Estimation of censored medical cost with dependent censoring using copula method

Choa Yun

The Graduate School
Yonsei University
Department of Biostatistics and Computing

Estimation of censored medical cost with dependent censoring using copula method

A Dissertation Submitted to the
Department of Biostatistics and Computing
and the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Biostatistics and Computing

Choa Yun

June 2023

This certifies that the dissertation of ***Choa Yun*** is approved.

_____

Chung Mo Nam: Thesis Supervisor

_____

Inkyung Jung: Thesis Committee Member #1

_____

Sohee Park: Thesis Committee Member #2

_____

Min Jin Ha: Thesis Committee Member #3

_____

Youn Nam Kim: Thesis Committee Member #4

The Graduate School

Yonsei University

June 2023

Contents

List of Tables

List of Figures

Abstract

Medical cost modeling presents many challenges because available data are often right-censored due to early termination or follow-up loss of data observations. The prevailing methods for estimating the average of cumulative costs fall into three categories, (a) the inverse probability weighted (IPW) estimators; (b) the generalized survival-adjusted estimators; (c) the joint-modeling methods. However, under violation of independent censoring assumption, traditional survival analysis methods have been shown to be biased when employed on medical data.

However, the prevailing methods were established under independent censoring. Generally, the medical costs for failure event and censoring time tend to be positively correlated (Etzioni et al. 1999; Lin 2003). Therefore, it is necessary to consider the joint distribution of survival and censoring time under dependent censoring. Copula function is attractive in statistical modeling because it gives a flexible and promising tool to modeling with dependence between survival and censoring time. For analysis of survival under dependent censoring, we use the copula-graphic estimator and copula-based univariate Cox regression employed an assumed copula.

In this study, we evaluate the prevailing methods for estimating the medical cost under independent censoring assumption. Also, we proposed new estimators that are expanding scheme on IPW and generalized survival-adjusted estimators under dependent censoring assumption. The purpose of this study is to adapt the copula method to estimation medical

cost with survival data including dependent censoring. We evaluate our proposed estimators with a series of simulations. We conduct a simulation study assuming various scenarios to appraise the performance of bias and S.E in estimation of medical costs for dependent censoring. In addition, we illustrated the prevailing and proposed method using National Health Insurance Service National Sample Cohort (NHIS-NSC) data.

Keywords: medical cost, dependent censoring, copula models, copula-graphic estimator, copula-based Cox regression, IPW method, generalized-adjusted survival estimation, joint-model

1. Introduction

In recent years, it has become common for hospitals, health insurers, or disease registries to collect data of medical cost to determine risk or economic burden and perform cost-effectiveness analyzes. However, medical cost modeling presents many challenges because available data are often right-censored due to early termination or follow-up loss of data observations. Considering medical cost as right-censored survival data, it is natural for researchers to use survival analysis technique for analysis with censored cost data such as the Kaplan Meier estimator, log-rank test, and Cox regression. However, under violation of independent censoring assumption, i.e., the medical cost at the censoring time points correlates with the cost of failure events, traditional survival analysis methods have been shown to be biased when employed on medical data. In order to solve this problem, much attention has been paid to the problem of estimating the average of cumulative costs. The prevailing methods fall into three categories, (a) the inverse probability weighted (IPW) estimators by Bang and Tsiatis (2000) and Lin (2003); (b) the generalized survival-adjusted estimators by Lin et al. (1997) and Basu and Manning; (c) the joint-modeling methods by Heitjan et al. (2004) and Liu (2009), Liu et al. (2007, 2008).

However, the prevailing methods were established under independent censoring. Generally, the medical costs for failure event and censoring time tend to be positively correlated (Etzioni et al. 1999; Lin 2003). Using the most standard survival analysis methods for calculation cumulative medical cost provide biased results under the

assumption of dependent censoring. Therefore, it is necessary to consider the joint distribution of survival and censoring time under dependent censoring. Copula function is attractive in statistical modeling because it gives a flexible and promising tool to modeling with dependence between survival and censoring time. For analysis of survival under dependent censoring, we use the copula-graphic estimator instead of Kaplan-Meier estimator. This method employs an assumed copula. To avoid the non-identifiability, copula function and its dependency parameter should be completely specified. Emura and Chend (2016) showed the biased estimation of Cox regression due to violation of independent censoring assumption. So, we use the copula-based univariate Cox regression proposed by Emura and Chen (2016) for correctly capturing the effect of covariate under dependent censoring.

In this paper, we evaluate the prevailing methods for estimating the medical cost under independent censoring assumption. Also, we proposed new estimators that are expanding scheme on IPW and generalized survival-adjusted estimators under dependent censoring assumption. We evaluate our proposed estimators with a series of simulations. The purpose of this study is to adapt the copula method to estimation of medical cost with survival data including dependent censoring. In Section 2, we briefly review the prevailing methods for estimating the medical cost under independent censoring assumption. In section 3, we illustrate the dependent censoring's issues arising from medical research. The Secsion 4 provides copula models' mathematical infrastructures for applications to survival analysis under dependent censoring. In Section 5, we propose a class of estimators under dependent

censoring using copula model which build on the censored medical cost estimators defined in section 2.2 and 2.3. In Section 6, we evaluate the performance of the proposed estimators adopted for censored medical data via a simulation study. Section 7 illustrated the application of our proposed estimators with the prevailing estimators to the analysis of real data example. We discuss our results and provide some conclusions in Section 8.

2. Estimating medical cost with censored data

Estimating medical costs have a common problem due to incompleteness of follow-up data. Naïve summary statistics including simple average on the collected data can mislead to statistical inference and introduce bias. So, this section focuses on the problem of estimating medical costs if the cost data are right-censored. The prevailing methods fall into three categories, (a) the inverse probability weighted (IPW) estimators; (b) the generalized survival-adjusted estimators; (c) the joint-modeling methods. In this section, we review the methods for estimating censored medical cost under independent censoring assumption.

2.1 Notation and assumptions

First define a general setting and notation for our problem. Let the random variable M represent the total medical cost over some specified period of time and denote it as a random variable T represented the survival time. If necessary, the time frame for evaluating each patient should be limited to τ. Therefore, we should consider the costs M paid by a patient up to a maximum of τ units of time. In that case, the variable T is bounded by τ. The distribution of T is assumed to be continuous from 0 to τ.

We can obtain the cost M for each patient and estimate the mean cost by computing simple average cost denoted by $\mu = E(M)$ when all patients have been observed for at

least τ units of time. However, in most studies, not all patient costs are completely observed due to different types of censoring. For instance, censoring may occur because patients enter the study at a time lag, in which case patients exiting the study who were not followed up for T units of time would be censored. This type of censoring is called administrative censoring. Also, censoring may occur when patients are lost to follow-up or leave the study. Let C be the censoring time. In this section, the censoring is assumed to continuous distribution and arise in a completely random.

$M(u)$ denotes the cumulative cost up to time u realizing that information about costs may be available at random points in time. Write $\mathrm{X} = \min(\mathrm{T}, \mathrm{C}), \Delta = \mathrm{I}(C \geq T)$, $S(t) = \Pr(T > t), K(t) = \Pr(C > t)$ where $I(\cdot)$ is the indicator function. Let $M_T = M(T)$ denotes the lifetime cost which cannot be observed for all patients due to the limitation of study duration. Thus, we focus on the total cost accumulated in interesting time period (0, τ]. In addition, we need to divide (0, τ] into K intervals $(t_{k-1}, t_k], (k = 1,2, \dots, K)$. Write T^*=min(T, τ) and $\Delta^* = I(C \geq T^*)$ where $I(\cdot)$ is the indicator function. Let $M_k = M(t_k)\ and\ m_k = M_k - M_{k-1}$. The subscript i indicates individual patients $i = 1,2, \dots, n$ and $M_i = M(T_i^*)$ denote the total cost for each patient i. In regression, $\boldsymbol{Z}$ is the set of $p \times 1$ covariates vector with parameter vector $\boldsymbol{\beta}$.

2.2 Inverse probability weighted estimator

The idea of weighing the complete observations by their inversed probabilities initiated by Horvitz and Thompson (1952) in sample survey. Bang and Tsiatis (2000) employed IPW method to account for informative censoring when estimating the mean of total medical cost. As regards independent censoring, the ith individual has a probability $K(T_i)$ of not being censored at the time of death T_i. Thus, each person observed to die uncensored represents, on average, $1/K(T_i)$ individuals who may have been censored. The simple weighted complete-case estimator (BT) of mean of total cost $E(M_i)$ is

$$\hat{\mu}_{BT} = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i^* M_i}{\widehat{K}(T_i^*)}.$$

They proposed to estimate the unknown survivor function $K(\cdot)$ by the Kaplan-Meier estimator (Kaplan and Meier, 1958). That estimator is

$$\widehat{K}(t) = \prod_{u \le t} \left\{1 - \frac{dN^c(u)}{Y(u)}\right\},$$

where $N^c(u) = \sum I(T_i^* \le u, \Delta_i = 0)$ and $Y(u) = \sum I(T_i^* \ge u)$.

This is an unbiased estimate of μ, which is a result of the following equality:

$$\mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i^* M_i}{K(T_i^*)}\right\} = E\left[\frac{1}{n}\sum_{i=1}^{n} E\left\{\frac{\Delta_i^* M_i}{K(T_i^*)}\middle| T_i^*, M_i(\cdot)\right\}\right]$$

$$= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{M_i}{K(T_i^*)} E\{I(C_i \ge T_i)|T_i^*, M_i(\cdot)\}\right]$$

$$= \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} M_i\right) = \mu.$$

The variance of $\hat{\mu}_{BT}$ is given by

$$\mathrm{Var}(\hat{\mu}_{BT}) = \frac{1}{n}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i^*(M_i-\hat{\mu}_{BT})^2}{\hat{K}(T_i^*)} + \frac{1}{n}\int_0^{\tau}\frac{dN^c(u)}{\hat{K}^2(u)}\{\hat{G}(M^2,u)-\hat{G}^2(M,u)\}\right],$$

where $\hat{G}(M,u) = \frac{1}{n}\frac{1}{\hat{S}(\mathrm{u})}\sum_{i=1}^{n}\frac{\Delta_i^* M_i I(T_i^*\geq u)}{\hat{K}(T_i^*)}$ and $\hat{S}(\mathrm{u})$ is the Kaplan-Meier estimator for $\mathrm{S(u)} = \mathrm{pr}(T_i^* \geq u)$. Then, given $T_i^{*k} = \min(T_i^*, t_k), \Delta_i^{*k} = I\{T_i^{*k} \leq C_i\}$, K sub intervals $(t_j, t_{j+1}]$ $(j = 1,2,\dots,K-1)$ of (0, τ], the partitioned version (BT_p) of more efficient estimator is

$$\hat{\mu}_{BT_p} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\Delta_i^{*k}\{M_i(t_k)-M_i(t_{k-1})\}}{\hat{K}(T_i^{*k})}$$

This partitioned estimator makes use of censored observation's cost history which are not used by the unpartitioned estimator.

The same scheme is used when building regressions of right censored medical costs on covariates. Lin (2000) modified linear regression form $\mathrm{E}(M_i|\boldsymbol{Z}_i) = \boldsymbol{\beta}'\boldsymbol{Z_i}$ for informative censoring using the IPW method. The resulting estimator for above equation is calculated by the weighted estimation function $\sum_{i=1}^{n}\frac{\Delta_i^*(M_i-\boldsymbol{\beta}'\boldsymbol{Z}_i)\boldsymbol{Z_i}}{\hat{K}(T_i^*)} = 0$, where $\hat{K}(T_i^*)$ is the Kaplan-Meier estimator for $K(T_i^*)$, is given by

$$\widehat{\boldsymbol{\beta}} = \left\{\sum_{i=1}^{n}\frac{\Delta_i^*}{\hat{K}(T_i^*)}\boldsymbol{Z_i}\boldsymbol{Z_i'}\right\}^{-1}\sum_{i=1}^{n}\frac{\Delta_i^*}{\hat{K}(T_i^*)}M_i\boldsymbol{Z_i}.$$

The partitioned estimator including the cost history of censored patients is used under the model of $\mathrm{E}(M_{k,i}|\boldsymbol{Z}_i) = \boldsymbol{\beta_k'}\boldsymbol{Z_i}$, where $\boldsymbol{\beta_k}$ is $p \times 1$ vector of unknown parameter for each interval k. Then the estimator for $\boldsymbol{\beta_k}$ in each partition is

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{k}} = \left\{ \sum_{i=1}^{n} \frac{\Delta_i^{*k}}{\widehat{K}\left(T_i^{*k} \middle| \boldsymbol{Z}_i\right)} \boldsymbol{Z}_{\boldsymbol{i}} \boldsymbol{Z}_{\boldsymbol{i}}' \right\}^{-1} \sum_{i=1}^{n} \frac{\Delta_i^{*k}}{\widehat{K}\left(T_i^{*k} \middle| \boldsymbol{Z}_i\right)} m_{k,i} \boldsymbol{Z}_{\boldsymbol{i}},$$

where $\widehat{K}(\mathrm{t}|\boldsymbol{Z}_i)$ could be derived via a proportional hazards model or use another consistent estimator that allows in that way for censoring dependent on covariates instead of using $\widehat{K}(T_i^*)$.

2.3 Generalized survival-adjusted estimator

In this section, the generalized survival-adjusted estimation method is performed by extending the Lin et al. (1997)'s work in a more direct way. To solve the informative censoring, we should divide time period into K intervals. The original estimator form of Lin et al. (1997) is

$$\hat{\mu}_{Lin97} = \sum_{k=1}^{K} \hat{S}_{k-1} \hat{E}_k$$

where $\hat{E}_k$ is the estimator of $E_k = E(m_k|T^* > t_{k-1})$ and $\hat{S}_k$ is the Kaplan-Meier estimator of $S_k = Pr(T^* \geq t_k)$.

Basu and Manning (2010) extended the estimator proposed by Lin (1997) to take covariates into account when the cost history data is available. First, they decomposed $E_k = h_k \mu_{1k} + (1 - h_k)\mu_{2k}$ which incorporate the different cost accumulation rates over the intervals where the patient dies and survives where $h_k = \Pr(t_{k-1} < T \leq t_k | T \geq t_{k-1})$ is the hazard rate of death during the kth interval, $\mu_{1k} = E(m_k|t_{k-1} < T \leq t_k)$

and $\mu_{2k} = E(m_k|T > t_k)$ are the mean of incremental costs for patients who died during or after the kth interval, respectively. Then, the mean estimator on covariates proposed by Basu and Manning (2010) is

$$\hat{\mu}(\boldsymbol{Z}) = \sum_{k=1}^{K} \hat{S}_k(\boldsymbol{Z}) \left[\hat{h}_k(\boldsymbol{Z})\hat{\mu}_{1k}(\boldsymbol{Z}) + \left(1 - \hat{h}_k(\boldsymbol{Z})\right)\hat{\mu}_{2k}(\boldsymbol{Z})\right],$$

where $\hat{S}_k(\boldsymbol{Z})$ and $\hat{h}_k(\boldsymbol{Z})$ can be derived from accelerated failure time (AFT) model and $\hat{\mu}_{1k}(\boldsymbol{Z})$ and $\hat{\mu}_{2k}(\boldsymbol{Z})$ from certain generalized linear model (GLM).

Under the assumption of random censoring, the process of estimation follows under three parts:

(a) Part-1: Estimate an individual's survival function after accounting for censoring using a flexible survival model, like an accelerated failure time model based on a generalized gamma distribution over time. Let $\hat{S}_k(\boldsymbol{Z})$ and $\hat{h}_k(\boldsymbol{Z})$ be the estimated survival and hazard functions for an interval (Notation for individuals has been suppressed for clarity.) We can get predictions obtained for all time periods for all patients.

(b) Part-2: Among those subject intervals, $(t_{k-1}, t_k]$, where we observe the subject dying, i.e., where $t_{k-1} < T \leq t_k$ & $\Delta_k = I(\min(C, T) = T) = 1$, we estimate through a generalized linear model (or two-part model which specification is necessary) for the observed costs after conditioning on covariates $\boldsymbol{Z}$ and U_k (as death can occur anywhere in the middle of the interval, so the time of death is continuous), where $U_k = t_k - t_{k-1}$ $if\ T = t_k$ or $U_k = T - t_{k-1}$ if $t_{k-1} < T < t_k$. Predict the costs $\hat{\mu}_{1k}(\boldsymbol{Z})$ for every

subject interval in the data using the parameter estimates from this model. To illustrate the stochastic nature of U within that interval (i.e., to account for costs if the patient died within that interval but at different times), we weighted the observed distribution of U between intervals observed that the patient died, and then averaged the conditional prediction for each value of U. That is, $\hat{\mu}_{1k}(\boldsymbol{Z}) = \int \hat{\mu}_{1k}(\boldsymbol{Z}, u) dF\left(U \middle| t_k < T^{obs} < t_{k+1}\right)$.

