Gray's Model and Cause-Specific Hazard Model

Chao Zhang<sup>1\*</sup>, Yuan Liu<sup>1,2\*</sup> and Yaqi Jia<sup>3</sup>

\*Correspondence: [chao.zhang2@emory.edu](mailto:chao.zhang2@emory.edu) and [yliu31@emory.edu](mailto:yliu31@emory.edu)



CrossMark

Click for updates

<sup>1</sup>Biostatistics and Bioinformatics Shared Resource, Winship Cancer Institute, Emory University, Atlanta GA.

<sup>2</sup>Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta GA.

<sup>3</sup>Division of Nephrology, School of Medicine, Johns Hopkins University, Baltimore.

## Abstract

SAS/STAT® 14.1 released in SAS® 9.4 TS1M3 can perform non-parametric method to calculate cumulative incidence function (CIF) and can create CIF plots by using PROC LIFETEST, but it can't directly show the information such as number at risk. And in PROC PHREG, both Fine and Gray's sub-distribution hazard model and cause-specific model are fitted when there is competing risk events. Up to date, there is no build-in variable selection procedure for the two kinds of models in SAS®. To address this limitation, we developed SAS® macros to apply Fine and Gray's approach and cause-specific method with backwards elimination. The two macros generate summary reports in the form of tables displaying results that include, Hazard Ratio in the case of cox regression with its 95% confidence intervals, sample size, and p-values. We also created the third SAS macro to show CIF plot. Comparing with the CIF plot generated by default SAS® procedure, our SAS® CIF Plot macro provides more information including Gray's test for the significant differences among the levels of a categorical covariate, a table with number at risk, accumulated number of events of interest, accumulated -events of competing for specified time-points at the bottom of the graph. And censored information also is shown in the CIF curve. The three macros used for competing risk survival data enable us to generate journal-quality summary tables and graphs, production of comprehension.

**Keywords:** Competing Risk, CIF, Number at Risk, Backward Elimination, Fine and Gray's Model, Cause-Specific Hazard Model

## Introduction

In analysis for the time to event data, competing risk arises when the event of interest cannot be observed due to a preceding event occurred before. For example, in a study with event of interest being a cause specific death, death from any other causes are competing events. Cumulative incidence function (CIF) is the cumulative probability of failure from a specific cause over time. Cumulative incidence functions were shown by cumulative incidence curves. These curves begin at 0 and calculate the probability of that event happening. The current SAS/STAT 14.1 can perform non-parametric approach to cal-

culate CIF and can create a CIF plot by using PROC LIFETEST. An important feature of the survival plots is the display of the subjects at risk at intervals along the time axis. But the current SAS procedure can't show this information. Fine and Gray's sub-distribution hazard model [1] is a popular method to directly model the effect of covariates on the cumulative incidence functions and it has been available in the PHREG procedure since SAS/STAT 13.1. While the current SAS procedure can't perform backward selection for a multivariate analysis. The purposes of the biostatisticians in our core are set out to produce high quality, professional looking analysis reports to

enhance communication and collaboration with investigators. We created the three macros that could produce tables and/or graphs in a rich text format (RTF) file with the goal of condensing the relevant information. It is easy to understand for clients or investigators [2]. The three SAS macros can also help generate summary results in a publishable format to enhance communication with non-statisticians and to save time for statisticians to process a set of independent variables.

The paper is organized as below: The next section briefly describes the parameters definition of the macros. The following sections taking bone marrow transplant (BMT) data as an example showed how to perform the three macros and what the outputs are respectively.

## Method

### The macro %Plots\_CIF

This macro is using PROC LIFETEST [3] to obtain nonparametric estimates of the cumulative incidence probabilities in the competing survival data. The macro has 13 parameters in which 6 parameters must be provided by the users when the macro is called: The parameters are listed and illustrated in the word file, i.e., Plots\_CIF V2, and we can download it from <https://doi.org/10.5281/zenodo.3332931>.

