Regression Analysis of Masked Competing Risks Data under Cumulative Incidence Function Framework

Yosra Yousif
IIUM

Faiz A. M. Elfaki[*]
Qatar University

Meftah Hrairi
IIUM

Abstract

In the studies that involve competing risks, somehow, masking issues might arise. That is, the cause of failure for some subjects is only known as a subset of possible causes. In this study, a Bayesian analysis is developed to assess the effect of risks factor on the Cumulative Incidence Function (CIF) by adopting the proportional subdistribution hazard model. Simulation is conducted to evaluate the performance of the proposed model and it shows that the model is feasible for the possible applications.

Keywords: competing risks, masked causes of failure, subdistribution hazard, Bayesian analysis.

1. Introduction

Assessing the effect of risk factors on the lifetime of targeted subjects is one of the interests when competing risks data under discussion. Regression analysis commonly involve modeling either cause-specific hazard functions or cumulative incidence functions. However, the researchers noted that the methods based on cause-specific hazard under proportional hazard formulation disallow the analyst direct assessment of the effect of a covariate on the cumulative incidence function. Henceforth, the attention turned into the modelling of the cumulative incidence function directly. For example, Fine and Gray (1999) suggested a novel semiparametric proportional hazards model for the cumulative incidence function by applying the partial likelihood principle and weighting techniques. Fine and Gray (1999) proposed a semiparametric regression model based on the cumulative incidence function and adopted a class of models that includes the proportional hazards model and the proportional odds model. Moreover, Klein and Andersen (2005) developed a method to model the cumulative incidence functions directly based on pseudo values from a jackknife of the cumulative incidence function in a generalized linear model that allows various link functions. Study conducted by Jeong and Fine (2006) presented a parametric regression analysis of cumulative incidence function that involved the maximum likelihood inferences and were derived to fit the parametric models of cumulative incidence functions for all causes simultaneously. Scheike and Zhang (2008) introduced a simple and flexible class of regression models that is easy to fit and contains ele-

ments of the Fine and Gray (F & G) model as a special case. Recently, Eriksson, Li, Scheike, and Zhang (2015) proposed a proportional odds cumulative incidence model for competing risks data. Delord and Génin (2016) extended Pan's multiple imputation approach to Cox regression for the case of interval-censored competing risks data. The coefficient of interest, its variance-covariance matrix, and the baseline cumulative incidence function are updated from multiple posterior estimation derived from the Fine and Gray (1999) sub-distribution hazards regression that provides augmented data. Noticeably, most of the studies conducted in this area were focused on the cases where causes of failure are known for all the subjects under observation. Nevertheless, there were exceptions where the masking issues was considered, such as, Do and Kim (2017) regarded the case of missing causes of failure. They applied a Klein?Andersen's pseudo-value approach based on the estimated cumulative incidence function and a regression coefficient is estimated through a multiple imputation. In this paper, a Bayesian analysis is suggested to assess the effect of covariates on the cumulative incidence function in presence of masking. The rest of this article is arranged as follows, Section 2 introduces the likelihood construction by utilizing proportional sub-distribution hazard model and describes the Bayesian analysis. Section 3 presents some results of simulated data to evaluate the performance of the model. A brief conclusion is presented in Section 4.

2. Model construction and the Bayesian analysis

As there is an interest in the cumulative incidence function, the proportional hazard model for the subdistribution suggested by Fine and Gray (1999) is favoured. The model has the form

$$\lambda_j(T, X) = \lambda_{0j}(T)e^{\beta_j' X}, \tag{1}$$

where j is the cause of interest and λ_{0j}, β_j are the baseline hazard and the regression coefficients vector specific to the jth cause of failure and X denotes the covariates vector. The jth cumulative incidence function can be written as,

$$F_j(t, X) = \mathbb{P}\left(T \le t, C = j | X\right) = 1 - e^{-\Lambda_{0j}(t)e^{\beta_j' X}}, \tag{2}$$

where $\Lambda_{0j}(t) := \int_0^t \lambda_{0j}(s)ds$ is the cumulative baseline hazard. It is considered that, there are N units under observation and they are subject to K competing risks. T_i is defined as the time until the ith subject fail due to one of the K causes while X_i is the corresponding vector of covariates ($i = 1, 2, \ldots, N$). The cause of failure is not observed for all subjects. However, it can be defined as a Minimum Random Subset (MRS) as $S \subseteq \{1, \ldots, K\}$. So, the likelihood contribution of the ith unit from the observed data is defined as $\mathbb{P}\left(T_i, S_i | X_i\right)$, Kuo and Yang (2000), and can be written as,

$$\mathbb{P}\left(T_i, S_i | X_i\right) = \sum_{j \in s_i} \mathbb{P}\left(T_i, C_i = j | X_i\right) \mathbb{P}\left(S_i | T_i, C_i = j, X_i\right), \quad j = 1, \ldots, K; i = 1, \ldots, N.$$