(c) PART-3: Next, among the subject intervals $(t_{k-1}, t_k]$, where no patients are observed to die but only costs are observed during a partial period due to censoring is excluded, estimate a generalized linear model (or model if a two-part specification is required) for the observed cost function, conditional on the covariate $\boldsymbol{Z}$. Parameter estimates of this model are used for all subject-intervals in the data to predict the costs $\hat{\mu}_{2k}(\boldsymbol{Z})$. As in the Bang and Tsiatis (2000) estimator, the estimation of this part does not use the subject-interval where censoring occurs, so continuous censoring time can be effectively allowed.

(d) The estimated cost function for interval k for any individual is given as

$$\hat{\mu}_k(\boldsymbol{Z}) = \hat{S}_k(\boldsymbol{Z})[\hat{h}_k(\boldsymbol{Z})\hat{\mu}_{1k}(\boldsymbol{Z}) + \left(1 - \hat{h}_k(\boldsymbol{Z})\right)\hat{\mu}_{2k}(\boldsymbol{Z})] \text{ and } \hat{\mu}(\boldsymbol{Z}) = \sum_{k=1}^{K} \hat{\mu}_k(\boldsymbol{Z}).$$

From the perspective of the approach used to estimate mean accumulated costs, it is necessary to re-emphasize the major differences between the IPW method described in Section 2.2 and the generalized survival adjustment estimates presented in this section. The former uses interval-based trajectories, cost histories, or total costs to accumulate individual specific costs and then calculate the average over patients. Thus, it is characterized by a minimal data case. Whereas, the methods described in this section are

averaged by the probability of survival after summing the cost over time intervals. To this end, they are generally characterized by interval data cases.

2.4 Joint-modeling method

The covariate effects on the total accumulated cost can also be realized through the accumulation intensity of the cost and survival. Therefore, it would be useful to integrate a regression model for cost with survival information. In this section, it is attempted by a joint modeling approach of both survival and medical costs.

Heitjan et al. (2004) considered the joint distribution of survival and medical cost under the assumptions of Weibull distribution for survival and gamma distribution for medical cost:

$$f^T(t) = \frac{\alpha\left(\frac{t}{\lambda}\right)^{\alpha-1} exp\left[-\left(\frac{t}{\lambda}\right)^{\alpha}\right]}{\lambda}, \qquad (t > 0, \alpha > 0, \lambda > 0) \quad and$$

$$f^C(c) = \frac{(v/\mu)^v c^{v-1}\exp(-vc/\mu)}{\Gamma(v)}, \qquad (c > 0, v > 0, \mu > 0),$$

where the mean cost μ and survival time T have a linear relationship, $\mu = a + bT$. In the presence of informative censoring, the joint distribution of cost and survival time is derived through Bayesian method. This estimation method had a assumption that censoring is independent of cost and survival. So, they derived the likelihood function of the joint distribution for estimation under ignore of the informative censoring.

Liu et al. (Liu et al. 2007, 2008; Liu 2009) implemented the idea of jointly modelling employing a shared random effects model. Let v_i be random effect which has a parametric distribution and affects both cost and hazard rate. The joint model of cost and death in interval k for subject i is written as

$$m_{k,i} = \boldsymbol{\beta}' \boldsymbol{Z}_{k,i} + \delta_1 v_i + e_{k,i}$$

$$\lambda_i(t) = \lambda_0(t)\exp\left(\boldsymbol{\gamma}' \boldsymbol{Z}_i + \delta_2 v_i\right)$$

where $\boldsymbol{\beta}, \boldsymbol{\gamma}, \delta_1$ and δ_2 are unknown parameters and $e_{k,i}$ is the error, and $\lambda_i(t)$ is the hazard for death with $\lambda_0(t)$ baseline hazard. Because the presence of a shared random effects term, v_i, makes it easy to see that survival and medical costs are correlated, the model should jointly derive estimators for both the cost and survival functions. Write $\underline{m_i} = \{m_i(1), \ldots, m_i(K)\}$ as the observed history vector of medical cost up to K for subject i. For estimation, we should construct the joint log-likelihood for new observed data $\mathbf{O}_i = \left\{\underline{m_i}, X_i, \Delta_i\right\}$ and v_i as

$$l^* = logL\left(\underline{m_i}, X_i, \Delta_i, v_i\right)$$

$$= \sum_{k=1}^{K}[logL(m_i(k)|X_i, \Delta_i, v_i) + logL(X_i, \Delta_i|v_i) + logp(v_i)],$$

where $p(v_i)$ is the density function. Assume that $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$ and use EM algorithm to obtain maximum likelihood estimation (MLE) for parameters $\theta \equiv \{\boldsymbol{\beta}, \boldsymbol{\gamma}, \delta_1, \delta_2, \sigma_v^2, \sigma_e^2\}$ because v_is' are unobserved. First M step, take the first derivative and second derivative

of l^* with respect to parameters θ. In the E-step, we can use Metropolis-Hastings (M-H) algorithm to generate M random numbers $v_i^{(m)}(m = 1, \dots, M)$ and then obtain estimated expectation value of the sufficient statistics involving frailties. For example, $\hat{E}(v_i|\mathbf{O}_i) = (1/M)\sum_{m=1}^{M} v_i^{(m)}$. And they used Louis's formula to calculate the information matrix for likelihood of observed data. The observed information matrix $I(\hat{\theta})$ is

$$I(\hat{\theta}) = -\hat{E}\left\{\frac{\partial^2 l^*}{\partial\theta\partial\theta'}\middle|\mathbf{O}_i, \hat{\theta}\right\} - \hat{E}\left\{\frac{\partial l^*}{\partial\theta}\frac{\partial l^*}{\partial\theta'}\middle|\mathbf{O}_i, \hat{\theta}\right\} + \hat{E}\left\{\frac{\partial l^*}{\partial\theta}\middle|\mathbf{O}_i, \hat{\theta}\right\}\hat{E}\left\{\frac{\partial l^*}{\partial\theta'}\middle|\mathbf{O}_i, \hat{\theta}\right\}.$$

All three terms are evaluated on the last iteration of the EM algorithm where the last term of the MLE is zero. The first two expectations can be calculated through averaging for the corresponding term containing the M-H values.

3. Dependent censoring

If the mechanism of censoring involves dropout or withdrawal due to worsening symptoms, censoring may introduce bias at the results of statistical analysis. This type of dropout is often mentioned to as informative dropout. Informative dropout is one of many causes of censoring. More generally, when the time of an event of interest is censored by a mechanism associated with the event, this phenomenon is referred to as dependent censoring. Most standard survival analysis methods provide unbiased results under the assumption of independent censoring. Therefore, if it is not independent censoring, it is necessary to pay attention to the survival analysis.

In a cancer follow-up study, survival time may be censored because of dropout due to tumor progression, toxicity, and initiation of next treatment, etc. So, overall survival and censoring time may be positively correlated because patients may usually die soon after dropout. This dropout leads to informative censoring and can have a detrimental effect on data analysis. For example, many terminally ill patients dropped out of clinical trial for stay in their home. This means that data collected from clinical trials not catch many observable deaths. As a result, Kaplan–Meier survival curves that treat these patients as censored may make upward bias.

Dependent censoring is applied to situations where the dependence between censoring and survival time is not accounted by observable covariate. That is, dependent censoring results from residual dependency which is not adjusted by covariates. In a sense, collecting

as many covariates as possible can reduce concerns about dependent censoring. For example, late-stage cancer patients are more likely to have shorter survival time and high possibility of dropout due to tumor progression, which confers a positive dependence between survival and dropout time. Therefore, cancer stage that is one of the covariates can achieve conditional independence between survival and dropout time.

4. Copula models for dependent censoring

In this section, we introduce a mathematical background to bivariate copula models that used in survival analysis. Let T is survival time, C is censoring time, and $\mathbf{z}$ is a vector of covariates. In addition, let $S_T(t|\mathbf{z}) = \Pr(T > t|\mathbf{z})$ and $S_C(c|\mathbf{z}) = \Pr(C > c|\mathbf{z})$ are the marginal survival functions given $\mathbf{z}$. A bivariate survival function

$$\Pr(\mathrm{T} > \mathrm{t}, \mathrm{C} > \mathrm{c}|\mathbf{z}) = \boldsymbol{C_\theta}\{S_T(t|\mathbf{z}), S_C(c|\mathbf{z})\},$$

where a function C_θ is a copula (Nelsen 2006) and parameter θ describes the degree of dependence between T and C.

4.1 Bivariate copula

This section provides a concise introduction to bivariate copulas. A bivariate copula is defined as a bivariate distribution whose marginal distribution is the uniform distribution on [0,1]. Let a bivariate copula, $\boldsymbol{C_\theta}: [0,1]^2 \mapsto [0,1]$, is indexed by a parameter θ. By the definition, any bivariate copula should be satisfying the following conditions

(C1) $\boldsymbol{C_\theta}(u,0) = \boldsymbol{C_\theta}(0,v) = 0$, $\boldsymbol{C_\theta}(u,1) = u$, and $\boldsymbol{C_\theta}(1,v) = v$ for $0 \le u \le 1$ and $0 \le v \le 1$.

(C2) $\boldsymbol{C_\theta}(u_2,v_2) - \boldsymbol{C_\theta}(u_2,v_1) - \boldsymbol{C_\theta}(u_1,v_2) + \boldsymbol{C_\theta}(u_1,v_1) \ge 0$ for $0 \le u_1 \le u_2 \le 1$ and $0 \le v_1 \le v_2 \le 1$.

(C1) requires the two marginal uniform distributions and (C2) requires that $\boldsymbol{C_\theta}$ produces a probability mass on the rectangular region $[u_1, u_2] \times [v_1, v_2]$.

For a copula $\boldsymbol{C_\theta}$, we can consider a pair of random variables (V, W) such that $\Pr(\mathrm{V} \leq \mathrm{u}, \mathrm{W} \leq \mathrm{v}) = \boldsymbol{C_\theta}(u, v)$. If one defines a pair of random variables (T, C) by setting $T = S_T^{-1}(V|\mathbf{z})$ and $C = S_C^{-1}(W|\mathbf{z})$, its bivariate survival function satisfies $\Pr(\mathrm{T} > \mathrm{t}, \mathrm{C} > \mathrm{c}|\mathbf{z}) = \boldsymbol{C_\theta}\{S_T(t|\mathbf{z}), S_C(c|\mathbf{z})\}$.

There are some copulas meet conditions (C1) and (C2):

(a) the independence copula is

$$\mathbf{C}(\mathrm{u}, \mathrm{v}) = \mathrm{uv}.$$

(b) the Clayton copula by Clayton 1978 is

$$\boldsymbol{C_\theta}(\mathrm{u}, \mathrm{v}) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}, \qquad \theta > 0.$$

(c) the Gumbel copula by Gumbel 1960 is

$$\boldsymbol{C_\theta}(\mathrm{u}, \mathrm{v}) = \exp\left[-\left\{(-logu)^{\theta+1} + (-logv)^{\theta+1}\right\}^{\frac{1}{\theta+1}}\right], \qquad \theta \geq 0.$$

(d) the Frank copula by Frank 1979 is

$$\boldsymbol{C_\theta}(\mathrm{u}, \mathrm{v}) = -\frac{1}{\theta} log\left\{1 + \frac{\left(e^{-\theta u} - 1\right)\left(e^{-\theta v} - 1\right)}{e^{-\theta} - 1}\right\}, \qquad \theta \neq 0.$$

(e) the Joe copula by Joe 1993 is

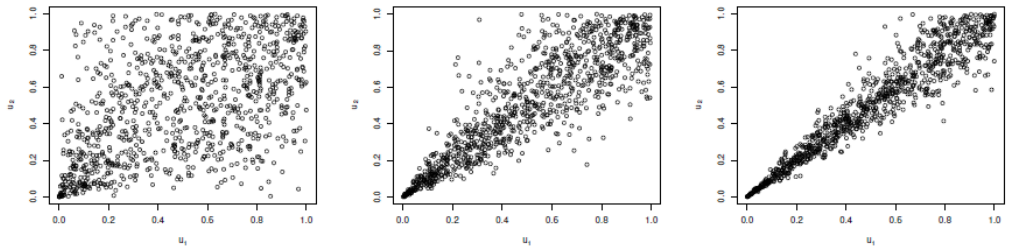
$$\boldsymbol{C_\theta}(\mathrm{u}, \mathrm{v}) = 1 - \left\{\left\{(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta\right\}\right\}^{\frac{1}{\theta}}, \qquad \theta \geq 1.$$

(f) The Farlie-Gumbel-Morgenstern (FGM) copula by Morgenstern (1956) is

$$\boldsymbol{C_\theta}(\mathrm{u}, \mathrm{v}) = uv\{1 + \theta(1-u)(1-v)\}, \qquad -1 \leq \theta \leq 1.$$

By Tovar Cuevas et al. (2019), the Clayton copula function models a highly dependent asymmetric data structure with the left tail indicating that the cloud is

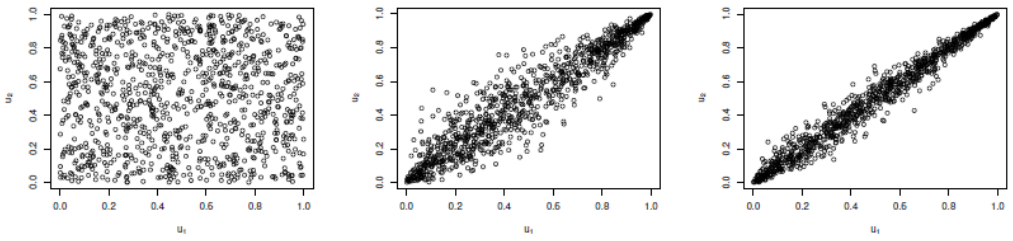
expanding.



(a) $\theta = 1$ (b) $\theta = 5$ (c) $\theta = 10$

Figure 1. Scatter plot of data under the Clayton copula with different θ.

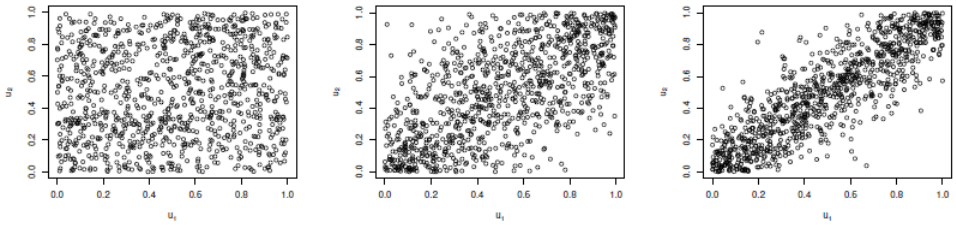
The Gumbel copula is useful for modeling data structures that have a strong dependency on the upper tail and a weak dependency on the lower tail, where we expect the upper data to be strongly correlated and the lower data to be weakly correlated.



(a) $\theta = 1$ (b) $\theta = 5$ (c) $\theta = 10$

Figure 2. Scatter plot of data under the Gumbel copula with different θ.

The Frank copula is appropriate to weak dependency with positive linear trend.



(a) $\theta = 1$ (b) $\theta = 5$ (c) $\theta = 10$

Figure 3. Scatter plot of data under the Frank copula with different θ.

For example, figure 3 gives the scatter plots for (T_i, U_i), $i = 1, \ldots, 500$ and T and U are with the standard exponential distribution under Clayton copula model.

$$\Pr(T_i > t, U_i > u) = \left\{(e^{-t})^{-\theta} + (e^{-u})^{-\theta} - 1\right\}^{-1/\theta}, \text{ for } \theta = 2 \text{ and } \theta = 8.$$

By letting $T_i = -logV_i$ and $U_i = -logW_i$ where (V_i, W_i), $i = 1, \ldots, 500$, the data set was generated from the Clayton copula. The plots show positive dependence, where the dependence's levels are different by θ.

