SOME METHODS FOR COMPETING RISKS WITH AND WITHOUT MASKED CAUSE OF FAILURE: A SHORT REVIEW.

Yosra Yousif[1], F. A. M. Elfaki[2,*], Meftah Hrairi[1] and M. Azram[2]
[1]Department of Mechanical Engineering, Faculty of Engineering, International Islamic University Malaysia,
Malaysia, 50728, Kuala Lumpur, Malaysia
[2]Department of Science in Engineering, Faculty of Engineering, International Islamic University
Malaysia, 50728, Kuala Lumpur, Malaysia
*Corresponding author e-mail: faizelfaki@yahoo.com; faizelfaki@iium.edu.my

ABSTRACT: Standard survival analysis focuses on failure-time data that has one type of failure. Competing risks arise in studies when subjects are exposed to more than one cause of failure and failure due to one cause excludes failure due to other causes. Examples of competing risks data can be found in many fields such as public health, industrial reliability, and demography. In some circumstances, the cause of failure in competing risks data could be masked. In such situations the standard survival analysis might not be applicable. There has been considerable work done discussing these issues in the statistical literature during the past several years, and this paper presents a review of some methods proposed for competing risks data with and without masking.

Keywords: Competing Risks; Masked Cause of Failure; Partly Interval-Censored; Bayesian analysis.

1. INTRODUCTION

The situation in which competing risks are present has been expressed in different ways. [1] described competing risks as the situation in which an individual can experience more than one type of event. [2] explained it as the failure to achieve independence between the time to an event and the censoring mechanism. [3] 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'.

Survival data in presence of competing risks arises in several areas such as public health, demography and industrial reliability. For example, in medical studies of treatment effect on death from cancer, death from other causes must be taken into account.

The standard methods for survival analysis might not be applicable in case of competing risks, for instance, the usual Kaplan-Meier estimates cannot be interpreted as probabilities in the presence of competing risks [4]. However, an alternative method, which is the cumulative probability of failure from a specific cause over time, was suggested and labeled the cumulative incidence function (CIF) that accounted for the competing risks and has a reasonable interpretation. A common mistake is to treat failures due to other causes as censored events and estimate the cumulative incidence function as the complement of the Kaplan-Meier estimate of the survival function. This leads to biased estimation unless it can be assumed that competing risks are independent.

It is often assumed that the failure time of the event of interest, the competing events failure times and the censoring times are all independent as this allows for the consistent estimation of the survival function. Although there are situations in which such independence assumption is of questionable validity, many methods (parametric or semi-parametric) under competing risks framework were introduced considering this assumption.

Most of the methods proposed for competing risks with or without masked cause of failure are based on right-censored failure time while less work has been done to study the case when the failure time is interval-censored. This paper reviews different methods that some of them develop statistical tests, some consider estimation of the joint survival function or cumulative incidence functions, and some discuss the regression problem.

2. COMPETING RISKS

2.1. METHODS FOR STATISTICAL TESTS

A considerable work has been done to develop tests to compare the cumulative incidence of different types of failure or the hazard rates in the presence of competing risks data. Not assuming that the underlying processes leading to failures of different types are acting independently for a given subject, and based on comparing weighted averages of the hazards of the subdistribution for the failure type of interest, [5] derived a class of tests to compare the cumulative incidence of a particular type of failure among different groups for right-censored competing risks data. While, [6] derived the likelihood ratio statistic for testing the null hypothesis of equality of cause- specific hazard rates, which are obtained under the restriction that these risks are uniformly ordered, against ordered alternatives. They also assumed a discrete time framework and allowed for right censoring. Unlike the two previous methods [7], considering uncensored case, provided a simple nonparametric hypothesis test (SNPHT) for comparing the cumulative incidence functions of a competing risks model when two causes of failure are possibly statistically dependent. The test statistic is the weighted sum of the differences of two cumulative incidence functions at system failure times.

