International Journal of Computer Science and Control Engineering
2019; 7(2): 17-24
http://www.openscienceonline.com/journal/csce

Open Science

Performance of Cause-specific and Subdistribution Hazard for Large Samples - A Simulation Study

Galappaththige Hasani Sandamali Karunarathna, Marina Roshini Sooriyarachchi

Department of Statistics, University of Colombo, Colombo, Sri Lanka

Email address

hasani@stat.cmb.ac.lk (G. H. S. Karunarathn)

To cite this article

Galappaththige Hasani Sandamali Karunarathna, Marina Roshini Sooriyarachchi. Performance of Cause-specific and Subdistribution Hazard for Large Samples - A Simulation Study. *International Journal of Computer Science and Control Engineering.* Vol. 7, No. 2, 2019, pp. 17-24.

Received: August 3, 2019; **Accepted**: October 8, 2019; **Published**: October 23, 2019

Abstract

The competing risks scenario is a complex setting for classical survival analysis when an individual is under risk of failing from various events. Since competing risk data are often found in many fields such as medicine, social science, biology etc., interest has been paid among researchers to focus towards the methodological competing risk setting. Additionally, it is not possible to have real data and thus to know about the real status, thus simulation studies lead to more advantages towards analyzing such responses. Hence, this paper focuses on investigating the performance of the most commonly used regression approaches for analyzing the competing risk responses namely, cause specific hazard model and sub-distribution hazard model by following pre-specified cause specific hazard ratio. A simulation study was carried out by varying the censoring distribution parameter and shape parameter while keeping the scale parameter constant, under nine scenarios. Summary statistics of cause specific hazard and sub-distribution hazard were different for the two methods and it showed that mean hazards of cause-specific hazard model decreases when the shape parameter of the censoring distribution is increased. As a conclusion, this simulation study reveals that cause specific and sub-distribution hazard ratios are monotonically increasing with all scenarios and all scenarios performed approximately equally with minor differences for the two types of regression models.

Keywords

Competing Risks, Cause-Speccific Hazard, Sub-distribution Hazard

1. Introduction

Classical survival analysis are more complex when there are more mutually exclusive cause of failure. Therefore, competing risk is such a complex setting since it arises when an individual is under the risk of failing from various events [1]. The occurrence of failure due to some specific cause may or may not prevents, occurrence of other causes in the competing risk background. Thus, classical time to event methods are not proper mechanism for the competing risk data since kaplan meier assumption is violated in the presence of multiple events [2, 3]. As a result, Kaplan-Meier generally overestimate the probability of the event of interest and hence it yields biased results in the presence of competing events [1-3].

Competing risk data are often found in many fields. In a demographic study where the leading causes such as heart attcack, cancer, etc are registered and interest has been paid off to analyze each of death seperately. In a reliability study, breakdown of a mechanical device from some special reason is event of interest when there are numerous causes are available. In a clinical trials, as an example, interest focuses to find out the benefits of a new drug to prevent myocardial infraction, patients who have coronary disease are followed during three years. The failure of interest is Myocardial Infraction though patient may die from other causes. So, competing risk setting is highly extendend concept.

Previously, latent failure time models were widely used and it had been heavily criticized in biomedical situation due to non-identibility of dependence structure between times to different types of event [4-6]. Recently, competing risk analysis has been determined on the cause, i.e event specific hazard, which are empirically distinguishable and entirely determine the competing risks process [2, 4, 5]. The most well-known approaches are cause specific hazard and subdistribution hazard for the analysis of competing risk data. In a cause specific hazard, modeling has been carried out by

applying Cox regression seperately for each event type/cause. In contrast, the proportional subdistribution hazard is directly linked to the cumulative incidence [7].

The aim of this article is explained how competing risks data can be simulated by following prespecified cause specific hazard as Bayersman et al, 2009 [8] and investigate the performance of most commonly used regression approach of competing risk under various realistic scenarios as an extension of the reference [9].