### The macro of %FINEGRAY\_SEL

The PHREG procedure since SAS/STAT 13.1 [4] enables to fit the sub-distribution hazard model and to predict the cumulative incidence function. But it can't conduct variables selection automatically. To address this limitation, the SAS® macro, %FINEGRAY\_SEL, applies Fine and Gray's approach with backwards elimination. This extends the capability of the Fine and Gray analysis in PROC PHREG which currently does not support any model selection capability. The %PSHREG macro [5] implements the method discussed in [6] for fitting the Fine and Gray model as a weighted Cox regression using the PHREG procedure. As such, this macro can perform model selection based on PROC PHREG's native model selection methods for the Cox regression including backward selection, etc. While the %FINEGARY\_SEL macro essentially implements the backward selection technique and utilizes the Fine and Gray model and its sandwich-type variance estimator as provided by the PHREG procedure, the %PSHREG macro uses instead the model-based variance from the Cox regression. There are 16 parameters in which 8 parameters must be provided by the users when the macro is called. The parameters of the macro %FineGray\_Sel are displayed in the word file, FineGray\_Sel V4. We can also find it in the same folder as 'Plots\_CIF V2' above.

### The macro of %CAUSE\_SPECIFIC\_SEL

This macro was developed by calling PROC PHREG to fit the cause-specific hazard models. For a particular cause of interest, we need to treat all the competing events as censored observations in the analysis [4]. The risk set at each time includes only those subjects who have not failed from competing events or

are truly censored [7]. The parameters of this macro are listed in the file, Cause\_Specific\_Sel V1.

## Case Study

**Study Background:** We use the bone marrow transplant (BMT) data [8] as a case study to illustrate the usage of the macros. The dataset contains 137 patients on bone marrow transplant, grouped into three disease types: acute lymphoblastic leukemia (ALL) (n=38), acute myelocytic leukemia (AML) low-risk (n=54), and AML high-risk (n=45). In this study, the goal is to detect whether disease free survival (days) is different among the three types of patients. Relapse is interest, and death during remission is competing-risk.

To perform the macro %FineGray\_Sel on multivariable back selection, we create two artificial predictors, i.e., sex, and race. The following SAS statements create the BMT dataset which is used to introduce the SAS macros in this paper:

```
%let dir= H:\fine_gray\output;
proc format;
value DiseaseGroup 1='ALL' 2='AML-Low Risk' 3='AML-High Risk';
value sex 0='F' 1='M';
run;

data bmt;
input Group T Status WaitTime @@;
logWaittime=log(WaitTime);
datalines;
1 2081 0 98 1 1602 0 1720 1 1496 0 127 1 1462 0 168 1 1433 0 93
*** more lines ***
3 625 1 150 3 48 1 2103 273 1 240 3 63 2360 3 76 1 330
3 113 1 2403 363 2 180
;
run;

data bmt; set bmt;
call streaminit(123); /* set random number seed */
u = rand("Uniform"); /* u ~ U[0,1] */
output;
run;

data bmt; set bmt;
if u>=0.7 then sex=1; elseif u<0.5 then sex=0;
if u>=0.4 then race='white'; else race='AA'; Dftime= t / 365;
label sex = 'gender' group = 'patient group'
T = 'Disease Free survival (days)';
format Group DiseaseGroup. SEX SEX.;
run;
```

The dataset BMT contains the following variables:  
T: the following-up time after marrow transplant in days.  
Status: indicator variable with the value 0 for censoring,

1 for relapse (interest), and 2 for death during remission (competing-risk).

Group: AML low-risk, AML high-risk, and ALL.

WaitTime: is the waiting time for transplant, in days. And log transformed Wait time.

The last two artificial predictors: race and gender.

**SAS code example of %PLOT\_CIF**

The following SAS statements use the macro % PLOT\_CIF to create a plot on displaying the cumulative incidence curves.