2.2. METHODS FOR ESTIMATING THE JOINT SURVIVAL AND CUMULATIVE INCIDENCE FUNCTIONS

Estimation of the joint survival function and the cumulative incidence function based on competing risks data discussed extensively in the literature. A lot of attempts to estimate the joint survival function (generally called the multiple decrement function) were presented based on the observed data (including the failure time and the cause of failure), however the question of the identifiability of the estimator immediately arises. To overcome identifiability problem [8] established the identifiability of the problem under the assumption that the random variables corresponding to the cause-specific failure times are mutually independent although this assumption of independence cannot be tested from the observed data. Bayesian methods, on the other hand, can provide unambiguous solutions to inference problems in the presence of nonidentifiability. Thus a number of studies in which Bayes estimates have been obtained were developed although the inconsistency of Bayes estimators was established in the presence of nonidentifiability. [30] stablished the foundations from which the efficacy of Bayesian updating in competing risks problems can be assessed, as the efficacy of Bayes estimators in such contexts (nonidentifiability) has received rather little attention. They use the term "efficacy" in comparing prior and posterior estimates, and refer to Bayesian updating as efficacious in circumstances in which the posterior estimate is better than the prior estimate.

They obtained a useful representation of the posterior distribution of the multiple decrement function assuming a Dirichlet process prior with a continuous prior measure $\boldsymbol{\alpha}$ and derive the limiting posterior distribution. Later in a companion paper, [9] developed a complementary analysis with a discrete $\boldsymbol{\alpha}$ to provide Bayes estimators of the multiple decrement function from discrete data.

As the usual Kaplan-Meier estimate cannot be interpreted as probability in the presence of competing risks, the cumulative incidence function (CIF) as an alternative, that accounted for the competing risks and has a reasonable interpretation, was first suggested by [10]. Since then many authors have studied the estimation of the CIF considering different situations. [11] proposed a simple version of the Gompertz distribution to parameterize the cumulative incidence function directly. They mentioned that in cases where the cumulative incidence function is of primary interest, it may be more natural to model the cumulative incidence function directly than it is to model the cumulative incidence function indirectly via the cause-specific hazards. While [12], using the stable distributions family of Hougaard, proposed a new four-parameter distribution by extending a two-parameter log-logistic distribution. They carried out a simulation study to compare the cumulative incidence estimated with this distribution with the estimates obtained using a non-parametric method. The results indicated that the four-parameter modeling of CIF was as efficient as the non-parametric method and sometimes led to better estimates of CIF. Recently, [13] considered parametric estimation of the cumulative incidence function (CIF) for competing risks data subject to interval censoring. In their article, they extended the [11] models to the general case of interval censored competing risks data and both maximum likelihood estimators (MLEs) and a naive estimator, which enables separate estimation of models for each cause, are considered. The naive likelihood is shown to be valid under mixed case interval censoring, but not under an independent inspection process model, in contrast with full maximum likelihood which is valid under both interval censoring models.

2.3. METHODS FOR REGRESSION

Considering the regression issue, a noticeable work has been presented based on competing risks data. For competing risks data with covariates, the analysis usually involves modeling the cause-specific hazard functions or cumulative incidence functions. As the cause-specific hazard function does not have a direct interpretation in terms of survival probabilities for a particular failure type, [14] proposed a novel semiparametric proportional hazards model for the subdistribution. The partial likelihood principle and weighting techniques were employed to derive estimation and inference procedures for the finite-dimensional regression parameter under a variety of censoring scenarios. Likewise, [15] modeled the cumulative incidence functions directly extending [39] method to the regression setting. They developed maximum likelihood inferences in which parametric models for the cumulative incidence functions for all causes are fit simultaneously. The baseline distribution was modeled using a Gompertz specification, which accommodates improper distributions like the cumulative incidence function. Furthermore, [16] considered a simple and flexible class of regression models that is easy to fit and contains the Fine–Gray model as a special case. One feature of this approach is that the regression modeling allows for non-proportional hazards. It is also allows for assessing the time-varying covariate effects in the data that cannot be covered by the Cox type model. In contrast, under the cause specific hazard rate framework [17] introduced a semiparametric Bayesian method assuming that each cumulative baseline cause-specific hazard rate function has a gamma prior distribution, and a marginal likelihood function based on data and the prior parameter values was proposed for the estimation of regression parameters by considering cumulative baseline cause-specific hazard rate functions as a nuisance parameter. Their method generalizes the work of [18] to the competing risks set up. Similarly, [19] modeled each cause-specific hazard with the additive hazard model that includes both constant and time-varying covariate effects employing [20] additive model. They developed the estimation methods for the covariate effects on the all-cause hazard, as well as for the survival and cumulative incidence functions, and derived the variance estimators of the estimated survival and cumulative incidence functions.