2. Methodological Background

In a competing risks scenario, an individual can fail from any of several (say K event types), but only the time to failure for the earliest of these is observed. Hence, the unique feature of a competing risks setting is that for each individual, take the value of failure time T and failure mode C, and a joint model for T and C is needed.

The joint distribution of (T,C) might be completely specified through the cause specific hazard function $h_k(t)$ which is the principal identifiable quantity in competing risks observation [10] and it represents the probability of failure due to cause k at time t, given that no failure of any kind has occurred so far. The cumulative cause specific hazard ($\Lambda_k(t)$) equals the cause specific hazard summed from start of observation to time t.

$$\Lambda(t) = \Lambda_k(t) + \Lambda_k(t) + \cdots + \Lambda_k(t) \quad (1)$$

Another important quantity is the cumulative incidence function ($F_k(t)$), which is

$$F_k(t) = \Pr(\text{failure time } T \le t, \text{cause} = k) = \int_0^t S(u)\lambda_k(u)du \quad (2)$$

$F_k(t)$ involves both the probability of having not failed from some other event first up to t ($S(u)$) and the cause specific hazard for the event of interest ($h_k(t)$) at that time [10].

Regression Models for Competing Risks Data

To summarize the effect of covariates in the competing risk settings, two well-known hazard based regression models; cause specific hazard regression model and subdistribution hazard regression models are reviewed by many authors [11 - 13].

2.1. Modeling Cause Specific Hazards

The Cox proportional hazard models is applied to model cause specific hazard [6]. Cause specific hazard ($h_k(t|X)$) is the function of baseline hazard rate $h_{k,,o}$ (t) and set of covariates by a vector X is given by,

$$h_k(t|X) = h_{k,0}(t)\exp(\beta^T{}_k X)\text{; Where, } \beta\text{-coefficient; X-matrix of covariates} \quad (3)$$

Here cause specific hazard is a function of some unspecified "baseline" cause specific hazard and set of covariates [10]. Since Cause specific hazard has been familiar with traditional cox proportional hazard, it can be modeled through standard statistical software by performing a classical regression, by considering the events at failure time of cause of interest and failures from other causes treated as censored observations [14]. Reference [8] indicates that cause specific hazards is "totally defined by the competing risk process".

2.2. Modeling Subdistribution Hazards

The subdistribution hazard ($h_k^*(t,X)$) for event k is defined as the probability for an individual to fail from cause k in an infinitesimal small interval Δt, given that the individual experienced no event until time t or experienced an event other than k before time t,

$$h_k^*(t,X) = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\,\text{pr}\,[t \le T \le t+\Delta t, C = k, |(T \ge t \text{ or } (T \le t, C \ne k), X)] \quad (4)$$

2.3. Relationship Between Cause Specific and Subdistribution Hazard Rate

The relationship is presented by [8] for the case of two possible endpoints and it can derived as,

$$h_1(t|X) = \left(1 + \frac{F_2(t|X)}{S(t|X)}\right).h_1^*(t|X) \quad (5)$$

Where, S (t|X) – probability of being free of any event up to time t given X, $h_1(t)$ and $h_1^*(t)$ denoting the cause specific hazard and subdistribution hazard for event of interest respectively. $F_2(t)$ is the cumulative incidence function for the cumulative incidence function for the competing event (k=2).

3. Simulation Study of Competing Risks Regression Models

Simulate the competing risk data with different scenario to understand the variation of the modeling approaches. Here we refer excellent articles [8, 10, 15-17] for planning the simulation study. The R software, which is an open source software was used and the suggested method is illustrated.

3.1. Simulating Survival Time Data Using the Inversion Method

If the convenience function is not available, then general simulation techniques will be useful. Reference [16, 18, 19] well described how time to event data depending on a covariate vector X can be generated for proportional hazard models using the inversion method when there is no available function for the cause specific hazard. R function "unitroot" can be used for the inversion method.

