Golestan University of Medical Sciences

Jorjani Biomedicine Journal

Online ISSN: 2645-3509

Analysis of competing risks in the CoxPH model for progressive censorship with binomial removal

Ahmad Pourdarvish [1], Reza Hashemi [2], Jabbar Azar [3], Solmaz Norouzi [4*]

1. Department of Statistics, University of Mazandaran, Babolsar, Iran
2. Department of Statistics, Faculty of Basic Science, Razi University, Kermanshah, Iran
3. University of Applied Science and Technology, Branch of Ravansar, Ravansar, Iran
4. Department of Statistics and Epidemiology, School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran

* Correspondence: Solmaz Norouzi. Department of Statistics and Epidemiology, School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran
Tel: +989102032985; Email: snorouzibiostatistics@gmail.com

Abstract

Background: In medical research and survival analysis, it is common for an individual or item's failure to be attributable to multiple causes, also known as competing risks. This article focuses on examining the competing risks model as the data increasingly becomes type II censored and randomly removed. The model assumes that the causes of failure are independent and that the lifetimes of individuals are described by the Cox model. At each failure time, the number of items or people removed follows a binomial distribution. The article derives estimators for the indefinite parameters in the model. The study presents a set of detailed data and includes a simulation study that also illustrates the results.

Methods: Different reasons, frequently known as competing risks, are frequently embroiled in an individual's or an item's failure in medical research survival analysis. The competing risks shown under sort II dynamic censoring with random removals are the subject of this research.
We get the maximum likelihood and inexact most extreme probability estimators of the obscure parameters. The asymptotic distribution of the maximum probability estimators is utilized to decide the CIs. Then, Monte Carlo simulations were applied to demonstrate the approach. The analyses were performed utilizing R 4.0.4 software.

Results: For stroke, systolic blood pressure (SBP) and hypertension status are the only significant variables. In contrast, gender, body mass index (BMI), smoking status, the logarithm of urinary albumin and creatinine ratio, and diabetes status are significant variables for coronary heart disease (CHD) and other cardiovascular diseases (CVDs). The results suggest that significant risk factors differ for different types of CVD events.

Conclusion: The outcomes of the simulation study indicate that progressively right-censored type II sampling designs outperformed the usual censored type II sampling designs. Therefore, the estimated parameters on the defined pattern setting are recommended. They can be used in many practical situations when competing risks occur, and progressive censoring could be considered.

Article History

Received: 16 June 2023
Received in revised form: 20 October 2023
Accepted: 14 December 2023
Published online: 21 December 2023
DOI: 10.29252/jorjanibiomedj.11.3.14

Keywords

Proportional Hazards Models
Binomial removals
Competing risks
Progressive censoring

Article Type: Original Article

Highlights

What is current knowledge?

Traditional approaches treat competing risks as censored data and analyze the lifetime distribution for a specific cause of failure.

What is new here?

The study focuses on survival analysis and competing risks where failures or events can have multiple causes.

The article introduces a specific focus on examining competing risks in the CoxPH model for progressive censorship with binomial removals.

It presents estimators for the model's indefinite parameters and provides a simulation study to illustrate the results.

Introduction

In certain medical investigations, the occurrence of individual or item failure can be attributed to multiple causes or factors that compete for the failure event of the subject under study. Competing risks emerge when an individual faces the possibility of failing in multiple ways, yet only one of these failure types can transpire (1-3). For instance, a person could pass away due to either cancer or a heart stroke but not both (even though they might have both ailments prior to their demise). Another instance involves a randomized clinical trial comparing treatments for lung cancer, where patients may succumb to lung cancer, heart disease, or other underlying reasons (4). Traditionally, researchers have been primarily concerned with analyzing the lifetime distribution under a specific cause of failure, treating all other competing causes as censored data (5). However, more recently, advanced models have emerged to assess the lifetimes of specific risk factors in the presence of competing risk factors. These models leverage datasets containing time-to-event information along with an indicator variable that signifies the specific cause of failure for the individual or item under observation. It is possible to assume that these distinct failure causes are either independent or dependent on each other (6). In many cases, the analysis of competing risk data assumes the independence of failure causes. For an in-depth exploration of various competing risk models, refer to Crowder and the comprehensive work by David and Moeschberger (7).

In some studies, models like Cox regression have been employed to detect survival with competing risks, aiming to investigate the impact of each covariate variable on the occurrence of distinct events and to estimate the hazard function. The Cox model is commonly used when assessing the influence of multiple variables on survival time concurrently. In essence, the observed distribution of survival times is associated with a particular failure factor, with other causes being treated as censored data. In this context, the Cox model is widely utilized as a semi-parametric model (8,9).