Here C_i is the actual cause of failure for the ith unit. The above likelihood can also be written as,

$$\mathbb{P}\left(T_i, S_i | X_i\right) = \sum_{j \in s_i} f_j(T_i | X_i) \mathbb{P}\left(S_i | T_i, C_i = j, X_i\right).$$

So, the full likelihood for the observed data by considering the right-censored observations can be expressed as,

$$L = \prod_{i=1}^{n_1} \sum_{j \in s_i} \mathbb{P}\left(S_i | T_i, C_i = j, X_i\right) f_j(T_i | X_i) \prod_{i=n_1+1}^{n_2} S(T_i | X_i).$$

Here, n_1 and $n_2(n_1 + n_2 = N)$ denote the numbers of failed and right-censored subjects respectively. It can be rewritten in terms of subdistribution hazard and cumulative incidence function as follows:

$$L = \prod_{i=1}^{n_1} \sum_{j \in s_i} \mathbb{P}\left(S_i | T_i, C_i = j, X_i\right) \lambda_j(T_i | X_i) \left(1 - F_j(T_i | X_i)\right) \prod_{i=n_1+1}^{n_2} \left[1 - \sum_{j=1}^{K} F_j(T_i | X_i)\right]. \quad (3)$$

Now, by substituting (1) and (2) in (3), the final full likelihood can be presented as,

$$L = \prod_{i=1}^{n_1} \sum_{j \in s_i} \mathbb{P}\left(S_i | T_i, C_i = j, X_i\right) \lambda_{0j}(T_i) e^{\beta_j' X_i} e^{-\Lambda_{0j}(t) e^{\beta_j' X_i}} \prod_{i=n_1+1}^{n_2} \left[1 - \sum_{j=1}^{K} \left(1 - e^{-\Lambda_{0j}(t) e^{\beta_j' X_i}}\right)\right].$$

To apply Bayesian approach, the likelihood function of the observed data and prior distributions for unknown parameters need to be provided. For prior distributions, common priors are utilized such as gamma distribution for baseline hazards and independent Dirichlet priors for masking probabilities while the regression coefficients are assumed as normal distribution. WinBUGS software is employed as it is a robust and flexible tool for Bayesian survival analysis. This software implements Gibbs sampler which is basically the Markov Chain Monte Carlo (MCMC) algorithm.

3. Simulation study

By considering the competing risks data under cumulative incidence function formulation is in interest, Fine and Gray (1999) simulation algorithms are adopted. They suggested that the sub-distribution hazard of the cumulative incidence function should follow the form as presented in Equation (1). A simple competing risks model with two causes of failure and one covariate is assumed. The preliminary results, where almost 50% of the observations are masked, presents that the developed model performs well, and it is comparable to Fine and Gray (F & G) model despite the presence of masking. Table 1 shows the posterior estimations of the regression coefficients (estimated coefficient, standard deviation, and Monte Carlo error (MCE)). The (MCE) values demonstrate the accuracy of the estimated coefficients. Figure 1 and 2 show the comparisons between Fine and Gray (F & G) model and the current model for two causes. Noticeably, the graphs are close to each other and highly consistent.

Table 1: Posterior summaries of the regression coefficients of the developed model.

Model	Coeff	SD	MCE	Coeff	SD	MCE
Current	-0.6789	0.3287	0.003447	1.2910	0.3804	0.004322
Fine & Gray	-0.8869	0.3441	-	1.9150	0.4221	-