Assume that, specified cause specific hazard is such that $h_0(t) > 0$, for all t. Then the cumulative all cause hazard $A_0(t) = \int_0^t h_0(u)du$ is strictly increasing as is the distribution function of T [18],

$$F(t) = P(T \le t) = 1 = \exp(-A_0(t)) \quad (6)$$

$F^{-1}(t)$ and ${A_0}^{-1}$ is the inverse of F and A_0 respectively. The important of the inversion method is that F (T), which is

transformed failure time, is uniformly distributed on [0, 1],

$$P(F(T) < u) = P\left(T \leq F^{-1}(u)\right) = F\left(F^{-1}(u)\right) = u;\ u \in [0,1] \quad (7)$$

If U is a random variable with uniform distribution on [0,1], then $F^{-1}(u)$ has the same distribution as T. hence, inversion method as follows [18],

1. Compute$F^{-1}(u) = {A_0}^{-1}(-\ln(1-u)), u \in [0,1]$.
2. Using the R function runif, generate random variable U.
3. $F^{-1}(u)$ is the chosen replicate of T.

3.2. Simulating the Competing Risk Data Following Pre-specified Cause Specific Hazards

Beyesmann et al. (2009) *presented an algorithm for generating competing risk data for two possible types of events as follows,*

1. Define cause specific hazard rates $h_1(t)$ and $h_2(t)$ for both types of event.
2. Simulate survival times T using the inversion method with overall hazard rates$h_{overall}(t) = h_1(t) + h_2(t)$.
3. For a simulated survival time T, run a binomial experiment with probabilities $h_1(t)/(h_1(t) + h_2(t))$ for an event type K=1.
4. Additionally, generate the censoring times C.
5. Simulate two predictors; one variable is categorical and other variable is continuous.

To meet the intended objective in this study, generate survival time and competing risk as above simulation algorithm. Then Cox cause specific hazard and sub-distribution/Fine-Gray hazard [10] for the event of interest (k=1) were computed for the 1000 data sets with 1000 observations each. These estimates are averaged due to the large number of simulated data sets to illustrate the behavior of the models under various scenarios. As a proposed methodology, two different parameters such that cause specific hazard for the event of interest ($h_1(t)$) and censoring distribution parameter were varied. Here event of interest ($h_1(t)$) was changed while competing risk event ($h_2(t)$) remains constant and assume that censoring distribution was Weibull distribution and its shape parameter (β) was varied while scale parameter remains constant (=1). Changing the parameters were chosen as follows,

Cause specific hazard [10]:

$$h_1(t)(=0.5) < h_2(t)(=1)$$

$$h_1(t)(=1) = h_2(t)(=1)$$

$$h_1(t)(=1.5) > h_2(t)(=1)$$

Censoring distribution-weibull:-
Scale parameter (α) =1 (constant)
Shape parameter – β
0 < β < 1 » β = 0.5 –Decreasing distribution
β = 1 – constant distribution
β > 1 » β = 3 – increasing distribution

4. Simulation Results and Discussion

A simulation study was conducted to evaluate the performance of the cause specific hazard and sub-distribution hazard for different settings under time constant hazard ratio. Summary statistics and hazard distribution for the estimated ratios for the different parameter of the censoring distribution with hazard values for the event of interest are shown in figures 1-2 and Tables 1-2.

Censoring Distribution Parameter – β
Hazard of Event of interest - h_1 (t)
Hazard of Competing event - h_2 (t).

