ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

COMPETING RISKS COX PROPORTIONAL HAZARDS MODEL THROUGH CAUSE SPECIFIC AND SUB-DISTRIBUTIONAL HAZARDS: A MODEL COMPARISON

Valarmathi S[1], Babu C Lakshmanan[2], Ponnuraja C[3]

1. Department of Epidemiology, The TamilNadu Dr.MGR Medical University, India.
2.Cognizant Technology Solutions (Chennai), India
3.Department of Statistics, National Institute for Research in Tuberculosis (ICMR), India

Manuscript Info

Manuscript History:

Received: 14 April 2015
Final Accepted: 22 May 2015
Published Online: June 2015

Key words:

Cox proportional Hazard (CoxPH), Competing Risks, Cause Specific Hazard, Sub-Distributional Hazard,

Corresponding Author

Valarmathi S

Abstract

Competing risks data emerge when the individuals under study can experience with multiple endpoints, and for each individual the time to failure and the type of failure will be observed. Consequently, the competing risks data is the extension to the ordinary survival time data which only concern with one endpoint. Sometimes the focus is not on the parameter estimates, but moderately on the probability of observing a failure from a specific cause for individuals with specified covariate values. The intention of this paper is to model the cause specific hazard and sub-distribution hazard for endpoint using Cox proportional hazard. This interpretation discusses and distinguishes between the two common types of competing risk analyses and cumulative incidence curves. It is concluded that the cause-specific hazards model is an advantageous approach than the sub-distribution hazards model. The application of the method is illustrated with an open source Bone Marrow Transplant data.

Introduction

Survival analysis is a class of statistical methods for studying the occurrence and timing of events and is useful for studying many kinds of events in medical sciences. Survival analysis is the analysis of data measured from a specific time of origin until an event of interest or a specified endpoint (Collett, 1994). A patient who dies due to kind of disease during the study period would be considered to have an 'event' at their date of death. A patient who is alive or not experience the event at the end of the study or at particular specified time would be considered to be 'censored'. Accordingly, all patients provide two pieces of information: time and status (i.e., event or censored). However, a patient can experience an event different from the event of interest. For example, a patient may die not because of the kind disease but due to causes unrelated to the kind of disease. Such events are termed competing risk events, discussed elaborately by (Satagopan et al 2004). The survival at a given time is the conditional probability of surviving to a specific time given that the individual is at risk for the event at that time. Survival data are often difficult to handle with traditional statistical methods since it takes into account two vital things censoring and time dependent covariates. Regression models for survival data have traditionally been based on the Cox regression model, which assumes that the underlying hazard functions for any two levels of covariates are proportional over the period of follow-up time. Competing Risks data are inherent to medical research in which response to treatment can be classified in terms of failure from one disease processes and/or non- disease related causes (Fine and Gray 1999).

Historically, the cumulative incidence function also known as sub-distribution has provided information that is secondary to that contained overall survival (Benichou and Gail 1990; Korn and Dorey 1992). The SAS Macros for cumulative incidence function by Rosthoj et al. (2004), is available but it is being calculated using SAS Oracle Virtual Box Manager

Material and Methods

1. Cox PH Model: Cause Specific Hazards

The Analysis of treatment and prognostic effects with censored survival data with an assumption of constant Hazard ratio Cox regression model proposed by Cox (1972) is most often cited in the articles. This model helps in estimating the relationship between the hazard rate and explanatory variables without assuming any particular form of probability distribution for the survival time. Hence it is sometimes referred as a semi parametric model.