3. COMPETING RISKS WITH MASKED CAUSE OF FAILURE

3.1. METHODS BASED ON SERIES SYSTEM FORMULATION

In competing risks data, sometimes information on the actual cause of failure for some subjects might be missing and can only be narrowed down to a set of potential causes. Such data is termed masked failure data, and it is quite common in the fields involving competing risks data. This phenomenon was widely studied during the three last decades. A substantial literature is available in work considering series system formulation in terms of latent failure times (observing the minimum of several lifetimes), as authors are often interested in estimating the reliability of components. For example, [21] discussed how to exploit all of the available information using a maximum likelihood approach extending and clarifying the useful work of [22]. They developed a likelihood approach for estimating component reliabilities from system life data when the cause of failure may be unknown. They adopted the symmetry assumption that entails an equal chance of observing the same masked subset of risk components irrespective of the true cause (within the masking subset) of failure. Unlike, [23] discussed reliability estimation when the masking probability is dependent on the particular cause of failure. They mentioned that when the number of masked failure is small relative to the number of observed failures the MLE's will often be relatively robust. They also illustrated how far off the maximum likelihood estimates (MLE's) of failure rates assuming independent masking can be from the correct MLE's under various degrees of dependent masking. Recently, [24] showed that the solution to the nonparametric maximum likelihood estimate (NPMLE) is not unique, and the NPMLE proposed in the current literature is inconsistent. Moreover, they constructed a consistent NPMLE and established its asymptotic normality. Their proofs do not rely on the symmetry assumption. In contrast, [25] developed methods for analyzing masked data from a Bayesian perspective. Under the assumptions of independent masking and exponentially distributed component life, they have shown that masked system life data can be effectively used in the estimation process. While [26] developed a Bayesian analysis for masked data from a general K component system. They assumed non-identical Weibull distributions for the failure times, and for more flexibility three different prior models are proposed. One crucial assumption in their method is the assumption of independence among the K competing risks. Furthermore, [27] investigated a general Bayesian formulation that includes most commonly used parametric lifetime distributions and that is sufficiently flexible to handle complex forms of censoring. They have provided a unified Bayesian modeling and inference framework under methodology based on Markov chain sampling. [28] unlike [27] have not subjected the masking probabilities to the symmetry assumption. They used independent Dirichlet priors to marginalize these nuisance parameters, and illustrated the developed methodology for the case where the components of a series system have independent log-Normal life distributions by employing independent Normal-Gamma priors for the component lifetime parameters. They mentioned that Gibbs sampling scheme developed for the required analysis can also be used to provide a Bayesian analysis of data arising from the conventional competing risks model of independent log-Normals. In addition, [29] considered a nonparametric Bayesian approach for masked data extending [30] results to series systems with partially masked competing risks. Relaxing the usual assumption of independence of failure causes they obtained the posterior distribution of the joint survival function, assuming a Dirichlet process prior, and derived the limiting posterior distribution. They showed that the posterior estimate of the reliability of the series system of interest in practice is consistent.