Table 1. *Summary Statistics of estimated average cause specific hazard.*

B	Specified Hazards		Mean	SE Mean	Variance	Q1	Median	Q3
	h_1 (t)	h_2 (t)						
	0.5		0.01101	0.000341	0.00306	0.00164	0.0026	0.00552
0.5	1		0.01095	0.000343	0.00323	0.00164	0.00257	0.00538
	1.5		0.01091	0.000321	0.00285	0.00164	0.00258	0.00537
	0.5		0.00865	0.000297	0.00221	0.00137	0.0021	0.00424
1	1	1	0.00673	0.000212	0.00111	0.00136	0.00202	0.00399
	1.5		0.00714	0.00027	0.00153	0.00132	0.00192	0.00377
	0.5		0.00584	0.000215	0.000981	0.00125	0.00174	0.00326
1.5	1		0.00562	0.000218	0.00107	0.00127	0.00177	0.00324
	1.5		0.00575	0.000217	0.001	0.00125	0.00175	0.00326

Table 2. *Summary Statistics of estimated average subdistribution hazard.*

β	Specified Hazards		Mean	SE Mean	Variance	Q1	Median	Q3
	h_1 (t)	h_2 (t)						
	0.5		0.13554	0.000251	0.00086	0.0546	0.12255	0.2074
0.5	1		0.20847	0.00029	0.01947	0.08647	0.19151	0.3192
	1.5		0.21276	0.000328	0.00321	0.04243	0.18277	0.35585
	0.5		0.12614	0.000221	0.0173	0.05337	0.11397	0.18914
1	1	1	0.20168	0.000222	0.01747	0.08809	0.18622	0.30354
	1.5		0.25106	0.000288	0.02597	0.11205	0.23501	0.3782
	0.5		0.24353	0.000255	0.02461	0.10917	0.22606	0.36633
1.5	1		0.20085	0.00022	0.01624	0.09204	0.18894	0.29933
	1.5		0.20575	0.000275	0.02982	0.03608	0.18284	0.34425