**Figure 1** shows the CIF for relapse for the three groups (AML low-risk, AML high-risk, and ALL). We added plus (“+”) sign in the CIF curves to indicate censored subjects. Gray’s test p-value is also reported. An important feature of the CIF plot is the display of the subjects at risk at intervals along the time axis. A table with number at risk, accumulative number of events of interest, accumulative events of competing for specified time-points is presented at the bottom of the graph. Taking AML-High Risk group as an example, at T=0 day, the number at risk was 45, the number of events of relapse was 0 and the number of patients on death was 0. At T=1 year, the number at risk is 17, and 16 patients of relapse had been occurred and 12 patients had been dead by 1 year. And there was no any censored patient from 0 to 1 year. In addition, A table with estimates CIF of event at the given time points with 95% CI was displayed. When multiple covariates were listed for the parameter, ‘GRPLIST’; this macro can repeatedly execute to

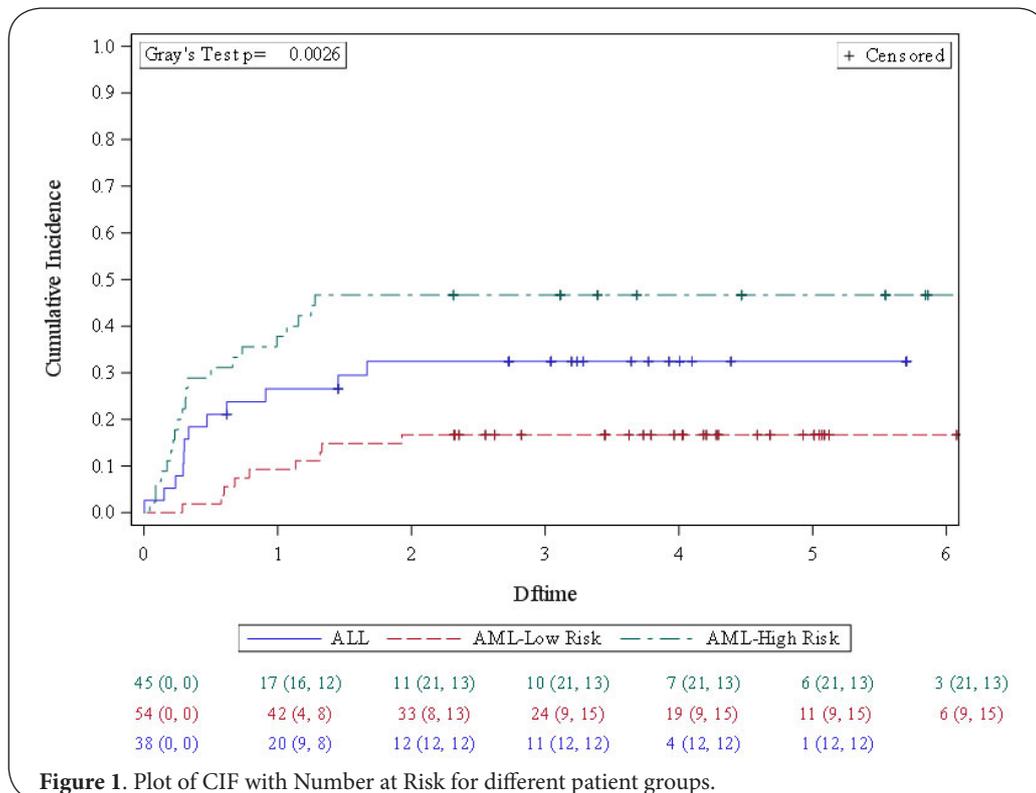
create plots and tables corresponding each covariate, respectively. We only show the plot for disease group here.

