

Paper 1424-2015  
**Competing risk survival analysis using SAS®**  
**When, why and how**

Lovedeep Gondara, University of Illinois Springfield

## ABSTRACT

Competing risk arise in time to event data when the event of interest cannot be observed because of a preceding event i.e. a competing event occurring before. An example can be of an event of interest being a specific cause of death where death from any other cause can be termed as a competing event, if focusing on relapse, death before relapse would constitute a competing event. It is well studied and pointed out that in presence of competing risks, the standard product limit methods yield biased results due to violation of their basic assumption. The effect of competing events on parameter estimation depends on their distribution and frequency. Fine and Gray's sub-distribution hazard model can be used in presence of competing events which is available in PROC PHREG with the release of version 9.4 of SAS® software.

## INTRODUCTION

Competing event can be considered as an alternative outcome that makes the observation of primary or outcome of interest impossible. Standard survival analysis concentrates on only one type of failure in time to event data i.e. it usually censors failures due to other causes which yields biased results. Cumulative incidence estimate proposed by Kalbfleisch and Prentice [10] can be used as a remedy.

$$I_j(t) = \int_0^t f_j(u) du = \Pr\{T \leq t \& J = j\}$$

Which represents the probability that an event of type  $j$  has occurred by time  $t$ .

Following two approaches can be used for introducing covariates in context of competing risks:

1. Apply a cox proportional hazard model to cause specific hazards
2. Use model proposed by Fine and Gray [1] that focuses on cumulative incident function

$$\bar{\lambda}_j(t, x) = \bar{\lambda}_{j0}(t) \exp\{x' \beta_j\}$$

Where

$\bar{\lambda}_{j0}(t)$  is the baseline sub hazard for events of type  $j$

$\exp\{x' \beta_j\}$  is the relative risk associated with covariates  $x$

The partial likelihood of the sub-distribution hazard model was defined by Fine and Gray as

$$L(\beta) = \prod_{j=1}^r \frac{\exp(x_j \beta)}{\sum_{i \in \mathcal{R}_j} w_{ji} \exp(x_i \beta)}$$

Where  $x_j$  is the covariate row vector of the subject experiencing an event of interest at  $t_j$ .

The risk set  $\mathfrak{R}_j$  is defined as

$$\mathfrak{R}_j = \{i; t_i \geq t \vee (t_i \leq t \wedge \epsilon_i = 2)\}$$

And it includes the individuals who at time  $t$  are at risk of event of interest and anyone who had a competing event before time  $t$ .

Time and subject specific weights are used when censoring occurs. Subjects at risk from event of interest (type 1) at time  $t$  and who have not witnessed a competing event before  $t$  have equal weights ( $w_i = 1$ ) and for subjects with competing event at  $t_i < t$  weights are given as  $w_i < 1$ .

Weights for Fine and Gray's model can be calculated as

$$w_{ji} = \begin{cases} 1 & \text{if } t_i \geq t \\ \frac{\hat{G}(t)}{\hat{G}(t_i)} & \text{if } \epsilon_i = 2 \wedge t_i < t \end{cases}$$

Where  $\hat{G}(t)$  denotes estimator of the censoring distribution i.e. cumulative probability of still being followed up at time  $t$ . It can be estimated by usual product limit method by reversing the meaning of censoring indicator.

## WHEN?

Competing risk survival analysis should be considered when the observation of event of interest is made impossible by a preceding competing event, e.g. In case of oncology competing risks are encountered when patients are followed after treatment, and their first failure event may be local recurrence, distant metastases, onset of second primary cancer, or death precluding these events. So if event of interest is any type of relapse, then the observation of this event is made impossible by death preceding relapse. It can be argued that competing risk models provide real world probabilities of death when competing events are present as opposed to standard survival models by allowing us to separate the probability of event into different causes.

