International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor (2018): 7.426

Competing Risk Regression Models

Vaishali Halani[1], **Manish Thaker**[2]

[1]Research Scholar, Department of Statistics, School of Science, Gujrat University, Ahmedabad, India

[2]Head of Statistics Department, M.G Science Institute, Ahmedabad, India

Abstract: ***Competing risks are common in clinical research, as patients are subject to multiple potential failure events, both diseases related and otherwise. Competing risks methodology is being increasingly applied to cause of mortality data as a way of obtaining real world probabilities of mortality broken down by specific causes. For example, cancer patients with cardiovascular and other comorbidities are at concurrent risk of multiple adverse events. Regression models are employed to understand and exploit the relationship between the lifetime variable and the covariates. The most widely used regression models in competing risks are proportional because specific hazard model and proportional sub distribution hazard model. These models are frequently used in literatures and many authors have tried to differentiate and interpret both the models in different way. Several modeling approaches are available to evaluate the relationship of covariates to cause-specific failures with competing risk. Depending on which model is used, a distinctly different picture of the relationship of covariates to outcomes may be seen. It is important to choose a modeling approach that addresses the question of interest and subsequently interpret the results accordingly. We compared cause specific hazard model and sub distribution hazard model with flexible regression model to analyze and predict competing risk data in clinical trial applications using R software. These models are useful for a detailed analysis of how covariate effects predict the cumulative incidence, and allows for a time-varying effect of the covariates. From the above comparison, we can say that the choice of method for competing risk data in clinical trial should be guided by the scientific question.***

Keywords: survival analysis**,** competing risks, regression model, flexible regression

1. Introduction

Survival analysis is predominantly used in medical or clinical research when the primary interest is in observing time to event for primary survival endpoint of interest and censoring is independent of the primary event of interest. For example, time till death, appearance of some disease, relapse etc. However, in many cases, patients are at concurrent risk of more than one event and the happening of one of these events will obscured the happening of any other event. These types of events are in some sense compete each other for occurrence and is referred as competing risk events.

Applying classical survival analysis to competing risk events is not appropriate and misguided us in a way that it treats competing events as censored and primary event is still possible and failure from the cause of interest is no longer possible or we just no longer observe it. In competing risk analysis, unlike the regular independent censoring, competing events are censored as occurrence of competing risks and the cause of interest are not independent. The standard competing risk endpoints of interest usually include overall survival from death or event from any cause, disease free survival i.e time to death or event, progression free survival i.e time to death or event when patient already suffered from disease or time to event i.e. time to the cause of interest. All of these endpoints are also known as composite endpoints that have all causes. Hence standard survival analysis can be applied to these composite endpoints as they are subject to independent censoring such as withdrawal or lost to follow up. However, analysis and interpretation of time to event endpoint is difficult as it can be censored by competing cause of failure in addition to independent censoring. For example, patients with atherosclerotic risk factors of myocardial infarction are at concurrent risk of venous thrombosis [1]. Cancer patients with cardiovascular and other comorbidities are at concurrent risk of multiple adverse outcomes. [2]. similarly, peritoneal dialysis patients are at risk of death associated with risk of renal transplantation or transfer to hemodialysis [3].

2. Methods

2.1 Competing Risk Framework

Survival data are generally presented as a pair of (T,C), where T is the time at which event occurred and C is the censoring variable. When T is the time at which the event of interest is occurred, the censoring variable C is 1 and when T is the time at which the observation is censored, the censoring variable C is 0. The definition can be extended to the competing risks situation where j≥2 types of events are possible. The data are again presented as a pair of (T,C_i), though C will take on value i, where i is the type of first event observed (i=1,2,…..j). When T is the time at which the event of type i occurred the censoring variable C=i otherwise it is time of censoring and the censoring variable C = 0 [4,5]. Scrutinio et al [6] reported the results of a randomized, double blinded multicenter trial on patients with myocardial infraction (MI) treated with either ticlopidine or aspirin. In this trial, T is defined as the time from randomization to the first failure. The types of failure and therefore choices for C are cardiovascular death (i=1), non-vascular death (i=2), non-fatal MI (i=3), non-fatal stroke (i=4) and angina (i=5).

