Page Proof Instructions and Queries

Journal Title: Statistical Methods in Medical Research
Article Number: 1220496

Thank you for choosing to publish with us. This is your final opportunity to ensure your article will be accurate at publication. Please review your proof carefully and respond to the queries using the circled tools in the image below, which are available in Adobe Reader DC* by clicking **Tools** from the top menu, then clicking **Comment**.

Please use *only* the tools circled in the image, as edits via other tools/methods can be lost during file conversion. For comments, questions, or formatting requests, please use T. Please do *not* use comment bubbles/sticky notes.

*If you do not see these tools, please ensure you have opened this file with **Adobe Reader DC**, available for free at get.adobe.com/reader or by going to Help > Check for Updates within other versions of Reader. For more detailed instructions, please see us.sagepub.com/ReaderXProofs.

No.	Query
GQ1	Please confirm that all author information, including names, affiliations, sequence, and contact details, is correct.
GQ2	Please review the entire document for typographical errors, mathematical errors, and any other necessary corrections; check headings, tables, and figures.
GQ3	Please confirm that the Funding and Conflict of Interest statements are accurate.
GQ4	Please ensure that you have obtained and enclosed all necessary permissions for the reproduction of artistic works, (e.g. illustrations, photographs, charts, maps, other visual material, etc.) not owned by yourself. Please refer to your publishing agreement for further information.
GQ5	Please note that this proof represents your final opportunity to review your article prior to publication, so please do send all of your changes now.
GQ6	Please note, only ORCID iDs validated prior to acceptance will be authorized for publication; we are unable to add or amend ORCID iDs at this stage.
AQ1	Please provide missing city for Refs. 26 and 28.

Original Research Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
1–20

DOI: 10.1177/09622802231220496
journals.sagepub.com/home/smm
Sage

Time-dependent receiver operating characteristic curve estimator for correlated right-censored time-to-event data

Kassu M Beyene[1] and **Ding-Geng Chen**[1,2]

GQ1

Abstract

In clinical trials, evaluating the accuracy of risk scores (markers) derived from prognostic models for prediction of survival outcomes is of major concern. The time-dependent receiver operating characteristic curve and the corresponding area under the receiver operating characteristic curve are appealing measures to evaluate the predictive accuracy. Several estimation methods have been proposed in the context of classical right-censored data which assumes the event time of individuals are independent. In many applications, however, this may not hold true if, for example, individuals belong to clusters or experience recurrent events. Estimates may be biased if this correlated nature is not taken into account. This paper is then aimed to fill this knowledge gap to introduce a time-dependent receiver operating characteristic curve and the corresponding area under the receiver operating characteristic curve estimation method for right-censored data that take the correlated nature into account. In the proposed method, the unknown status of censored subjects is imputed using conditional survival functions given the marker and frailty of the subjects. An extensive simulation study is conducted to evaluate and demonstrate the finite sample performance of the proposed method. Finally, the proposed method is illustrated using two real-world examples of lung cancer and kidney disease.

GQ2 GQ4 GQ5

Keywords

Correlated, predictive accuracy, right-censored, receiver operating characteristic curve, time-to-event

1 Introduction

The time-to-event data arises when an interest lies in the time from a certain origin to the occurrence of a specific event of interest. As an example, time-to-event is often used as an endpoint in clinical trials, refers to the time between randomization and the occurrence of an event of interest. The event, for example, can be death due to certain disease, progression, treatment failure or the recurrences of a disease. In such follow-up studies, the event of interest is not necessarily experienced by all study participants at the end of the study, so the actual event times for some subjects are unknown. This loss of information on time-to-event is known as censoring, which may occur when a subject withdraws from the study, lost to follow-up, or the study ends before the event has occurred. Survival analysis, or more generally, time-to-event analysis, is a standard tool to analyze the event time data taking the unique censoring feature into account.

In time-to-event analysis, prognostic model, which is a class of clinical prediction models, is an important tool to evaluate the association between one or more risk factors (biomarkers) and the outcomes of interest (time-to-event) and to predict the risk of an individual developing a particular state of health or experiencing a future outcome. This plays an increasingly

[1]Arizona State University, College of Health Solutions, AZ, USA
[2]Department of Statistics, University of Pretoria, Pretoria, South Africa

Corresponding author:
Kassu M Beyene, Arizona State University, College of Health Solutions, AZ, USA.
Email: kbeyene1@asu.edu

important role in assisting health professionals in making clinical decisions and improving health outcomes of patients. To handle censored time-to-event data, regression models have been developed in the field of survival analysis. The standard or most widely used regression model to analyze time-to-event data is the proportional hazards model, which was introduced by Cox in 1972 and is commonly known as the Cox regression model.[1] The risk score (marker) derived from these models should, however, be evaluated for its accuracy before being used to predict the future outcome that is important for better clinical decision. In this regard, receiver operating characteristic (ROC) curve and the associated area under the ROC curve (AUC) are commonly used to assess predictive accuracy. Unlike the traditional ROC curve analyses which typically assume that event status of a subject is fixed and known, in prognostic studies such as time-to-event analysis, the disease status of subject can change over time. In such situations, the event of interest is time-dependent and is defined by considering a specific fixed time point of interest. This gives rise to concepts like time-dependent sensitivity, time-dependent specificity, and consequently, the time-dependent ROC and AUC. Several researchers, such as Heagerty et al.,[2] Heagerty and Zheng,[3] Etzioni et al.,[4] and others, have proposed various definitions and extensions to adapt classical methodologies for handling time-to-event data in prognostic studies. In literature, taking various censoring mechanisms into account, several time-dependent ROC curve and AUC estimation methods have been proposed; see, for example, Heagerty et al.,[2] Heagerty and Zheng,[3] Li et al.,[5] Blanche et al.,[6] Martìnez-Camblor et al.,[7] Martìnez-Camblor and Pardo- Fernández,[8] Beyene et al.,[9] Díaz-Coto,[10] Wu and Cook,[11] Beyene and El Ghouch[12,13] and the references given in these papers.