Let T be a random time and $\underline{Z}$ be a vector of covariates in R^p. The proportional hazards model assumes that conditional on $\underline{Z}$, the hazard rate function of T, is given by $h(t;\underline{Z}) = h_0(t)\exp(\underline{\beta}^T\underline{Z}),\ t \in R^+$, where $\underline{\beta} \in R^p$ is an indefinite regression parameter vector, and $h_0(t)$ is an indefinite baseline hazard rate function (4).

Censorship represents an inherent element of medical and reliability studies, stemming from the practical challenge of acquiring comprehensive data about the complete lifespan of every individual. For instance, participants in a follow-up study might withdraw or the study might need to conclude at a predetermined time point. The 2 primary forms of censoring are type I and type II censoring (10-12). The present paper primarily delves into the realm of competing risk data within the context of progressively type II censoring. Progressive censoring is of particular significance in designing duration experiments, especially in the domain of reliability studies.

We employ competing risk data analysis within the framework of progressively type II censoring, which assumes elevated importance in shaping the design of duration experiments, particularly in the context of medical research. Both the conventional and progressive censoring structures share similar objectives, yet the latter is strategically designed to alleviate information loss by effectively managing the number of observed failures relative to the broader sample approach.

Let $(T_1, \underline{Z}_1), \ldots, (T_n, \underline{Z}_n)$ be n independent of (T, $\underline{Z}$). A progressive type II censored sample is attained using the following method. First, suppose that we

are given integers $r_1, r_2, \dots, r_m$, chosen a priori, such that $r_1 + \cdots + r_m + m = n$. Consider the lifetimes $T_1, T_2, \dots, T_n$ as the failure times of n items that are placed on the test at time zero. We denote by $X_{(1)}$ (the first failure time) and i_1 (the number of the unit that failed). Immediately after the occurrence of the first failure, a certain number of items r_1 labeled $i_2, \dots, i_{r_1+1}$ are randomly chosen and removed from the test. This set of removed items is denoted as $I_1 = [i_1, \dots, i_{r_1+1}]$. Then, at the second observed failure time $X_{(2)}$, the corresponding item number is i_{r_1+2}, and a new set of r_2 surviving items labeled $i_{r_1+3}, \dots, i_{r_1+r_2+2}$ is randomly selected and removed from the test. We then denote $I_2 = [i_{r_1+2}, \dots, i_{r_1+r_2+2}]$. This process continues until, at the time $X_{(m)}$ of the m-the observed failure of unit number $i_{r_1+\cdots+r_{m-1}+m}$, the surviving items $i_{r_1+\cdots+r_{m-1}+m+1}, \dots, i_n$ are all removed from the experiment, and we denote $I_m = [i_{r_1+\cdots+r_{m-1}+m}, \dots, i_n]$. Note that this censoring structure leads to a subsample $X_{(1)} < \cdots < X_{(m)}$ of the order statistics $T_{(1)} < \cdots < T_{(n)}$ obtained from $T_1, \dots, T_n$. Note also that the sets of unit numbers $I_1, \dots, I_m$ satisfy

$$\cup_{k=1}^{m} I_k = [1, \dots, n] \quad and \quad I_k \cap I_l = \emptyset \quad for \quad 1 \le k < l \le m.$$

For easiness, we present the notations $\alpha_1 = 1$ and $\alpha_k = \sum_{j=1}^{k-1} r_j + k$ for $2 \le k \le m+1$; then, we can write for $1 \le k \le m$

$$I_k = [i_{\alpha_k}, \dots, i_{\alpha_{k+1}-1}].$$

Note that the complete and type right-censored samples are single of the above scheme when $r_1 = r_2 = \cdots = r_m = 0$ and $r_1 = r_2 = \cdots = r_{m-1} = 0, r_m = n - m$, respectively. It is important to note that the order statistics that result from a progressive type II censoring scheme are a specific example of generalized order statistics, which were introduced by Kamps (13,14). For full instructions and more details on the review, the reader is referred to the book by Balakrishnan and Aggarwala (15-17) or other studies (18,19).

It is worth noting that within this structure, the variables $R_1, R_2, \dots, R_m$ are predetermined. However, practical scenarios may arise where the count of items or individuals being dropped or excluded from a study becomes a random variable. Yuen and Tse mention, for instance, that the number of patients excluded at different stages of clinical trials is subject to random selection and cannot be anticipated beforehand. Similarly, in certain reliability tests, testers might opt not to examine specific units, deeming it unnecessary or risky even if those units could potentially fail. Consequently, the patterns of exclusions become stochastic, as all instances of exclusion are subject to randomness.

Let's assume that each test unit failing the lifetime assessment is independent of other units but shares the same probability of being excluded, denoted as p. Following this, Xie et al indicate that the count of instances being excluded in each fault interval adheres to a binomial distribution (20).