The Cox model assumes that the hazard rate for a given patient can be factored into a baseline hazard rate (common to all patients) and a parametric function of the covariates which describes the patients' characteristics. For the ith subject, i=1, 2, … ,n, let X_i , δ_i , and $\mathbf{Z}_i$ (t) be the observed time, cause of failure, and covariate vector at time t, respectively. Assume that K causes of failure are observable ($\delta i \varepsilon\{1,2,\ldots,K\}$) δ= 0 indicates a censored observation. Consider failure from cause 1 to be the event of interest, and consider failures from other causes to be competing events. Thus the hazard rate for patient i can be expressed as

$$h_i(t) = h_0\ (t)exp(\beta Z_i)$$

Where $h_0(t)$ is a baseline hazard rate at time t; **Zi**, is the ith patient's covariate and β is the risk or regression coefficient. Coefficients of the covariates are estimated using a maximum likelihood (ML) procedure. ML estimates are obtained by maximizing a (partial) likelihood function (L) (Collett, 1994). This can be expressed as

$$L\ \ (\beta)\ = \prod_i \left(\frac{\exp(\beta Z_i)}{\sum_{j\varepsilon Ri} \exp\left(\beta Z_j\right)} \right)^{\delta_i = 1}$$

In this model the major assumption is proportional hazards which imply that the hazard ratio is constant overtime. This means that the hazard for an individual is proportional to the hazard for any other individual. When the value of the exponential expression for the estimated hazard ratio is a constant that does not depend on time, the proportional hazards assumption is satisfied. Here R_i is the risk set of patients who do not fail or are not censored before X_i

2. Cox PH Model: Sub-Distribution Hazards

In survival analysis individuals are often at risk of more than one event. For example, individuals with bone marrow transplantation are at risk of death due to transplantation or at risk of death due to other causes. The probability of dying from transplantation will depend upon the mortality rate due to transplantation and the mortality rate due to other causes. This is a classic competing risks situation. Based on the relationship between the hazard and survival functions, Putter et al. (2007) defined a sub-distribution function. Also Andersen et al. (2012) mentioned the key point about the Sub-distribution hazard that it has no resemblance to an epidemiological rate as individuals that die from other causes remain in the risk set. But Fine and Gray (1999) define the sub distribution hazard, which is the hazard of the cumulative incidence function is the probability of failure due to cause k prior to time t. The cumulative incidence function is referred to as the sub distribution function, because it is not a true probability distribution. In order to express the proportional hazard model making an assumption on the sub distribution hazards model

$$h_1(t) = h_0\ (t)exp(\beta Z_1)$$

Where the $h_1(t)$ is the baseline sub-distribution hazard model of cause 1. The partial likelihood of these proportional hazards model is given by

$$L\ \ (\beta)\ = \prod_i \left(\frac{\exp(\beta Z_i)}{\sum_{j\varepsilon Ri}\ \mathrm{w_{ij}} \exp\left(\beta Z_j\right)} \right)^{\delta_{i=1}}$$

In the presence of competing risks, the sub distribution hazard is not the same as the cause-specific hazard, in terms of estimating these quantities; the difference is in the risk set. For the cause-specific hazard, the risk set (*Ri*) decreases at each time point when there is a failure of a different cause. However, for the sub-distribution hazard, individuals who fail from a competing cause remain in the risk set (*Ri*) until their potential censoring time. The weights w_{ij} are needed as soon as censoring occurs. Patients who experience no event of interest before X_i are given a weight w_{ij} = 1, whereas patients who experience competing events before X_i are given a weight w_{ij} that reduces with time.

$$w_{ij} = \frac{G(X_i)}{G\left(min\{X_i X_j\}\right)}$$

where G(t) is the Kaplan-Meier estimate of the survival function of the censoring distribution, which is the cumulative probability that a patient is still being followed at time t. The regression coefficients β are obtained by maximizing the partial likelihood L(β) and the covariance matrix of the parameter estimator is computed as a sandwich estimate.

Caplan and colleagues (1994) had commenced a critique of the Kaplan–Meier curve for estimating rates of cause-specific failures and also suggested to use cumulative incidence curves for cause-specific failure instead of using Kaplan–Meier analysis. Gray (1988) proposed a class of tests for comparing the cumulative incidence curves of a particular type of failure among different groups in the presence of competing risks and the Gray test compares weighted averages of the hazards of cumulative incidence function using the cumulative incidence estimation equation. Kim (2012) discussed methods to calculate the cumulative incidence of an event of interest in the presence of competing risks, to compare cumulative incidence curves in the presence of competing risks, and to perform competing risks regression analysis