One of the basic assumptions common to the aforementioned methods is that the survival times of different subjects are independent of each other given observed values of covariates. In many practical applications, this assumption, however, may not hold true since all relevant covariates cannot be observed. In many clinical trials, for example, event times of different subjects maybe clustered or correlated because of certain common features such as genetic traits or shared environmental factors, or repeated events. In this situation, there are many unobserved characteristics that seem to be shared between observations from a cluster, so event times of different subjects are presumed to be correlated if they came from the same cluster. When survival analysis is performed without taking the correlated nature of the data into account, parameter estimates and their standard errors will be incorrect.[14,15] For correlated time-to-event data analysis, a number of models widely referred to as frailty models have been proposed, see, for example,[14,16–21] and frailty models were extensively studied, for example, in.[22–26] In spite of the fact that these methods have been active area of research for the past several decades, and many applications have been published, almost no literature has been published about evaluating the predictive ability of these models, specifically the time-dependent ROC curve and its corresponding AUC estimation method that considers the correlated nature of the data.

In this paper, we propose a new time-dependent ROC curve and time-dependent AUC estimation methods for right-censored time-to-event data taking the correlated nature into account. In this regard, the proposed method is a generalization of the time-dependent ROC curve introduced by Beyene and El Ghouch[12] which assumed that individual event times are independent. As in Beyene and El Ghouch,[12] the unknown event status of censored individuals is imputed with conditional survival function, the conditionality in our approach, however, is both on the marker and frailty of the subjects. This conditional survival function can be estimated using a frailty model, such as, a parametric model.

The rest of this article is organized as follows. The following section introduces some important notations and definitions and describes the proposed estimators for time-dependent ROC curve and its associated AUC in the presence of correlation. In the "Simulation" section, the finite sample performances of the proposed method are evaluated through a simulation study. The practical use of the proposed method is illustrated with a real-data application in the "Real data analysis" section. Finally, some remarks and discussions are presented in the "Discussion" section.

2 Methods

In this section, we first introduce some important notations and definitions, followed by the estimator for the time-dependent ROC curve and its associated summary measure.

2.1 Notations and definitions

Let us consider that there are $i = 1, 2, \ldots, B$ different clusters (centers, families, litters, etc.) in a clustered time-to-event data set, where each cluster have $j = 1, 2, \ldots, n_i$ subjects resulting in total of $n = \sum_{i=1}^{B} n_i$ observations. Let T_{ij} be a non-negative survival time corresponding to jth individual from the ith cluster and let C_{ij} be the non-informative right censoring time. Assuming that the jth individual in the ith cluster is observed until either a survival time T_{ij} or a right censoring time C_{ij}, the observed time (Y_{ij}) is defined as $Y_{ij} = \min(T_{ij}, C_{ij})$. Let $\Delta_{ij} = I(T_{ij} \leq C_{ij})$ be the censoring indicator variable, where $I(.)$ is the indicator function, and X_{ij} be the vector of covariates. Let ν_i denote frailties that are independent and identically distributed random sample from a density $f_V(\nu; \alpha)$ with support $[0, \infty]$, where α is a parameter that control the

shape of the frailty distribution. To avoid model identifiability issues, the distribution is assumed to have mean 1 and some unknown finite variance α.[27] In the literature, several distributions have been proposed for the frailty distribution, including the gamma, lognormal, and positive stable distributions.[28] However, due to its mathematical and computational advantages, the gamma distribution is the most commonly used distribution.

A frailty model generalizes the classical survival model by allowing within-cluster correlations. According to this model, observations share a common frailty, resulting in correlated outcomes within the group. Assuming a proportional hazards frailty model, the hazard can be written as

$$\lambda_{ij}(t|\nu_i, X_{ij}) = \lambda_0(t)\nu_i \exp(\beta X_{ij}) \tag{1}$$

where $\lambda_0(.)$ is a baseline hazard function, and β is the vector of unknown regression coefficient associated to the covariate vector X_{ij}. From this, the survival function can be given by

$$S_{ij}(t|\nu_i, X_{ij}) = \exp\big[-\Lambda_0(t)\nu_i \exp(\beta X_{ij})\big] = S_0(t)^{\nu_i \exp(\beta X_{ij})} \tag{2}$$

where $\Lambda_0(t) = \int \lambda_0(s)ds$, and $S_0(t) = \exp(-\Lambda_0(t))$. In the following, for notational simplicity, we suppress the subscript that used to indicate clustering or repetition.

Suppose we have a quantitative (bio)marker, denoted by M, and assume that larger marker values are associated with higher risk of getting the event. The (bio)marker, also known as the risk score, can take two forms: A simple marker that considers a single risk factor, or a composite marker that combines more than one risk factors. Despite the fact that simple markers can offer a straightforward way to predict risk, composite markers often provide a more accurate prediction of individual's risk to experience the event of interest. For a given prediction time t, the event status of a subject is defined as $D_t = I(T \leq t)$, with 1 indicating the presence of the event before t and 0 otherwise. Using these, to define the time-dependent ROC curve, we first define the time-dependent true positive rate (TPR_t) and false positive rate (FPR_t), respectively, as follows

$$\begin{aligned} \text{TPR}_t(m) &= P(M > m|D_t = 1) \\ \text{FPR}_t(m) &= P(M > m|D_t = 0) \end{aligned} \tag{3}$$

where $m \in \mathbb{R}$ is a fixed cutoff value.

The time-dependent ROC curve (ROC_t) can be obtained by plotting FPR_t versus TPR_t for all possible values of m. Alternatively, as a function, it can be written as

$$\text{ROC}_t(u) = \text{TPR}_t(\text{FPR}_t^{-1}(u)) \tag{4}$$

where $u \in [0, 1]$ and $\text{FPR}_t^{-1}(u) = m$ is a threshold value such that $u = \text{FPR}_t(m)$. Quantitatively, the time-dependent ROC curve can be summarized using are the area under the ROC curve (AUC_t), defined as

$$\text{AUC}_t = \int_0^1 \text{ROC}_t(s)ds \tag{5}$$

Generally, AUC_t values range from 0 to 1, with a higher value indicating a better predictive performance, meaning it better differentiates between subjects with and without the event of interest.