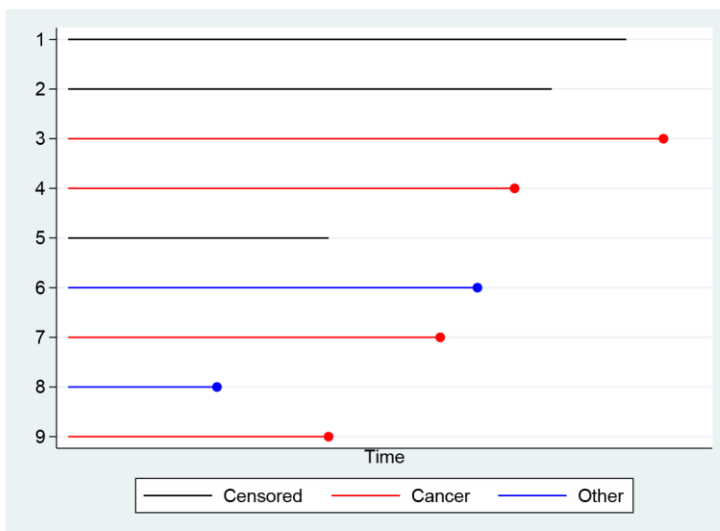
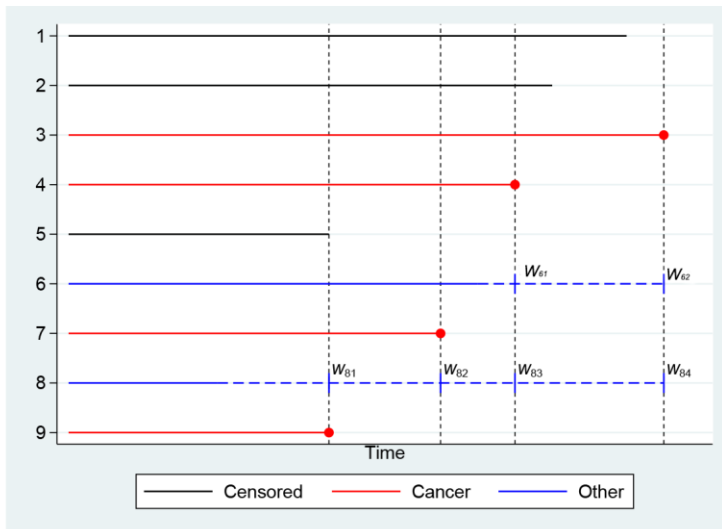


Figure1. Competing risk scenario

Figure 1 simulates a cohort including competing events, where event of interest can be considered as cancer but its observation is made impossible in case 6 & 8 by a preceding competing event. Figure 2 shows systematic application of diminishing weights ( $w_i$ ) to these cases after occurrence of competing events, which are calculated by reversing the indicator definition using standard product limit methods.



**Figure 2. Application of weights for competing events**

### WHY?

It has been frequently pointed out that in presence of competing events, standard product limit method of estimating survivor function for event of interest yields biased results as the probability of occurrence is modified by an antecedent competing event. We use examples below to support this point.

### DATA

We use the BMT dataset presented by Klein and Moeschberger(1997) [6] which contains data for bone marrow transplant for 137 patients, grouped into three risk categories based on their status at the time of transplantation: acute lymphoblastic leukemia (ALL), acute myelocytic leukemia (AML) low-risk, and AML high-risk. During the follow-up period, some patients might relapse or some patients might die while in remission. Consider relapse to be the event of interest. Death is a competing risk because death impedes the occurrence of leukemia relapse.

### SCENARIO 1. RARE COMPETING EVENTS

Data was altered to change the frequency and distribution of competing events with most competing events occurring at a later time period. Summary of events of interest, competing events and censoring is given in table below.