Results

1. Data

A Dataset of 137 patients who underwent bone marrow transplant was used for the study from Klein and Moeschberger (1997). The Patients were categorized at the time of transplant into one of three risk categories: ALL (acute lymphoblastic leukemia) is coded as 1, AML (acute myelocytic leukemia)-Low Risk is coded as 2 and AML-High Risk is coded as 3.
The endpoint of interest is the disease-free survival time, which is the time to death or relapse or to the end of the study in days. In this data set, the variable Group represents the patient's risk category, the variable T represents the disease-free survival time, and the variable status is the censoring indicator, with the value 1 indicating an event time, value 2 indicating patients die before experiencing the event, the value 0 as a censored time. This data was analyzed using SAS Oracle Virtual Box Manager online software.

2. Comparison of the Cause specific and Competing Risks Models

In Cause Specific model patients who die in remission are considered as censored observations. 70% of the patients are considered as censored where as in the competing risks model it is only 39% (in **Table1**)

Table.1 - Summary of the Number of Event and Censored Values

Cause Specific				**Competing Risks**			
Total	**Event**	**Censored**	**Percent Censored**	**Event of Interest**	**Competing Event**	**Censored**	**Percent Censored**
137	41	96	70.07	41	42	54	39.41

Table 2 provides the deviance to assess the better model among all. The lowest deviance is the indicator for better model.

Table 2 - Model Fit Statistics

cause specific			**competing risks**	
Criterion	**Without Covariates**	**With Covariates**	**Without Covariates**	**With Covariates**
-2 LOG L	371.726	**370.665**	385.503	385.410

Illustrating with the **table 3** of the first part of Cause-Specific Hazards Ratios for group, the hazard of relapse for the AML high-risk patients is 1.4 times that for the AML low-risk patients (95% CI 0.67 to 2.97) and is 1.09 times that for the ALL patients (95% CI 0.48 to 2.46). The hazard of relapse for the ALL patients is 1.29 times that for the AML low-risk patients (95% CI 0.57 to 2.96).

Table 3 : Comparison Between Cause-Specific Hazards and Sub-distribution Hazards

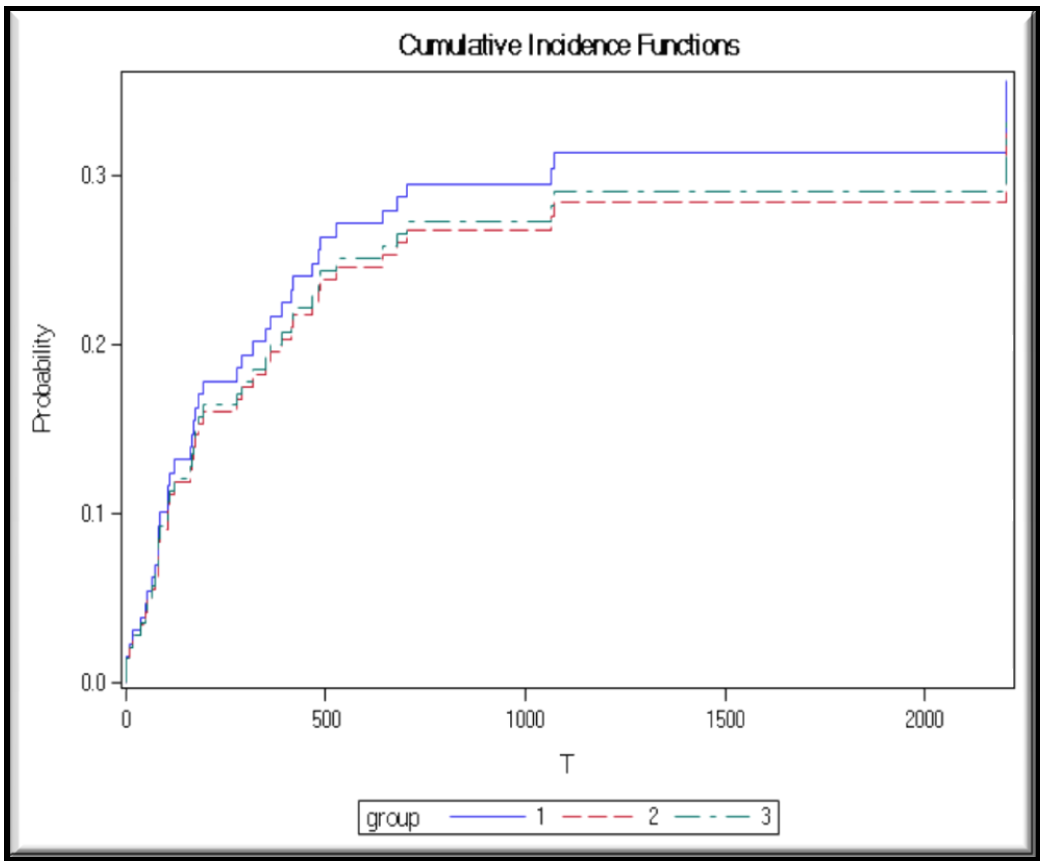
Cause-Specific Hazards: Hazard Ratios for group				**Sub-distribution Hazards: Hazard Ratios for group**		
Description	**Estimate**	**95% CI**		**Estimate**	**95% CI**	
group 2 vs 3	0.712	0.337	1.505	1.006	0.478	2.117
group 3 vs 2	**1.404**	**0.665**	**2.965**	**0.994**	**0.472**	**2.090**
group 2 vs 1	0.774	0.338	1.771	0.972	0.429	2.204
group 1 vs 2	**1.293**	**0.565**	**2.960**	**1.029**	**0.454**	**2.331**
group 3 vs 1	**1.086**	**0.479**	**2.459**	**0.966**	**0.419**	**2.228**
group 1 vs 3	0.921	0.407	2.086	1.035	0.449	2.387