2.2 Estimators

In this section, we first derive theoretical formulas for the time-dependent TPR, FPR and the time-dependent ROC curve defined above which is the basis for our proposal. Assume that the event time T and then censoring time C are conditionally

independent given the marker M and the frailty ν, with some simple algebra, the time-dependent TPR given by (3) can be written as

$$\begin{aligned}\mathrm{TPR}_t(m) = P(M > m|D_t = 1) &= \frac{E(I(M > m, T < t))}{E(I(T < t))} \\ &= \frac{E\{I(M > m)E(I(T < t)|\Delta, Y, M, \nu)\}}{E\{E(I(T < t)|\Delta, Y, M, \nu)\}} \\ &= \frac{E\{I(M > m)P(T < t|\Delta, Y, M, \nu)\}}{E\{P(T < t|\Delta, Y, M, \nu)\}} \\ &= \frac{E\{I(M > m)W_t(\Delta, Y, M, \nu)\}}{E\{W_t(\Delta, Y, M, \nu)\}}\end{aligned} \tag{6}$$

Similarly, the FPR can be given by

$$\begin{aligned}\mathrm{FPR}_t(m) &= \frac{E\{I(M > m)E(I(T > t)|\Delta, Y, M, \nu)\}}{E\{E(I(T > t)|\Delta, Y, M, \nu)\}} \\ &= \frac{E\{I(M > m)P(T > t|\Delta, Y, M, \nu)\}}{E\{P(T > t|\Delta, Y, M, \nu)\}} \\ &= \frac{E\{I(M > m)(1 - W_t(\Delta, Y, M, \nu))\}}{E\{(1 - W_t(\Delta, Y, M, \nu))\}}\end{aligned} \tag{7}$$

where $W_t(\Delta, Y, M, \nu) = P(T < t|\Delta, Y, M, \nu)$ is a conditional distribution function, and $I(.)$ is indicator function. The $W_t(\Delta, Y, M, \nu)$ is nothing but the event status of an individual which is unknown when it is censored and has the observed time Y value less than the prediction time t. In this case, it is imputed with the ratio $\frac{S(t|M,\nu)}{S(Y|M,\nu)}$. In this sense, the intuition of this approach is similar to the idea of imputation in the context of missing data. The conditional distribution $W_t(\Delta, Y, M, \nu)$ is equal to D_t when the subject is uncensored (the event observed), i.e. when $T = Y$, and $P(T \leq t|T > Y, M, \nu) = \left[1 - \frac{S(t|M,\nu)}{S(Y|M,\nu)}\right] I(Y \leq t)$ when the subject is censored. Thus, the random variable $W_t(\Delta, Y, M, \nu)$ denoted by W_t can be written as

$$W_t = \left[1 - (1 - \Delta)\frac{S(t|M, \nu)}{S(Y|M, \nu)}\right] I(Y \leq t) \tag{8}$$

where $S(.|M, \nu)$is the conditional survival probability of survival time T given marker M and frailty ν. The interesting point here is that when the study subjects are from homogeneous populations with unit frailty value, meaning that all individuals have the same vulnerability to adverse outcomes, the conditional survival function $S(.|M, \nu)$ reduces to $S(.|M)$. Thus, the event status W_t become $\left[1 - (1 - \Delta)\frac{S(t|M)}{S(Y|M)}\right] I(Y \leq t)$ which coincides with the formula given in.[5,7,9,12]

Using the same approach, the theoretical formula for the time-dependent ROC curve given in equation (4) can be written as

$$\begin{aligned}\mathrm{ROC}_t(u) = P(\mathrm{FPR}_t(M) \leq u|D_t = 1) &= \frac{P(Z_t \leq u, T \leq t)}{P(T \leq t)} \\ &= \frac{E(I(Z_t \leq u, T \leq t))}{E(I(T \leq t))} = \frac{E\{E(I(Z_t \leq u, T \leq t)|Y, \Delta, M, \nu)\}}{E\{E(I(T \leq t)|Y, \Delta, M, \nu)\}} \\ &= \frac{E\{I(Z_t \leq u)P(T \leq t|Y, \Delta, M, \nu)\}}{E\{P(T \leq t|Y, \Delta, M, \nu)\}} = \frac{E\{I(Z_t \leq u)W_t\}}{E\{W_t\}} \\ &= E\{I(Z_t \leq u)\mathcal{W}_t\}\end{aligned} \tag{9}$$

where $\mathcal{W}_t$ denotes $W_t/E(W_t)$ and Z_t denotes the FPR evaluated at M, i.e. $\mathrm{FPR}_t(M)$.

The empirical estimator for the TPR (6) and FPR (7), respectively, can be obtained as follows

$$\begin{aligned}\widehat{\mathrm{TPR}}_t(m) &= \frac{E\{\widehat{I(M > m)}W_t\}}{\widehat{E\{W_t\}}} \\ &= \frac{\sum_{j=1}^{n} I(M_j > m)\hat{W}_{tj}}{\sum_{j=1}^{n} \hat{W}_{tj}}\end{aligned} \tag{10}$$

$$\widehat{\text{FPR}}_t(m) = \frac{E\{I(M > \widehat{m})(1 - W_t)\}}{E(\widehat{1 - W_t})} = \frac{\sum_{j=1}^{n} I(M_j > m)(1 - \hat{W}_{tj})}{\sum_{j=1}^{n}(1 - \hat{W}_{tj})} \tag{11}$$

where $\hat{W}_t$ is a consistent estimator of $W_t = [1 - (1 - \Delta)\frac{S(t|M,\nu)}{S(Y|M,\nu)}]I(Y \leq t)$. To estimate W_t, we need first to estimate the unknown conditional survival function $S(.|M, \nu)$. The maximum likelihood estimator of the quantity $S(.|M, \nu)$ can be done by fitting the frailty model given in (2). This has been implemented in many R packages including the *frailtypack*.[29]