This paper aims to analyze competing risk models in the presence of progressively type II censored data with binomial random removals arising from the proportional hazards model (11,21). The Cox model is a commonly used statistical technique to study the relationship between a patient's survival and various risk factors. The purpose of the model is to simultaneously delve into the effects of several variables on survival (12).

Methods

The Model's Assumptions

The paper assumes that there are q independent causes of failure directed toward each unit (22-24). The model studied in the paper contains the following assumptions:

1. In this study, we put n independent and identical units on the life test. The test is terminated when $m \le n$, m is pre-specified, and units fail.

2. The lifetime of the i-th unit is denoted by $X_i, i = 1,2, \dots, n$, and X_{ij} denotes the time of failure of the i-th unit by the cause j where $j = 1,2, \dots, q$, so $X_i = min[X_{i1}, X_{i2}, \dots, X_{iq}]$.

3.

- $F(.)$: Cumulative Distribution Function of X_i,
- $F_j(.)$: Cumulative Distribution Function of X_{ij},
- $\bar{F}_j(.)$: Survival Function of X_{ij}, $\bar{F}_j(.) = 1 - F_j(.)$,
- δ_i: Indicator Variable Representing the Reason for Failure of the i-th Unit.

4. The lifetime of individuals is considered as the Cox proportional hazard model with K explanatory variables, that for the i-th individual, the sub-hazard functions for the $j = 1,2, \cdots, q$; causes are specified as

$$h(t, \mathbf{z}) = h_0(t)\exp(\sum_{i=1}^{k} \beta_{ij} z_i) \quad [1]$$

where the $h_0(t)$ forms a set of baseline sub-hazards, $\underline{z_i}$ denotes the vector of explanatory variables for the i-th individual, and $\underline{\beta_j}$ is the associated vector of regression coefficients corresponding to cause j. Thus, the parameter vector (i.e., the vector of coefficients in the full model) is

$$\underline{\beta} = (\beta_{11}, \beta_{12}, \dots, \beta_{1k}, \beta_{21}, \beta_{22}, \dots, \beta_{2k}, \dots, \beta_{q1}, \beta_{q2}, \dots, \beta_{qk})'$$

of length $k \times q$, where k represents the count of explanatory variables. Given that the same explanatory variable might exert distinct effects on various risks, it is reasonable to postulate that the $\underline{\beta_j}$, $j = 1,2, \dots, q$, vectors are independent of each other. In addition, the survival function of *the* i-th individual is as follows:

$$\bar{F}_j(x_i) = [S_0(j, x_i)]^{\exp(\underline{\beta_j' z_i})} \quad [2]$$

where

$$S_0(j, x_i) = \exp\left(-\int_0^{x_i} h_0(j, u)du\right)$$

5. When the i-th failure occurs, $i = 1,2, \dots, m-1$, [1] we observe 2 values $X_{(i)}$ and $\delta_i \in [1,2, \dots, q]$, where $X_{(i)}$ denotes the i-th order statistics out of the m failed items, which in turn denotes the statistics from the whole sample and [2] R_i of surviving units is randomly selected and removed, where R_i follows binomial distribution with parameters $n - m - \sum_{l=1}^{i-1} R_l$ and p.

Finally, this experiment terminates when the m-th failure occurs, and [1] we observe 2 values $X_{(m)}$ and $\delta_m \in [1,2, \dots, q]$ and [2] the rest $R_m = n - m - \sum_{i=1}^{m-1} R_i$ surviving units are all removed from the test. Here, $\delta_i = j$, $j = 1,2, \dots, q$, means the unit i has failed at the time $X_{(i)}$ due to cause j. The parameter p for binomial distribution is assumed to be the same for all removals.

Given the assumptions outlined above, the available data are a progressively type II censored sample that includes the following:

$$(X_{(1)}, \delta_1, R_1), (X_{(2)}, \delta_2, R_2), \dots, (X_{(m)}, \delta_m, R_m)$$

where $X_{(1)} < X_{(2)} < \cdots < X_{(m)}$ denote the m observed failure times, $\delta_1, \delta_2, \dots, \delta_m$ denote the causes of failures, and $R_1, R_2, \dots, R_m$ denote the number of units removed from the test at the failure time $x_{(1)} < x_{(2)} < \cdots < x_{(m)}$.

To simplify the notation, we will use henceforth X_i instead of $x_{(xi)}$ (22, 23).