Group:1=ALL, 2=AML-Low Risk &3=AML-High Risk

The second part of table 3 of Sub-Distributional Hazards Ratios for group, the hazard of relapse for the AML high-risk patients is 0.99 times that for the AML low-risk patients (95% CI 0.47 to 2.09) and is 0.97 times that for the ALL patients (95% CI 0.42 to 2.22). The hazard of relapse for the ALL patients is 1.029 times that for the AML low-risk patients (95% CI 0.45 to 2.33). According to the results presented in table 3, the cause-specific and the sub-distribution hazards ratios are merely distinguished each other. Since they are different and it could take precaution before interpreting the same. In addition, the nature of the data has also to be considered. The main difference between cause-specific and sub-distribution hazards is the risk set. For the cause-specific hazard the risk set decreases each time there is a death from another cause which is included as censoring. With the sub-distribution hazard subjects that die from another cause remain in the risk set and are given a censoring time that is larger than all event times. However, as explained above, the cause-specific and the sub-distribution hazards do not have the same interpretation. Another advantage of the cause-specific approach is that it is easier to handle than with the sub-distribution hazards model (Beyersmann and Schumacher (2008)). In summary, applying the sub-distribution hazards model is recommended for any predictive research, and for etiological research such as of medical science concerned with the causes and origins of diseases, the cause-specific hazards model is more appropriate (Noordzil et al, 2013).

The cumulative incidence of three groups is depicted through the below **figure1**. The Figure depicts that at any given time after the transplant group 2 and 3 is more likely to relapse than group 1 patients.

Figure1: Cumulative Incidence Function of three groups



Discussion

This paper compares the two familiar models for competing risks data. The model for cause specific who did not have experienced the event of interest can be fitted using Cox regression by censoring all individuals but it is difficult to plot the cumulative incidence. Also it is evident from this paper emphasizes that the models could be easily fitted using Oracle Virtual Box Manager by SAS which is accessed only through online. These two models yield almost similar results between the disease groups for the bone marrow transplant which is not in all the other data. Each of the model might provide useful insights about the covariates (Pintilie, 2006; Dignam, Zhang, and Kocherginsky 2012). Covariate effects in the cause-specific hazard model pertain to the event of interest only, without regard to how the covariates act on the competing risks. In the real world health perspective this model might be a little use to patients, where death from other risk factors plays a major role. It also concludes that for any predictive research, applying the sub-distribution hazards model is recommended, and for this kind of etiological research such as of medical science concerned with the causes and origins of diseases, the cause-specific hazards model is more appropriate and it provides quantities that are easy to interpret. Indeed, the sub-distribution hazards resulting from the sub-distribution method may not be interpreted as cause specific hazards, because patients who are in fact no longer at risk of the event of interest remain in the risk set. An advantage of the cause-specific approach is that the estimated hazards ratio can be interpreted as easier as than the sub-distribution hazards model.