The empirical estimator for ROC_t given in equation (9) is

$$\widehat{\text{ROC}}_{emp}(u) = \frac{n^{-1}\sum_{j=1}^{n} \hat{W}_{tj} I(\hat{Z}_{tj} \leq u)}{n^{-1}\sum_{j=1}^{n} \hat{W}_{tj}} = n^{-1}\sum_{j} \hat{\mathcal{W}}_{tj} I(\hat{Z}_{tj} \leq u) \tag{12}$$

where $\hat{\mathcal{W}}_{tj} = \hat{W}_{tj}/n^{-1}\sum_{j=1}^{n} \hat{W}_{tj}$ and $\hat{Z}_{tj} = \widehat{\text{FPR}}_t(M_j)$. This empirical time-dependent ROC estimator can be smoothed by replacing the indicator function $I(.)$ in equation (12) with $K_b(.)$, where b is a smoothing parameter and $K(x) = \int_0^1 k(s)ds$ is a kernel distribution function with a density k. As demonstrated in many studies, see for example,[8,12,13,30] the smooth ROC curve estimators tend to exhibit smaller mean integrated squared errors MISEs as compared to the empirical estimators. Following the smoothing ROC estimator of the uncorrelated (classical) right-censored time-to-event data introduced by Beyene and El Ghouch,[12] the empirical time-dependent ROC curve estimator in equation (12) can be smoothed as $\widehat{\text{ROC}}_{t,b}(u) = n^{-1}\sum_j \hat{\mathcal{W}}_{tj} K((Q(u) - Q(\hat{Z}_{tj}))/b)$, where b is a smoothing parameter and Q is a quantile transformation function introduced to overcome the boundary problem that arise due to the fact that $\hat{Z}_t$ has bounded support in unit interval $[0, 1]$. To choose the smoothing bandwidth parameter b, the authors proposed the normal reference, plug-in and cross-validation methods. For more details, the reader is referred to Beyene and El Ghouch.[12]

The estimate of the time-dependent AUC (5) can be obtained by approximating the integral using a numerical integration method, i.e. $\widehat{\text{AUC}}_t = \int_0^1 \widehat{\text{ROC}}_t(v)dv$. Alternatively, an analytic expression of the time-dependent AUC can be easily derived as follows

$$\widehat{\text{AUC}}_t = 1 - n^{-1}\sum_{j=1}^{n} \hat{\mathcal{W}}_{tj}\hat{Z}_{tj} \tag{13}$$

3 Simulation

In this section, we conduct an extensive simulation study with various scenarios to investigate the finite sample performance of the proposed time-dependent ROC curve and the time-dependent AUC. In addition, the performance of the proposed methods will be compared with the ROC curve and AUC estimators proposed by Beyene and El Ghouch,[12] hereafter referred by Beran, that was developed for classical uncorrelated survival data. The Beran method is implemented in R package cenROC. To this end, we first present the data-generating process, followed by the simulation results.

3.1 Data generating process

In order to perform the simulations, we first need to generate the data. To this end, the following procedure is used in order to generate survival times from the frailty model. The failure times are assumed to be independent and follow a proportional hazards model, given the frailty. The survival times were generated from a frailty model given by

$$S(t|X, \nu) = exp(-\Lambda_0(t)\nu\exp(\beta X))$$

where $\Lambda_0(.)$ is the cumulative baseline hazard function, ν is a random frailty term, X is a covariate, and β is an unknown regression parameter. From this, following the cumulative hazard inversion method of generating survival time of the classical Cox model introduced by Bender et al (2005), the survival time T can be expressed as

$$T = \Lambda_0^{-1}\left\{\frac{-\log(U)}{\nu\exp(\beta X)}\right\}$$

where U is a uniform distributed random variable that generated from $Unif[0,1]$, and $\Lambda_0^{-1}(.)$ is an inverted cumulative hazard function. To generate the event time T, we considered the following two distributional assumptions for the baseline function.

Scenario I: The baseline hazard function is assumed to be Weibull distributed, i.e. $\Lambda_0(t) = (t/\lambda)^\eta$, with $\lambda = 2$ and $\eta = 2$ are the scale and shape parameters, respectively. From this, the survival time is given by $T = \lambda^{-1}[-\log(U)/(\nu\exp(\beta X))]^\eta$. The censoring time ($C$) is generated from a Weibull distribution with shape parameter $\eta = 2$ and scale parameter λ_C, where the value of λ_C is determined to achieve the desired censoring proportions of 20%, 40%, and 60%. The (bio)marker is defined as $M = -\sqrt{\rho}T - \sqrt{1-\rho}\epsilon$, where ρ is a correlation between event time and the marker and ϵ is a Weibull distributed random variable with shape = 2 and scale = 2. In this simulation, we generated the marker in this way because we wanted to make it similar to that inthe work of Beyene and El Ghouch,[12] the paper against which our method is compared. A similar approach is also used in other studies, such as the work of Martínez-Camblor and Pardo-Fernández.[8]

Scenario II: Considering a log-normal distribution with $\mu = 1.65$ and $\sigma = 1$ for the baseline function, the event time is generated as $T = \exp(\sigma * \text{qnorm}(1 - \exp(\log(U)/\nu\exp(\beta X))) + \mu)$, where qnorm is an inverse of the standard normal distribution function. The censoring time (C) is generated from a log-normal distribution with mean μ_C and standard deviation $\sigma = 1$, where the value of μ_C is selected to achieve 20%, 40%, and 60% censoring proportion. The marker is given as $M = -\sqrt{\rho}T - \sqrt{1-\rho}\epsilon$, where ϵ is an independent log-normal distributed random variable with $\mu = 1.65$ and $\sigma = 1$.