The Likelihood Function of the Model and Estimators

The Likelihood Function

Using the mentioned assumptions in section 2 and conditional on $I_1, I_2, \dots, I_m$, the likelihood function for the type II progressively censored model under competing risk is as follows (10):

$$L(\underline{\beta}; x, \delta, R) = L_1(\underline{\beta}; x, \delta | R = r)P(R, p) \quad [3]$$

where

$$L_1(\underline{\beta}; x, \delta | R = r) = c \prod_{k=1}^{m} h(\delta_{i_{\alpha_k}}, x_k, \underline{z_{i_{\alpha_k}}}) \left[\prod_{j=1}^{q} [S_0(j, x_k)]^{\sum_{l \in I_k} \exp(\underline{\beta_j' z_l})}\right] \quad [4]$$

where $c = n(n - r_1 - 1) \cdots (n - r_1 - r_2 - \cdots - r_{m-1} - m + 1)$,

$h(\delta_{i_{\alpha_k}}, x_k, \underline{z_{i_{\alpha_k}}}) = h_0(\delta_{i_{\alpha_k}}, x_k)\exp(\underline{\beta'_{\delta_{i_{\alpha_k}}}}\ \underline{z_{i_{\alpha_k}}})$.

Assuming that a test unit extracted from the life test is independent of the others but with a consistent probability P, the count of units removed at each failure time adheres to a binomial distribution. This distribution can be formulated as:

$$P(R_1 = r_1) = n - m_{r_1}\, p^{r_1}(1 - p)^{n-m-r_1}$$

where $0 \le r_1 \le n - m$, and

$$P(R_i = r_i | R_{i-1} = r_{i-1}, \dots, R_1 = r_1) = n - m - \sum_{l=1}^{i-1} r_l\, {}_{r_i}\, p^{r_i}(1 - p)^{n-m-\sum_{l=1}^{i} r_l}$$

where $0 \le r_i \le n - m - \sum_{l=1}^{i-1} r_l, i = 2, \dots, m-1$. The remaining items, if there are some, are all removed from the test at the m-th failure with probability 1. Suppose further that R_i is independent of X_i for all i; thus, we can write:

$$P(R, p) = P(R_m = r_m | R_{m-1} = r_{m-1}, \dots R_1 = r_1) \cdots P(R_2 = r_2 | R_1 = r_1)P(R_1 = r_1)$$

Therefore,

$$P(R, p) = \frac{(n-m)!}{\prod_{i=1}^{m-1} r_i!(n-m-\sum_{i=1}^{m-1} r_i)!}\ p^{\sum_{i=1}^{m-1} r_i}\ (1 - p)^{(m-1)(n-m)-\sum_{i=1}^{m-1}(m-i)r_i} \quad [5]$$

substituting [4] and [5] into [3], the likelihood function takes the following form:

$$L(\underline{\beta}; x, \delta, R) = c^* \prod_{k=1}^{m} h(\delta_{i_{\alpha_k}}, x_k, \underline{z_{i_{\alpha_k}}}) \left[\prod_{j=1}^{q} [S_0(j, x_k)]^{\sum_{l \in I_k} \exp(\underline{\beta_j' z_l})}\right]$$
$$\times\, p^{\sum_{i=1}^{m-1} r_i}\ (1 - p)^{(m-1)(n-m)-\sum_{i=1}^{m-1}(m-i)r_i} \quad [6]$$

where (10)

$$c^* = \frac{(n - m)!\, c}{\prod_{i=1}^{m-1} r_i!\, (n - m - \sum_{i=1}^{m-1} r_i)!}$$

Estimators of Unknown Parameters

Note that $P(R, p)$ does not depend on the parameters $\underline{\beta}$, and hence, independently, the maximum likelihood estimators (MLEs) of parameter p can be obtained by maximizing [5]. Thus, we find immediately

$$\hat{p} = \frac{\sum_{i=1}^{m-1} R_i}{(m-1)(n-m)-\sum_{i=1}^{m-1}(m-i)R_i+\sum_{i=1}^{m-1} R_i}. \quad [7]$$

In the previous section of the introduction, we mentioned that the progressive type II censoring scheme results in a stochastic partition in $I_1, \dots, I_m$ of $[1, \dots, n]$, such that $Card(I_i) = r_i + 1$ for $1 \le i \le m$. Once the experiment is completed, we can observe the actual partition $I1$, ..., Im. This observed partition reflects how the data points in the interval [1, ..., n] were divided based on the censoring scheme. Then, we can calculate the probability of observing the partition $I_1, \dots, I_m$ as