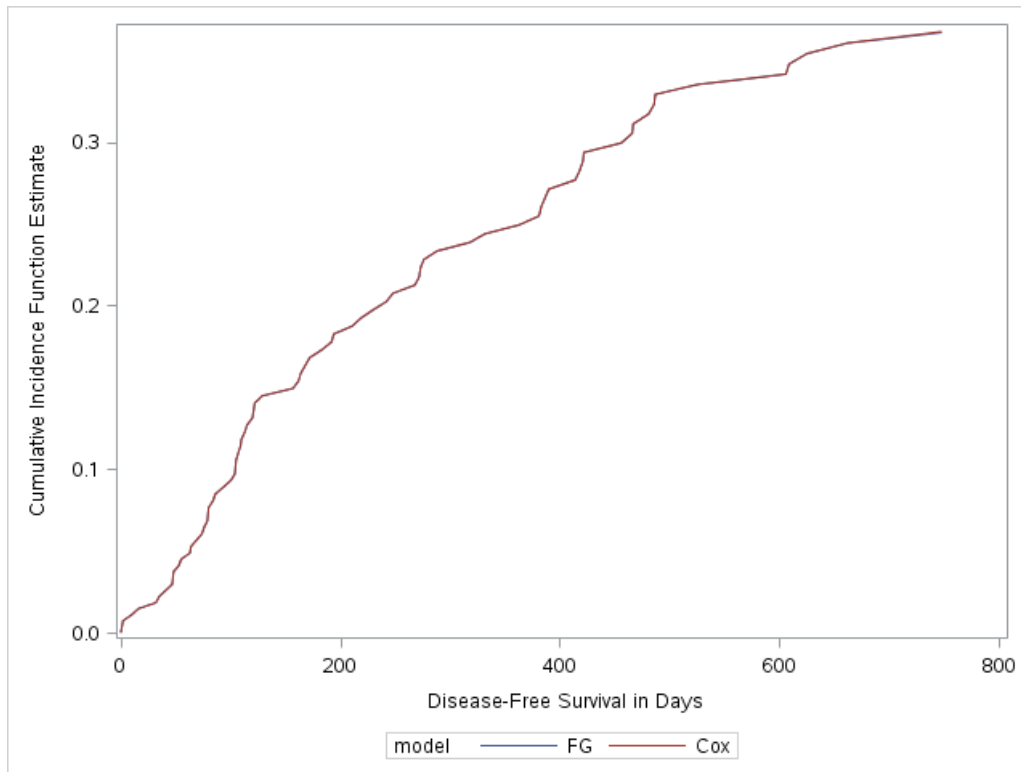
Event	Frequency
1 (Event of interest)	77
2 (Competing event)	6
0 (Censored)	54

**Table 1. Event distribution for dataset with sparse competing events**

Two models were fit to this data, one being the Cox proportional hazard model in which all competing events are censored and second using Fine and Gray's sub-distribution hazard model. Comparisons of both models and CIF plot are given below.

Covariate	Cox Parameter Estimate	FG Parameter Estimate	Cox P-value	FG P-value	Cox Hazard ratio	FG Hazard ratio
Disease-All	0.76	0.76	0.0099	0.0098	2.13	2.13
Disease-High risk	1.13	1.13	<0.0001	<0.0001	3.08	3.08

**Table 2. Results of cox proportional hazard model and Fine and Gray's competing risk model**



**Figure 3. CIF when competing events are rare**

Results in Table 2 and Figure 3 make it clear that when competing events are rare and distributed towards end of follow up, using cox proportional hazard model with censored competing events generates similar estimates to Fine and Gray's sub-distribution hazard model.

## SCENARIO 2. FREQUENT COMPETING EVENTS

We use the original unaltered data for this example, censoring and event distribution is shown below.