In our simulation, the covariate X is assumed to follow a standard normal distribution and the regression parameter associated with the covariate is $\beta = -0.75$. For identifiability reasons, the random frailty term ν is assumed to follow one-parameter gamma distribution with mean 1 and variance α, i.e. $\nu \sim \Gamma\left(\frac{1}{\alpha}, \frac{1}{\alpha}\right) = \Gamma(1.75, 1.75)$. For the correlation between the marker and the event time variables, we considered two values: $\rho = 0.25$ and $\rho = 0.50$, the latter corresponds to a stronger association between T and M. The cluster sizes (K) considered are $K = 10$ and $K = 20$. The intra-cluster correlations of the clustered event times for the first and for the second scenario are 0.4 and 0.1, respectively. For the estimation, we consider the prediction t corresponding the quantile values (i.e. Q1, Q2, Q3) of the survival time T. The true time-dependent ROC curves and the associated AUC values computed at the considered times for scenarios I and II are presented in Figure 1.

The simulations are conducted by generating 1000 datasets with two different sample sizes $n = 200$ and $n = 400$. From the 1000 replications, we computed the mean integrated bias (MIB) and the MISE for the time-dependent ROC estimators and the percent bias (%Bias) and the mean squared errors (MSEs) for the time-dependent AUC estimators, which, respectively, defined as

$$\text{MIB} = 1000^{-1}\sum_{s=1}^{1000}\int_{-\infty}^{\infty}\{\widehat{\text{ROC}}_s(u) - \text{ROC}(u)\}du$$

$$\text{MISE} = 1000^{-1}\sum_{s=1}^{1000}\int_{-\infty}^{\infty}\{\widehat{\text{ROC}}_s(u) - \text{ROC}(u)\}^2du$$

$$\%\text{Bias} = (1000^{-1}\sum_{s=1}^{1000}\widehat{\text{AUC}}_s - \text{AUC})/\text{AUC}, \quad \text{MSE} = 1000^{-1}\sum_{s=1}^{1000}\{\widehat{\text{AUC}}_s - \text{AUC}\}^2$$

3.2 Simulation results

In this section, we compare the performance of the proposed empirical (non-smoothed) method under various data generating process, namely under the first scenario we considered a Weibull survival frailty model with $\beta = -0.75$, and gamma frailty term, and a log-normal survival frailty model was considered in the second scenario. In order to estimate the unknown conditional survival function $S(.|M, \nu)$ needed to estimate the unknown weight given in equation (8), a frailty survival model given in equation (2) with a Weibull baseline survival function was considered.

The estimation of the parameters for this model was done using the R package frailtypack, which uses the robust Marquardt algorithm which combines a Newton-Raphson algorithm and a steepest descent algorithm.[29] To assess the consequence of estimating the time-dependent ROC curve and the corresponding AUC using a method that developed for the classical uncorrelated right censored data, we compare the performance of the proposed empirical estimator with an empirical method proposed by Beyene and El Ghouch.[12] This method is proposed for right-censored data where the event times of individuals are assumed independent, given the observed covariates. It uses the popular Beran (also called weighted Kaplan–Meier) estimator[31] to estimate the conditional survival probability, $S(.|M)$, which is a key component of

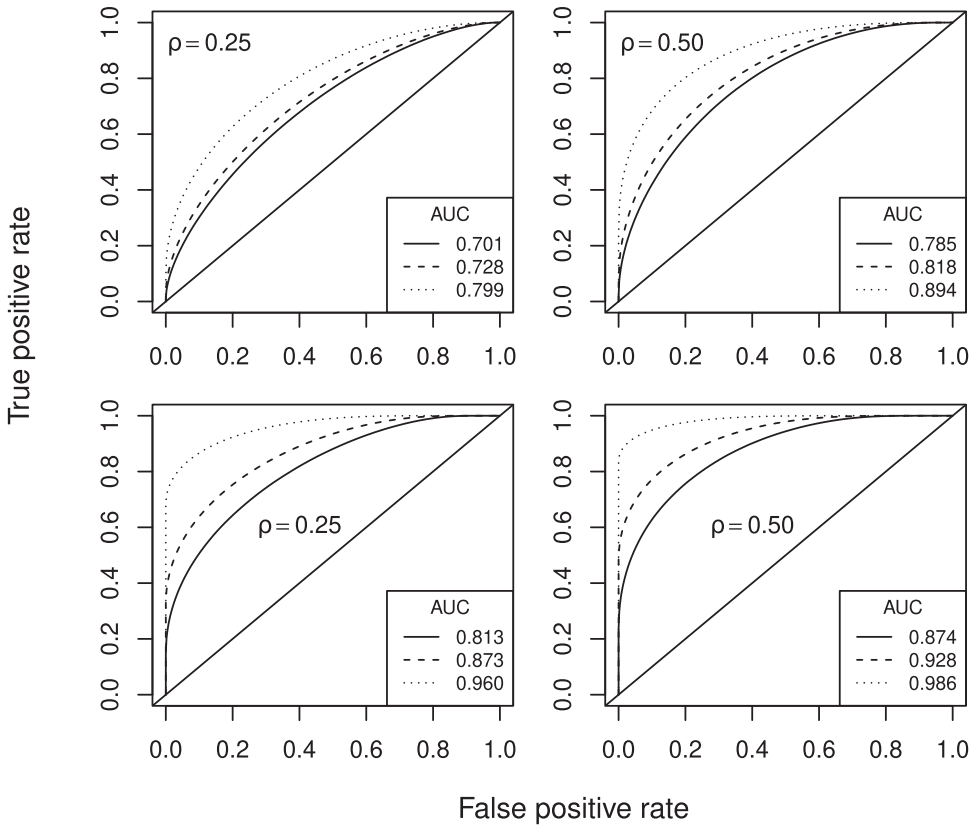

Figure 1. True time-dependent ROC curves and the corresponding AUC values of Scenario I (top row) and Scenario II (bottom row) computed at prediction time Q1 (solid line), Q2 (dashed line) and Q3 (dotted line).
ROC: receiver operating characteristic; AUC: area under the ROC curve.

their ROC and AUC estimators. We chose the Beran-based estimator because it is well-studied (see, for example,[5,9,12]) and has been shown to have good performance than most existing methods proposed for right-censored data.