$$p(I_1 = I_1, \dots, I_m = I_m)$$
$$= \int_0^{+\infty} h(\delta_{i_{\alpha_1}}, x_1, \underline{z_{i_{\alpha_1}}})W_1 \int_{x_1}^{+\infty} h(\delta_{i_{\alpha_2}}, x_2, \underline{z_{i_{\alpha_2}}})W_2 \cdots$$
$$\cdots \int_{x_{m-1}}^{+\infty} h(\delta_{i_{\alpha_m}}, x_m, \underline{z_{i_{\alpha_m}}})W_m\ dx_1 dx_2 \cdots dx_m$$
$$= \prod_{k=1}^{m} \frac{\exp(\underline{\beta'_{\delta_{i_{\alpha_k}}}}\ \underline{z_{i_{\alpha_k}}})}{\sum_{j \in I_k^m} \exp(\underline{\beta'_{\delta_{i_{\alpha_k}}}}\ \underline{z_j})} = C(\underline{\beta}),$$

here $I_k^m = I_k \cup I_{k+1} \cup \cdots \cup I_m$ for $1 \le k \le m$, and

$$W_k = \prod_{j=1}^{q} [S_0(j, x_k)]^{\sum_{l \in I_k} \exp(\underline{\beta_j' z_l})}.$$

Therefore, it is logical to estimate the unknown regression parameter $\underline{\beta}$ by the value of $\underline{\hat{\beta}}$ that makes the above probability of observing the partition $I_1, \ldots, I_m$ maximal since this probability depends no longer on the unknown functional parameters $h_0(j,.)$, $j = 1,2,\ldots,q$. Large-sample inference can be conducted by treating the logarithm of this probability in the usual way. Thus, we have

$$\log C(\underline{\beta}) = \sum_{k=1}^{m} \left\{ \underline{\beta}'_{\delta_{i_{\alpha_k}}} \underline{z}_{i_{\alpha_k}} - \log d_k \right\}$$

where

$$d_k = \sum_{j \in I_k^m} \exp(\underline{\beta}'_{\delta_{i_{\alpha_k}}} \underline{z}_j)$$

with first and second derivatives, respectively,

$$U(\underline{\beta}) = \frac{\partial \log C(\underline{\beta})}{\partial \underline{\beta}} = \sum_{k=1}^{m} U_k(\underline{\beta}),$$

$$V(\underline{\beta}) = -\frac{\partial^2 \log C(\underline{\beta})}{\partial^2 \underline{\beta}} = \sum_{k=1}^{m} V_k(\underline{\beta}),$$

where the vector $U_k(\underline{\beta})$ has l-th component, $l = 1,2,\ldots,q$,

$$\frac{\partial \log(\exp(\underline{\beta}'_{\delta_{i_{\alpha_k}}} \underline{z}_{i_{\alpha_k}})/d_k)}{\partial \underline{\beta}_l} = \left\{ \underline{z}_{i_{\alpha_k}} - \underline{v}_k \right\}_0 \; 1cm\delta_{i_{\alpha_k}} = l_{o.w}$$

and the matrix $V_k(\underline{\beta})$ has ll-th entry, $l = 1,2,\ldots,q$,

$$-\frac{\partial^2 \log(\exp(\underline{\beta}'_{\delta_{i_{\alpha_k}}} \underline{z}_{i_{\alpha_k}})/d_k)}{\partial \underline{\beta}_l \partial \underline{\beta}'_l} = \left\{ Z_k - \underline{v}_k \; \underline{v}_k{}^T \right\}_0 \; 1cm\delta_{i_{\alpha_k}} = l_{o.w}$$

where

$$\underline{v}_k = d_k^{-1} \sum_{j \in I_k^m} \underline{z}_j \exp(\underline{\beta}'_{\delta_{i_{\alpha_k}}} \underline{z}_j) \quad and \quad Z_k = d_k^{-1} \sum_{j \in I_k^m} \underline{z}_j \, \underline{z}_j{}^T \exp(\underline{\beta}'_{\delta_{i_{\alpha_k}}} \underline{z}_j)$$

Under some standard regularity conditions, $V^{-1}(\underline{\hat{\beta}})$ provides an estimate for the variance-covariance matrix of $\underline{\hat{\beta}}$ and so follows hypothesis tests and confidence intervals for $\underline{\beta}$.

Finally, the Breslow-type estimator (24) of the baseline type-specific cumulative hazard function for the j-th failure type can be written as

$$\hat{H}_0(j,t) = \sum_{1 \le i \le m; X_{(i)} \le t} \frac{1}{\sum_{l \in I_l^m} \exp(\underline{\hat{\beta}}'_j \underline{z}_l)}, \quad t \ge 0 \quad for \quad j = 1,2,\ldots,q, \; [8]$$

where (25)

$$H_0(j,t) = \int_0^t h_0(j,u)du.$$

A Simulation Study

In this subsection, we provide the outcomes of a simulation study conducted over 500 replications for progressively type II censored, type II censored, and uncensored samples, each having a size of n = 100. We consider a scenario where there exist 2 independent causes of failure attributed to each unit. The covariates are treated as random variables, following a Bernoulli distribution with a parameter of 1/2. The conditional distribution of a random lifetime t, given a covariate Z, is characterized by the sub-hazard rate functions $(j,t;z) = (t/3)^2 \exp(\beta_j z)$ for $t \ge 0$, where the regression parameters $(\beta_1, \beta_2) = [(0.2,0.5), (0.4,1), (0.6,1.5)]$ have to be estimated. For each set of simulated samples (Table 1), we calculate the estimators of β_1, β_2, and the standard deviation of estimators (within parentheses). We consider the 3 following sampling schemes:

(A) Complete data.