In the traditional analysis of competing risks data, the events due to all other causes except the event due to cause of interest are combined and treated as censored under the assumption that the causes of events are independent of each other. Recently, many different models have been developed to assess the lifetimes of a specific risk in presence of the other competing risk factors with assumption for the causes of failures to be dependent or independent. Two important concepts that are used to specify the distribution of the observable random pair (T, C) in competing risks set up are cause specific hazard rate functions ($\lambda_j(t)$) and cause specific sub distribution functions (cumulative incidence functions). The cause specific hazard may be better applicable for studying etiology of disease whereas sub distribution hazard are used in individual risk.

2.2 Competing Risk Regression

In survival studies, the focus of interest is to establish the relationship between failure time variable and covariates. The covariates or explanatory variables plays an important role in describing heterogeneity among failure time data in a population. Regression models are generally applicable to understand and determine the association between covariates and the failure time variable. For example, in a breast cancer study, factors such as age, tumor size, number of positive nodes can be considered as covariates. In some practical situations, the effect of covariates on failure time variable changes over time and such covariates are referred to as time-dependent or time varying covariates.

Several different regression models have been developed for failure time data to evaluate time invariant and time dependent covariate that affect the survival of patient from the competing risk event. The statistical analysis and inferences related to competing risk data has been analyzed in different ways by different authors. In the model of cause-specific hazard, there is no direct relation between the regression coefficients and the incidence of events as the effect of covariates on the competing event(s) is ignored, Prentice et al [7] proposed Cox-type regression on the cause-specific hazard where competing events are treated as censored observations and assumptions and extension as known from classical Cox regression. Benichou and Gail [8] and Dorey [9] have derived the estimates of the cause-specific hazard functions based on absolute risk regression. Unfortunately, hazard ratios as obtained by cause specific Cox regression analyses do not directly quantify the ability of the single markers to predict the unconditional absolute risk of an event of interest. Fine and Gray [10] introduced a regression approach focusing on sub distribution hazard. In the Fine and Gray model the regression coefficients are monotonously linked to the cumulative incidence function and the occurrence of competing events has an influence on the coefficients. Modified standard survival models can be fit to estimate the influence of the investigated covariates on the sub distribution hazard. Bryan et al [12] have discussed sub distribution hazard and cumulative incidence function for treating competing risks and their applications in regression settings. They have compared assumptions, uses and advantages of three different regression approaches viz, cause specific proportional hazard model, sub distribution proportional hazard model and parametric mixture model. The usage, interpretation and influence of covariate effect evaluated in of common competing risk regression models in relation to cause-specific hazard or on the cumulative incidence of the failure types have been discussed by James et al [12]. They also illustrate how covariate effects differ between these approaches in simulation studies. The difference, odds ratio and ratio between two cumulative incidence function has been investigated by Zhang et al. [13] Logistic risk regression is another useful model which can extend odds ratio to multiple regression in competing risks. However, relative absolute risks are easier to understand.

Recently, there is a boom in predictive modelling in biomarker research. There is high demand for statistical techniques to quantify the predictive capability of genotype, phenotype, environmental factors and treatment in future disease course of patients. For example, a patient diagnosed with diabetes may be interested in the risk of mortality related to cardiovascular disease. It is of interest in a larger perspective to quantify how multiple risk factors change the predicted risk of death caused by cardiovascular diseases [14] Practical properties of different regression models specifically for predicting the individual risks of cancer patients have been reviewed and compared by Thomas et al. These models are not new, and the mathematical properties are well studied in the framework of the linear transformation model [15]

In this article, we have reviewed and compared different competing risk regression approaches for estimation and assessment of covariate effects. We have chosen widely used cause specific hazard regression and sub distribution hazard regression and compared it with flexible hazard regression.

2.3 Cause Specific Hazard Rate:

The cause specific hazard functions plays an important role in competing risk framework as hazard rates can be estimated in the presence of censored observations. The cause-specific hazard rate for event type k provides an individual's probability for failing from an event of type k in an infinitesimal small time interval t to t+Δt given that failure has not occur from any event up to time t. For the cause specific hazard for event type k at time t individuals that failed from an event other than k prior to t removed from the remaining risk set.

$$\lambda_k\,(t) = \lim_{\Delta t\to 0} \frac{P(t \le T \ < t + \Delta t, D = k|T \ge t)}{\Delta t}$$

Cause specific hazard functions are well criticize in literature for its assumptions, interpretation and identifiability problems. Prentice et al. had proposed cause specific hazard function with time-dependent covariates for observable quantities where competing events are treated as censored observations.