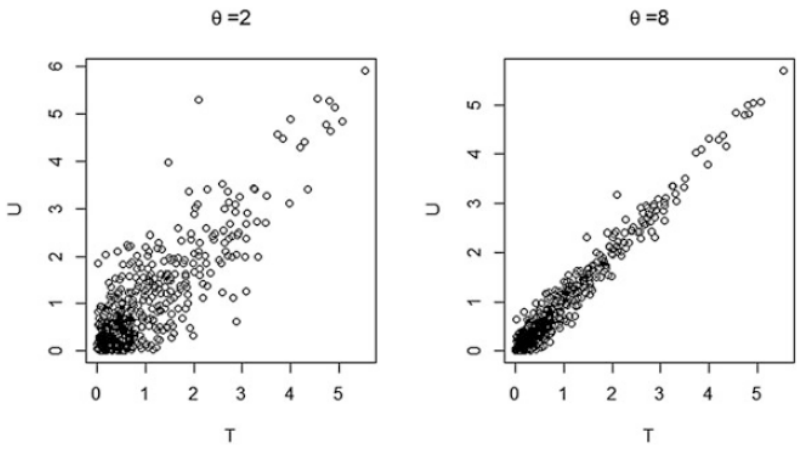


Figure 4. Scatter plot of data ($n = 500$ pairs) generated under the Clayton copula with $\theta = 2$ ($\tau_\theta = 0.5$) and $\theta = 8$ ($\tau_\theta = 0.8$) form the standard exponential distribution

An Archimedean copula is defined as

$$\boldsymbol{C_\theta}(u, v) = \phi_\theta^{-1}\{\phi_\theta(u) + \phi_\theta(v)\},$$

where $\phi_\theta: [0,1] \mapsto [0, \infty]$ is called a generator of the copula that is continuous and strictly decreasing function from $\phi_\theta(0) > 0$ to $\phi_\theta(1) = 0$. If $\phi_\theta(0) \equiv \lim_{t \downarrow 0} \phi_\theta(t) = \infty$, the generator is called a strict generator and has the inverse function ϕ_θ^{-1}: $[0, \infty] \mapsto [0,1]$. The Clayton, Gumbel, Frank, and Joe copulas have a strict generator but, FGM copula does not have a generator as it is not an Archimedean copula.

Let (V, W) be a pair of random variables that satisfy $\Pr(\mathrm{V} \leq \mathrm{u}, \mathrm{W} \leq \mathrm{v}) = C_\theta(u, v)$. To measure of dependence between V and W, Kendall's tau is defined as

$$\tau_\theta = Pr\{(V_2 - V_1)(V_2 - V_1) > 0\} - Pr\{(V_2 - V_1)(V_2 - V_1) < 0\},$$

where (V_1, V_2) and (W_1, W_2) also have the same distribution as (V, W). It can be expressed that

$$\tau_\theta = 4 \int_0^1 \int_0^1 C_\theta(u, v) C_\theta(du, dv) - 1 = 4 \int_0^1 \int_0^1 C_\theta(u, v) C_\theta^{[1,1]}(u, v) du dv - 1,$$

where $C_\theta^{[1,1]}(u, v) = \frac{\partial^2}{\partial u \partial v} C_\theta(u, v)$. Table 1 summarizes τ_θ for copulas and τ_θ increases with $\tau_\theta \to 1$ as $\theta \to \infty$.

Table 1. Examples of copulas

Copula	Parameter	Generator: ϕ_θ	Kendall's tau: τ_θ
Clayton	$\theta > 0$	$(t^{-\theta} - \theta)/\theta$	$\theta/(\theta + 2)$
Gumbel	$\theta \geq 0$	$\{-log(t)\}^{\theta+1}$	$1 - 1/\theta$
Frank	$\theta \neq 0$	$-log\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right)$	$1 - \frac{4}{\theta}\left(1 - \frac{1}{\theta}\int_0^\theta \frac{t}{e^t - 1} dt\right)$
Joe	$\theta \geq 1$	$-log\left(1 - (1 - t)^\theta\right)$	$1 - 4\int_0^\infty \frac{t(1 - e^t)^{2/\theta - 2} e^{-2t}}{\theta^2} dt$
FGM	$-1 \leq \theta \leq 1$	None	$2\theta/9$

4.2 The Copula-Graphic (CG) estimator

Zheng and Klein (1995) proposed the idea assumed copula while analyzing survival data subjected to dependent censoring. They saw a bivariate distribution function of survival and censoring time with completely specified forms of copula functions, including parameter values. This strong assumption about the copula is imposed to make the model identifiable.

They estimated the marginal survival function with a copula-graphic (CG) estimator under assumed copula. The survival function estimated by the CG estimator is similar to that estimated by the Kaplan-Meier estimator. It is reduced to a Kaplan–Meier estimator under independence copulas. In practical applications, the CG estimator is computed assuming one of the Archimedean copulas. Rivest and Wells (2001) proposed a simple

expression of the CG estimator when the assumed copula belongs to the Archimedean copula. Today, CG estimators are indispensable tools for survival analysis with dependent censoring (Braekers and Veraverbeke 2005; Emura and Chen 2018).

Under dependent censoring, Kaplan-Meier estimator may introduce biased information but a survival curve calculated form CG estimator gives unbiased information if copula function between death and censoring time is rightly specified. We introduce the CG estimator proposed by Rivest and Wells (2001). Consider random variables defied as T is survival time and C is censoring time and an Archimedean copula model

$$\Pr(\mathrm{T} > \mathrm{t}, \mathrm{C} > \mathrm{c}) = \phi_\theta^{-1}[\phi_\theta\{S_T(t)\} + \phi_\theta\{S_C(t)\}],$$

where $\phi_\theta: [0,1] \mapsto [0,\infty]$ is generator function, which is strictly decreasing and continuous from $\phi_\theta(0) = \infty$ to $\phi_\theta(1) = 0$; $S_T(t) = \Pr(T > t)$ and $S_C(c) = \Pr(C > c)$ are the marginal survival functions.

Let $(x_i, \Delta_i), \mathrm{i} = 1, \ldots, \mathrm{n},$ be survival data without covariates, where $x_i = \min\{T_i, C_i\}$, $\Delta_i = I(T_i \le C_i)$, $\mathrm{I}(\cdot)$ is the indicator function. All the observed times are assumed to distinct ($x_i \neq x_j$ whenever $i \neq j$). The CG estimator is defined as

$$\hat{S}_T(t) = \phi_\theta^{-1}\left[\sum_{x_i \le t, \Delta_i = 1} \phi_\theta\left(\frac{n_i - 1}{n}\right) - \phi_\theta\left(\frac{n_i}{n}\right)\right], \qquad 0 \le t \le \max_i(x_i)$$

where $n_i = \sum_{\ell=1}^{n} I(t_\ell \ge x_i)$ is the number at risk at time x_i; $\hat{S}_T(t) = 1$ if no death occurs up to time t; $\hat{S}_T(t)$ is undefined for $t > \max_i(x_i)$.

The derivation of the CG estimator can be obtained as follows. Assume that $S_T(t)$ is

decreasing step function with jumps at death times. Then, $\Delta_i = 1$ implies $S_T(x_i) \neq S_T(x_i - dt)$ and $S_C(x_i) = S_C(x_i - dt)$. Let's set $t = c = x_i$ in $\Pr(\mathrm{T} > \mathrm{t}, \mathrm{C} > \mathrm{c}) = \phi_\theta^{-1}[\phi_\theta\{S_T(t)\} + \phi_\theta\{S_C(t)\}]$, we have

$$\phi_\theta \Pr(\mathrm{T} > x_i, \mathrm{C} > x_i) = \phi_\theta\{S_T(x_i)\} + \phi_\theta\{S_C(x_i)\}.$$

In the left-side of the preceding equation, estimate $\Pr(\mathrm{T} > x_i, \mathrm{C} > x_i)$ by $(n_i - 1)/n$, where $n_i - 1 = \sum_{\ell=1}^{n} I(t_\ell > x_i)$ is the number of survivors at time x_i. Accordingly,

$$\phi_\theta\left(\frac{n_i - 1}{n}\right) = \phi_\theta\{S_T(x_i)\} + \phi_\theta\{S_C(x_i)\}, \ \Delta_i = 1.$$

Meanwhile, we set $t = c = x_i - dt$ in equation $\Pr(\mathrm{T} > \mathrm{t}, \mathrm{C} > \mathrm{c}) = \phi_\theta^{-1}[\phi_\theta\{S_T(t)\} + \phi_\theta\{S_C(t)\}]$ and then estimate $\Pr(\mathrm{T} > x_i - dt, \mathrm{C} > x_i - dt)$ by n_i/n. That is,

$$\phi_\theta\left(\frac{n_i}{n}\right) = \phi_\theta\{S_T(x_i - dt)\} + \phi_\theta\{S_C(x_i)\}, \ \Delta_i = 1.$$

The result in the system of difference equations is

$$\phi_\theta\left(\frac{n_i - 1}{n}\right) - \phi_\theta\left(\frac{n_i}{n}\right) = \phi_\theta\{S_T(x_i)\} - \phi_\theta\{S_T(x_i - dt)\}, \ \Delta_i = 1.$$

When x_i is the smallest, we can impose the constraint that $S_T(x_i - dt) = 1$. Then, the solution of different equations is

$$\phi_\theta\{S_T(t)\} = \sum_{x_i \le t, \Delta_i = 1} [\phi_\theta\{S_T(x_i)\} - \phi_\theta\{S_T(x_i - dt)\}]$$

$$= \sum_{x_i \le t, \Delta_i = 1} \left[\phi_\theta\left(\frac{n_i - 1}{n}\right) - \phi_\theta\left(\frac{n_i}{n}\right)\right],$$

which is equivalent to the CG estimator.

When $\phi_\theta(t) = -\log(t)$ under independence copula, the CG estimator is same to the Kaplan-Meier estimator and given by $\phi_\theta(t) = (t^{-\theta} - 1)/\theta$ for $\theta > 0$ under Clayton copula, the CG estimator is

$$\hat{S}_T(t) = \left[1 + \sum_{x_i \le t, \Delta_i = 1} \left\{ \left(\frac{n_i - 1}{n}\right)^{-\theta} - \left(\frac{n_i}{n}\right)^{-\theta} \right\} \right]^{-1/\theta}.$$

4.3 Copula-based univariate Cox regression

Let T survival time, C censoring time, and $\boldsymbol{Z} = (z, \dots, z_p)'$ covariate vector. The joint distribution of T and C can have an arbitrary pattern of dependence for any given covariate z_j. Skala's theorem by Skalar 1959 and Nelsen 2006 assures that the joint survival function can expressed as

$$\Pr(\mathrm{T} > \mathrm{t}, \mathrm{C} > \mathrm{c} | z_j) = \boldsymbol{C_j}\{S_T(t|z_j), S_C(c|z_j)\}, \ j = 1, \dots, p,$$

where $\boldsymbol{C_j}$ is a copula. Under independent censoring assumption, $\boldsymbol{C_j}(t, c) = tc$ for $j = 1, \dots, p$, that is

$$\Pr(\mathrm{T} > \mathrm{t}, \mathrm{C} > \mathrm{c} | z_j) = \Pr(\mathrm{T} > \mathrm{t} | z_j) \times \Pr(\mathrm{C} > \mathrm{c} | z_j).$$

Emura and Chen (2016) proposed a one-parameter copula model under relaxing assumption of independent censoring:

$$\Pr(\mathrm{T} > \mathrm{t}, \mathrm{C} > \mathrm{c} | z_j) = \boldsymbol{C_\theta}\{\Pr(\mathrm{T} > \mathrm{t} | z_j), \Pr(\mathrm{C} > \mathrm{c} | z_j)\}, \qquad j = 1, \dots, p.$$

The assumption that every j has one copula $\boldsymbol{C}$ may be strong. Nonetheless, the copula relaxes the independent censoring assumption by allowing the user to choose the dependency parameter θ flexibly. For example, Clayton copula is

$$\boldsymbol{C_\theta}(\mathrm{u},\mathrm{v}) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}, \qquad \theta > 0,$$

where θ is related to Kendall's tau, $\tau = \frac{\theta}{\theta+2}$. By letting $\theta \to 0$, the Clayton copula model reduces to independent censoring model.

They assumed the Cox models for marginal distribution,

$$\Pr\left(\mathrm{T} > \mathrm{t}\middle|z_j\right) = exp\{-\Lambda_{0j}(t)e^{\beta_j z_j}\}, \qquad \Pr\left(\mathrm{C} > \mathrm{c}\middle|z_j\right) = exp\{-\Gamma_{0j}(c)e^{\gamma_j z_j}\},$$

where β_j and γ_j are regression coefficient and Λ_{0j} and Γ_{0j} are baseline cumulative hazard functions.

The objective parameter is β_j the univariate effect of the jth covariate on survival and other parameters which are γ_j, Λ_{0j} and Γ_{0j} are nuisance. However, under the copula model, the estimator of β_j through the partial likelihood method is not satisfied to consistency. For consistently estimation of parameters, the estimation method should be full likelihood under copula and Cox models. Let

$$D_{\theta,1}(u,v) = \frac{\partial \boldsymbol{C_\theta}(u,v)/\partial u}{\boldsymbol{C_\theta}(u,v)} = -\frac{\partial \Phi_\theta(u,v)}{\partial u},$$

$$D_{\theta,2}(u,v) = \frac{\partial \boldsymbol{C_\theta}(u,v)/\partial v}{\boldsymbol{C_\theta}(u,v)} = -\frac{\partial \Phi_\theta(u,v)}{\partial v},$$

where $\Phi_\theta(u,v) = -log\boldsymbol{C_\theta}(u,v)$. Denote $\{(x_i\Delta_i z_{ij}), i = 1,\dots,n\}$, where $x_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \le C_i)$, $I(\cdot)$ is the indicator function. Let Λ_{0j} and Γ_{0j} are

increasing step functions which have jump sizes $d\Lambda_{0j}(x_i) = \Lambda_{0j}(x_i) - \Lambda_{0j}(x_i - dt)$ for $\Delta_i = 1$ and $d\Gamma_{0j}(x_i) = \Gamma_{0j}(x_i) - \Gamma_{0j}(x_i - dt)$ for $\Delta_i = 0$ as by Chen (2010). Define the log-likelihood function for any given θ.

$$\begin{aligned}
&\ell(\beta_j, \Lambda_{0j}, \Gamma_{0j}|\theta) \\
&= \sum_i \Delta_i \left[\beta_j z_{ij} + log\eta_{1ij}(x_i; \beta_j, \gamma_j, \Lambda_{0j}, \Gamma_{0j}|\theta) + logd\Lambda_{0j}(x_i)\right] \\
&+ \sum_i (1 - \Delta_i)\left[\gamma_j z_{ij} + log\eta_{2ij}(x_i; \beta_j, \gamma_j, \Lambda_{0j}, \Gamma_{0j}|\theta) + logd\Gamma_{0j}(x_i)\right] \\
&- \sum_i \Phi_\theta\left[exp\{-\Lambda_{0j}(t)e^{\beta_j z_{ij}}\}, exp\{-\Gamma_{0j}(x_i)e^{\gamma_j z_{ij}}\}\right],
\end{aligned}$$

where,

$$\begin{aligned}
&\eta_{1ij}(t; \beta_j, \gamma_j, \Lambda_{0j}, \Gamma_{0j}|\theta) \\
&= exp\{-\Lambda_{0j}(t)e^{\beta_j z_{ij}}\} D_{\theta,1}\left[exp\{-\Lambda_{0j}(t)e^{\beta_j z_{ij}}\}, exp\{-\Gamma_{0j}(c)e^{\gamma_j z_{ij}}\}\right], \\
&\eta_{2ij}(t; \beta_j, \gamma_j, \Lambda_{0j}, \Gamma_{0j}|\theta) \\
&= exp\{-\Gamma_{0j}(t)e^{\gamma_j z_{ij}}\} D_{\theta,2}\left[exp\{-\Lambda_{0j}(t)e^{\beta_j z_{ij}}\}, exp\{-\Gamma_{0j}(c)e^{\gamma_j z_{ij}}\}\right].
\end{aligned}$$

The maximum likelihood estimator given θ is denoted as $\left(\hat{\beta}_j(\theta), \hat{\gamma}_j(\theta), \hat{\Lambda}_{0j}(\theta), \hat{\Gamma}_{0j}(\theta)\right)$. The standard error $SE\{\hat{\beta}_j(\theta)\}$ is computed from the information matrix by Chen (2010). The log-likelihood function is maximized by optimization algorithms.

For example, log-likelihood function can be easily computed. Under the Clayton copula,

$\Phi_\theta(u,v) = \alpha^{-1}\log\left(u^{-\theta}+v^{-\theta}-1\right)$, $D_{\theta,1}(u,v) = u^{-\theta-1}\left(u^{-\theta}+v^{-\theta}-1\right)^{-1}$, and $D_{\theta,1}(u,v) = v^{-\theta-1}\left(u^{-\theta}+v^{-\theta}-1\right)^{-1}$.

