

Regression on Cause-Specific Hazard in Competing Risks for Survival Analysis

V. Sujatha^{*1}, V. Dharanidharan²

¹*School of Advanced Sciences, VIT University, Vellore.*

²*Department of Computer Science, Apollo Engineering College, Chennai.*

*Corresponding Author: V. Sujatha

Abstract: In the study of survival analysis, Kaplan–Meier (KM) estimates of survival curves and Cox proportional hazard models were widely used to describe survival trends and identify significant prognostic factors. All these statistical analyses deal with only one type of event like death, independently of its cause. Most part of the analysis are all based on the hazards. In a competing risks setting, There are two types of hazard namely, the cause-specific hazard and the sub distribution hazard. The cause-specific hazard is the instantaneous risk of progression to a specific type of event, conditionally on being at risk for experiencing that event, generally hazards are not the probabilities, they are the rates., and used to quantify the probability of the crude risk or cause-specific cumulative incidence to progress to a specific cause by time t , When estimating the cause-specific hazard, the individuals experienced a competing event leaving behind the risk set. The cause-specific regression model was applied on to the real data sets and try to shed light on causal mechanisms and prediction at the population level.

Keywords: *Competing Risks, Cause-specific hazard, Weibull estimate, Marginal distribution hazard, Cox-Regression Model.*

Introduction

Survival analysis in medical research has been seen by two components. one the Kaplan-Meier a nonparametric approach. It is used to estimate the time-to-event distribution for the total population that the sample represents. A log-rank test is used to test whether the subgroups have a different time-to-event distribution, the effect of covariables are quantified by a regression model generally the Cox proportional hazards model allows us to investigate whether variables act differently on each of the event types

A competing risks analysis describes progression to the different end points. Whenever there is right or left censoring the estimation is usually based on the hazard. In competing risk settings hazards are of two types the cause-specific hazard and the sub distribution hazard. The cause-specific hazard is the instantaneous risk of progression to a specific type of event, conditionally on being at risk for

experiencing that event. The sub distribution hazard is the instantaneous risk of progression to a specific type of event, conditionally on not having experienced that event. Individuals that experienced a competing event remain in the risk set

The cause-specific hazard is estimated by removing individuals from the risk set when they experience a competing event. The techniques are all same as the standard Cox proportional hazards model the difference is in the interpretation, by quantifying effects on the cause-specific hazard, not on the marginal hazard. Only on non-informative censoring the estimates are interpreted as the effects on marginal hazards

Kalbfleisch and Prentice described competing risks as the situation in which an individual can experience more than one type of event [1]. Gelman et al. and Kaplan et al. explain it as the failure to achieve independence between the time to an event

and the censoring mechanism [2]. Gooley et al. defined the concept of competing risks as the situation where one type of event 'either precludes the occurrence of another event under investigation or fundamentally alters the probability of occurrence of this other

event [3] Which helps the clinical researcher or for the statistician to face the possibility of many types of failure that compete with each other. Table.1. identifies possible competing risks associated with each of these endpoints [4,5].

Table1: On end points

Endpoint	Measurement	Event of interest	Possible competing risks
Survival	Time to death from any cause	Death from any cause	No competing risks
Disease-free survival	Time to first failure: relapse or death	Relapse or death	No competing risks
Cause specific survival	Time to death due to disease under study	Death due to disease under study	Death due to other causes

Materials and Methods

To analyse the multiple causes of death and to estimate the cause specific failure probabilities by treating the subjects dying from extraneous as censored and then computing the ordinary Kaplan-Meier results in biased survival estimates. When events are independent, each event is analysed by censoring on all other events as well as censoring on loss to follow up, which yields the unbiased estimate of cause-specific hazard $\lambda(t)$, the cause-specific hazards yield the valid results even when the competing events are not independent.

The hazard function for failure from cause j , in the presence of all risks 1 top is defined and is given by

$$\begin{aligned}
 h(j, t) &= \lim_{\delta \rightarrow 0} pr(C = j, T \leq t + \delta \setminus T > t) \\
 &= \lim_{\delta \rightarrow 0} \delta^{-1} (\bar{F}(j, t) - \bar{F}(j, t + \delta)) / \bar{F}(t) \\
 &= f(j, t) / \bar{F}(t)
 \end{aligned}$$

thus,

$$h(t) = \sum_{j=1}^p h(j, t)$$

Maximum likelihood Estimators

Maximum likelihood estimates based on large numbers of observations have approximate normal distributions. And it is optimal in the sense that it usually had its smaller variance than competing variates. The estimated variance of a maximum likelihood estimates is necessarily calculated as a part of the estimation process. it provides the parameter of the regression model given for the cause-specific hazard $h(j, t)$.