$$\lambda_k(t|X) = \lambda_{k,0}\,(t) \exp(\ \beta_k^T X\ )$$

Where $\lambda_{k,0}$ (), is arbitrary and β_k, k=1,…..,m are cause specific regression coefficients to be estimated from data using standard aymptotic likelihood methods. Here inference on the effects of treatments or exposure variables X required no strong modelling assumptions under same set of conditions as causes of failures are assumed to be independent Thus $\lambda_{k,0}$() functions can be estimated with assumptions and extension as known from classical Cox regression. [7]

2.4 Sub distribution Hazard Rate

Cause specific hazard function is considered as standard analysis for competing risk data with the assumption that hazard rates are proportional but it does not provide direct interpretation of survival probabilities for a specific type of event. Thus, covariate effects testing on the sub distribution hazard function is not possible under cause specific hazard formulation and model selection issues and efficient prediction cannot be addressed directly. In order to define a “hazard-type” quantity that is directly linked to the cumulative incidence function, the marginal failure probabilities for a particular cause is intuitively appealing in the presence of competing risk data. Sub distribution hazard rate is introduced by Gray. For the sub distribution hazard rate, individuals for event type k at time t that failed from an event other than k prior to t remain in the risk set.[16]

$$\lambda_k^*(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \ < t + \Delta t, D = k | T \geq t \cup \{T < t, D \neq k\})}{\Delta t}$$

Fine and Gray proposed a semi-parametric Cox-type regression on the sub distribution hazard. Assumptions known from standard models are translated to sub distribution hazards (e.g. proportionality). The proportional sub distribution hazard is given as:

$$\lambda_k^*(t|X) = \lambda_{k,0}^* \,(t) \exp(\ \beta_k^{*T} X \ )$$

where $\lambda_{k,0}^* \ (t)$ is a completely unspecified, nonnegative function in t, the log(-log) transformation model results with $h_0(t)$ =log{$\int_0^t \lambda_{k,0}(s)ds$ }.Thus the baseline hazard and regression coefficients have a straightforward interpretation and does not depend on the probabilistic structure of the sub distribution hazard from the Cox transformation model In applications, we anticipate time x covariate interactions. To address this issue, they have extended the model to the case of time varying covariates X(t), which are functions of the original, time-independent covariates X and t. [10]

Direct link between regression coefficients and cumulative incidence:

$$F_k \,(t|X) = 1 - \exp(- \int_0^t \lambda_{0k}^*(s) \exp \ (\beta_k^{*T} X \ ) ds = 1 - \exp(- \ \Lambda_k^*(t|X))$$

2.5 Flexible Risk Regression

The cause specific hazard and cumulative incidence function for all causes are the standard approach and it contains the same information represented in different ways, thus leading to a different understanding of the subject matter. Scheike & Zhang consider a simple and flexible class of regression models that is easy to fit and also allows non-proportional hazards. It contains the Fine-Gray model as a special case which leads to a new simple goodness-of-fit procedure for the proportional sub distribution hazards assumption that aims in particular at representing time varying effects in the data that was not covered by the Cox type model. It fits a non-parametric, semi-parametric and parametric model for the cause-specific quantities. [17]

A class of flexible models represented as:

$$h\{P_1\,(t; x, z)\} = x^T \,\alpha(t) + g\,(z, \gamma, t)$$

where h and g are known link functions and $\alpha(t)$ and γ are unknown regression coefficients. Where h() is a known link-function h() and g(t, x, z) is known prediction-function for the probability of dying from cause 1 in a situation with competing causes of death.

In this article we considered two classes of flexible models from timereg package 19.2 of R software [17]:

1) The additive model where,
h (x) = 1-exp(-x) and g(t, x, z) = xTA(t) + (diag(tp)z)Tβ.

2) The proportional setting that includes the Fine & Gray (FG) "prop" model and some extensions where,
h(x) =1-exp(-exp(x)) and g(t, x, z) = xTA(t) + (diag(tp)z)Tβ.

The FG model is obtained when x = 1, but the baseline is parameterized as exp(A(t)).

2.6 Example of Bone Marrow Transplantation Study:

We analyzed data from 177 patients who received a stem cell transplant for acute leukemia. The aim of the analysis was to estimate the cumulative incidence of relapse in the presence of transplant-related death as competing events. The effect of predictive factors on relapse and its corresponding covariates such as Age, Sex, Disease (lymphoblastic or myeloblastic leukemia), Source of stem cells (bone marrow (BM) and peripheral blood (PB), or peripheral blood (PB)), and Phase at transplant (Relapse, CR1, CR2, CR3) were evaluated.