Hence,

$$\eta_{1ij}(t;\beta_j,\gamma_j,\Lambda_{0j},\Gamma_{0j}|\theta) = \frac{\left[exp\{-\Lambda_{0j}(t)e^{\beta_j z}\}\right]^{-\theta}}{\left[exp\{-\Lambda_{0j}(t)e^{\beta_j z_{ij}}\}\right]^{-\theta}+\left[exp\{-\Gamma_{0j}(c)e^{\gamma_j z_{ij}}\}\right]^{-\theta}-1},$$

$$\eta_{2ij}(t;\beta_j,\gamma_j,\Lambda_{0j},\Gamma_{0j}|\theta) = \frac{\left[exp\{-\Gamma_{0j}(t)e^{\beta_j z_{ij}}\}\right]^{-\theta}}{\left[exp\{-\Lambda_{0j}(t)e^{\beta_j z_{ij}}\}\right]^{-\theta}+\left[exp\{-\Gamma_{0j}(c)e^{\gamma_j z_{ij}}\}\right]^{-\theta}-1}.$$

5. Proposed method

5.1 Inverse probability weighted estimator under dependent censoring

In section 2.2, the simple weighted complete-case estimator (BT) of mean of total cost $E(M_i)$ and the partitioned version (BP_p) of more efficient estimator are

$$\hat{\mu}_{BT} = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i^* M_i}{\widehat{K}(T_i^*)} \;\; and \;\; \hat{\mu}_{BP_p} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K} \frac{\Delta_i^{*k}\{M_i(t_k) - M_i(t_{k-1})\}}{\widehat{K}(T_i^{*k})}.$$

They proposed to estimate the unknown survival function $K(\cdot)$ by the Kaplan-Meier estimator under independent censoring. However, the consistency of the Kaplan-Meier estimator is not ensured since T and C will be in general dependent (de Uña-Álvarez and Veraverbeke 2013). Thus, this study proposes to estimate the unknown survivor function $K(\cdot)$ by the CG estimator. That estimator is

$$\widehat{K}(\mathrm{t}) = \phi_\theta^{-1}\left[\sum_{t_i^* \le t, \Delta_i = 0} \phi_\theta\left(\frac{n_i - 1}{n}\right) - \phi_\theta\left(\frac{n_i}{n}\right)\right], \qquad 0 \le t \le \max_i(t_i^*)$$

where $n_i = \sum_{\ell=1}^{n} I(t_\ell \ge t_i^*)$ is the number at risk at time t_i^*.

We can derive the asymptotic properties of $\hat{S}(\cdot)$ using martingale techniques for the dependent censoring model and do not assume the Archimedean copula for the joint distribution of T and C. The proof process is derived from the survival function of survival time, but it can be transformed into censoring time if $\Delta_i = 0$. Instead, we assume $\Pr(\mathrm{T} > \mathrm{t}, \mathrm{C} > \mathrm{c}) = \boldsymbol{C_\theta}\{S_T(t), S_C(c)\}$ which is a general copula model between T and C

and define that the data is made up n independent replications of $X_1 = C_1 \wedge T_1$ and $\Delta_1 = I[X_1 = T_1]$. So, CG estimator is biased under this model.

Let $N_i(t) = I[X_i \leq t,\ \Delta_i = 1]$, $Y_i(t) = I[X_i \geq t]$ $(i = 1, \dots, n)$, $\bar{N}(t) = \sum_{i=1}^{n} N_i(t)$ and $\bar{Y}(t) = \sum_{i=1}^{n} Y_i(t)$. Then,

$M_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda^{\#}(s)ds$ and $\bar{M}(t) = \bar{N}(t) - \int_0^t \bar{Y}(s)\lambda^{\#}(s)ds$

are martingales w.r.t σ-algebras $\mathcal{F}_t^i = \sigma\{I[X_i \leq t,\ \Delta_i = 1],\ I[X_i \leq t,\ \Delta_i = 0] : 0 < u < t\}$ and $\mathcal{F}_t = \bigvee_{i=1}^{n} \mathcal{F}_t^i$, where $\lambda^{\#}(t)$, the crude hazard rate, is

$$\lambda^{\#}(t) = \frac{-\frac{\partial}{\partial c} P[T \geq c, C \geq t]|_{c=t}}{P[T \geq t, C \geq c]}.$$

Define the dependency between T and C by $\Pr(\mathrm{T} > \mathrm{t}, \mathrm{C} > \mathrm{c}) = C\{S(t), S(c))\}$ and then,

$$\lambda^{\#}(t) = \lambda(t) \frac{S(t) C_{10}\big(S(t), S(c)\big)}{C\big(S(t), S(c)\big)},$$

where $C_{10}(\mathrm{u}, \mathrm{v})$ is the partial derivative of $C(u, v)$ w.r.t u and $\lambda(t)$ is the net hazard rate which is defined by $\lambda(t) = \lim_{h \downarrow 0} \frac{1}{h} P[t \leq T \leq t + h | T \geq t]$. The CG estimator for Archimedean copula using counting process notation is given by

$$\hat{S}(\mathrm{t}) = \phi^{-1}\left[\int_0^t I[\bar{Y}(u) > 0]\left\{\phi\left(\frac{\bar{Y}(u) - 1}{n}\right) - \phi\left(\frac{\bar{Y}(u)}{n}\right)\right\} d\,\bar{N}(u)\right].$$

Because of $\phi\left(\frac{\bar{Y}(u)-1}{n}\right) - \phi\left(\frac{\bar{Y}(u)}{n}\right) \approx -\phi'[\frac{\bar{Y}(u)}{n}]/n$, one has

$$\hat{S}(\mathrm{t}) \approx \phi^{-1}\left[-\frac{1}{\mathrm{n}}\int_0^{\mathrm{t}} \mathrm{I}[\bar{Y}(\mathrm{u}) > 0]\phi'[\frac{\bar{Y}(\mathrm{u})}{n}]\mathrm{d}\,\bar{N}(\mathrm{u})\right].$$

This equation is the first estimate obtained by Zheng and Klein (1994) as the solution of the differential equation. For independent copula, $\phi(\cdot) = -ln(\cdot)$, the Zheng and Klein estimate is reduced by Fleming and Harrington's proposal, which is asymptotically equivalent to Kaplan Meyer estimate. Thus, the CG estimator for Archimedean Copulas and Zheng and Klein's proposal have the same asymptotic behavior.

Now, Rivest and Wells (2001) will deduce the large sample properties of CG estimator for Archimedean copula. Because the CG estimator and Zheng and Klein's proposal have the same asymptotic distribution, $\hat{S}(\mathrm{t})$ denotes Zheng and Klein's estimator in this section. They do not assume that the copula for the dependence between T and C is Archimedean copula correspond to $\phi(\cdot)$ used to calculate the CG estimator. Therefore $\hat{S}(\mathrm{t}) = \phi^{-1}\left[\int_0^t I[\bar{Y}(u) > 0]\left\{\phi\left(\frac{\bar{Y}(u)-1}{n}\right) - \phi\left(\frac{\bar{Y}(u)}{n}\right)\right\} d\,\bar{N}(u)\right]$ estimates a survival distribution S^* which is defined by

$$S^* = \phi^{-1}\left[-\int_0^t \phi'(\pi(u))\pi(u)d\Lambda^{\#}(u)\right],$$

where $\Lambda^{\#}(t) = \int_0^t \lambda^{\#}(u)du$ is the cumulative crude hazard function and $\pi(t) = E(\bar{Y}(t)/n)$, $\pi(t) = C(S(t), C(t))$. If the copula for the dependency between T and C is Archimedean, $S^* = S$ with the dependence given by ϕ. The proofs involve analysis of the martingale $\bar{M}(u)$ and the empirical process $X_n(u) \equiv (1/\sqrt{n})\sum_{i=1}^n\{I(X_i \le u) - \pi(u)\}$.

First, we investigate the consistency of $\hat{S}(\mathrm{t}) = \phi^{-1}\left[\int_0^t I[\bar{Y}(u) > 0]\left\{\phi\left(\frac{\bar{Y}(u)-1}{n}\right) -\right.\right.$

$\phi\left(\frac{\bar{Y}(u)}{n}\right)\Big\} d\,\bar{N}(u)\Big]$ and it suffices to consider $\phi(\hat{S}(\mathrm{t}))$. Let $\psi(s) = -s\phi'(s)$. One has

$$\phi\left(\hat{S}(\mathrm{t})\right) - \phi(S^*(t))$$

$$= -\frac{1}{n}\int_0^t I[Y(u) > 0]\,\phi'\left(\frac{\bar{Y}(u)}{n}\right) d\bar{M}(u)$$

$$+ \int_0^t \left\{ I[Y(u) > 0]\left[\psi\left(\frac{\bar{Y}(u)}{n}\right) - \psi(\pi(u))\right]\right\} d\,\Lambda^{\#}(u).$$

In the proof, $-\frac{1}{n}\int_0^t I[Y(u) > 0]\,\phi'\left(\frac{\bar{Y}(u)}{n}\right) d\bar{M}(u)$ goes to zero in probability by theorem 3.4.2 in Fleming and Harrington (1984). Let $t_0 \in (0, \infty)$ and $\pi(t_0) > 0$. When n is large, $I[Y(u) > 0] = 1$ for $u \in (0, t_0)$ except on a set with al very small probability. The $sup_{0<u<t_0}\left|\frac{\bar{Y}(u)}{n} - \pi(u)\right| \to 0$ as $n \to \infty$ by Clivenko-Cantelli theorem. Hence, $\int_0^t \left\{ I[Y(u) > 0]\left[\psi\left(\frac{\bar{Y}(u)}{n}\right) - \psi(\pi(u))\right]\right\} d\,\Lambda^{\#}(u)$ converges in probability to zero uniformly in t if the derivative of $\psi(t)$ is bounded in $(\pi(t_0), 1)$.

Theorem 1 *in Rivest and Wells* (2001). Let $t_0 \in (0, \infty)$ *be such that* $\pi(t_0) > 0$. *Under the dependent censoring model given by* $\mathrm{H}(\mathrm{t}, \mathrm{c}) = \boldsymbol{C}\{S(t), S(c)\}$ *where* $H(\cdot)$ *denotes the joint survival function of* (t, c) *and assuming that the derivatives of* $\phi(s)$ *and of* $\psi(s)$ *are bounded for* $s \in (\pi(t_0), 1)$, *then estimate* $\hat{S}(\mathrm{t})$ *is uniformly consistent estimate of the marginal survival function* $S^*(t)$ *on* $[0, t_0)$.

Theorem 4.1 *in Zheng and Klein* (1995). Suppose that two marginal distribution functions F, G, are continuous and strictly increasing on $(0, \infty)$, and the assumed copula has density function u > 0 on $[0,1] \times [0,1]$. Then $\hat{F}_n$ and $\hat{G}_n$ *are strongly consitent for* F *and* G. *That is with probability* 1 *as* $n \to \infty, \hat{F}_n(t) \to F(t)$ *and* $\hat{G}_n(t) \to G(t)$ *for all* $t \in [0, \infty)$.

The Theorem 4.1 in Zheng and Klein (1995) about consistency assumed that copula has a strictly positive density on $[0,1] \times [0,1]$. It is a restrictive condition that many Archimedean copulas do not meet. For example, while Clayton copula satisfies the assumption in Theorem 1 in Rivest and Wells (2001), this condition fails. Under Archimedean copula, the crude hazard rate is given by

$$\lambda^{\#}(t) = \lambda(t) \frac{S(t)\phi'\big(S(t)\big)}{\pi(t)\phi'\big(\pi(t)\big)}.$$

The assumption in Theorem 1 in Rivest and Wells (2001) means that the crude ratio $\frac{\lambda^{\#}(t)}{\lambda(t)}$ of the net hazard rate is bounded at zero. The most Archimedean copulas by $\phi(\cdot)$ function meet this condition.

When censoring and survival are not independent, that is $\phi(t) \neq -\ln(t)$, the first term of

$$\phi\left(\hat{S}(\mathrm{t})\right) - \phi(S^*(t))$$

$$= -\frac{1}{n}\int_0^t I[Y(u) > 0]\,\phi'\left(\frac{\bar{Y}(u)}{n}\right) d\bar{M}(u)$$

$$+ \int_0^t \left\{ I[Y(u) > 0]\left[\psi\left(\frac{\bar{Y}(u)}{n}\right) - \psi(\pi(u))\right]\right\} d\,\Lambda^{\#}(u)$$

is a martingale and the second term is not asymptotically null. Hence, the asymptotic distribution of CG estimator depends, in the general case, on $Cov(\bar{M}(u), X_n(s))$ where empirical process $X_n(u) \equiv (1/\sqrt{n})\sum_{i=1}^{n}\{I(X_i \le u) - \pi(u)\}$.

Using martingale's elementary properties, possibly evaluate $Cov(\bar{M}(u), X_n(s))$. Because $\bar{M}(u)$ and $X_n(s)$ are summation of independent random variables,

$$Cov(\bar{M}(u), X_n(s)) = n^{1/2} Cov(M_1(u), I[X_1 > s]).$$

When $u > s$ as in Theorem 1.3.2 by Fleming and Harrington (1984),

$$E\{I[s < X_1 < u, \Delta_1 = 1]\} = \int_s^u P[X_1 > v]\,\lambda^{\#}(v)dv.$$

Hence $E\{M_1(u)I[X_1 > s]\} = -\pi(s)\Lambda^{\#}(s)$ in this case, while when $u \le s$, $E\{M_1(u)I[X_1 > s]\} = -\pi(s)\Lambda^{\#}(u)$. Thus we have proved which

$$Cov(\bar{M}(u), X_n(s)) = -n^{1/2}\pi(s)\Lambda^{\#}(\mathrm{s} \wedge \mathrm{u}).\backslash$$

It is used to prove the following result.

Theorem 2 in Rivest and Wells (2001). Let $t_0, t_0 > 0$, *be such that* $\pi(t_0) > 0$. *Under the dependent censoring model given by* $\mathrm{H(t,c)} = \boldsymbol{C}\{S(t), S(c)\}$ and *assuming that the first two derivatives of* $\phi(s)$ *and* $\psi(s)$, *where* $\psi(s) = -s\phi'(s)$, *are bounded for* $s \in (\pi(t_0), 1)$, *the process* $\sqrt{n}\left\{\hat{S}(\mathrm{t}) - S^*(t)\right\}$

converges weakly on $D[0,t_0)$ *to a mean zero Gaussian process with variance function,*

$$
\begin{aligned}
& v(t) \\
&= \frac{1}{\phi'\left(S^*(t)\right)^2}\left\{\int_0^t \pi(s)\left[\phi'\left(\pi(s)\right)\right]^2 d\Lambda^{\#}(s)\right. \\
&+ 2\int_0^t \int_0^s \pi(u)[1-\pi(u)]\,\psi'\left(\pi(u)\right)\psi'\left(\pi(s)\right) d\Lambda^{\#}(u) d\Lambda^{\#}(s) \\
&\left. + 2\psi'\left(\pi(s)\right) d\Lambda^{\#}(u) d\Lambda^{\#}(s)\right\}.
\end{aligned}
$$

Now, we can calculation the asymptotic distribution of new estimator using δ-method.

$$
\begin{aligned}
& \hat{\mu}_{NEW} \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i^* M_i}{\widehat{K}(T_i^*)} = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i^* M_i}{\phi^{-1}\left[\sum_{t_i^* \le t, \Delta_i = 0} \phi\left(\frac{n_i - 1}{n}\right) - \phi\left(\frac{n_i}{n}\right)\right]}, \quad 0 \le t \le \max_i(t_i^*)
\end{aligned}
$$

where $n_i = \sum_{\ell=1}^{n} I(t_\ell \ge t_i^*)$ is the number at risk at time t_i^*.

Let $\theta = CG\ estimator$ and $g(\theta) = \frac{c}{\theta}, where\ c\text{: }constnat = 1/n\sum_{i=1}^{n}\Delta_i^* M_i$. Then

$$
\sqrt{n}\left(g\left(\hat{\theta}\right) - g(\theta)\right) \xrightarrow{d} N(0, \sigma^2(\theta)[g'(\theta)]^2),
$$

where

$$
\sigma^2(\theta)[g'(\theta)]^2 = v(t)[1/n\sum_{i=1}^{n}\Delta_i^* M_i]^2\left[\phi^{-1}\left[\sum_{t_i^* \le t, \delta_i = 0} \phi\left(\frac{n_i - 1}{n}\right) - \phi\left(\frac{n_i}{n}\right)\right]\right]^{-4}.
$$

5.2 Generalized survival-adjusted estimators under dependent censoring

In section 2.3, Basu the and Manning (2010)'s mean estimator on covariates is

$$\hat{\mu}(\boldsymbol{Z}) = \sum_{k=1}^{K} \hat{S}_k(\boldsymbol{Z})[\hat{h}_k(\boldsymbol{Z})\hat{\mu}_{1k}(\boldsymbol{Z}) + \left(1 - \hat{h}_k(\boldsymbol{Z})\right)\hat{\mu}_{2k}(\boldsymbol{Z})],$$

where $\hat{S}_k(\boldsymbol{Z})$ and $\hat{h}_k(\boldsymbol{Z})$ can be derived from accelerated failure time (AFT) model and $\hat{\mu}_{1k}(\boldsymbol{Z})$ and $\hat{\mu}_{2k}(\boldsymbol{Z})$ from certain generalized linear model (GLM).