4. Concluding remarks

A regression analysis for competing risks data under cumulative incidence function (CIF) framework is developed in this paper by considering the cause of failure which could be masked for some subjects. Bayesian approach is adopted to estimate the regression coefficients and implemented by using Markov Chain Monte Carlo technique. The preliminary results obtained from the simulated data demonstrate that the developed model shows better performance, and it is applicable for the possible usages. More improvements in the proposed model will be introduced with proper illustration in the next paper where real data set will be used for analysis.
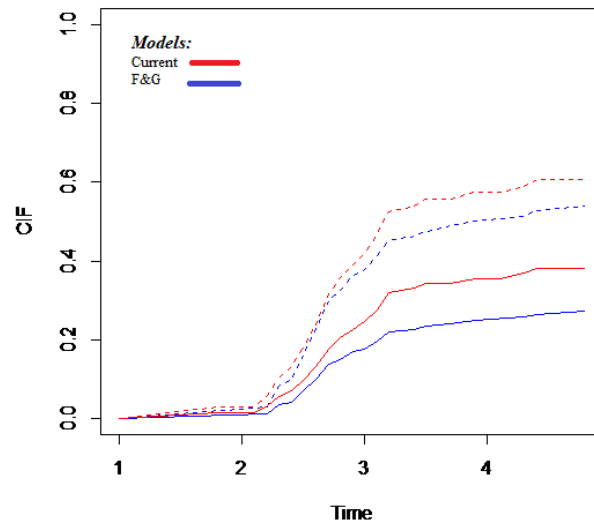
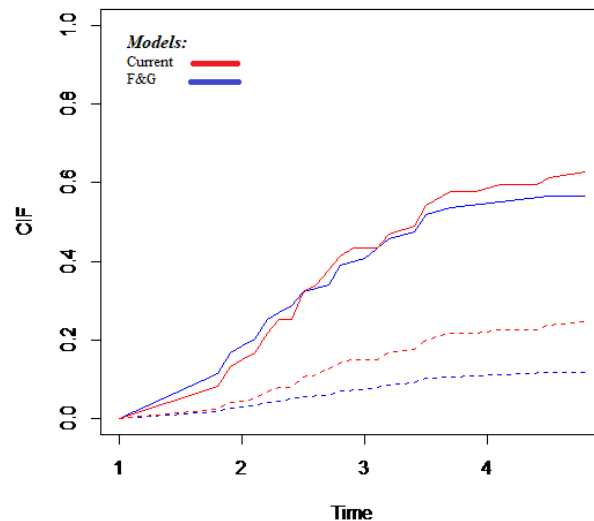Figure 1: Comparison of the CIF's of current model with F & G model for cause 1.

Figure 2: Comparison of the CIF's of current model with F & G model for cause 2.

References

Delord M, Génin E (2016). "Multiple Imputation for Competing Risks Regression with Interval-censored Data." *Journal of Statistical Computation and Simulation*, **86**(11), 2217–2228. ISSN 0094-9655. doi:10.1080/00949655.2015.1106543. URL https://doi.org/10.1080/00949655.2015.1106543.

Do G, Kim YJ (2017). "Analysis of Interval Censored Competing Risk Data with Missing Causes of Failure Using Pseudo Values Approach." *Journal of Statistical Computation and Simulation*, **87**(4), 631–639. ISSN 0094-9655. doi:10.1080/00949655.2016.1222530. URL https://doi.org/10.1080/00949655.2016.1222530.

Eriksson F, Li J, Scheike T, Zhang MJ (2015). "The Proportional Odds Cumulative Incidence

Model for Competing Risks." *Biometrics*, **71**(3), 687–695. ISSN 0006-341X. doi:10.1111/biom.12330. URL https://doi.org/10.1111/biom.12330.

Fine JP, Gray RJ (1999). "A Proportional Hazards Model for the Subdistribution of a Competing Risk." *Journal of the American Statistical Association*, **94**(446), 496–509. ISSN 0162-1459. doi:10.2307/2670170. URL https://doi.org/10.2307/2670170.

Jeong JH, Fine JP (2006). "Parametric Regression on Cumulative Incidence Function." *Biostatistics*, **8**(2), 184–196. ISSN 1465-4644. doi:10.1093/biostatistics/kxj040. http://oup.prod.sis.lan/biostatistics/article-pdf/8/2/184/621557/kxj040.pdf, URL https://doi.org/10.1093/biostatistics/kxj040.

Klein JP, Andersen PK (2005). "Regression Modeling of Competing Risks Data Based on Pseudovalues of the Cumulative Incidence Function." *Biometrics*, **61**(1), 223–229. ISSN 0006-341X. doi:10.1111/j.0006-341X.2005.031209.x. URL https://doi.org/10.1111/j.0006-341X.2005.031209.x.

Kuo L, Yang TY (2000). "Bayesian Reliability Modeling for Masked System Lifetime Data." *Statistics & Probability Letters*, **47**(3), 229–241. ISSN 0167-7152. doi:10.1016/S0167-7152(99)00160-1. URL https://doi.org/10.1016/S0167-7152(99)00160-1.

Scheike TH, Zhang MJ (2008). "Flexible Competing Risks Regression Modeling and Goodness-of-fit." *Lifetime Data Analysis*, **14**(4), 464–483. ISSN 1380-7870. doi:10.1007/s10985-008-9094-0. URL https://doi.org/10.1007/s10985-008-9094-0.

Affiliation:

Yosra Yousif
Faculty of Engineering
International Islamic University Malaysia
P.O. Box 10, 50728 Kuala Lumpur, Malaysia
E-mail: yosra_ar@yahoo.com

Faiz A.M. Elfaki (Corresponding Author)
Department of Mathematics, Statistics and Physics
College of Arts and Sciences, Qatar University
P.O. Box 2713, Doha, Qatar
E-mail: felfaki@qu.edu.qa

Meftah Hrairi
Faculty of Engineering
International Islamic University Malaysia
P.O. Box 10, 50728 Kuala Lumpur, Malaysia
E-mail: meftah@iium.edu.my