3. Results

The data set is available at http://www.stat.unipg.it/luca/R in the file ‘bmtcrr.csv’ and the contained variables are summarized in Table 1

Table 1: Variables in Bone Marrow transplant study data

Variable	Statistics	
Age	N	177
	Mean (SD)	30.5(13.04)
	Min-Max	04-64
Sex		
Male	N (%)	100(56.5)
Female	N (%)	77(43.5)
Disease		
ALL	N (%)	73(41.2)
AML	N (%)	104(58.8)
Phase		
CR1	N (%)	47(26.6)
CR2	N (%)	45(25.4)
CR3	N (%)	12(6.8)
Relapse	N (%)	73(41.2)

Status		
0	N (%)	46(26.0)
1	N (%)	56(31.6)
2	N (%)	75(42.4)
Source		
BM+PB	N (%)	21(11.9)
PB	N (%)	156(88.1)
ftime	N	177
	Median	6.6
	Min-Max	0.13-131.8

Table 2: Cause Specific Hazard

Cause 1: Death from treatment related causes				
	β	Exp(β)	Se(β)	p-value
platelet	-0.51987	0.59460	0.18721	0.00549 **
age	0.40836	1.50435	0.08903	4.51e-06 ***
tcell	-0.65169	0.52116	0.27634	0.01836 *
Cause 2: Relapse				
platelet	-0.2346	0.7909	0.2321	0.312
age	0.1425	1.1532	0.1118	0.202
tcell	0.3015	1.3519	0.2827	0.286

Table 3: Fine & Gray Sub distribution Hazard

Cause 1: Death from treatment related causes				
	β	**Exp(β)**	**Se(β)**	**p-value**
platelet	-0.426	0.653	0.1810	1.9e-02***
age	0.331	1.393	0.0799	3.4e-05***
tcell	-0.583	0.558	0.2699	3.1e-02***
Cause 2: Relapse				
platelet	-0.0587	0.943	0.230	0.800
age	-0.0212	0.979	0.121	0.860
tcell	0.5225	1.686	0.282	0.064

Table 4: Flexible Risk Regression-Additive Model

Cause 1: Death from treatment related causes		
Variable	Non-Parametric Model Test for constant effect	Parametric Model B (SE; P)
platelet	0.01	-0.00667(0.00254;8.72e-03***)
age	0.00	0.00617(0.00117;1.34e-07***)
tcell	0.04	-0.00905(0.00307;3.18e-03***)
Cause 2: Relapse		
platelet	0.26	-1.10e-04(0.001810; 0.952)
age	0.37	5.26e-05(0.000884; 0.953)
tcell	0.16	4.90e-03(0.003320; 0.141)

Table 5: Flexible Risk Regression-Multiplicative Model

Cause 1: Death from treatment related causes		
Variable	Non-Parametric Model Test for constant effect	Parametric Model B (SE; P)
platelet	0.16	-0.561 (0.2040;0.00605**)
age	0.13	0.324 (0.0918;0.00041***)
tcell	0.10	-0.694 (0.3030;0.02170*)
Cause 2: Relapse		
platelet	0.05	-0.0926 (0.241;0.701)
age	0.07	0.0294 (0.126;0.816)
tcell	0.04	0.4170 (0.296; 0.159)

4. Discussion

Cause specific hazard and sub distribution hazard regression approach has been widely used in most of the published articles for competing risk analysis but has its own limitations and relationships. It is known from the literature that cause specific hazard regression method are to be used for etiology and sub distribution hazard regression method are to be used for prognosis.

Both the methods uses different assumptions and may give different results. Flexible models for the cumulative incidence are very useful and also allows time varying covariates. It provides flexibility of non-parametric effects and it may lead to give different predictions for the cumulative incidence functions for the different causes. In our example, cause specific hazard and sub distribution hazard regression shows that out of all covariates, platelet, age & tcell are significantly affect the cause of death from treatment related causes and none of the covariates affect cause of relapse. Both the methods gives the same results with different degree of significance. Fine & gray sub distribution hazard is very sensitive and shows that platelet, age & tcell are highly significant. We have then used flexible risk regression nonparametric and parametric additive and multiplicative models and compare the results with cause specific hazard and sub distribution hazard regression. Flexible risk regression nonparametric and parametric additive models show the same as cause specific hazard regression and sub distribution hazard regression respectively for death from treatment related causes while with nonparametric and parametric multiplicative models gives different significance.