$h_k = \Pr\left(t_{k-1} < T \leq t_k | T \geq t_{k-1}\right)$ is the hazard rate of death during the kth interval, $\mu_{1k} = E(m_k | t_{k-1} < T \leq t_k)$ and $\mu_{2k} = E(m_k | T > t_k)$ are the mean of incremental costs for patients who died during or after the kth interval, respectively.

Under the assumption of dependent censoring between survival and censoring time, the process of new estimation follows under three parts:

(a) Part-1: Let $\hat{S}_k(\boldsymbol{Z})$ and $\hat{h}_k(\boldsymbol{Z})$ be the estimated survival and hazard functions for an interval. Estimate an individual's survival function using copula-based Cox regression.

The Cox models for marginal distribution,

$$\Pr\left(\mathrm{T} > \mathrm{t} | z_j\right) = exp\{-\Lambda_{0j}(t)e^{\beta_j z_j}\}, \qquad \Pr\left(\mathrm{C} > \mathrm{c} | z_j\right) = exp\{-\Gamma_{0j}(c)e^{\gamma_j z_j}\},$$

where β_j and γ_j are regression coefficient and Λ_{0j} and Γ_{0j} are baseline cumulative hazard functions. The cause-specific hazard is defined as

$h^{\#}\left(t | z_j\right) = \Pr\left(t < T < t + dt, T \leq C | T \geq t, C \geq c, z_j\right)/dt$. If independent censoring

holds, then

$$h^{\#}(t|z_j) = h(t|z_j) \equiv \frac{\Pr(t < T < t + dt | T \geq t)}{dt}.$$

Skala's theorem by Skala 1959 and Nelsen 2006 assures that the joint survival function can expressed as

$$\Pr(\mathrm{T} > \mathrm{t}, \mathrm{C} > \mathrm{c}|z_j) = \boldsymbol{C_\theta}\{S_T(t|z_j), S_C(c|z_j)\}, \ j = 1, \dots, p.$$

Rivest and Wells (2001) indicated the cause-specific hazard becomes $h_\theta^{\#}(t|z_j) = r_\theta(t|z_j)h(t|z_j)$, where

$$r_\theta(t|z_j) = \frac{\boldsymbol{C_{\theta,1}}\{S_T(t|z_j), S_C(c|z_j)\}S_T(t|z_j)}{\boldsymbol{C_\theta}\{S_T(t|z_j), S_C(c|z_j)\}},$$

and $\boldsymbol{C_{\theta,1}}(u, v) = \frac{\partial C_\theta(u,v)}{\partial u}$. Emura and Chen (2014) defined the apparent effect of covariate z_j as

$$\beta_\theta^{\#} \equiv log\frac{h_\theta^{\#}(t|z_j = 1)}{h_\theta^{\#}(t|z_j = 0)} = log\frac{h(t|z_j = 1)}{h(t|z_j = 0)} + log\frac{r_\theta(t|z_j = 1)}{r_\theta(t|z_j = 0)}.$$

This equation shows that the apparent effects can be divided into true (net) effects and bias because of dependent censoring. Here, the copula structure is entered only in the bias term.

(b) Part-2: Among those subject intervals, $(t_{k-1}, t_k]$, where we observe the subject dying, i.e., where $t_{k-1} < T \leq t_k$ & $\Delta_k = I(\min(C, T) = T) = 1$, we estimate through a generalized linear model (or two-part model which specification is necessary) for the observed costs after conditioning on covariates $\boldsymbol{Z}$ and U_k (as death can occur anywhere in the middle of the interval, so the time of death is continuous), where $U_k = t_k -$

t_{k-1} $if\ T = t_k$ or $U_k = T - t_{k-1}$ if $t_{k-1} < T < t_k$. Predict the costs $\hat{\mu}_{1k}(\boldsymbol{Z})$ for every subject interval in the data using the parameter estimates from this model. To illustrate the stochastic nature of U within that interval (i.e., to account for costs if the patient died within that interval but at different times), we weighted the observed distribution of U between intervals observed that the patient died, and then averaged the conditional prediction for each value of U. That is, $\hat{\mu}_{1k}(\boldsymbol{Z}) = \int \hat{\mu}_{1k}(\boldsymbol{Z}, u) dF\left(U \middle| t_k < T^{obs} < t_{k+1}\right)$.

(c) PART-3: Next, among the subject intervals $(t_{k-1}, t_k]$, where no patients are observed to die but only costs are observed during a partial period due to censoring is excluded, estimate a generalized linear model (or model if a two-part specification is required) for the observed cost function, conditional on the covariate $\boldsymbol{Z}$. Parameter estimates of this model are used for all subject-intervals in the data to predict the costs $\hat{\mu}_{2k}(\boldsymbol{Z})$. As in the Bang and Tsiatis (2000) estimator, the estimation of this part does not use the subject-interval where censoring occurs, so continuous censoring time can be effectively allowed.

(d) The estimated cost function for interval k for any individual is given as

$$\hat{\mu}_k(\boldsymbol{Z}) = \hat{S}_k(\boldsymbol{Z})[\hat{h}_k(\boldsymbol{Z})\hat{\mu}_{1k}(\boldsymbol{Z}) + \left(1 - \hat{h}_k(\boldsymbol{Z})\right)\hat{\mu}_{2k}(\boldsymbol{Z})] \text{ and } \hat{\mu}(\boldsymbol{Z}) = \sum_{k=1}^{K} \hat{\mu}_k(\boldsymbol{Z}).$$

6. Simulation study

In this section, we conduct various simulation studies to evaluate the performance of the proposed new estimators in dependent censoring data by varying type of copulas, dependency parameters, the relation of between covariate and censoring distribution, and censoring rate.

6.1 Simulation setting

We start by using Lin's (2003) and Basu and Manning (2010) simulation design points to carry out extensive simulations to evaluate our proposed estimators and to compare it with the prevailing methods. Following Lin (2003) and Basu and Manning (2010), the survival times are generated from the exponential distribution with mean m and censoring times are generated from the uniform $(0, c)$ distribution, respectively. The maximum follow-up time is set to 10, (0, 10], at equal intervals and all survival time and cost cumulative processes are censored at the end. The medical costs for individual i in the k th interval are generated using:

$$y_{ki} = \left[I(k=1)u_i^d + I(T_i > t_k)(\eta_i + u_{ki}) + I(t_{k-1} < T_i \leq t_k)\{(\eta_i + u_{ki})(T_i - t_{k-1}) + u_i^f\}\right]e^{\beta' X_i},$$

where $k = 1, \ldots, 10$, $i = 1, \ldots, n$, η_i, u_{ki}, u_i^d and u_i^f are independent random variables with η_i, $u_{ki} \sim$ uniform (0,1) distribution and $u_i^d \sim$ uniform $(0,5)$, $u_i^f \sim$ uniform (0,10), respectively.

The scheme creates a J-shaped time pattern; each time interval in which the subject is

alive has some basic cost. In addition, the first interval has a relatively high diagnostic cost, and the interval at which the subject dies has a much higher final cost. In our simulation, z was set as the treatment indicator with n/2 in each of the two groups. We chose $n = 1{,}000$ and β was set to 1. The true value was calculated under empirical distribution with $n = 100{,}000$. Standard errors are computed from the summary statistics across the replicates.

We focus on the average incremental effect of the treatment on the cost. Therefore, interest lies in the incremental effect parameter:

$$\Delta = \sum_{k=1}^{10}\big(\mu_k(Z=1) - \mu_k(Z=0)\big), \text{ where } \mu_k(Z) = E(y_{ki}|Z).$$

Table 2 show the details of the simulation scenarios. For each scenario, we generated 1,000 random datasets consisting of 1,000 random subjects. In the scenario 1-12 and 25-36, censoring and survival time were generated regardless of covariate but, in the scenario 13-24 and 37-48, the data was generated so that the censoring time was longer in the treatment group. Each scenario consisted of approximately 20, 30, 40, and 50% censored survival times through combination of survival and censoring time and $\theta = 0, 2, 10$ of Clayton copula.

Table 2. Information of simulation scenarios

Scenario	Relation of Z and (T, C)	θ levels of Clayton copula	Percentage of censoring	Survival time	Censoring time
1	Independent	0	20	$Exp(5)$	$U(0,21)$
2	Independent	0	30	$Exp(5)$	$U(0,14)$
3	Independent	0	40	$Exp(5)$	$U(0,11)$
4	Independent	0	50	$Exp(5)$	$U(0,8)$
5	Independent	2	20	$Exp(5)$	$U(0,12.5)$
6	Independent	2	30	$Exp(5)$	$U(0,11)$
7	Independent	2	40	$Exp(5)$	$U(0,9)$
8	Independent	2	50	$Exp(5)$	$U(0,7.5)$
9	Independent	10	20	$Exp(5)$	$U(0,10)$
10	Independent	10	30	$Exp(5)$	$U(0,9)$
11	Independent	10	40	$Exp(5)$	$U(0,8)$
12	Independent	10	50	$Exp(5)$	$U(0,7)$
13	Dependent	0	20	$Exp(5)$	$U_{z=1}(0,24)$,$U_{z=0}(0,19)$
14	Dependent	0	30	$Exp(5)$	$U_{z=1}(0,17)$,$U_{z=0}(0,12)$
15	Dependent	0	40	$Exp(5)$	$U_{z=1}(0,14)$,$U_{z=0}(0,9)$
16	Dependent	0	50	$Exp(5)$	$U_{z=1}(0,11)$,$U_{z=0}(0,6)$
17	Dependent	2	20	$Exp(5)$	$U_{z=1}(0,16)$,$U_{z=0}(0,11)$
18	Dependent	2	30	$Exp(5)$	$U_{z=1}(0,13.5)$,$U_{z=0}(0,8.5)$
19	Dependent	2	40	$Exp(5)$	$U_{z=1}(0,12)$,$U_{z=0}(0,7)$
20	Dependent	2	50	$Exp(5)$	$U_{z=1}(0,10.5)$,$U_{z=0}(0,5.5)$
21	Dependent	10	20	$Exp(5)$	$U_{z=1}(0,13)$,$U_{z=0}(0,8)$
22	Dependent	10	30	$Exp(5)$	$U_{z=1}(0,11.5)$,$U_{z=0}(0,6.5)$
23	Dependent	10	40	$Exp(5)$	$U_{z=1}(0,11)$,$U_{z=0}(0,6)$
24	Dependent	10	50	$Exp(5)$	$U_{z=1}(0,10.5)$,$U_{z=0}(0,5.5)$

(Continued on next page)

Table 2. Information of simulation scenarios (continued)

Scenario	Relation of Z and (T, C)	θ levels of Clayton copula	Percentage of censoring	Survival time	Censoring time
25	Independent	0	20	$Exp(10)$	$U(0,31)$
26	Independent	0	30	$Exp(10)$	$U(0,21)$
27	Independent	0	40	$Exp(10)$	$U(0,16)$
28	Independent	0	50	$Exp(10)$	$U(0,13.5)$
29	Independent	2	20	$Exp(10)$	$U(0,19)$
30	Independent	2	30	$Exp(10)$	$U(0,16)$
31	Independent	2	40	$Exp(10)$	$U(0,13.5)$
32	Independent	2	50	$Exp(10)$	$U(0,12.5)$
33	Independent	10	20	$Exp(10)$	$U(0,14.5)$
34	Independent	10	30	$Exp(10)$	$U(0,13.5)$
35	Independent	10	40	$Exp(10)$	$U(0,12.5)$
36	Independent	10	50	$Exp(10)$	$U(0,12)$
37	Dependent	0	20	$Exp(10)$	$U_{z=1}(0,34), U_{z=0}(0,29)$
38	Dependent	0	30	$Exp(10)$	$U_{z=1}(0,24), U_{z=0}(0,19)$
39	Dependent	0	40	$Exp(10)$	$U_{z=1}(0,18.5), U_{z=0}(0,13.5)$
40	Dependent	0	50	$Exp(10)$	$U_{z=1}(0,15.5), U_{z=0}(0,10.5)$
41	Dependent	2	20	$Exp(10)$	$U_{z=1}(0,22), U_{z=0}(0,17)$
42	Dependent	2	30	$Exp(10)$	$U_{z=1}(0,19), U_{z=0}(0,14)$
43	Dependent	2	40	$Exp(10)$	$U_{z=1}(0,16.5), U_{z=0}(0,11.5)$
44	Dependent	2	50	$Exp(10)$	$U_{z=1}(0,15.5), U_{z=0}(0,10.5)$
45	Dependent	10	20	$Exp(10)$	$U_{z=1}(0,18), U_{z=0}(0,13)$
46	Dependent	10	30	$Exp(10)$	$U_{z=1}(0,17), U_{z=0}(0,12)$
47	Dependent	10	40	$Exp(10)$	$U_{z=1}(0,16), U_{z=0}(0,11)$
48	Dependent	10	50	$Exp(10)$	$U_{z=1}(0,15), U_{z=0}(0,10)$

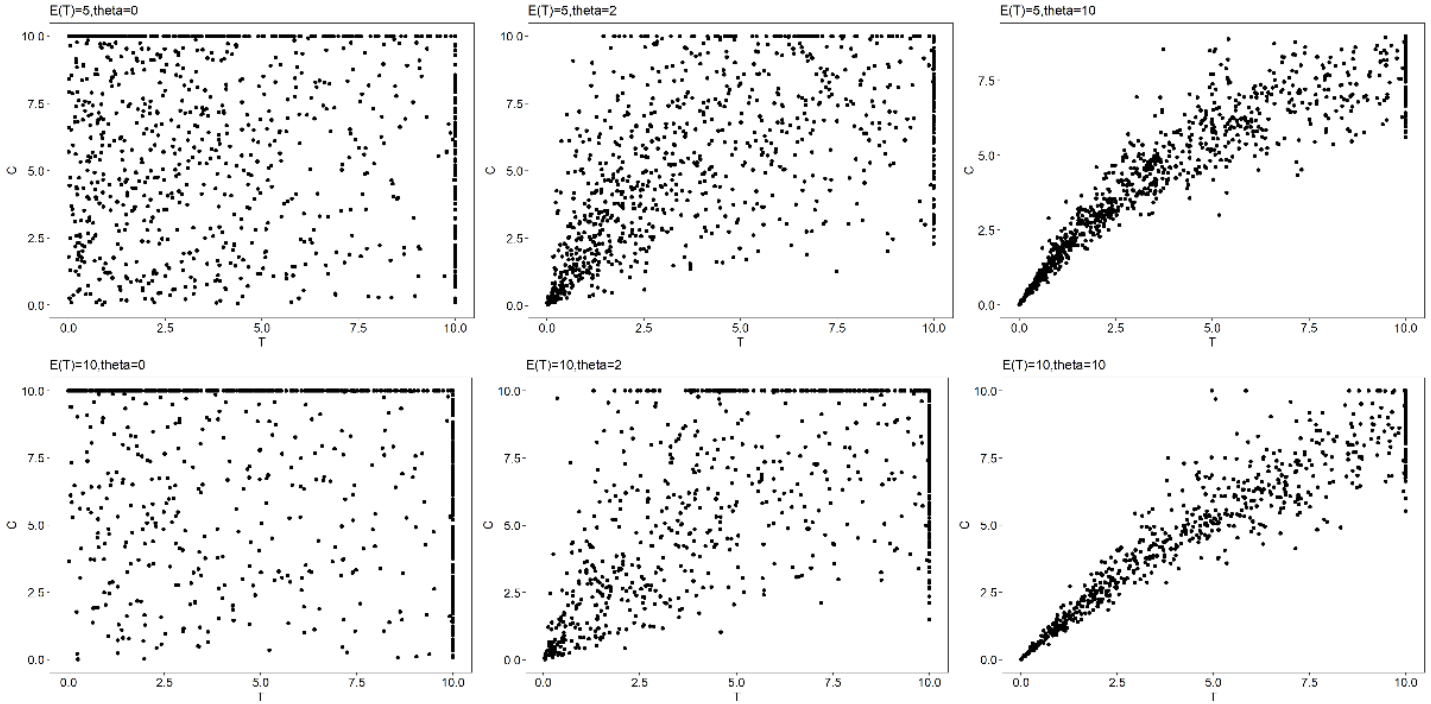


Figure 5. Scatter plot of simulation data ($n = 500$ pairs) generated under the Clayton copula, $T \sim Exp(5)$ or $T \sim Exp(10)$, and independence between Z and (T, C) according to $\theta = 0, 2,$ and 5