3.2. METHODS BASED ON CAUSE-SPECIFIC HAZARD FORMULATION

Using the cause-specific hazard formulation, many methods were developed to make inference for competing risks with masked data. [31] showed how stage1 and stage 2 information can be combined to provide statistical inference in the presence of competing risks data with masked cause of failure. They proposed maximum likelihood estimation using a model with nonparametric proportional hazards. Since in many practical situations the assumption of proportional hazards cannot be justified, [32] proposed a model with completely parametric cause-specific hazards, and discussed in detail the case where the failure times for the competing risks have a Weibull distribution. Moreover, based on the EM algorithm [33] proposed inference methods for estimating the parameters of a weakly parameterized competing risks model with masked causes of failure and second-stage data. Their method is flexible enough to handle cases with only one-stage data or grouped data, and to allow the masking probabilities to depend on time. Further, it can be used to conduct robust likelihood ratio tests for some assumptions such as the proportional hazards assumption and the symmetry assumption. In addition, [34] proposed an EM-based approach which allows for dependent competing risks and produces estimators for the subdistribution functions (CIFs). They showed that models with piecewise constant hazards can be used for estimation in a general situation in which the competing risks do not act independently. Besides, they discussed identifiability of parameters if none of the masked items have their cause of failure clarified in a second stage analysis.

3.3. METHODS FOR REGRESSION

To assess the covariates effect in competing risks data with masked cause of failure, noticeable methods were developed. For example, [35] proposed a method to estimate the regression coefficients in a competing risks model when cause of failure is missing for some subjects, and where the cause-specific hazard for the cause of interest is related to covariates through a proportional hazards relationship. They used parametric models to model the probability that a missing cause is that of interest, and employed multiple imputation procedures to impute missing cause of failure. In addition, when cause of failure is missing at random [36] considered the problem of estimating the regression coefficient, where the relationship between the cause-specific hazard for the cause of interest and covariates is described using linear transformation models. They derived augmented inverse probability weighted complete-case estimators for the regression coefficient. [37] developed a semiparametric Bayesian approach under the framework of cause-specific hazard function, where the partial information about the cause of death is incorporated by means of latent variables. Their approach does not need to rely on assumptions such as symmetry and second-stage data. They discussed three different models using a variety of priors, and exploited the simulation-based MCMC technique to implement the Bayesian methodology. While [38] explored the use of multiple imputations for cause-specific hazard modeling of cumulative incidence when causes of death data are incomplete. The effects of covariates on cause-specific hazard functions were assumed to satisfy proportional hazards models. They proved the asymptotic normality of the multiple imputation estimates and derived a consistent variance estimate for them.

4. CONCLUSION

In this paper we tried to review some methods involving competing risks data with and without masked cause of failure. The literature on such data is abundant, since interesting and varied methods have been presented considering such data in order to tackle inference issues. As can be noticed many of these methods were developed under some common assumptions although these assumptions are of questionable validity and cannot be tested using the observed data. For instance, the independence of the competing risks and censoring mechanism assumption is crucial for some methods to work out. Besides, the symmetry assumption which has been extensively employed when the case is competing risks with masked cause of failure. Moreover, the majority of authors discussed the case when the failure time is right-censored. However, much less work has been done considering other types of censoring, and no one according to our knowledge studied the partly interval-censored case under competing risks framework. There are many topics can be investigated for partly interval-censored data based on competing risks with and without masking. For instance, one can study the estimation of the cumulative incidence functions, or discuss the regression case either parametrically or non-parametrically for both cases with and without masking.

Extending some methods reviewed above we are going to study a Bayesian analysis to competing risks with masked data in the presence of partly interval-censoring. Our work is based on cause-specific hazard formulation, and utilizes the Markov Chain Monte Carlo (MCMC) technique. Furthermore, our interest is in engineering applications as most of the approaches that have been developed in engineering field were based on series system formulation.

REFERENCES

[1] Kalbfleisch D. J, and Prentice. L. R. (2002)." *The Statistical Analysis of Failure Time Data*." Edition: 2. New York: john wiley & sons.

[2] Caplan, Richard J., Thomas F. Pajak, and James D. Cox., (1994). "*Analysis of the Probability and Risk of Cause-Specific Failure.*" International Journal of Radiation Oncology Biology Physics 29(5):1183–86.