Title “Figure 1 Plot of CIF for patient groups “;

```
%Plots_CIF(dsn= bmt,
grplist= group race ,
time_event=Dftime,
censor=status,
eventcode=1,
xaxisvalue=0 1 2 3 4 5 6 ,
yaxisvalue= ,
timelist= 0 2 4 6 ,
NatRisk =T,
units= Years,
filename= CIF PLOTS,
outpath= &dir.\,
debug=T)
Title;
```

**SAS<sup>®</sup> code example of %FineGray\_SEL**

The following code was used to perform the backwards selection on Fine and Gray’s model. To designate relapse (Status=1) as the event of interest, we specify EVENTCODE=1 in the macro. **Tables 1** and **2** shows the condense results including subdistribution hazard ratio and its 95% CI, sample size, and p-values.



**Table 1. CIF Estimate for different patient groups.**

patient group	TIME (Years)	CIF Estimate (95% CI)
ALL	2	0.324 (0.179, 0.479)
	4	0.324 (0.179, 0.479)
	6	NA (NA, NA)
AML-High Risk	2	0.467 (0.314, 0.606)
	4	0.467 (0.314, 0.606)
	6	0.467 (0.314, 0.606)
AML-Low Risk	2	0.148 (0.069, 0.257)
	4	0.167 (0.081, 0.278)
	6	0.167 (0.081, 0.278)

In addition, number of observations in the original data set, number of observations used, removed variables, and selection criteria are reported in the footnote.

**Title** "Table 3.2 Multivariable Survival Analysis for Fine and Gray's Model";

`%finegray_sel` (dsn=bmt,  
 event=t,

```

censor=Status,
var= race Group logWaitTime sex ,
cVar= Group(desc)* race *sex,
event_code=1,
inc = 1,
alpha=0.2,
Type3=t,
debug=F,
outpath = &dir.\,
filename = Multivariable Fine and Gray);
Title;
    
```

**SAS<sup>®</sup> code example of %Cause\_Specific\_SEL**

For a particular cause of interest, we can fit cause-specific hazard model as regular Cox model. Here we need to treat all the competing events as censored observations in the analysis. For example, 0=true censored, 1=relapse, 2=death, if relapse is the event of interest, specify 0 and 2 as the value for the parameter of CENSRED\_VALUE. **Table 3** shows the condense results. In the footnote of the table, it shows that relapse is the event of interest and true censoring and death are treated as censored. If the death is the event of interest,

**Table 2. Multivariable Survival Analysis for Fine and Gray's Model.**

Covariate	Level	N	disease-free survival time (days)		
			Hazard Ratio (95% CI)	HR P-value	Type3 P-value
Race	Black	51	1.32 (0.69-2.53)	0.403	0.403
	White	86	-	-	
Patient group	AML-High Risk	45	1.59 (0.77-3.28)	0.208	<b>0.002</b>
	AML-Low Risk	54	0.38 (0.16-0.94)	<b>0.036</b>	
	ALL	38	-	-	
logWaittime		137	0.75 (0.51-1.10)	0.146	0.146

\* Number of observations in the original data set = 137. Number of observations used = 137.

\*\* Backward selection with an alpha level of removal of 0.2 was used. The following variables were forced in the model: race. The following variables were removed from the model: gender.

**Table 3. Multivariable Cause-Specific Hazard Regression for Relapse.**

Covariate	Level	N	disease-free survival time (days)		
			Hazard Ratio	HR P-value	Global P-value
Group	ALL	34	0.61 (0.27-1.35)	0.221	<b>0.013</b>
	AML-Low Risk	44	0.28 (0.12-0.66)	<b>0.003</b>	
	AML-High Risk	36	-	-	
logWaittime		114	0.80 (0.53-1.21)	0.290	0.290

\*Number of observations in the original data set = 137. Number of observations used = 114.

\*Backward selection with an alpha level of removal of .3 was used. The following variables were removed from the model: gender.

\*Cause-Specific Hazard Model: Type of Censored= 0 2; Type of Interest= 1.

we just need to set CENSORED VALUE=0 1.