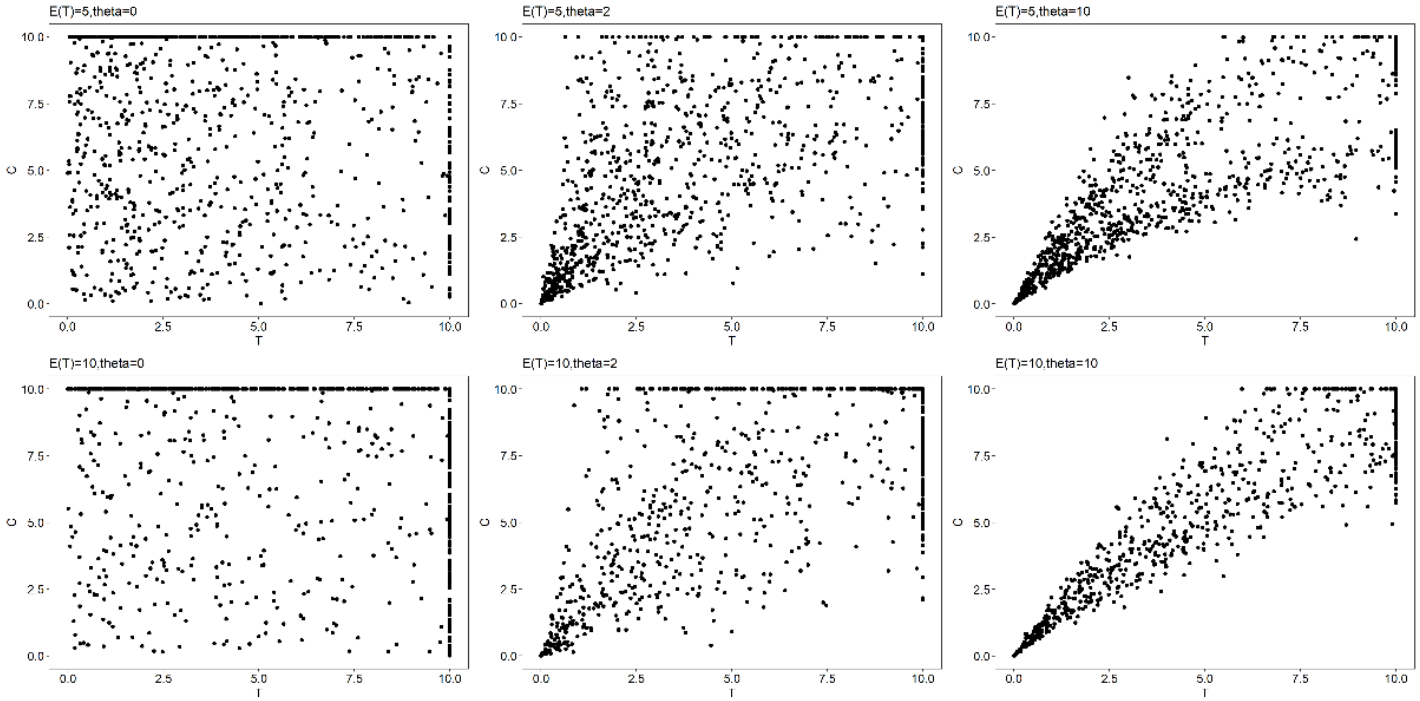


Figure 6. Scatter plot of simulation data ($n = 500$ pairs) generated under the Clayton copula, $T \sim Exp(5)$ or $T \sim Exp(10)$, and dependence between Z and (T, C) according to $\theta = 0,2,$ and 5

6.2 Results

The results of all scenarios are summarized in Table 3-14 with bias and S.E. Table 3-6 provides the results of simulation studies using survival time $Exp(5)$ and independence between Z and (T, C) (Table 4 and 5) and independence between Z and (T, C) (Table 5 and 6). Table 4 showed that the proposed method using IPW scheme had smaller bias under $\tau = 0.5$ and 20 and 30 % of censoring compared to BT_p. But, S.E. of proposed method had smaller value in all conditions. In generalized survival-adjusted estimator, bias using proposed method is smaller than BM under 40% censoring. When the dependency between Z and (T, C) existed, BT_p had higher bias. The proposed method outperformed the existing method in estimation both in terms of bias and S.E. Among the estimation method using the CG estimator, the Gumbel function worked best (Table 5). Table 6 showed that generalized survival-adjusted estimator was more effective estimation method than IPW method regardless of τ. The generalized survival-adjusted estimator copula-based Cox regression had lower bias and S.E. compared to BM estimator. Regardless of the τ and censing percentage, the estimator calculated using the Clayton function had the smallest bias value.

Table 7-10 provides the results of simulation studies using survival time $Exp(10)$ and independence between Z and (T, C) (Table 7 and 8) and independence between Z and (T, C) (Table 9 and 10). In table 7, BT_p was not good at estimating when expectation value of mean survival time is closer to τ. The proposed methods have the smaller bias and S.E than BT when τ was above 0.5. Among the estimation method using the CG

estimator, the Gumbel function worked best. Table 8 showed that proposed method under generalized survival-adjusted estimator only had smaller bias when $\tau = 0.5$. The proposed method was not work for estimating cost when expectation value of mean survival time is closer to τ and τ is above the 0.5. When the dependency between Z and (T, C) existed, BT_p had higher bias and S.E. The proposed method outperformed the existing method in estimation both in terms of bias and S.E. Among the estimation method using the CG estimator, the Gumbel function worked best (Table 9). Table 10 showed that generalized survival-adjusted estimator was good at bias compared to IPW method in most settings. The generalized survival-adjusted estimator copula-based Cox regression had lower bias compared to BM estimator. Regardless of the τ and censing percentage, the estimator calculated using the Clayton function had the smallest bias value only except $\tau = 0.83$ and % of censoring$= 50$.

In previous simulation, we used $\tau = 0.5$ for estimating to proposed method regardless of the type of copula function because we generated survival time and censoring time using copula function at $\tau = 0.5$. However, we investigated the simulation according to different τ (Table 11-14). In survival time $Exp(5)$, we found that we should choose the higher τ in proposed method using IPW method for aspect of bias under independence and dependence between Z and (T, C). But the choosing the τ in the proposed method using copula-based Cox regression was suggested to censoring %. Under 30% of censoring, we should choose the lower τ for reduction of bias. Above 40 % of censoring, we should choose the higher τ (Table 11 and Table 12). In survival time $Exp(10)$, we found that we

should choose the higher τ in proposed method using IPW method for aspect of bias under independence and dependence between Z and (T, C) (Table 13 and Table 14). But the choosing the τ in the proposed method using generalized-adjusted survival was suggested to the type of copula function. Under Clayton and Frank copula, we should choose the lower τ for reduction of bias. Under Gumbel copula, we should choose lower τ under 30 % of censoring and choose the same τ which was used to generating T and C above 40% of censoring (Table 13). In Table 14, under Clayton and Frank copula, we should choose the lower τ under 20% censoring, choose the same $\tau = 0.5$ under 30-40% censoring, and choose the higher τ above the 50% censoring. Whereas, we could not find some specific trend the choosing τ for reduction of bias under Gumbel copula.

Table 3. Simulation scenarios 1-12 results of the incremental effect (Δ) using IPW scheme under independence between Z and (T, C)

Clayton Copula	% of cens	True value of Δ	Δ under IPW estimator									
			BT		BT_p		CG Clayton		CG Frank		CG Gumbel	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0$ ($\theta = 0$)	20	19.094	5.926	0.720	-2.698	2.801	-	-	-	-	-	-
	30	18.385	6.112	0.753	-3.008	3.051	-	-	-	-	-	-
	40	18.092	6.797	0.695	-3.389	3.790	-	-	-	-	-	-
	50	17.205	7.155	0.630	0.413	1.940	-	-	-	-	-	-
$\tau = 0.5$ ($\theta = 2$)	20	19.683	6.599	0.719	-5.654	3.398	5.104	0.627	4.965	0.622	4.099	0.619
	30	19.305	7.026	0.696	-6.812	3.866	5.273	0.606	5.127	0.599	4.292	0.598
	40	18.837	7.892	0.608	-1.558	2.506	6.190	0.485	6.028	0.477	5.290	0.480
	50	18.171	7.669	0.616	-1.852	1.681	6.331	0.436	6.159	0.428	5.523	0.435
$\tau = 0.83$ ($\theta = 10$)	20	19.836	8.508	0.541	-0.793	1.938	5.615	0.549	5.478	0.557	5.243	0.631
	30	19.615	8.808	0.561	-1.260	2.145	5.961	0.531	5.781	0.546	5.446	0.682
	40	19.125	8.873	0.548	-1.916	1.893	6.142	0.469	5.888	0.498	5.384	0.695
	50	18.549	8.898	0.593	-2.362	1.849	6.410	0.420	6.079	0.449	5.551	0.490

% of cens: percentage of censoring; SE: standard deviation of estimates across 1000 replicates
When $\tau = 0.5$, θ of Clayton copula=2, , θ of Frank copula=5.736, and , θ of Gumbel copula=2.
When $\tau = 0.83$, θ of Clayton copula=10, , θ of Frank copula=22.224, and , θ of Gumbel copula=6.

Table 4. Simulation scenarios 1-12 results of the incremental effect (Δ) using generalized survival-adjusted estimator under independence between Z and (T, C)

Clayton Copula	% of cens	True value of Δ	Δ under generalized survival-adjusted estimator							
			BM		Copula based Cox regression under Clayton		Copula based Cox regression under Frank		Copula based Cox regression under Gumbel	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0$ ($\theta = 0$)	20	19.094	2.890	0.804	-	-	-	-	-	-
	30	18.385	1.908	0.898	-	-	-	-	-	-
	40	18.092	1.373	1.023	-	-	-	-	-	-
	50	17.205	1.172	0.988	-	-	-	-	-	-
$\tau = 0.5$ ($\theta = 2$)	20	19.683	2.955	0.794	-0.519	1.330	-0.412	1.324	-0.240	1.320
	30	19.305	2.331	0.884	-0.348	1.624	-0.230	1.604	0.001	1.592
	40	18.837	1.866	0.886	1.201	1.380	1.316	1.334	1.626	1.333
	50	18.171	1.510	0.862	2.727	1.255	2.791	1.225	3.074	1.217
$\tau = 0.83$ ($\theta = 10$)	20	19.836	3.958	0.752	0.911	1.405	0.935	1.388	0.938	1.386
	30	19.615	3.729	0.748	2.211	1.376	2.235	1.355	2.239	1.354
	40	19.125	3.132	0.781	3.155	1.329	3.184	1.314	3.189	1.311
	50	18.549	2.170	0.769	4.044	1.372	4.084	1.350	4.094	1.347

% of cens : percentage of censoring ; SE : standard deviation of estimates across 1000 replicates
When $\tau = 0.5$, θ of Clayton copula=2, , θ of Frank copula=5.736, and , θ of Gumbel copula=2.
When $\tau = 0.83$, θ of Clayton copula=10, , θ of Frank copula=22.224, and , θ of Gumbel copula=6.

Table 5. Simulation scenarios 13-24 results of the incremental effect (Δ) using IPW estimator under dependence between Z and (T, C)

Clayton Copula	% of cens	True value of Δ	Δ under IPW estimator									
			BT		BT_p		CG Clayton		CG Frank		CG Gumbel	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0$ $(\theta = 0)$	20	19.418	5.850	0.759	-2.897	2.792	-	-	-	-	-	-
	30	19.142	5.618	0.730	-2.959	2.903	-	-	-	-	-	-
	40	19.073	5.625	0.730	-4.976	2.924	-	-	-	-	-	-
	50	18.884	6.519	0.666	-6.387	3.575	-	-	-	-	-	-
$\tau = 0.5$ $(\theta = 2)$	20	20.353	5.634	0.710	-3.175	3.013	4.463	0.632	4.399	0.631	3.533	0.651
	30	20.425	5.533	0.696	-7.177	3.010	4.233	0.623	4.128	0.618	3.146	0.616
	40	20.528	6.141	0.725	-8.815	3.471	4.359	0.625	4.233	0.619	3.204	0.617
	50	20.521	7.445	0.696	-11.251	4.039	4.703	0.603	4.572	0.591	3.528	0.590
$\tau = 0.83$ $(\theta = 10)$	20	20.955	4.708	0.664	-5.367	2.695	3.337	0.648	3.315	0.661	3.414	0.702
	30	21.344	5.559	0.683	-6.845	2.951	3.236	0.649	3.265	0.657	3.275	0.690
	40	21.398	6.157	0.693	-9.436	3.188	3.098	0.663	3.086	0.668	2.972	0.708
	50	21.391	7.104	0.698	-11.150	3.856	3.212	0.648	3.137	0.653	2.914	0.700

% of cens: percentage of censoring; SE: standard deviation of estimates across 1000 replicates
When $\tau = 0.5$, θ of Clayton copula=2, , θ of Frank copula=5.736, and , θ of Gumbel copula=2.
When $\tau = 0.83$, θ of Clayton copula=10, , θ of Frank copula=22.224, and , θ of Gumbel copula=6.

Table 6. Simulation scenarios 13-24 results of the incremental effect (Δ) using using generalized survival-adjusted estimator under dependence between Z and (T, C)

Clayton Copula	% of cens	True value of Δ	Δ under generalized survival-adjusted estimator							
			BM		Copula-based Cox regression under Clayton		Copula-based Cox regression under Frank		Copula based Cox regression under Gumbel	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0$ $(\theta = 0)$	20	19.418	3.310	0.805	-	-	-	-	-	-
	30	19.142	2.861	0.835	-	-	-	-	-	-
	40	19.073	2.448	0.919	-	-	-	-	-	-
	50	18.884	1.302	0.970	-	-	-	-	-	-
$\tau = 0.5$ $(\theta = 2)$	20	20.353	4.440	0.752	-0.669	1.279	-0.674	1.277	-0.694	1.278
	30	20.425	4.005	0.756	-2.443	1.393	-2.390	1.392	-2.382	1.391
	40	20.528	3.676	0.875	-2.926	1.349	-2.839	1.345	-2.747	1.341
	50	20.521	2.807	0.902	-2.953	1.717	-2.886	1.675	-2.678	1.661
$\tau = 0.83$ $(\theta = 10)$	20	20.955	5.227	0.678	-2.862	1.223	-2.875	1.224	-2.874	1.224
	30	21.344	5.764	0.705	-3.058	1.285	-3.065	1.285	-3.065	1.285
	40	21.398	6.155	0.735	-3.022	1.304	-3.030	1.302	-3.031	1.301
	50	21.391	6.191	0.757	-2.789	1.468	-2.803	1.463	-2.809	1.461

% of cens : percentage of censoring; SE: standard deviation of estimates across 1000 replicates
When $\tau = 0.5$, θ of Clayton copula=2, , θ of Frank copula=5.736, and , θ of Gumbel copula=2.
When $\tau = 0.83$, θ of Clayton copula=10, , θ of Frank copula=22.224, and , θ of Gumbel copula=6.

Table 7. Simulation scenarios 25-36 results of the incremental effect (Δ) using IPW scheme under independence between Z and (T, C)

Clayton Copula	% of cens	True value of Δ	Δ under IPW estimator									
			BT		BT_p		CG Clayton		CG Frank		CG Gumbel	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0$ ($\theta = 0$)	20	22.248	3.668	0.877	-23.421	5.105	-	-	-	-	-	-
	30	21.598	3.698	1.018	-21.925	5.533	-	-	-	-	-	-
	40	20.829	3.869	1.021	-21.338	5.326	-	-	-	-	-	-
	50	20.206	4.861	1.027	-20.436	5.770	-	-	-	-	-	-
$\tau = 0.5$ ($\theta = 2$)	20	23.063	3.694	0.903	-29.860	5.650	3.833	0.765	3.761	0.761	3.075	0.758
	30	22.821	3.730	0.953	-31.156	5.981	4.359	0.762	4.269	0.760	3.636	0.756
	40	22.102	3.670	1.042	-33.134	6.571	4.708	0.736	4.607	0.731	4.061	0.726
	50	21.812	3.880	1.133	-34.122	6.830	4.981	0.710	4.880	0.706	4.386	0.703
$\tau = 0.83$ ($\theta = 10$)	20	23.293	3.251	0.886	-43.958	6.003	0.809	0.696	0.656	0.694	0.134	0.735
	30	23.106	3.219	0.972	-50.640	6.522	1.416	0.714	1.261	0.706	0.728	0.726
	40	22.967	3.366	1.063	-56.703	6.629	2.581	0.711	2.415	0.698	1.937	0.709
	50	22.621	3.160	1.150	-60.294	6.815	2.992	0.673	2.828	0.668	2.411	0.679

% of cens : percentage of censoring ; SE : standard deviation of estimates across 1000 replicates
When $\tau = 0.5$, θ of Clayton copula=2, , θ of Frank copula=5.736, and , θ of Gumbel copula=2.
When $\tau = 0.83$, θ of Clayton copula=10, , θ of Frank copula=22.224, and , θ of Gumbel copula=6.