[3] Gooley, T. A., W. Leisenring, J. Crowley, and B. E. Storer., (1999). "*Estimation of Failure Probabilities in the Presence of Competing Risks: New Representations of Old Estimators.*" Statistics in medicine 18(6):695–706.

[4] Pintilie, Melania., (2006). "Competing Risks: A Practical Perspective". John Wiley& Sons.

[5] Gray, Robert J., (1988). "*A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risks.*" The Annals of Statistics 16(3):1141–54.

[6] Dykstra, Richard, Subhash Kochar, and Tim Robertson., (1995). "*Likelihood Based Inference for Cause Specific Hazard Rates under Order Restrictions.*" Journal of multivariate analysis 163–74.

[7] Sun, Yanqing, and Ram C. Tiwari., (1997). "*Comparing Cumulative Incidence Functions of a Competing-Risks Model.*" IEEE Transactions on Reliability 46:247–53.

[8] Berman, Simeon M., (1963). "*Note on Extreme Values, Competing Risks and Semi-Markov Processes.*" The Annals of Mathematical Statistics 34(3):1104–6.

[9] Neath, Andrew a., and Francisco J. Samaniego., (1997). "*On Bayesian Estimation of the Multiple Decrement Function in the Competing Risks Problem, II: The Discrete Case*." Statistics & Probability Letters 35(4):345–54.

[10] Kalbfleisch D. J, and Prentice. L. R, (1980). "*The Statistical Analysis of Failure Time Data"*. New York: John wiley & sons,inc.

[11] Jeong, Jong Hyeon, and Jason Fine., (2006). "*Direct Parametric Inference for the Cumulative Incidence Function.*" Journal of the Royal Statistical Society. Series C: Applied Statistics 55:187–200.

[12] Shayan, Zahra, Seyyed Mohammad Taghi Ayatollahi, and Najaf Zare., (2011). "*A Parametric Method for Cumulative Incidence Modeling with a New Four-Parameter Log-Logistic Distribution.*" Theoretical biology & medical modelling 8:43.

[13] Hudgens, Michael G., Chenxi Li, and Jason P. Fine., (2014). "*Parametric Likelihood Inference for Interval Censored Competing Risks Data*." Biometrics 70(1):1–9.

[14] Fine, J. P., and Gray, R. J. (1999). "*A Proportional Hazards Model for the Subdistribution of a Competing Risk.*" American Statistcal Association 94(446):496–509.

[15] Jeong, Jong-Hyeon, and Jason P. Fine., (2007). "*Parametric Regression on Cumulative Incidence Function.*" Biostatistics *(Oxford, England)* 8(2):184–96.

[16] Scheike, Thomas H., and Mei-Jie Zhang., (2008). "*Flexible Competing Risks Regression Modeling and Goodness-of-Fit."* Lifetime data analysis 14(4):464–83.

[17] Sreedevi, E. P., and P. G. Sankaran., (2009). "*A Semiparametric Bayesian Approach for the Analysis of Competing Risks Data.*" Communications in Statistics - Theory and Methods 41(15):2803–18.

[18] Kalbfleisch, John., (1978). "*Non-Parametric Bayesian Analysis of Survival Time Data.*" Journal of the Royal Statistical Society 40(2):214–21.

[19] Zhang, Xu, Haci Akcin, and Hyun J. Lim., (2011). "*Regression Analysis of Competing Risks Data via Semi-Parametric Additive Hazard Model.*" Statistical Methods & Applications 20(3):357–81.

[20] Mckeague, Ian W., and Peter D. Sasieni., (1994). "*A Partly Parametric Additive Risk Model.*" Biometrika. 81(3):501–14.

[21] Guess, Frank M., John S. Usher, and Thom J. Hodgson., (1992). "*Estimating System and Component Reliabilities under Partial Information on Cause of Failure.*" Journal of Statistical Planning and Inference 29:75–85.