(B) Type II censoring sampling plan: $m = 65$.

(C) Progressive type II censoring sampling plan: $m = 65$, $p = 0.2$ and $R = (9,8,5,1,4,2,2,0,0,1,0,0,1,2,0, \ldots, 0)$.

Results

Employed in a Real-World Dataset

Consider a scenario where demographic, personal, clinical, and laboratory data are gathered through interviews and physical examinations conducted on a sample of 200 individuals participating in a study on cardiovascular disease (CVD). These participants, between the ages of 50 and 79, who had no CVD at baseline, will be followed for 10 years. To further demonstrate the use of the proportional hazards model in competing risks, a subset of 68 participants from the simulated data is used. The event of interest, denoted as " t," represents the period of time during which participants remain free from CVD. This time duration is defined in terms of years, starting from the baseline and ending at the earliest occurrence of either a participant being diagnosed with CVD or being confirmed to have passed away due to CVD-related causes. Cardiovascular disease, in this case, includes coronary heart disease (CHD) and stroke. Covariates of interest include age (AGE), sex (SEX = 1 for males and SEX = 0 for females), smoking status (SMOKE = 1 for current smokers and SMOKE = 0 otherwise), and body mass index ([BMI = weight] . in kilograms divided by height in meters squared), systolic blood pressure (SBP), logarithm of the ratio of urine albumin to creatinine (LACR), logarithm of triglycerides (LTG), hypertension status (HTN = 1 if SBP 140 mm Hg or DBP 90 mm Hg or on treatment hypertension and otherwise HTN = 0) and diabetes status (DM = 1 at fasting glucose of 126 mg/dL or on treatment for diabetes and otherwise DM = 0).

To evaluate the CVD outcome of interest, we use DG to denote the type of CVD. Specifically, DG = 0 if the CVD-free time is censored, DG = 1 if the participant had a stroke, DG = 2 if the participant had CHD, and DG = 3 if the participant had other types of CVD.

Results for Stroke

For stroke, SBP and hypertension status are the only significant variables. In contrast, gender, BMI, smoking status, body, the logarithm of urinary albumin and creatinine ratio, and diabetes status are significant variables for CHD and other CVDs. The outcomes suggest that the significant risk factors vary for different types of CVD events.

Simulation Results

Generally, in the simulated data, we can observe that the results for type II progressive right-censored sampling designs produce better results than results based on the usual type II censored sampling designs (Table 1).

Table 1. Simulation's results

Scheme	$\beta_1, \beta_2 = 0.2,0.5$	$\beta_1, \beta_2 = 0.4,1$	$\beta_1, \beta_2 = 0.6,1.5$
(A)	0.247(0.264), 0.375(0.330)	0.401(0.284), 0.741(0.358)	0.621(0.314), 1.027(0.289)
(B)	0.138(0.321), 0.292(0.394)	0.276(0.353), 0.279(0.398)	0.695(0.348), 1.043(0.371)
(C)	0.177(0.382), 0.301(0.344)	0.347(0.361), 0.475(0.308)	0.667(0.389), 1.032(0.356)

To fit the competing risk model, the data are generated from a progressively type II censored sample with m = 32, p = 0.2, and a specific censoring scheme.

$R = (7,6,3,5,2,2,2,0,2,1,2,2,0,0,0,1,0,0,0,0,1,0, \ldots, 0)$.

Table 2 presents the outcome of applying the backward selection technique for fitting the proportional hazards model.

Table 2. Results on cardiovascular disease event time data from the fitted competing risks model

Variable	Parameter Estimate	Standard Error	Hazards Ratio	95% Confidence Interval of Hazard Ratio
β_{1SBP}	0.05978	0.02020	1.062	[1.020, 1.104]
β_{1HTN}	-2.52146	1.18298	0.080	[0.008, 0.816]
β_{2SEX}	1.50344	0.59055	4.497	[1.413, 14.309]
β_{2SMOKE}	1.27235	0.50990	3.569	[1.314, 9.696]
β_{2BMI}	0.08072	0.03750	1.084	[1.007, 1.167]
β_{2LACR}	0.32808	0.11426	6.360	[1.916, 21.116]
β_{2DM}	1.85004	0.61226	6.360	[1.916, 21.116]

Discussion

This article provides a comprehensive exploration of both theoretical concepts and practical applications involved in survival modeling when dealing with competing risks within the framework of the Cox proportional hazards model, particularly focusing on progressive censorship with binomial distances. The results of a simulation study highlight the superiority of type II progressive right-censored sampling plans over conventional type II censored sampling plans. The analysis revealed shifting significant risk factors across various CVD events.