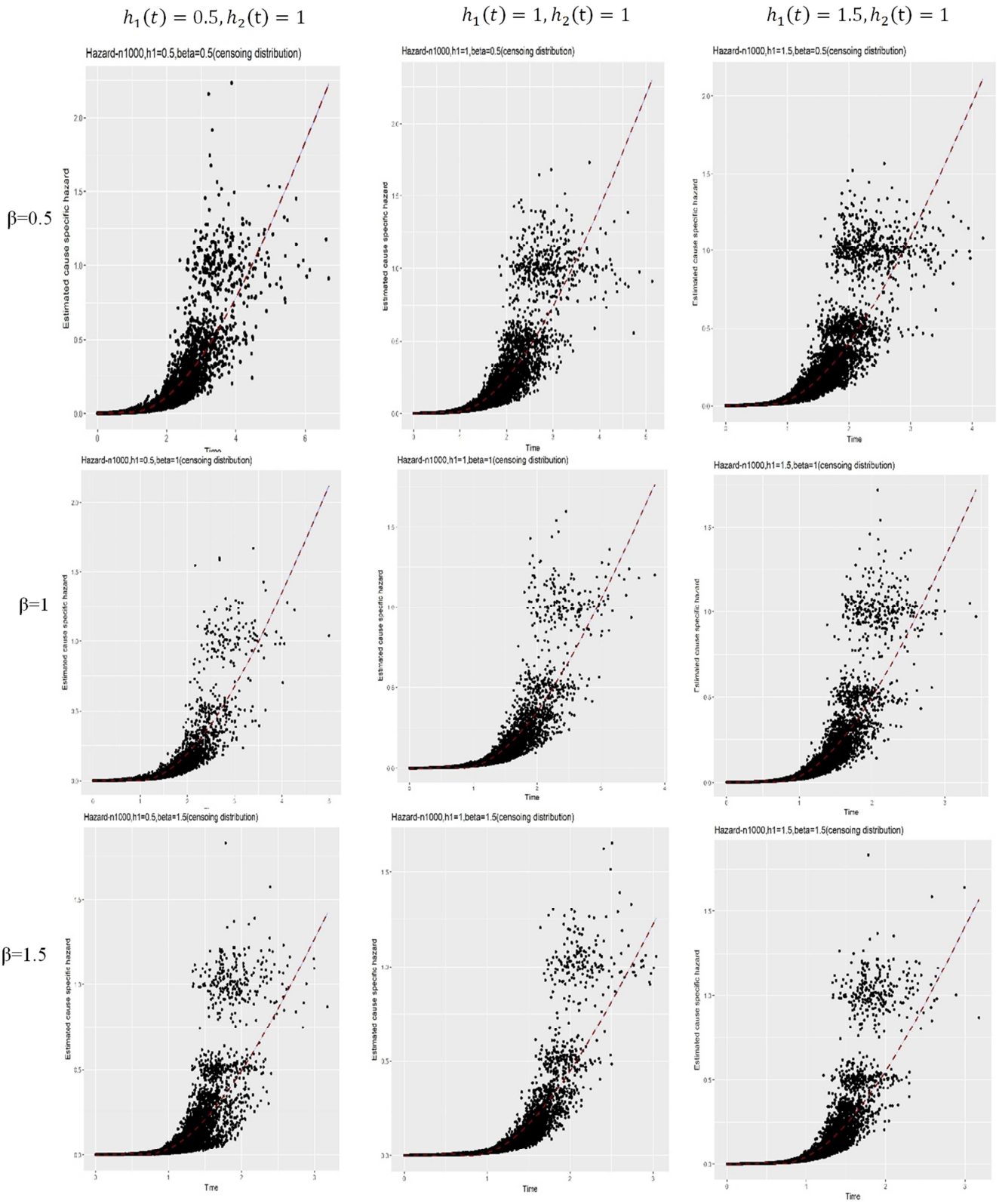

Figure 1. *Estimated Cause Specific hazard ratio.*

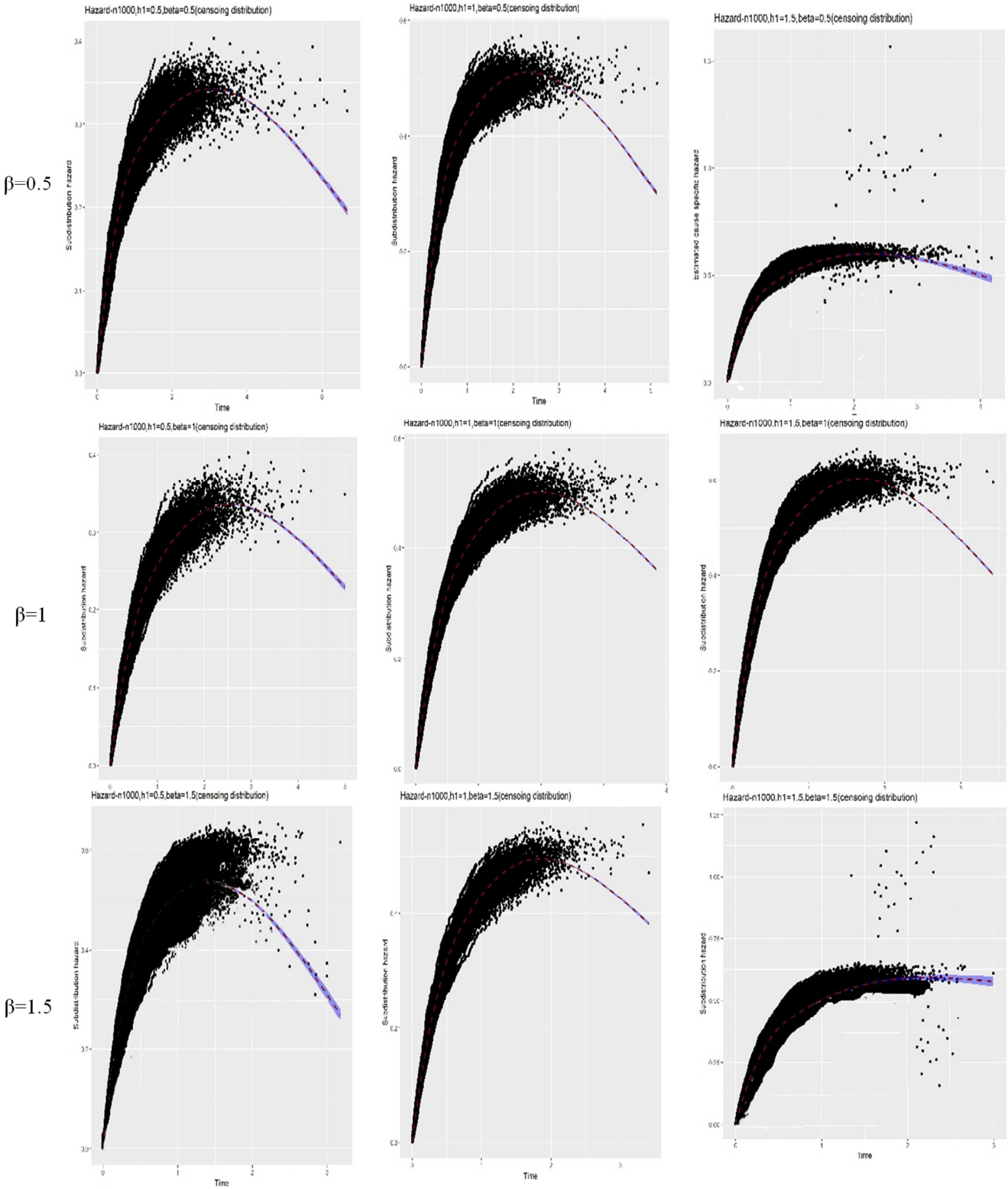

Figure 2. *Estimated sub distribution hazard ratio.*

Regarding the average cause specific hazard ratios, mean hazards are decreased when the shape parameter of the censoring distribution is increased. However, it seems, hazard ratios are similar to each other when the hazard of event of interest are varied while the shape parameter remains constant. In addition, Table 1 clearly reveals that interquartile ranges almost similar in each scenarios, but it indicates slight difference in variability.

Table 2 represents estimated sub-distribution hazard under nine different scenarios. When hazard of event of interest are 1 and 1.5, mean sub-distributions are identical to each other. However, there is a minor difference at the event of interest is

0.5 when compare to the 1 and 1.5 for the mean sub-distribution hazard. According to Table 2 and Figure 2, variance of estimated hazard ratios is lower for β=0.5 and 1.5 with $h_1(t) = 1.5$ than the other scenarios.

Figure 1 and Figure 2 depicts the cause specific hazard ratios are monotonically increasing and sub-distribution hazard ratios are increasing in each scenarios.

5. Conclusions

In this article, proposed methodology was a simulation study to investigate the performance of cause specific hazard and sub-distribution hazard, which are the most popular competing risk regression models among the researchers. Results are illustrated by the graphical display of summary statistics of the estimated cause specific and sub-distribution hazard obtained from the nine different scenarios by changing the censoring distribution shape parameter and hazard of event of interest while the hazard of competing risk remains constant.

As presented in the simulation results, all nine scenarios performed equally with a minor difference within the regression method. However when comparing the two regression methods, cause specific and sub-distribution hazard's summary statistics were different to each other. But, both Cause specific hazard and sub-distribution hazard showed increasing pattern. However, there was a significant variability difference in sub-distribution hazard when the hazard of event of interest is 1.5 and shape parameter is 0.5 and 1.