The MIB and MISE of the time-dependent ROC curve were computed to evaluate the finite sample performance, and for the time-dependent AUC the percent bias (%Bias) and MSE were used as performance measure. Since the behavior of the estimator depends on various data aspects, its performance was examined considering different sample sizes (n), censoring rates ($\%cen$), correlations between marker and survival time (ρ), number of subjects within cluster (K), and prediction time (t).

3.3 Results of time-dependent ROC curve estimator

The MIB and the MISE of both the proposed empirical time-dependent ROC estimator and the empirical Beran method for the first and second scenarios obtained with sample sizes $n = 200, 400$, $K = 10, 20$, censoring rates 20%, 40%, and 60%, $\rho = 0.25, 0.50$, $t = Q1$, Q2, Q3 are presented in Tables 1 and 2. From the results, we can highlight that both the MIB and the MISE of the proposed estimation method are negligibly small in all scenarios. To assess the performance of the proposed estimator under misspecified frailty distributional assumption, we estimated the ROC curve using gamma (the true distribution) and log-normal(the misspecified distribution) frailty term. The results using the gamma frailty are similar to the log-normal frailty with slightly better overall performance from the former. Similarly, the result also reveled that the method performed well with small MIB and MISE under misspecified baseline distribution. Also in this case, the proposed estimator performs better than the Beran method. Moreover, the MISE, in general, decrease with the sample size and increases with both the censoring percentages and cluster size. As compared to the Beran estimator, which does not acknowledge the presence of correlated observations, the proposed time-dependent ROC curve estimation method has, in general, better finite sample performance than the Beran estimator, because the latter generally has smaller MIB and MISE. This is true for all censoring rates, sample sizes, K, ρ and t values. Furthermore, the Beran method tends to underestimate the ROC curve in all cases as the computed MIB is consistently negative in all scenarios.

Table 1. Scenario I: MIB(×1000) and MISE(×1000) of the proposed empirical time-dependent ROC estimator with Weibull and Log-normal frailty, and the Beran method computed for different sample sizes (n), cluster sizes (K), right censoring rates (% cen), correlation values (ρ), and t values.

			Proposed approach											
			Gamma				Log Normal				Beran			
			K=10		K=20		K=10		K=20		K=10		K=20	
%cen	t	ρ	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE
n = 200														
20	1.31	0.25	−0.010	4.179	−0.974	4.377	−0.953	4.168	−0.721	4.342	−0.528	4.272	−1.431	4.442
20	2.55	0.25	1.282	3.033	−0.237	3.131	0.855	3.041	−0.525	3.114	−0.551	3.171	−2.029	3.262
20	5.83	0.25	1.245	3.187	−1.177	3.539	2.766	3.302	0.452	3.606	−3.906	3.741	−5.382	4.176
40	1.31	0.25	0.684	4.086	−0.418	4.272	0.227	4.107	−0.273	4.227	−1.429	4.321	−2.075	4.521
40	2.55	0.25	2.924	2.974	1.185	3.070	3.134	3.035	1.913	3.017	−2.955	3.384	−3.978	3.454
40	5.83	0.25	4.321	2.966	1.161	3.244	10.65	3.196	7.937	3.406	−8.428	5.304	−9.837	5.995
60	1.31	0.25	1.675	3.955	0.274	4.217	1.563	3.946	0.524	4.205	−3.901	4.550	−4.274	4.822
60	2.55	0.25	5.449	3.149	2.934	3.193	6.706	3.201	4.283	3.206	−8.035	4.345	−10.68	4.449
60	5.83	0.25	8.807	3.439	5.533	3.302	21.97	4.006	18.55	3.747	−18.71	21.37	−22.66	24.70
20	1.31	0.50	0.271	3.610	−0.977	3.764	−0.267	3.695	−1.024	3.757	−0.379	3.680	−1.420	3.814
20	2.55	0.50	1.123	2.494	−0.489	2.555	1.041	2.481	−0.758	2.558	−0.811	2.605	−2.247	2.658
20	5.83	0.50	0.732	1.936	−1.302	2.137	1.314	1.952	−0.669	2.210	−3.464	2.341	−5.026	2.596
40	1.31	0.50	0.672	3.497	−0.403	3.648	0.194	3.541	−1.061	3.643	−1.595	3.712	−2.499	3.860
40	2.55	0.50	3.074	2.392	1.549	2.440	3.030	2.407	1.109	2.460	−3.768	2.748	−5.028	2.812
40	5.83	0.50	2.340	1.590	0.017	1.779	5.629	1.668	3.356	1.843	−8.450	3.473	−10.31	3.951
60	1.31	0.50	1.996	3.379	0.625	3.569	2.202	3.367	0.538	3.529	−4.672	3.911	−5.492	4.082
60	2.55	0.50	6.419	2.439	3.974	2.426	8.387	2.472	4.830	2.426	−10.11	3.595	−12.80	3.700
60	5.83	0.50	3.661	1.504	1.088	1.560	9.661	1.594	7.278	1.532	−19.65	13.56	−24.59	16.27
n = 400														
20	1.31	0.25	1.237	2.046	−1.485	2.210	0.533	2.050	−1.924	2.188	0.503	2.089	−2.006	2.244
20	2.55	0.25	1.210	1.476	−0.319	1.687	0.762	1.485	−0.612	1.689	−0.545	1.544	−1.884	1.757
20	5.83	0.25	1.245	1.509	−1.882	1.799	2.491	1.564	−0.489	1.835	−3.029	1.810	−5.487	2.131
40	1.31	0.25	1.443	2.011	−1.157	2.158	0.854	2.021	−1.553	2.155	−0.294	2.138	−2.770	2.287
40	2.55	0.25	2.386	1.440	0.203	1.671	2.341	1.464	0.187	1.662	−2.800	1.662	−4.541	1.924
40	5.83	0.25	3.977	1.433	−0.954	1.692	10.88	1.619	5.359	1.749	−7.183	2.569	−11.10	3.037

(continued)