Competing risk analysis in the CoxPH model for progressive censorship with binomial removal has been studied by Chacko and Mohan (26). They consider the analysis of competing risk data under progressive type II censoring, assuming the number of units removed at each stage follows a binomial distribution. They obtain Bayes estimators assuming a Weibull distribution for the population under consideration. On the other hand, Lodhi et al (27) discuss a competing risks model using Gompertz distribution under progressive type II censoring. Singh et al (28) discuss the inference for the competing risks model when the failure times follow Chen distribution, with partially observed causes of failures considered as independent. They obtain maximum likelihood estimates for model parameters under generalized progressive hybrid censoring. Also, a study examined the influence of unaccounted causes of failure on inference in the Bayesian "index of local sensitivity to non-ignorability" within the proposed competing risks model (28). They concluded that the missing data mechanism should be given special consideration when using the suggested model in cases where the causes of failure are potentially missing.

In some studies, the model parameters have been thoroughly examined and compared in terms of classical and Bayesian inferences, using extensive simulation studies to evaluate their respective performances (29). They showed that Bayes estimators perform better than the MLE in terms of Mean Squared Error (MSE), and non-informative prior performs better than estimators based on informative priors.

One aspect of the current study involved considering an independent failure causes model with exponential lifetimes. The removal of items or subjects at each failure point followed a binomial distribution. Maximum likelihood estimators were derived for the unknown parameters within this model, accompanied by the calculation of their asymptotic distributions. It was observed that varying binomial distances with different probabilities lead to estimations characterized by varying levels of precision. Future simulation studies are recommended to explore the impact of the parameter "p" on the accuracy of MLE estimates. Additionally, other studies have also explored Bayesian estimators under similar conditions, assuming a Weibull distribution for the underlying population (27).

Also, a simulation study was conducted to evaluate the performance of the different estimators derived in these contexts (23,29).

The present study introduces a hierarchical Bayes methodology and devises a Metropolis-Hastings sampling algorithm to facilitate intricate posterior computation. Furthermore, the efficacy of the proposed techniques is demonstrated through comprehensive simulation experiments and practical data analysis.

Conclusion

To sum up, the estimated parameters on the defined scheme setting are recommended. They can be used in many practical situations when competing risks occur, and progressive censoring could be considered.

Acknowledgement

We would like to thank the Department of Biostatistics and Epidemiology, the School of Medicine, Zanjan University of Medical Sciences, for providing the environment for modeling data and manuscript writing.

Funding sources

This study did not have any funds.

Ethical statement

The Ethics Committee of Zanjan University of Medical Sciences approved the protocol of this study (Approval No. IR.ZUMS.REC.1399.221).

Conflicts of interest

The authors declared no conflicts of interest.

Author contributions

All authors studied and confirmed the manuscript. Conceptualization: AP, RH, JA, SN. Data curation: AP. Formal analysis: JA, RH. Funding acquisition: None. Methodology: AP, RH, SN. Writing original draft: AP, RH, JA, SN. Writing, review, and editing: AP, RH, JA, SN.