Table 8. Simulation scenarios 25-36 results of the incremental effect (Δ) using generalized survival-adjusted estimator under independence between Z and (T, C)

Clayton Copula	% of cens	True value of Δ	Δ under generalized survival-adjusted estimator							
			BM		Copula-based Cox regression under Clayton		Copula-based Cox regression under Frank		Copula-based Cox regression under Gumbel	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0$ $(\theta = 0)$	20	22.248	1.201	0.835	-	-	-	-	-	-
	30	21.598	0.308	0.953	-	-	-	-	-	-
	40	20.829	-0.770	1.015	-	-	-	-	-	-
	50	20.206	-1.700	1.134	-	-	-	-	-	-
$\tau = 0.5$ $(\theta = 2)$	20	23.063	1.226	0.857	1.043	1.371	1.111	1.372	1.226	1.373
	30	22.821	0.417	0.924	0.853	1.471	0.960	1.471	1.135	1.472
	40	22.102	-0.988	0.995	0.385	1.617	0.543	1.616	0.798	1.617
	50	21.812	-1.714	1.065	0.185	1.730	0.371	1.727	0.670	1.724
$\tau = 0.83$ $(\theta = 10)$	20	23.293	-2.192	0.829	-13.170	1.760	-13.133	1.761	-13.114	1.760
	30	23.106	-3.704	0.933	-13.238	2.268	-13.182	2.267	-13.153	2.266
	40	22.967	-5.946	1.048	-13.289	3.093	-13.207	3.088	-13.168	3.085
	50	22.621	-7.684	1.181	-12.974	3.646	-12.881	3.636	-12.839	3.630

% of cens : percentage of censoring ; SE : standard deviation of estimates across 1000 replicates
When $\tau = 0.5$, θ of Clayton copula=2, , θ of Frank copula=5.736, and , θ of Gumbel copula=2.
When $\tau = 0.83$, θ of Clayton copula=10, , θ of Frank copula=22.224, and , θ of Gumbel copula=6.

Table 9. Simulation scenarios 37-48 results of the incremental effect (Δ) using IPW estimator under dependence between Z and (T, C)

Clayton Copula	% of cens	True value of Δ	Δ under IPW estimator									
			BT		BT_p		CG Clayton		CG Frank		CG Gumbel	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0$ $(\theta = 0)$	20	22.430	3.664	0.909	-23.636	5.218	-	-	-	-	-	-
	30	22.257	3.737	0.989	-22.804	5.486	-	-	-	-	-	-
	40	21.971	3.577	0.979	-22.075	5.325	-	-	-	-	-	-
	50	21.689	2.630	1.023	-21.071	5.788	-	-	-	-	-	-
$\tau = 0.5$ $(\theta = 2)$	20	23.602	4.067	0.865	-27.427	5.604	3.482	0.774	3.443	0.771	2.730	0.759
	30	23.534	3.810	0.920	-28.071	5.759	3.451	0.775	3.395	0.771	2.639	0.764
	40	23.578	3.129	1.001	-28.568	6.236	3.471	0.787	3.390	0.782	2.621	0.771
	50	23.466	2.008	1.020	-28.730	6.503	3.342	0.743	3.245	0.739	2.489	0.732
$\tau = 0.83$ $(\theta = 10)$	20	24.068	4.212	0.868	-15.390	5.567	2.546	0.788	2.473	0.792	2.200	0.824
	30	24.279	4.333	0.942	-13.876	6.236	2.236	0.807	2.132	0.804	1.736	0.829
	40	24.460	4.187	0.982	-14.788	6.142	1.780	0.795	1.637	0.792	1.090	0.821
	50	24.557	0.813	0.977	-48.048	5.716	0.890	0.758	0.708	0.753	0.020	0.792

% of cens: percentage of censoring; SE: standard deviation of estimates across 1000 replicates
When $\tau = 0.5$, θ of Clayton copula=2, , θ of Frank copula=5.736, and , θ of Gumbel copula=2.
When $\tau = 0.83$, θ of Clayton copula=10, , θ of Frank copula=22.224, and , θ of Gumbel copula=6.

Table 10. Simulation scenarios 37-48 results of the incremental effect (Δ) using using generalized survival-adjusted estimator under dependence between Z and (T, C)

Clayton Copula	% of cens	True value of Δ	Δ under generalized survival-adjusted estimator							
			BM		Copula-based Cox regression under Clayton		Copula-based Cox regression under Frank		Copula-based Cox regression under Gumbel	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0$ $(\theta = 0)$	20	22.430	1.490	0.878	-	-	-	-	-	-
	30	22.257	1.138	0.970	-	-	-	-	-	-
	40	21.971	0.767	0.987	-	-	-	-	-	-
	50	21.689	0.207	1.064	-	-	-	-	-	-
$\tau = 0.5$ $(\theta = 2)$	20	23.602	2.609	0.844	1.545	1.393	1.556	1.395	1.575	1.399
	30	23.534	2.405	0.875	1.382	1.441	1.389	1.444	1.404	1.450
	40	23.578	2.267	0.959	1.178	1.523	1.178	1.526	1.185	1.530
	50	23.466	1.945	0.996	0.798	1.644	0.808	1.646	0.816	1.648
$\tau = 0.83$ $(\theta = 10)$	20	24.068	4.751	0.830	1.648	1.390	1.614	1.391	1.596	1.392
	30	24.279	5.643	0.875	1.510	1.547	1.467	1.547	1.446	1.548
	40	24.460	6.490	0.872	1.041	1.662	0.996	1.662	0.977	1.662
	50	24.557	6.622	0.909	0.140	1.805	0.106	1.805	0.094	1.805

% of cens : percentage of censoring; SE: standard deviation of estimates across 1000 replicates
When $\tau = 0.5$, θ of Clayton copula=2, , θ of Frank copula=5.736, and , θ of Gumbel copula=2.
When $\tau = 0.83$, θ of Clayton copula=10, , θ of Frank copula=22.224, and , θ of Gumbel copula=6.

Table 11. Simulation scenarios 1-24 results of the incremental effect (Δ) using IPW and generalized survival-adjusted estimator under independence between Z and (T, C) according to different τ

			Δ											
Clayton Copula	% of cens	True value of Δ	CG Clayton ($\tau = 0.3$)		CG Frank ($\tau = 0.3$)		CG Gumbel ($\tau = 0.3$)		Copula-based Cox reg. under Clayton ($\tau = 0.3$)		Copula-based Cox reg. under Frank ($\tau = 0.3$)		Copula-based Cox reg. under Gumbel ($\tau = 0.3$)	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0.5$ ($\theta = 2$)	20	19.683	5.209	0.626	5.149	0.624	4.344	0.620	-0.807	1.382	-0.718	1.376	-0.338	1.365
	30	19.305	5.367	0.605	5.301	0.603	4.492	0.599	-0.786	1.635	-0.685	1.617	-0.137	1.591
	40	18.837	6.330	0.490	6.257	0.486	5.508	0.481	0.632	1.411	0.731	1.371	1.443	1.360
	50	18.171	6.473	0.431	6.392	0.427	5.705	0.426	2.265	1.285	2.318	1.259	2.961	1.241

			CG Clayton ($\tau = 0.5$)		CG Frank ($\tau = 0.5$)		CG Gumbel ($\tau = 0.5$)		Copula-based Cox reg. under Clayton ($\tau = 0.5$)		Copula-based Cox reg. under Frank ($\tau = 0.5$)		Copula-based Cox reg. under Gumbel ($\tau = 0.5$)	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0.5$ ($\theta = 2$)	20	19.683	5.104	0.627	4.965	0.622	4.099	0.619	-0.519	1.330	-0.412	1.324	-0.240	1.320
	30	19.305	5.273	0.606	5.127	0.599	4.292	0.598	-0.348	1.624	-0.230	1.604	0.001	1.592
	40	18.837	6.190	0.485	6.028	0.477	5.290	0.480	1.201	1.380	1.316	1.334	1.626	1.333
	50	18.171	6.331	0.436	6.159	0.428	5.523	0.435	2.727	1.255	2.791	1.225	3.074	1.217

Clayton Copula	% of cens	True value of Δ	CG Clayton ($\tau = 0.83$)		CG Frank ($\tau = 0.83$)		CG Gumbel ($\tau = 0.83$)		Copula-based Cox reg. under Clayton ($\tau = 0.83$)		Copula-based Cox reg. under Frank ($\tau = 0.83$)		Copula-based Cox reg. under Gumbel ($\tau = 0.83$)	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0.5$ ($\theta = 2$)	20	19.683	4.572	0.612	4.072	0.604	3.421	0.623	-0.171	1.394	-0.143	1.393	-0.134	1.392
	30	19.305	4.657	0.609	4.174	0.601	3.624	0.622	0.068	1.571	0.097	1.568	0.103	1.566
	40	18.837	5.578	0.482	5.157	0.482	4.751	0.506	1.772	1.445	1.800	1.434	1.802	1.433
	50	18.171	5.709	0.405	5.367	0.415	5.088	0.436	3.245	1.248	3.266	1.243	3.270	1.243

Table 12. Simulation scenarios 1-24 results of the incremental effect (Δ) using IPW and generalized survival-adjusted estimator under dependence between Z and (T, C) according to different τ

			Δ											
Clayton Copula	% of cens	True value of Δ	CG Clayton ($\tau = 0.3$)		CG Frank ($\tau = 0.3$)		CG Gumbel ($\tau = 0.3$)		Copula-based Cox reg. under Clayton ($\tau = 0.3$)		Copula-based Cox reg. under Frank ($\tau = 0.3$)		Copula-based Cox reg. under Gumbel ($\tau = 0.3$)	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0.5$ ($\theta = 2$)	20	20.353	4.497	0.656	4.475	0.656	3.771	0.664	-0.577	1.266	-0.581	1.265	-0.643	1.268
	30	20.425	4.281	0.644	4.241	0.643	3.406	0.643	-2.216	1.340	-2.174	1.338	-2.204	1.335
	40	20.528	4.459	0.607	4.407	0.605	3.500	0.600	-2.995	1.463	-2.927	1.458	-2.765	1.451
	50	20.521	4.808	0.563	4.759	0.559	3.796	0.557	-3.379	1.675	-3.335	1.633	-2.867	1.611

			CG Clayton ($\tau = 0.5$)		CG Frank ($\tau = 0.5$)		CG Gumbel ($\tau = 0.5$)		Copula-based Cox reg. under Clayton ($\tau = 0.5$)		Copula-based Cox reg. under Frank ($\tau = 0.5$)		Copula-based Cox reg. under Gumbel ($\tau = 0.5$)	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0.5$ ($\theta = 2$)	20	20.353	4.463	0.632	4.399	0.631	3.533	0.651	-0.669	1.279	-0.674	1.277	-0.694	1.278
	30	20.425	4.233	0.623	4.128	0.618	3.146	0.616	-2.443	1.393	-2.390	1.392	-2.382	1.391
	40	20.528	4.359	0.625	4.233	0.619	3.204	0.617	-2.926	1.349	-2.839	1.345	-2.747	1.341
	50	20.521	4.703	0.603	4.572	0.591	3.528	0.590	-2.953	1.717	-2.886	1.675	-2.678	1.661

Clayton Copula	% of cens	True value of Δ	CG Clayton ($\tau = 0.83$)		CG Frank ($\tau = 0.83$)		CG Gumbel ($\tau = 0.83$)		Copula-based Cox reg. under Clayton ($\tau = 0.83$)		Copula-based Cox reg. under Frank ($\tau = 0.83$)		Copula-based Cox reg. under Gumbel ($\tau = 0.83$)	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0.5$ ($\theta = 2$)	20	20.353	4.197	0.642	3.727	0.648	2.802	0.681	-0.687	1.308	-0.684	1.308	-0.680	1.307
	30	20.425	3.822	0.598	3.257	0.589	2.357	0.614	-2.425	1.382	-2.406	1.382	-2.394	1.382
	40	20.528	3.825	0.559	3.224	0.560	2.386	0.580	-2.781	1.363	-2.754	1.362	-2.740	1.362
	50	20.521	4.088	0.557	3.471	0.551	2.738	0.570	-2.627	1.538	-2.602	1.534	-2.594	1.534

Table 13. Simulation scenarios 25-36 results of the incremental effect (Δ) using IPW and generalized survival-adjusted estimator under independence between Z and (T, C) according to different τ

			Δ											
Clayton Copula	% of cens	True value of Δ	CG Clayton ($\tau = 0.3$)		CG Frank ($\tau = 0.3$)		CG Gumbel ($\tau = 0.3$)		Copula-based Cox reg. under Clayton ($\tau = 0.3$)		Copula-based Cox reg. under Frank ($\tau = 0.3$)		Copula-based Cox reg. under Gumbel ($\tau = 0.3$)	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0.5$ ($\theta = 2$)	20	23.063	3.878	0.780	3.847	0.777	3.232	0.761	0.973	1.363	1.027	1.363	1.234	1.365
	30	22.821	4.413	0.795	4.373	0.792	3.774	0.778	0.790	1.502	0.874	1.502	1.195	1.503
	40	22.102	4.745	0.716	4.697	0.713	4.154	0.703	0.028	1.635	0.155	1.634	0.652	1.633
	50	21.812	5.054	0.685	5.004	0.683	4.499	0.678	-0.129	1.709	0.020	1.705	0.617	1.695

			CG Clayton ($\tau = 0.5$)		CG Frank ($\tau = 0.5$)		CG Gumbel ($\tau = 0.5$)		Copula-based Cox reg. under Clayton ($\tau = 0.5$)		Copula-based Cox reg. under Frank ($\tau = 0.5$)		Copula-based Cox reg. under Gumbel ($\tau = 0.5$)	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0.5$ ($\theta = 2$)	20	23.602	3.833	0.765	3.761	0.761	3.075	0.758	1.043	1.371	1.111	1.372	1.226	1.373
	30	23.534	4.359	0.762	4.269	0.760	3.636	0.756	0.853	1.471	0.960	1.471	1.135	1.472
	40	23.578	4.708	0.736	4.607	0.731	4.061	0.726	0.385	1.617	0.543	1.616	0.798	1.617
	50	23.466	4.981	0.710	4.880	0.706	4.386	0.703	0.185	1.730	0.371	1.727	0.670	1.724

Clayton Copula	% of cens	True value of Δ	CG Clayton ($\tau = 0.83$)		CG Frank ($\tau = 0.83$)		CG Gumbel ($\tau = 0.83$)		Copula-based Cox reg. under Clayton ($\tau = 0.83$)		Copula-based Cox reg. under Frank ($\tau = 0.83$)		Copula-based Cox reg. under Gumbel ($\tau = 0.83$)	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0.5$ ($\theta = 2$)	20	23.602	3.606	0.775	3.335	0.761	2.698	0.770	1.196	1.348	1.224	1.348	1.239	1.349
	30	23.534	4.067	0.738	3.794	0.726	3.286	0.742	1.186	1.448	1.226	1.449	1.247	1.450
	40	23.578	4.330	0.714	4.087	0.707	3.715	0.722	0.607	1.689	0.664	1.689	0.690	1.690
	50	23.466	4.687	0.694	4.465	0.688	4.148	0.702	0.601	1.665	0.664	1.665	0.692	1.665

Table 14. Simulation scenarios 37-48 results of the incremental effect (Δ) using IPW and generalized survival-adjusted estimator under dependence between Z and (T, C) according to different τ

			Δ											
Clayton Copula	% of cens	True value of Δ	CG Clayton ($\tau = 0.3$)		CG Frank ($\tau = 0.3$)		CG Gumbel ($\tau = 0.3$)		Copula-based Cox reg. under Clayton ($\tau = 0.3$)		Copula-based Cox reg. under Frank ($\tau = 0.3$)		Copula-based Cox reg. under Gumbel ($\tau = 0.3$)	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0.5$ ($\theta = 2$)	20	23.602	3.472	0.794	3.457	0.793	2.858	0.782	1.457	1.408	1.466	1.408	1.501	1.413
	30	23.534	3.491	0.765	3.469	0.763	2.807	0.752	1.395	1.376	1.402	1.377	1.425	1.382
	40	23.578	3.484	0.735	3.452	0.733	2.762	0.721	1.275	1.408	1.278	1.411	1.264	1.424
	50	23.466	3.407	0.740	3.364	0.738	2.660	0.722	0.932	1.581	0.941	1.583	0.912	1.586