Appendix

```
# This ensures that the package is loaded requires(cmprsk)
factor2ind<-function(x, baseline){
xname<-deparse(substitute(x))
n<-length(x)
x<-as.factor(x)
if(!missing(baseline))
x<-relevel(x, baseline)
X<-matrix(0, n, length(levels(x)))
X[(1:n)+n*(unclass(x) -1)]<-1
dimnames(X)<-list(names(x), paste(xname, levels(x),
sep=":"))

return(X[,-1,drop=FALSE])
}

#Generating data using a function
gen.data<-function(h1,h2,n,c1,c2){
 #definition of cause sppecific hazard functions for the
event of interest
h1<-h1

#cumulative cause specific hazard function
H1<-function(t){
 h1*t
}
#definition of cause specific hazard function for the
competing event
h2<-h2

#cumulative hazard -competing event
H2<-function(t){
 h2*t
}

#determination of event types
ev.type<-c()
for(i in 1:n){
ev.type[i]<-sample(1:2,1,prob=c(h1,h2))
}

#generating event times using inversion method(when
there is no way to find out inverse of A0)
s.fct<-function(t,y){
 return(H1(t)+H2(t) +y)
}
ev.time<-c()
for(i in 1:n){
uz<-runif(1)
ev.time[i]<-uniroot(s.fct,c(0.000000000001,500),y=log(1-
uz))$root
}

 #generation of censoring times and determination of status
variable
censtime<-rweibull(n,c1,c2)
obs.time<-pmin(censtime,ev.time)
stat<-ev.type*as.numeric(ev.time<censtime)

#Generation of covariates
x1<-sample(0:1,n,rep=TRUE,prob=c(0.5,0.5))
x2<-runif(n)

cov<-cbind(x1,x2)

#generation of data frame
data<-data.frame(obs.time,stat,x1,x2)
data$x1<-
factor(data$x1,levels=c(0,1),labels=c("male","female"))
return(data)
}

#example for generating data for the
h1=0.5,h2=1,n=1000,c1=0.5,c2=1
data_1<-gen.data(0.5,1,1000,0.5,1)
x1<-data_1$x1
x2<-data_1$x2
obs.time<-data_1$obs.time
stat<-data_1$stat
covariate<-cbind(factor2ind(x1),x2)
write.table(data_1,file="file path with data file
name",sep="\t", col.names = TRUE, row.names=FALSE)
```

```
#cause specific hazard regression
mod_cause_1<-
coxph(Surv(obs.time,stat==1)~x1+x2,data_1)
capture.output(summary(mod_cause_1),file=" file path
with cause specific model file name ")

#baseline hazard
baseh_1<-basehaz(mod_cause_1,centered=TRUE)
basehv_1<-data.frame(baseh_1["hazard"])
baseh_11<-data.frame()
baseh_11[1,1]<-basehv_1[1,1]
for(i in 2:nrow(basehv_1)){
baseh_11[i,1]<-basehv_1[i,1]-basehv_1[i-1,1]
}
coefmod<-coef(mod_cause_1)
expvalue_1<-
matrix(exp(coefmod["x1female"]*factor2ind(x1)+coefmod["
x2"]*x2))
d_1<-matrix(,nrow(baseh_11),nrow(data_1))
for(i in 1:nrow(expvalue_1)){
 d_1[,i]<-expvalue_1[i,1]*baseh_11[,1]
}

haz_cau_1<-data.frame(baseh_1["time"],d_1)
write.table(haz_cau_1,file=" file path with file ",sep="\t",
col.names = TRUE, row.names=FALSE)

#subdistributional hazard
mod_sub_1 <- crr(obs.time,stat,covariate)
capture.output(summary(mod_sub_1),file=" file path with
file name for merging cause specific hazard ")
write.table(predict(mod_sub_1,covariate),file=" file path
with file name ",sep="\t", col.names = TRUE,
row.names=FALSE

#simulate 1000 datasets for one set of parameter. i.e
gen.data function have to be run 1000 times for generating
1000 data sets.
#merge data, calculate the average hazard and plotting
setwd("file directory for data merging/")
files<-list.files(getwd())
DF<-read.csv(files[1])
time<-DF[,1]
hazard<-rowMeans(DF[,-1])
DF_mean<-data.frame(time,hazard)
for(i in 1:length(files)){
df<-read.csv(files[i])
time<-df[,1]
hazard<-rowMeans(df[,-1])
df_mean<-data.frame(time,hazard)
hazard<-rbind(DF_mean,df_mean)
DF_mean<-hazard
}
write.csv(hazard,"hazard_ratio.csv",row.names=FALSE,qu
ote=FALSE)
myplot<-
ggplot(hazard,aes(x=time,y=hazard))+ggtitle("Hazard-
n1000,h1=0.5,beta=0.5(censoing distribution)")
+labs(x="Time",y="Estimated cause specific
hazard")+geom_point()+geom_smooth(linetype="dashed",co
lor="darkred",fill="blue")
ggsave(filename="myPlot.jpeg", plot=myplot)
```

References

[1] B. Haller, "The Analysis of Competing Risks Data with a Focus on Estimation of Cause-Specific and Subdistribution Hazard Ratios from a Mixture Model," 2014.