5. Conclusion

The selection of appropriate method shall be based on scientific question. Comparing the results of cause specific hazard, sub distribution hazard and flexible risk regression for competing risk analysis, the results reveal similar causes i.e. platelet, age & tcell in all the analysis with difference in level of significance. Hence it is quite difficult to say which link function should be preferred in flexible risk regression as per our understanding and experience while dealing with different models, the additive link function has the best numerical and small sample performance and is therefore to be preferred. Future research should continue to explore the differences in approaches and expand the tools to understand and implement competing risk methods for clinical trials.

6. Acknowledgement

The research leading these results have received no funding and no conflict of interest declared.

References

[1] Elizabeth A. Bayliss, Liza M. Reifler, Chan Zeng, Deanna B. McQuillan, Jennifer L.Ellis, John F. Steiner (2014). Competing risks of cancer mortality and cardiovascular events in individuals with multimorbidity; Journal of Comorbidity,4,29–36.

[2] Sigrid K. Brækkan, Erin M. Hald, Ellisiv B. Mathiesen, Inger Njølstad, Tom Wilsgaard, Frits R. Rosendaal, John-Bjarne Hansen (2012). Competing Risk of

Atherosclerotic Risk Factors for Arterial and Venous Thrombosis in a General Population: The Tromsø Study; Arterioscler Thromb Vasc Biol, 487-491.

[3] Laetitia Teixeira, Anabela Rodrigues, Maria J Carvalho, António Cabrita and Denisa Mendonça (2013). Modelling competing risks in nephrology research: an example in peritoneal dialysis; BMC Nephrology,14:110

[4] Kalbfleish, J. G. & Prentice, R. L. (1980). The Statistical Analysis of Failure Time Data. New York: Wiley.

[5] Martin J. Crowder (2001). Classical competing risks. New York: Chapman and Hall/CRC.

[6] Scrutinio, D., Cimminiello, C., Marubini, E., Pitzalis M. V., Di Biase, M. and Rizzon, P. (2001). Ticlopidine versus aspirin after mayocardial infraction (STAMI) trial. Journal of the American College of Cardiology, 37, 1259-1265.

[7] R. L. Prentice, J. D. Kalbfleisch, A. V. Peterson, Jr., N. Flournoy, V. T. Farewell, N. E. Breslow (1978). The Analysis of Failure Times in the Presence of Competing Risks, 34, 541-554

[8] Benichou, J and Gail, M.H. (1990). Estimates of absolute cause-specific risk in cohort studies, Biometrics, 46, 813-826

[9] Korn EL, Dorey FJ. Applications of crude incidence curves (1992). Statistics in Medicine 11(6), 813–829.

[10] Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the sub distribution of competing risk. Journal of American Statistical Association, 94,496-509

[11] Bryan Lau, Stephen R. Cole, and Stephen J. Gange (2009). Competing Risk Regression Models for Epidemiologic Data. American Journal of Epidemiology,170, 244–256

[12] Dignam, James & Zhang, Qiang & Kocherginsky, Masha. (2012). The Use and Interpretation of Competing Risks Regression Models. Clinical cancer research: an official journal of the American Association for Cancer Research, 18, 2301-2308

[13] Zhang Mei-Jie, Fine Jason. Summarizing differences in cumulative incidence functions. (2008) Statistics in Medicine,27, 4939–4949.

[14] Thomas A. Gerds, Thomas H. Scheike and Per K. Andersen. Absolute risk regression for competing risks: interpretation, link functions, and prediction. (2012) Statistics in Medicine. DOI: 10.1002/sim.5459

[15] Thomas H. Scheike and Mei-Jie Zhang (2008). Flexible competing risks regression modeling and goodness-of fit. Lifetime Data Analysis, 14, 464-483

[16] Gray, R.J (1988). A class of k-sample tests for comparing the cumulative incidence of competing risks. Annals of Statistics, 16, 133-143

[17] Thomas H. Scheike and Mei-Jie Zhang (2011). Analyzing Competing Risk Data Using the R timereg Package. Journal of Statistical Software, 38 464-483

Author Profile

Vaishali Thakkar received Msc. And Mphill degrees in Statistics from Gujarat University in 2000 and 2002 respectively. She has 15 years of industry experience with Synchron Research Services, Torrent Pharmaceuticals and Karmic Life Sciences LLP. She is now Founder and Director of Statiza Statistical Services, Ahmedabad.

Dr. Manish Thaker received his PhD degree from School of Science, Gujarat University, Ahmedabad. She is currently working as a head at Department of Statistics, M.G. Science Institute, Ahmedabad. His expertise is Logistic Regression and many more.