[22] Miyakawa, Masami., (1984). "*Analysis of Incomplete Data in Competing Risks Model.*" IEEE Transactions on Reliability R-33(4):293–96.

[23] Dennis, L. K. J., (1994). "*System Life Data Analysis with Dependent Partial Knowledge on the Exact Cause of System Failure.*" Microelectronics Reliability 34:535–44.

[24] Yu, Qiqing, and Jiahui Li., (2012). "*The NPMLE of the Joint Distribution Function with Right-Censored and Masked Competing Risks Data.*" Journal of Nonparametric Statistics 24(3):753–64.

[25] Reiser, B., I. Guttman, Dennis K. J. Lin, Frank M. Guess, and John S. Usher., (1995). "*Bayesian Inference for Masked System Lifetime Data.*" Royal Statistical Society 44(No. 1):pp. 79–90.

[26] Basu, Sanjib, P. Basu Asit, and Chiranjit Mukhopadhyay, (1999). "*Bayesian Analysis for Masked System Failure Data Using Non-Identical Weibull Models.*" Journal of Statistical Planning and Inference 78:255–75.

[27] Basu, Sanjib, Ananda Sen, and Mousumi Banerjee. (2003). "*Bayesian Analysis of Competing Risks with Partially Masked Cause of Failure.*" Journal of the Royal Statistical Society: Series C (Applied Statistics) 52(1):77–93.

[28] Mukhopadhyay, Chiranjit, and Sanjib Basu., (2007). "*Bayesian Analysis of Masked Series System Lifetime Data.*" Communications in Statistics - Theory and Methods 36(2):329–48.

[29] Xu, Ancha, and Yincai Tang., (2011). "*Nonparametric Bayesian Analysis of Competing Risks Problem with Masked Data.*" Communications in Statistics - Theory and Methods 40(13):2326–36.

[30] Neath, Andrew a., and Francisco J. Samaniego., (1996). "*On Bayesian Estimation of the Multiple Decrement Function in the Competing Risks Problem*." Statistics & Probability Letters 31(2):75–83.

[31] Flehinger, B., (1998). "*Survival with Competing Risks and Masked Causes of Failures.*" Biometrika 85(1):151–64.

[32] Flehinger, Betty J., Benjamin Reiser, and Emmanuel Yashchin., (2002). "*Parametric Modeling for Survival with Competing Risks and Masked Failure Causes.*" Lifetime data analysis 8(2):177–203.

[33] Craiu, Radu V., and Thierry Duchesne., (2004). "*Inference Based on the Em Algorithm for the Competing Risks Model with Masked Causes of Failure.*" Biometrika 91:543–58.

[34] Craiu, Radu V, and Benjamin Reiser., (2006). "*Inference for the Dependent Competing Risks Model with Masked Causes of Failure.*" Lifetime data analysis 12(1):21–33.

[35] Lu, K., and a a Tsiatis., (2001). "*Multiple Imputation Methods for Estimating Regression Coefficients in the Competing Risks Model with Missing Cause of Failure.*" Biometrics. 57(4):1191–97.

[36] Gao, Guozhi, and Anastasios A.Taiatis., (2005). "*Semiparametric Estimators for the Regression Coefficients in the Linear Transformation Competing Risks Model with Missing Cause of Failure."* Biomertrika 92(4):875–91.

[37] Sen, Ananda, Mousumi Banerjee, Yun Li, and Anne-Michelle Noone., (2010). "*A Bayesian Approach to Competing Risks Analysis with Masked Cause of Death.*" Statistics in medicine 29(16):1681–95.

[38] Lee, Minjung, Kathleen a Cronin, Mitchell H. Gail, James J. Dignam, and Eric J. Feuer., (2011). "*Multiple Imputation Methods for Inference on Cumulative Incidence with Missing Cause of Failure.*" Biometrical Journal. 53(6):974–93.

[39] Jeong, Jong-hyeon., (2006). "*Direct Parametric Inference for the Cumulative.*" Appl. Statist. 55(2):187–200.