[2] H. Putter, M. Fiocco and R. B. Geskus, "Tutorial in biostatistics: competing risks and multi state models," *Stat Med,* vol. 26, no. 11, pp. 2389-2430, 2007.

[3] P. K. Anderson and N. Keiding, "Interpretability and importance of functional in competing risk and multistate models," *Stat Med,* vol. 31, pp. 1074-1088, 2012.

[4] J. D. Kalbfleisch and R. L. Prentice, The Statistical Analysis of Failure Time Data (2nd ed.), New York: Wiley, 2002.

[5] P. Andersen, S. Abildstrom and S. Rosthoj, "Competing risks as a multi state model," *Statisical methods in Medical Research,* vol. 11, no. 2, pp. 203-215, 2002.

[6] R. Prentice, J. Kalbfleish, A. Peterson, N. Flournoy, V. Farewell and N. Breslow, "The Analysis of Failure Times in the presense of Competing Risks," *Biometrics,* vol. 34, pp. 541-554, 1978.

[7] J. P. Fine and R. J. Gray, "A Proportional hazard model for the subdistribution of a competing risk," *Journal of the American Statistical Association,* pp. 496-509, 1999.

[8] J. Beyersmann, A. Latouche, A. Buchholz and M. Schumacher, "Simulating Competing risks data in survival analysis," *Statistics in Medicine,* pp. 956-971, 2009.

[9] G. H. S. Karunarathna and M. R. Sooriyarachchi, "Investigating Hospital Discharge and Mortality: Contribution of Competing Risk Regression Model," 2017.

[10] J. J. Dignam, Q. Zhang and M. N. Kocherginsky, "The Use and Interpretion of Competing Risks Regression Models," 2012.

[11] H. T. Kim, "Cumulative incidence in competing risk data and competing risk regression analysis," *Clin Cancer Res,* vol. 13, pp. 559-565, 2007.

[12] B. Tai, R. Grundy and D. Machin, "On the importance of accounting for competing risks in pediatric brain cancer: II. Regression modeling and sample size," *Radiat Oncol Biol Phys,* vol. 79, pp. 1139-1146, 2011.

[13] P. C. Austin and J. P. Fine, "Practical Recommendation for reporting Fine-Gray model analyses for competing risk data," *Statistics in Medicine,* vol. 36, pp. 4391-4400, 2017.

[14] B. Haller, G. Schmidt and K. Ulm, "Applying competing risk regression models: An overview," Springer Science Business media, 2012.

[15] A. Burton, D. Altman, P. Royston and R. Holder, "The design of simulation studies in medical statistics," *Statistics in Medicine,* vol. 25, pp. 4279-4292, 2006.

[16] R. Bender, T. Augustin and M. Blettner, "Generating survival times to simulate cox proportional hazard models," *Statistics in Medicine,* vol. 24, pp. 1713-1723, 2005.

[17] R. J. Madachy and D. Houston, What every engnieer should know about Modeling and Simulation, CRC Press, 2017.

[18] J. A. Bayersmann and M. Schumacher, "Competing Risks and Multistate Models with R," *Springer,* 2011.

[19] I. Poguntke, M. Schumacher, J. Bayersmann and M. Wolkewitz, "Simulation shows undesirable results for competing risks analysis with time-dependent covariates for clinical outcomes," *BMC Medical Research Methodologyvolume,* vol. 18, 2018.