Table 1. Continued

			Proposed approach											
			Gamma				Log Normal				Beran			
			K=10		*K*=20		*K*=10		*K*=20		*K*=10		*K*=20	
%cen	*t*	ρ	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE
60	1.31	0.25	2.356	1.965	−0.256	2.138	2.036	1.942	−0.443	2.143	−2.491	2.284	−4.756	2.485
60	2.55	0.25	5.677	1.533	2.992	1.718	7.075	1.579	4.158	1.728	−7.033	2.190	−8.353	2.458
60	5.83	0.25	10.01	1.783	4.134	1.901	23.71	2.417	18.55	2.270	−17.66	13.66	−23.46	14.22
20	1.31	0.50	0.880	1.749	−1.383	1.913	1.045	1.760	−1.273	1.908	0.278	1.786	−1.819	1.947
20	2.55	0.50	0.971	1.211	−0.292	1.378	1.075	1.207	−0.072	1.372	−0.793	1.270	−1.816	1.442
20	5.83	0.50	0.269	0.956	−1.580	1.115	1.476	0.968	−0.723	1.142	−3.167	1.176	−4.755	1.376
40	1.31	0.50	1.125	1.717	−1.024	1.858	0.750	1.699	−1.825	1.868	−0.804	1.823	−2.829	1.980
40	2.55	0.50	2.371	1.175	0.606	1.317	2.445	1.190	0.916	1.349	−3.591	1.371	−4.883	1.585
40	5.83	0.50	1.800	0.824	−1.250	0.950	4.769	0.856	2.461	0.947	−7.598	1.771	−10.08	2.085
60	1.31	0.50	2.230	1.655	0.088	1.824	2.174	1.678	0.238	1.816	−3.478	1.937	−5.427	2.138
60	2.55	0.50	6.152	1.209	4.423	1.325	7.620	1.273	5.988	1.347	−8.670	1.836	−9.684	2.037
60	5.83	0.50	3.921	0.811	1.020	0.889	9.653	0.925	7.968	0.917	−16.55	9.150	−20.56	9.263

MIB: mean integrated bias; MISE: mean integrated squared error; ROC: receiver operating characteristic.

Table 2. Scenario II: MIB(×1000) and MISE(×1000) of the proposed empirical time-dependent ROC estimator with Weibull and Log-normal frailty, and the Beran method computed for different sample sizes (n), cluster sizes (K), right censoring rates (% cen), correlation values (ρ), and t values.

			Proposed approach											
			Gamma				Log Normal				Beran			
			$K = 10$		$K = 20$		$K = 10$		$K = 20$		$K = 10$		$K = 20$	
%cen	t	ρ	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE
$n = 200$														
20	1.31	0.25	−2.970	3.157	−4.448	3.606	−2.750	3.173	−4.760	3.680	−2.791	3.204	−4.196	3.676
20	2.55	0.25	−3.042	1.682	−6.036	1.978	−3.042	1.691	−6.086	1.947	−3.183	1.731	−5.746	2.017
20	5.83	0.25	−1.911	0.835	−3.773	0.912	−1.669	0.838	−3.487	0.906	−2.803	0.945	−4.328	0.999
40	1.31	0.25	−2.995	3.102	−4.855	3.551	−2.942	3.108	−4.134	3.500	−4.946	3.331	−6.133	3.795
40	2.55	0.25	−1.501	1.744	−4.153	1.961	−1.458	1.735	−2.861	1.921	−7.473	2.090	−9.116	2.317
40	5.83	0.25	0.989	0.966	0.369	0.977	2.099	0.945	2.006	0.949	−4.959	1.505	−5.574	1.530
60	1.31	0.25	−2.116	3.105	−3.892	3.670	−2.313	3.035	−5.757	3.622	−9.655	3.851	−11.68	4.491
60	2.55	0.25	4.053	2.201	1.341	2.471	5.199	2.157	2.002	2.443	−12.58	3.129	−14.55	3.537
60	5.83	0.25	5.312	1.080	5.054	1.084	7.145	1.046	6.900	1.053	−8.628	3.598	−5.869	2.951
20	1.31	0.50	−1.962	2.370	−3.297	2.651	−1.807	2.395	−3.126	2.668	−2.047	2.412	−2.900	2.697
20	2.55	0.50	−2.220	1.032	−4.169	1.109	−2.129	1.036	−4.109	1.176	−2.302	1.073	−3.826	1.167
20	5.83	0.50	−2.001	0.592	−2.941	0.627	−1.884	0.580	−2.793	0.663	−1.624	0.616	−2.427	0.645
40	1.31	0.50	−3.262	2.355	−3.999	2.616	−3.443	2.352	−3.848	2.606	−4.454	2.539	−5.195	2.791
40	2.55	0.50	−1.577	1.124	−2.712	1.188	−1.295	1.123	−2.236	1.180	−6.315	1.459	−7.192	1.505
40	5.83	0.50	−1.770	0.701	−2.038	0.693	−1.283	0.668	−1.913	0.767	−2.689	0.925	−3.036	0.941
60	1.31	0.50	−2.279	2.331	−3.494	2.663	−1.407	2.315	−4.455	2.725	−9.563	3.020	−11.48	3.449
60	2.55	0.50	3.978	1.292	2.765	1.366	4.524	1.302	2.893	1.378	−11.15	2.341	−12.14	2.538
60	5.83	0.50	−1.341	0.688	−1.326	0.679	−1.065	0.673	−1.054	0.752	−4.485	1.584	−3.505	1.446
$n = 400$														
20	1.31	0.25	−3.474	1.684	−2.477	1.711	−2.871	1.675	−3.046	1.716	−3.249	1.700	−2.709	1.720
20	2.55	0.25	−3.312	0.886	−2.829	0.903	−2.340	0.885	−3.779	0.909	−3.038	0.923	−3.243	0.934
20	5.83	0.25	−1.949	0.593	−2.687	0.626	−1.456	0.589	−2.689	0.621	−2.474	0.599	−3.148	0.642
40	1.31	0.25	−3.600	1.651	−2.771	1.662	−2.959	1.680	−2.659	1.649	−4.458	1.802	−4.001	1.784
40	2.55	0.25	−1.230	0.972	−1.208	0.987	−0.768	0.958	−0.490	0.977	−4.984	1.133	−5.375	1.132
40	5.83	0.25	1.340	0.716	0.819	0.735	2.357	0.698	2.321	0.714	−3.428	0.842	−3.581	0.897