**title** 'Table 3.3 Cause-Specific Hazard Regression for Relapse';

```
%CauseSpecific_sel(dsn=bmt,
```

```
event=t,
```

```
censor=status,
```

```
censored_value=02,
```

```
var=Group logWaitTime sex,
```

```
cvar=Group sex,
```

```
inc=0,
```

```
slstay=.3,
```

```
ORIENTATION = portrait,
```

```
outpath= &dir.\,
```

```
filename=Multivariable cause specific hazard model,
```

```
debug=F);
```

```
title";
```

## Conclusion

Backward elimination has an advantage over other selection methods since it is possible for a set of variables to have considerable predictive capability even though a subset of them does not. The three macros for the competing risk survival fill in the limitations by the current version SAS<sup>®</sup> procedures. And they enable users to generate tables and plots that may be directly used for publications, and presentations. They provide results in a format easy for researchers to understand thus enabling efficient communication and collaboration with investigators. However, there are always some gaps that need users' treatment. Any feedback or bug reporting is greatly appreciated.

## Data and Software availability

For the example dataset BMT, we can get it from SAS official [7]. Source code available from [https://github.com/chaozhangEmory/SASMacros\\_CompetingRisk](https://github.com/chaozhangEmory/SASMacros_CompetingRisk).

Archived source code at time of publication: <https://doi.org/10.5281/zenodo.3332931>.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

Authors' contributions	CZ	YL	YJ
Research concept and design	✓	✓	--
Collection and/or assembly of data	--	--	--
Data analysis and interpretation	✓	✓	✓
Writing the article	✓		
Critical revision of the article	✓	✓	✓
Final approval of article	✓	✓	✓
Statistical analysis	✓	✓	✓

## Acknowledgement

Research reported in this publication was supported in part by the Biostatistics and Bioinformatics Shared Resource of Winship Cancer Institute of Emory University and NIH/NCI under award number P30CA138292.

## Publication history

EIC: Jimmy Efird, East Carolina University, USA.

Received: 28-June-2019 Final Revised: 17-Aug-2019

Accepted: 21-Aug-2019 Published: 25-Sep-2019

## References

1. Fine J.P and R.J. Gray. **A proportional hazards model for the subdistribution of a competing risk.** *Journal of the American Statistical Association.* 1999; **94**:496-509. | [Article](#)
2. Nickleach D, L.Y., Shrewsbury A, Ogan K, Kim S and Wang Z. **Macros to Conduct Common Biostatistical Analyses and Generate Reports.** SESUG. 2013.
3. SAS Institute Inc, 2017 SAS/STAT<sup>®</sup> 14.3 User's Guide. Cary, NC: SAS Institute Inc. User's Guide the LIFETEST Procedure.
4. SAS Institute Inc, 2017 SAS/STAT<sup>®</sup> 14.3 User's Guide. Cary, NC: SAS Institute Inc. User's Guide the PHREG Procedure.
5. Kohl M, Plischke M, Leffondre K and Heinze G. **PSHREG: a SAS macro for proportional and nonproportional subdistribution hazards regression.** *Comput Methods Programs Biomed.* 2015; **118**:218-33. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
6. Geskus RB. **Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring.** *Biometrics.* 2011; **67**:39-49. | [Article](#) | [PubMed](#)
7. Changbin G. and S. Ying. **Cause-Specific Analysis of Competing Risks Using the PHREG Procedure.** *Paper SAS2159-2018.* 2018. | [Pdf](#)
8. So, Y., G. Lin and G. Johnston. **Using the PHREG procedure to analyze competing-risks data.** *Paper presented at the SAS Global Forum 2015.* 2015. | [Pdf](#)

## Citation:

Zhang C, Liu Y and Jia Y. **SAS<sup>®</sup> Macros on Performing Competing Risks Survival Data: CIF Plot, Backward Elimination Fine & Gray's Model and Cause-Specific Hazard Model.** *J Med Stat Inform.* 2019; 7:2.

<http://dx.doi.org/10.7243/2053-7662-7-2>