			CG Clayton ($\tau = 0.5$)		CG Frank ($\tau = 0.5$)		CG Gumbel ($\tau = 0.5$)		Copula-based Cox reg. under Clayton ($\tau = 0.5$)		Copula-based Cox reg. under Frank ($\tau = 0.5$)		Copula-based Cox reg. under Gumbel ($\tau = 0.5$)	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0.5$ ($\theta = 2$)	20	23.602	3.482	0.774	3.443	0.771	2.730	0.759	1.545	1.393	1.556	1.395	1.575	1.399
	30	23.534	3.451	0.775	3.395	0.771	2.639	0.764	1.382	1.441	1.389	1.444	1.404	1.450
	40	23.578	3.471	0.787	3.390	0.782	2.621	0.771	1.178	1.523	1.178	1.526	1.185	1.530
	50	23.466	3.342	0.743	3.245	0.739	2.489	0.732	0.798	1.644	0.808	1.646	0.816	1.648

Clayton Copula	% of cens	True value of Δ	CG Clayton ($\tau = 0.83$)		CG Frank ($\tau = 0.83$)		CG Gumbel ($\tau = 0.83$)		Copula-based Cox reg. under Clayton ($\tau = 0.83$)		Copula-based Cox reg. under Frank ($\tau = 0.83$)		Copula-based Cox reg. under Gumbel ($\tau = 0.83$)	
			Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$\tau = 0.5$ ($\theta = 2$)	20	23.602	3.300	0.769	3.058	0.751	2.226	0.757	1.530	1.367	1.535	1.367	1.538	1.368
	30	23.534	3.262	0.736	2.975	0.720	2.203	0.731	1.389	1.395	1.394	1.397	1.399	1.398
	40	23.578	3.216	0.723	2.893	0.711	2.204	0.725	1.203	1.532	1.211	1.534	1.220	1.534
	50	23.466	3.019	0.734	2.689	0.723	2.048	0.738	0.793	1.630	0.808	1.632	0.823	1.633

7. Application

7.1 National Health Insurance Service National Sample Cohort (NHIS-NSC) data

In this session, we used real data examples to demonstrate the performance of the proposed method. Analysis was performed using the National Health Insurance Service National Sample Cohort (NHIS-NSC) database. The NHISNSC (2002-2010) database is a cohort data that connects the same subjects until 2010 by sampling about 1 million people, 2% of the total population, as of 2002. The NHIS records have garnered academic interest due to the effectiveness of the system and relevance to public health and medical research. To meet this interest, a population database has been developed, the 'National Health Information Database' (NHID) containing personal information, demographics, and medical treatment data for Korean citizens, who were categorized as insured employees, insured self-employed individuals or medical aid beneficiaries. The NHID was generated using participants' medical bill expenses claimed by medical service providers.

Data were rearranged according to date of medical treatment rather than date of claim. To prevent the effects of other existing diseases, the period 2002–2003 was designated as a washout period. In addition, to identify newly diagnosed lung cancer cases in 2004, those who were diagnosed with lung cancer in 2002–2003 were excluded. After exclusion (International Classification of Diseases 10th revision codes: C34), the total population of this study was 528 individuals. This section aimed to analyze gender differences in the medical costs associated with lung cancer disease within 5 years after diagnosis (2004) in

the South Korean population. We apply both prevailing estimation method including IPW, generalized survival-adjusted estimator, and also our proposed estimator to this data. We used bootstrapping for the computation of SE of median and non-parametric bootstrap confidence interval for median.

7.2 Results

The analysis results of real data are summarized in Table 16 and Figure 7. The censoring percentage is approximately 40%. Also, we use the tree type of copula function and $\tau = 0.5$. We compared the 5-years censored medical costs estimated by prevailing method and our proposed estimator. We find that compared with our estimator under IPW scheme, the BT and BT_p estimates for the 5-year medical costs produces lower estimates of mean costs for all gender. In generalized survival-adjusted method, our estimates of the incremental costs between gender are almost two times in magnitude that those of the BM estimator. Overall, it can be seen that generalized survival-adjusted method estimates the cost higher than the IPW method. In the IPW scheme, there is not much difference in the estimated value depending on the type of copula function. However, in generalized survival-adjusted method, it is found that there is a large difference in estimates depending on the type of copula function, and that the estimate is the smallest in Clayton and the largest in Gumbel function.

Figure 7 summarized the results about mean cost profiles by gender predicted using prevailing method and proposed estimator. Regardless of gender, The IPW method

estimates the mean cost lower than the simple mean, and the generalized survival-adjusted method estimates higher. Within the same scheme, the estimated mean cost value of the proposed method is higher than that of prevailing methods. Looking at the results of copula function in both schemes, the estimated mean value is high in the order of Clayton, Frank, and Gumbel. In addition, in male, there is a large difference in estimated mean value depending on the methods used for estimation, while in female, there is little difference depending on the copula method within IPW or within the generalized survival-adjusted method. Among the prevailing methods, BM is estimated to be slightly higher only in male than the value obtained by simple means, and almost similar in female.

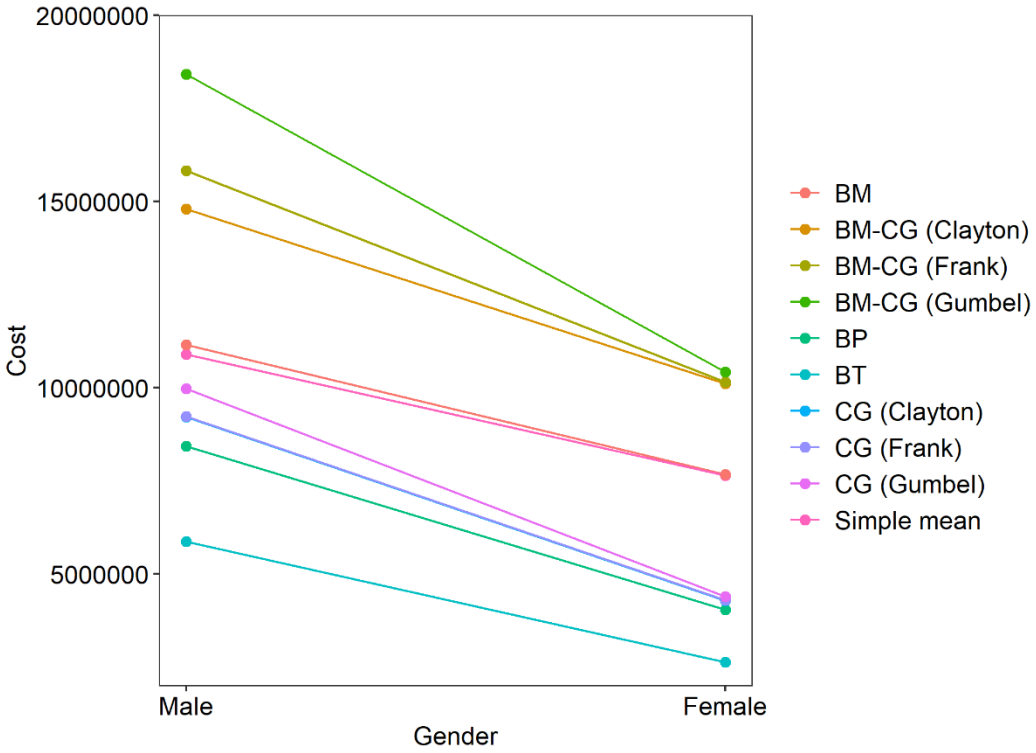


Figure 7. Mean cost profiles by gender predicted using prevailing method and proposed estimator

Table 15. Comparison of estimated 5-year difference costs by gender

	Prevailing estimator			Proposed estimator					
	BT	BT_p	BM	CG Clayton ($\tau = 0.5$)	CG Frank ($\tau = 0.5$)	CG Gumbel ($\tau = 0.5$)	Cox reg. under Clayton ($\tau = 0.5$)	Cox reg. under Frank ($\tau = 0.5$)	Cox reg. under Gumbel ($\tau = 0.5$)
Male	5867805	8433645	11162712	9224106	9228097	9982585	14806947	15825979	18417896
	5992956 (4396459, 7102816)	8539487 (6653472, 9942999)	11180358 (9800741, 12414497)	9452226 (6911917, 11108349)	9403784 (7013897, 10999194)	10111471 (7576327, 11858258)	15470449 (11729269, 18786894)	16612162 (11809259, 20522707)	19016224 (12217231, 24776160)
Female	2633623	4040209	7673656	4283604	4302063	4394687	10108290	10148977	10417704
	2650223 (1541813, 3641067)	4065847 (2788864, 5244381)	7713941 (6128939, 9118899)	4219184 (2849337, 5512662)	4236564 (2854903, 5542561)	4302418 (2878411, 5660696)	10111007 (7930769, 12086961)	10158370 (7964339, 12160822)	10443635 (8107332, 12527803)
Δ	3234182	4393436	3489056	4940502	4926034	5587898	4698657	5677002	8000192
	3270505 (1383101, 4937956)	4431180 (2218429, 6423156)	3468239 (1527922, 5505879)	5085377 (2245649, 7417462)	5039567 (2317969, 7334838)	5683354 (2804836, 8129725)	5215300 (934484, 9307677)	6293935 (981612, 10997095)	8450981 (1106202, 15008831)

Note: Values are expressed as estimated mean value and median (95% CI)

8. Conclusion and Discussion

This study examined how to extend the estimation of censored medical cost in dependent censoring data. There are the prevailing methods for estimating the right censored medical cost which fall into three categories, (a) the IPW estimators; (b) the generalized survival-adjusted estimators; (c) the joint-modeling methods. However, the prevailing methods were established under independent censoring. However, the medical costs for failure event and censoring time tend to be generally positively correlated. Using the prevailing methods for calculation mean medical cost provide biased results under the assumption of dependent censoring. The proposed estimators using copula method can reduce bias in inferences and return better results than the prevailing method in data under dependent censoring.

Our simulation study revealed that the proposed method was either comparable or superior to the prevailing method in most scenarios. Especially, the proposed method using IPW method reduced bias and S.E. under dependence between Z and (T, C). And the proposed method using generalized survival-adjusted method certainly reduced bias under dependence between Z and (T, C) and all censoring scenarios. Also, we find that prevailing methods showed a significant increase in bias under dependent censoring, and BM method showed a smaller bias than the IPW method regardless of the dependency between survival and censoring time.

This study's simulations also examined the effects of the type of copula function and dependency parameter θ. In IPW method using copula graphic estimator, we should

choose the high θ regardless of copula's function to have the smallest bias. However, in generalized survival-adjusted method using copula-based Cox regression, we should choose the high θ under 20-30% censoring rate and the low θ under 40-50% censoring rate regardless of the type of copula function. When the expected mean value of survival time is similar to limitation of study time, we should choose the low θ in Clayton and Frank function under 20-50% censoring rate under independence between Z and (T, C).

We confirmed through real data example that the value differs significantly between the estimator considering the dependency of the survival and the censoring time and the estimator that does not. Therefore, if dependent censoring exists, the censored medical cost should be calculated using the proposed method considering it.

As a result, we confirm the performance of the proposed method compared to the prevailing estimator based on IPW and generalized survival-adjusted method. The proposed estimation method showed good performance in most cases, especially in the context of dependence between Z and (T, C). This method is expected to be a useful tool to estimation aspect to bias censored medical cost on dependent censoring data. Further works should be carried out on study results to confirm the generalizability of these results. There are further works, (1) when the censoring percentage in the IPW scheme is 50% and the mean of survival time is similar to study limitation time, the bias of estimators rapidly decreases, (2) survival and censoring time were generated with a clayton copula, but rather, the bias of the IPW estimator using the Gumbel function was smaller, and (3) Result of incorrect assumption of copula function and dependency parameter θ.

References

Bang, H., & Tsiatis, A. A. (2000). Estimating medical costs with censored data. Biometrika, 87(2), 329-343.

Basu, A., & Manning, W. G. (2010). Estimating lifetime or episode-of-illness costs under censoring. Health economics, 19(9), 1010-1028.

Braekers R, Veraverbeke N (2005). A copula-graphic estimator for the conditional survivalfunction under dependent censoring. Can J Stat 33:429–447.

Chen YH (2010). Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. J R Stat Soc Ser B Stat Methodol, 72:235–251.

Clayton DG (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65 (1):141–151.

de Uña-Álvarez, J., & Veraverbeke, N. (2013). Generalized copula-graphic estimator. Test, 22, 343-360.

Deng, L., Lou, W., & Mitsakakis, N. (2019). Modeling right-censored medical cost data in regression and the effects of covariates. Statistical Methods & Applications, 28, 143-155.

Emura, T., & Chen, Y. H. (2018). Analysis of survival data with dependent censoring: copula-based approaches. Singapore: Springer.

Etzioni, R. D., Feuer, E. J., Sullivan, S. D., Lin, D., Hu, C., & Ramsey, S. D. (1999). On the use of survival analysis techniques to estimate medical care costs. Journal of health economics, 18(3), 365-380.

Frank MJ (1979). On the simultaneous associativity of f(x, y) and x + y - f(x, y). Aequationes Matbematicae, 19:194–226.

Gumbel EJ (I960). Distributions de valeurs extremes en plusieurs dimensions. PubL Inst Statist. Parids 9: 171–173.

Heitjan DF, Kim CY, Li H (2004). Bayesian estimation of cost-effectiveness from censored data. Stat Med 23(8):1297–1309.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American statistical Association, 47(260), 663-685.

Joe H (1993). Parametric families of multivariate distributions with given margins. J Multivar Anal, 46:262–282.

Korea Health Panel Study [Internet]. [cited June, 06, 2023]. https://www.khp.re.kr:444/eng/main.do.

Lin, D. Y., Feuer, E. J., Etzioni, R., & Wax, Y. (1997). Estimating medical costs from incomplete follow-up data. Biometrics, 419-434.

Lin, D. Y. (2000). Linear regression analysis of censored medical costs. Biostatistics, 1(1), 35-47.

Morgenstern D (1956). Einfache Beispiele zweidimensionaler Verteilungen. Mitteilungsblatt für Mathematishe Statistik, 8:234–235

Nelsen RB (2006). An introduction to copulas, 2nd edn. Springer, New York.

Rivest LP and Wells MT (2001). A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. J Multivar Anal 79:138–155.

Tovar Cuevas, J. R., Portilla Yela, J., & Achcar, J. A. (2019). A method to select bivariate copula functions. Revista Colombiana de Estadística, 42(1), 61-80.

Zheng, M., & Klein, J. P. (1994). A self-consistent estimator of marginal survival functions based on dependent competing risk data and an assumed copula. Communications in Statistics-Theory and Methods, 23(8), 2299-2311.

Zheng M, Klein JP (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. Biometrika 82(1):127–138.

국 문 요 약

의존적 중도절단 데이터에서의
코퓰라 방법을 이용한 의료비 추정

중도절단 된 의료비를 추정하기 위하여 기존에 제안된 모델링 방법은 크게 세 가지 범주로 나뉜다: (1) 역확률 가중치 추정, (2) 일반화된 생존 확률 조정법, (3) 조인트 모델링이다. 그러나 이 방법들은 독립 중도절단 가정하에 성립된 모델들이기 때문에 이러한 가정이 위반되게 되면 그 추정량들은 편향을 발생시키게 된다.

하지만 일반적으로 의료비 추정 시 관심 사건과 중도절단 시간은 양의 상관관계를 가진다 (Etzioni et al. 1999; Lin 2003). 따라서 의존적 중도절단 하에서는 생존 시간과 중도절단 시간의 결합확률분포를 고려하여 의료비를 추정할 필요가 있다. 이 문제를 고려하여 의료비를 모델링하기에 코퓰라 방법은 아주 유용한 도구이다. 의존적 중도절단 데이터의 생존 분석을 하기 위해 이 연구에서는 가정된 코퓰라 방법 중 코퓰라-그래픽 추정량과 코퓰라에 근거한 콕스 회귀를 적용하였다.

이 연구에서는 독립적 중도절단과 의존적 중도절단 데이터에서 기존 제안된 추정 방법들과 이 연구에서 제안한 추정 방법들을 평가하였다. 일련의 시뮬레이션과 다양한 시나리오를 가정하여 추정 방법들의 편향과 표준 오차를 통하여 성능을

평가하였다. 또한 국민 건강 보험의 국가 표본 코호트를 사용하여 실제 데이터에 적용해 보았다.

핵심되는 말: 의료비, 의존적 중도절단, 코플라 방법, 코플라-그래픽 추정량, 코플라에 근거한 콕스 회귀, 역할률 가중치 추정, 일반화된 생존 확률 조정 방법, 조인트 모델링