References

1. Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. Int J Epidemiol. 2012;41(3):861-70. [View at Publisher] [DOI] [PMID] [Google Scholar]
2. Pintilie M. Competing risks: a practical perspective. England: John Wiley & Sons; 2006. [View at Publisher] [DOI] [Google Scholar]
3. Pintilie M. An introduction to competing risks analysis. Rev Esp de Cardiol. 2011;64(7):599-605. https://doi.org/10.1016/j.rec.2011.03.016 [View at Publisher] [DOI] [Google Scholar]
4. Kleinbaum DG, Klein M. Survival Analysis: A Self-Learning Text. New York: Spinger; 2012. [View at Publisher] [Google Scholar]
5. Crowder MJ. Classical competing risks. New York: CRC Press; 2001. [View at Publisher] [DOI] [Google Scholar]
6. Dutz A, Löck S. Competing risks in survival data analysis. Radiother Oncol. 2019;130:185-9. [View at Publisher] [DOI] [PMID] [Google Scholar]
7. Basu AP, Ghosh JK. Competing risks theory and identifiability problems. The Exponential Distribution: Routledge; 2019. p.489-95. [View at Publisher] [DOI] [PMID] [Google Scholar]
8. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American statistical association. 1958;53(282):457-81. [View at Publisher] [DOI] [Google Scholar]
9. Fox J. Cox proportional-hazards regression for survival data. An R and S-PLUS companion to applied regression. London: SAGE; 2002. [View at Publisher] [Google Scholar]
10. Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. New York: Springer; 2003. [View at Publisher] [DOI] [Google Scholar]
11. Klein JP, Van Houwelingen HC, Ibrahim JG, Scheike TH. Handbook of survival analysis. New York: CRC Press; 2016. [View at Publisher] [DOI] [Google Scholar]
12. Kleinbaum DG, Klein M. Survival analysis: A Self-Learning Text. New York: Springer; 2010. [View at Publisher] [DOI] [Google Scholar]
13. Kamps U. A concept of generalized order statistics. Journal of Statistical Planning and Inference. 1995;48(1):1-23. [View at Publisher] [DOI] [Google Scholar]
14. Razmkhah M, Simriz S. Statistical inferences based on INID progressively type II censored order statistics. Annals of the Institute of Statistical Mathematics. 2018;70(3):583-604. [View at Publisher] [DOI] [Google Scholar]
15. Balakrishnan N, Aggarwala R. Progressive censoring: theory, methods, and applications. US: Springer Science & Business Media; 2000. [View at Publisher] [DOI] [Google Scholar]
16. Balakrishnan N, Cramer E. The art of progressive censoring: Applications to Reliability and Quality. New York: Birkhäuser; 2014. [View at Publisher] [DOI] [Google Scholar]
17. Balakrishnan N. Progressive censoring methodology: an appraisal. Test. 2007;16(2):211-59. [View at Publisher] [DOI] [Google Scholar]
18. Salemi U, Rezaei S, Si Y, Nadarajah S. On optimal progressive censoring schemes for normal distribution. Annals of Data Science. 2018;5(4):637-58. [View at Publisher] [DOI] [Google Scholar]
19. Burkschat M, Cramer E, Kamps U. On optimal schemes in progressive censoring. Statistics & probability letters. 2006;76(10):1032-6. [View at Publisher] [DOI] [Google Scholar]
20. Tse SK, Yang C, Yuen H-K. Statistical analysis of Weibull distributed lifetime data under Type II progressive censoring with binomial removals. Journal of Applied Statistics. 2010;27(8):1033-43. [View at Publisher] [DOI] [Google Scholar]
21. Miller Jr RG. Survival analysis. 2nd ed. John Wiley & Sons; 2011. [View at Publisher] [Google Scholar]
22. Pareek B, Kundu D, Kumar S. On progressively censored competing risks data for Weibull distributions. Computational Statistics & Data Analysis. 2009;53(12):4083-94. [View at Publisher] [DOI] [Google Scholar]
23. Sarhan AM, Alameri M, Al-Wasel I. Analysis of progressive censoring competing risks data with binomial removals. Int J Math Analysis. 2008;2(20):965-76. [View at Publisher] [Google Scholar]
24. Lin DY. On the Breslow estimator. Lifetime data analysis. 2007;13:471-80. [View at Publisher] [DOI] [PMID] [Google Scholar]
25. Hashemi R, Amiri L. Analysis of progressive Type-II censoring in the Weibull model for competing risks data with binomial removals. Appl Math Sci. 2011;5(21-24):1073-87. [View at Publisher] [Google Scholar]
26. Lodhi C, Tripathi YM, Bhattacharya R. On a progressively censored competing risks data from Gompertz distribution. Communications in Statistics-Simulation and Computation. 2023;52(4):1278-99. [View at Publisher] [DOI] [Google Scholar]
27. Wang L, Wu SJ, Lin H, Tripathi YM. Inference for block progressive censored competing risks data from an inverted exponentiated exponential model. Quality and Reliability Engineering International. 2023;39(7):2736-64. [View at Publisher] [DOI] [Google Scholar]
28. Azizi F, Haghighi F, Tabibi Gilani N. Statistical inference for competing risks model under progressive interval censored Weibull data. Communications in Statistics-Simulation and Computation. 2020;49(7):1931-44. [View at Publisher] [DOI] [Google Scholar]
29. Chacko M, Mohan R. Bayesian analysis of Weibull distribution based on progressive type-II censored competing risks data with binomial removals. Computational Statistics. 2019;34:233-52. [View at Publisher] [DOI] [Google Scholar]

How to Cite:
Pourdarvish A, Hashemi R, Azar J, Norouzi S. Analysis of competing risks in the CoxPH model for progressive censorship with binomial removal. *Jorjani Biomedicine Journal.* 2023;11(3):14-7.