Event	Frequency
1 (Event of interest)	42
2 (Competing event)	41
0 (Censored)	54

**Table 3. Event and censoring distribution**

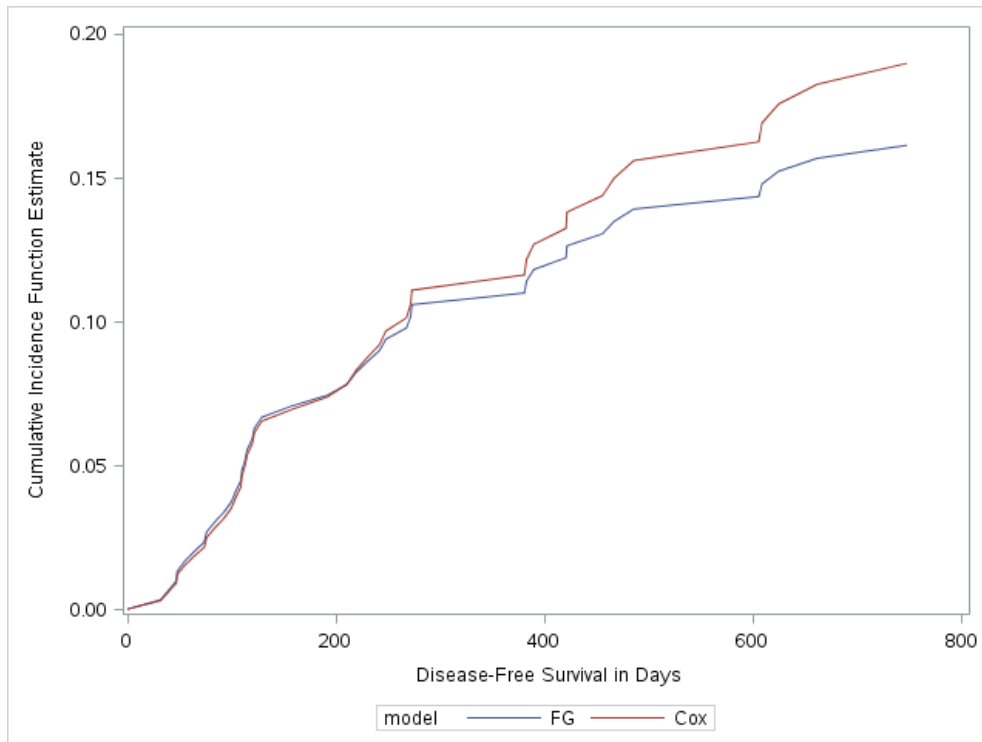
Following results were obtained for Fine and Gray's competing risk model and ordinary Cox proportional hazard model (censoring all events except event of interest).

Covariate	Cox Parameter Estimate	FG Parameter Estimate	Cox P-value	FG P-value	Cox Hazard ratio	FG Hazard ratio
Disease-All	0.89	0.80	0.04	0.06	2.45	2.23
Disease-High risk	1.50	1.31	<0.0001	0.0007	4.5	3.71

**Table 4. Results of Cox proportional hazard model and Fine and Gray's sub-distribution hazard model**

Results in Table 4 show that in presence of competing events, using Cox proportional hazard model can yield biased estimates affecting inference.

CIF plot from both models can illustrate this point better, as seen in Figure 4 Cox proportional hazard model overestimates hazard in presence of competing events, degree of overestimation will depend on frequency and distribution of competing events. Accounting for competing risks in an appropriate model gives more realistic and unbiased estimates.



**Figure 4. CIF estimates for Cox and Fine and Gray's model**

## HOW?

With the release of version 9.4 of SAS software, Fine and Gray's sub-distribution hazard model can be fitted by specifying eventcode option in PROC PHREG.

The following DATA step creates the data set BMT used in examples. The variable Disease represents the risk group of a patient, which is either ALL, AML-Low Risk, or AML-High Risk. The variable T represents the disease-free survival in days, which is the time to relapse, time to death, or censored. The variable Status has three values: 0 for censored observations, 1 for relapsed patients and 2 for patients who die before experiencing a relapse.

```

data Bmt;
  input Disease T Status @@;
  label T='Disease-Free Survival in Days';
  format Disease DiseaseGroup.;
  datalines;
1 2081 0 1 1602 0 1 1496 0 1 1462 0 1 1433 0
1 1377 0 1 1330 0 1 996 0 1 226 0 1 1199 0
1 1111 0 1 530 0 1 1182 0 1 1167 0 1 418 2
1 383 1 1 276 2 1 104 1 1 609 1 1 172 2
1 487 2 1 662 1 1 194 2 1 230 1 1 526 2
1 122 2 1 129 1 1 74 1 1 122 1 1 86 2
1 466 2 1 192 1 1 109 1 1 55 1 1 1 2