There are many statistical packages used today in medical statistics which implemented the CIF for competing risk events, and for even fewer provide tools for comparing the cumulative incidence of a particular type of failure among different groups. A 'competing risks' analysis is provided by an add-on package of R.6 R is an open source software for statistical computing and graphics, which is freely available at www.r-project.org. R performs many statistical analyses needed in practical applications linear and generalized linear models, nonlinear regression models, time-series analysis, parametric and nonparametric tests, clustering, smoothing and survival analysis. R also provides additional packages for specific purposes. R version 3.3.2 software with the package 'competing risk' was used to estimate the parameter and to predict the effect of covariate on the population level. We perform a competing risk analysis in R by an example using a sample dataset containing survival times of anonymous patients who attained the stokes (once, twice, thrice) and underwent for treatments

Results and Discussions

A sample of 2422 sets on stoke (heart attack) were taken, and there are 42 covariates. The data set describes the various covariates which influences the heart failure rates. Death by heart attack in hospital is the event, and the might be at first or second or third attack, this paper describes how the post event preventive surgery influence the event, it precludes all the events which will affect the cardiac-diseases. Here all other

events which affect the stroke as taken as censoring. A cause-specific hazard model is applied on to this data and the effects of the covariates are analysed. To validate this result chi-square test is performed.

The function `cuminc` calculates the CIF estimates for type of event together with the variance based on Aalen's work

The Estimates and Variances are:

\$est		500	1000
1	1	0.143	0.279
1	3	0.084	0.084
1	5	0.201	0.201

\$var		500	1000
1	1	4.691e-04	1.105e-03
1	3	5.664e-05	5.664e-05
1	5	9.984e-05	9.983e-05

For this data set the estimates for the whole group are calculated, and since there is just one stratum, it is automatically marked as 1. The second column represents the type of event according to the codes used in the `cens` variable. First time attack was coded as 1, relapse as 2 and death as 3. Hence, the first line gives the probability for the heart attack for the first time at 0.5, 1, 1.5, 2, 2.5 years. The second line provides the estimates for the probability of relapse and the third line gives the estimates for the probability of death without relapse (after taken the treatment). Thus the CIF of stroke for 1 years is 0.00 this is because of the first time attack and hence no CIF is there and the CIFs of relapse is 0.08407757 and death at 1 year is 0.20052552. In the same way \$var contains the variance estimates for the three types of events at each time point. The CIF estimates at 3, 6, 9 and 12 months (0.25, 0.5, 0.75 and 1 year) can be obtained using time points:

\$var	0.25	0.5	0.75	1
1	0.00e+00	0.00e+00	0.00e+00	1
1	5.66e05	5.66e-05	5.66e-05	3
1	9.98e05	9.98e-05	9.98e-05	5

We assume cause-specific baseline hazards to be proportional. In our sample we quantify the parameter which describes the progression rates. The proportional baseline hazards are fitted based on the type of end point" which plays the role of a covariable.

In R we fit the model through

```
Coxph (formula = Surv (Time.to. first.event.
post. attack, First. event. post. attack == 4)
~ Post.event.preventative.surgery, data = h)
```

coef	exp(coef)	se(coef)	z	p
-0.4178	0.6585	0.0683	6.12	9.3e-10
Likelihood ratio test = 39.9 on 1 df, p=2.62e-10				
n= 2421, number of events= 584				

This explains a robust standard error is not needed because every individual has at most one row with an event.

Using the stacked data format also allows us to assume effects to be equal over event types. If the effects of all covariables are allowed to differ by stratum, the partial likelihood of the combined model using the stacked data set is the product of the partial likelihoods of the separate analyses. The stacked format allows us to compute them for all direct transitions

Post. event. preventative. surgery

coef	exp(coef)	se(coef)	z	p
0.679	0.507	0.138	-4.93	8.4e-07
Likelihood ratio test= 58.4 on 4 df, p=6.27e-12				
n= 2421, number of events= 584				

By applying the Pepe–Mori test for conditional probability via observation of relapse in the older population is hindered by the occurrence of death, using the function `comp CP` the p-value 0.0001750457 is found and the chi square value is found to be 14.08162

To estimate the probability of first attack using the Weibull distribution. After the first attack, when there is no death but it constitutes the competing risk (non-informative censoring). A Weibull distribution is fitted to the marginal distribution.

Survreg (formula = Surv (Time.to. first.
event. post. attack, evcens) ~ 1, data = h)

Loglik(model)= -18826.5 , Loglik(intercept
only)= -18826.5

Coefficients:		
(Intercept)	Scale	n
7.045656	0.5623381	2421

The estimation of the parameter on weibull
distribution is $\lambda = \frac{1}{7.045656}$ and $\alpha = \frac{1}{0.5623381}$

References

- 1 Kalbfleisch JD, Prentice RL (2002) The Statistical Analysis of Failure Time Data. Hoboken, NJ: John Wiley & Sons, Inc.
- 2 Gelman R, Gelber R, Henderson IC, Coleman CN, Harris JR (1990) Improved methodology for analyzing local and distant recurrence. Journal of Clinical Oncology, 8:548-555.
- 3 Caplan RJ, Pajak TF, Cox JD (1994) Analysis of probability and risk of cause-specific failure. International Journal of Radiation Oncology, Biology, Physics, 29:1183-1186.
- 4 Gooley TA, Leisenring W, Crowley J, Storer BE (1999) Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Statistics in Medicine, 18:695-706.
- 5 Pepe MS, Mori M (1993) Kaplan–Meier, marginal or conditional probability curves in summarizing competing risks failure time data Statistics in Medicine, 12:737-751.