References

1. Andersen, P.K., Geskus, R. B, de Witte, T., & Putter, H. (2012): Competing risks in epidemiology: possibilities and pitfalls. International Journal of Epidemiology Vol.41, No.3, pp.861-70
2. Benichou, J., and Gail, MH. (1990): Estimates of absolute cause-specific risk in cohort studies Biometrics Vol.46, No.3, pp.813-26.

3. Beyersmann, J., and Schumacher, M. (2008): Time-dependent covariates in the proportional subdistribution hazards model for competing risks Biostatistics; Vol.9, No.4, pp. 765–776.
4. Caplan, RJ, Pajak, TF, Cox, JD.(1994): Analysis of the probability and risk of cause-specific failure Int J Radiat Oncol Biol Phys Vol.29, No.5, pp.1183–6.
5. Collett, D. (1994):. Modelling Survival Data in Medical Research Chapman & Hall: London,
6. Cox, D.R. (1972): Regression analysis and life tables J R Stat Soc. Vol.34, No.2, pp.187-220.
7. Dignam, J. J., Zhang, Q., and Kocherginsky, M. (2012): The Use and Interpretation of Competing Risks Regression Models Clinical Cancer Research Vol.18, No.8, pp.2301–2308.
8. Fine, J. P. and Gray, R. J. (1999): A Proportional Hazards Model for the Sub-distribution of a Competing Risk Journal of the American Statistical Association Vol.94, No.446, pp.496–509.
9. Gray, RJ. (1988): A class of K-sample tests for comparing the cumulative incidence of a competing risk Ann Stat Vol.16, No.3, pp.1140-54.
10. Kim, HT.(2007): Cumulative Incidence in Competing Risks Data and Competing Risks Regression Analysis Clin Cancer Res Vol.13, No.2, pp.559-565
11. Klein J.P., Rizzo, J.D. Zhang, M-J., and Keiding, N.(2001): Statistical methods for the analysis and presentation of the results of bone marrow transplants. Part 2: Regression modeling Bone Marrow Transplantation Vol.28,No.11, pp. 1001–1011 - 2001 Nature Publishing Group
12. Klein, JP. and Moeschberger, ML. (1997): Survival Analysis, Techniques for Censored and Truncated Data Springer.
13. Korn, E.L., and Dorey, F.J. (1992): Applications of crude incidence curves Statistics in Medicine Vol.11, No.6, pp.813-29.
14. Noordzij, M., Leffondre, K., Stralen, KJV., Zoccali, C., Dekker, FW., et al. (2013): When do we need competing risks methods for survival analysis in nephrology? Nephrol Dial Transplant Vol.28, No.11, pp. 2670–2677.
15. Pintilie, M. (2006): Competing Risks: A Practical Perspective Statistics in Practice, Chichester, UK: John Wiley & Sons.
16. Prentice, R. L., Kalbfleish, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978): The Analysis of Failure Times in the Presence of Competing Risks Biometrics Vol.34, No.4, pp.541–544.
17. Putter, H., Fiocco, M., and Geskus, R. B. (2007): Tutorial in biostatistics: competing risks and multi-state models Statistics in Medicine, Vol.26, No.11, pp.2389-2430.
18. Rosthoj, S., Andersen, P. K., and Abidstrom, S. Z. (2004): SAS Macros for Estimation of the Cumulative Incidence Functions Based on a Cox regression Model for Competing Risks Survival Data Computer Method and Programs in Biomedicine Vol.74, No.1, pp.69–75.
19. Satagopan, JM, Ben-Porat, L, Berwick, M, Robson, M, Kutler, D, Auerbach, AD. (2004): A note on competing risks in survival data analysis Br J Cancer Vol.91, No.7, pp.1229–35.