```

```

1    107    2    1    110    1    1    332    2    2    2569    0    2    2506    0
2   2409    0    2   2218    0    2   1857    0    2   1829    0    2   1562    0
2   1470    0    2   1363    0    2   1030    0    2    860    0    2   1258    0
2   2246    0    2   1870    0    2   1799    0    2   1709    0    2   1674    0
2   1568    0    2   1527    0    2   1324    0    2    957    0    2    932    0
2    847    0    2    848    0    2   1850    0    2   1843    0    2   1535    0
2   1447    0    2   1384    0    2    414    2    2   2204    2    2   1063    2
2    481    2    2    105    2    2    641    2    2    390    2    2    288    2
2    421    1    2     79    2    2    748    1    2    486    1    2     48    2
2    272    1    2   1074    2    2    381    1    2     10    2    2     53    2
2     80    2    2     35    2    2    248    1    2    704    2    2    211    1
2    219    1    2    606    1    3   2640    0    3   2430    0    3   2252    0
3   2140    0    3   2133    0    3   1238    0    3   1631    0    3   2024    0
3   1345    0    3   1136    0    3    845    0    3    422    1    3    162    2
3     84    1    3    100    1    3     2    2    3     47    1    3    242    1
3    456    1    3    268    1    3    318    2    3     32    1    3    467    1
3     47    1    3    390    1    3    183    2    3    105    2    3    115    1
3    164    2    3     93    1    3    120    1    3     80    2    3    677    2
3     64    1    3    168    2    3     74    2    3     16    2    3    157    1
3    625    1    3     48    1    3    273    1    3     63    2    3     76    1
3    113    1    3    363    2

```

;

PROC PHREG can be used with eventcode option specifying the code for event of interest (1 in this dataset) and "CIF" in plot option generates a plot for Cumulative Incidence Function.

```

proc phreg data=Bmt plots(overlay=stratum)=cif;
  class Disease (order=internal ref=first);
  model T*Status(0)=Disease / eventcode=1;
run;

```

All eventcodes apart from censor and event of interest are considered as competing events in the model, it is worth noting that all categories of competing events are considered equal i.e. no distinction is made during parameter estimation for different competing risk types.

## CONCLUSION

It has been shown that in presence of competing events ordinary product limit survival methods yield biased estimates where amount of bias depends on frequency and distribution of competing events. This bias can be reduced by the using a model which takes competing events into account such as Fine and Gray's sub-distribution hazard model which has been made available with release of version 9.4 of SAS software.

## REFERENCES

1. Fine, Jason P., and Robert J. Gray. "A proportional hazards model for the subdistribution of a competing risk." *Journal of the American statistical association* 94.446 (1999): 496-509.
2. Kleinbaum, David G., and Mitchel Klein. "Competing risks survival analysis." *Survival Analysis: A self-learning text* (2005): 391-461.
3. Gray, Robert J. "A class of K-sample tests for comparing the cumulative incidence of a competing risk." *The Annals of statistics* (1988): 1141-1154.
4. Satagopan, J. M., et al. "A note on competing risks in survival data analysis." *British journal of cancer* 91.7 (2004): 1229-1235.
5. Allison, Paul D. *Survival analysis using SAS: a practical guide*. Sas Institute, 2010.
6. Klein, John P., and Melvin L. Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2003.
7. Lin, Guixian, Ying So, and Gordon Johnston. "Analyzing survival data with competing risks using SAS® software." *SAS Global Forum*. Vol. 2102. 2012.
8. Fürstová, Jana, and Zdeněk Valenta. "Statistical analysis of competing risks: overall survival in a group of chronic myeloid leukemia patients." *European Journal for Biomedical Informatics* 7.1 (2011): 2-10.
9. P.C. Lambert. "Estimating and modelling cumulative incidence functions using time dependent weights" *UK Stata users group 2013*.
10. Kalbfleisch, John D., and Ross L. Prentice. *The statistical analysis of failure time data*. Vol. 360. John Wiley & Sons, 2011.
11. Kohl, Maria, et al. "PSHREG: A SAS macro for proportional and nonproportional subdistribution hazards regression." *Computer methods and programs in biomedicine* 118.2 (2015): 218-233.

## CONTACT INFORMATION

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

Lovedeep Gondara  
Dept. Of Computer Science  
University of Illinois Springfield  
Lgond2@uis.edu



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.