(continued)

Table 2. Continued

			Proposed approach											
			Gamma				Log Normal				Beran			
			$K = 10$		$K = 20$		$K = 10$		$K = 20$		$K = 10$		$K = 20$	
%*cen*	*t*	ρ	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE
60	1.31	0.25	−2.636	1.723	−2.345	1.733	−3.044	1.745	−3.319	1.775	−8.607	2.146	−7.762	2.145
60	2.55	0.25	4.538	1.409	3.773	1.428	5.810	1.401	4.712	1.378	−7.478	1.679	−7.875	1.689
60	5.83	0.25	5.932	0.856	5.037	0.850	7.643	0.865	6.865	0.838	−4.128	1.558	−4.633	1.574
20	1.31	0.50	−2.502	1.265	−2.447	1.291	−2.348	1.286	−2.099	1.281	−2.405	1.281	−2.035	1.289
20	2.55	0.50	−2.422	0.553	−2.815	0.575	−1.965	0.550	−2.842	0.585	−1.969	0.569	−2.156	0.581
20	5.83	0.50	−2.135	0.578	−2.504	0.594	−2.023	0.570	−2.556	0.587	−1.439	0.500	−1.714	0.523
40	1.31	0.50	−3.687	1.263	−3.507	1.265	−3.048	1.270	−3.480	1.234	−3.811	1.367	−3.516	1.354
40	2.55	0.50	−1.303	0.667	−1.593	0.677	−1.191	0.647	−1.227	0.680	−3.874	0.795	−4.259	0.815
40	5.83	0.50	−1.794	0.642	−1.868	0.648	−1.406	0.614	−1.666	0.710	−2.074	0.630	−1.917	0.661
60	1.31	0.50	−3.025	1.355	−2.583	1.367	−2.878	1.337	−2.112	1.355	−7.955	1.706	−7.452	1.697
60	2.55	0.50	4.465	0.842	4.098	0.862	5.045	0.854	5.037	0.854	−6.116	1.240	−6.645	1.278
60	5.83	0.50	−1.228	0.621	−1.463	0.627	−0.811	0.594	−0.854	0.606	−2.207	0.868	−2.414	0.878

MIB: mean integrated bias; MISE: mean integrated squared error; ROC: receiver operating characteristic.

3.4 Results of time-dependent AUC estimator

The empirical percent bias (%Bias) and MSE of both the proposed empirical time-dependent AUC estimator and the empirical Beran method for the first and second scenario are presented in Tables 3 and 4, respectively. The general conclusion of these results is consistent with the ROC curve findings of the previous simulations. The AUC estimator shows good performance with small bias and MSE. In order to study how well the proposed empirical estimator performs under misspecification of the frailty distribution, the data are generated under a gamma distribution for both scenarios, and the estimation was done under a log-normal frailty distribution. The results presented in Tables 3 and 4 show that the AUC estimated with true frailty distribution and misspecified frailty distribution are similar with slight better performance from the latter. Results presented in Table 4 revealed that the proposed method also performs well with small bias and MSE when the estimation is done with a misspecified baseline survival function. Furthermore, as expected, the MSE decrease as the sample size increases. In contrast, the MSE, in general, increases with both the censoring percentage and cluster size K. Compared to the Beran method, the proposed time-dependent AUC estimator, globally, performs well with small bias and MSE.

4 Real data analysis

In this section, two different real-world examples are provided to illustrate the proposed time-dependent ROC curve and associated time-dependent AUC estimation methods. The first data is the lung cancer data from the North Central Cancer Treatment Group (NCCTG). The second data set is infections in kidney patients data. In the upcoming subsections, we will provide details about the data sets and results of our analyses.

4.1 NCCTG lung cancer data

Our analysis in this section uses the popular NCCTG lung lancer data which contains 228 patients, of whom 63 are right-censored (i.e. patients left the study before experiencing the event of interest). The NCCTG lung cancer data was collected from 18 different health institutions (clusters). The number of subjects per institution ranges from 2 to 36. In this data, patients from the same institution (cluster) are likely to have correlated event times because they share the same facility. This correlation among event times within the same cluster should be considered in the analysis to ensure that the results are valid and accurate. The NCCTG data set records survival times together some important predictor variables such as sex (Male = 1 and Female = 2), age (in years), ph.ecog (Eastern Cooperative Oncology Group (ECOG) performance status assessed by the physician, on a scale ranges from 0 (asymptomatic) to 5 (dead)), and pat.karno (Karnofsky performance status, assessed by the patient). Originally, Loprinzi et al.[32] analyzed this data in an attempt to determine whether descriptive information gathered from a patient-completed questionnaire could provide prognostic information independently of that previously gathered from the patient's physician. This data set is available in the R package survival as the *lung* data set.[33] In this analysis, we derive a prognostic score using three variables: sex, age and ph.ecog. To this end, we fit frailty model given in equation (2) with gamma frailty, and the marker is given by $M = \hat{v}\exp(\hat{\beta}_1 sex + \hat{\beta}_2 age + \hat{\beta}_3 ph.ecog)$, where $\hat{v}$ is frailty estimate, and $\hat{\beta}_i, i = 1, 2, 3$, are the estimated regression coefficients. In the final analysis, only 227 patients were used since one patient had missing ph.ecog measure. The objective of this analysis is to determine the accuracy of the marker in predicting the risk of lung cancer death over time.

In this analysis, we estimated the time-dependent ROC curves $\widehat{\mathrm{ROC}}_t$ and the corresponding $\widehat{\mathrm{AUC}}_t$ using the proposed empirical method (assuming both gamma and log-normal distributions for the frailty term) and the Beran approach at time points $t = 60$, 120, and 180 days. The results of the estimated $\widehat{\mathrm{ROC}}_t$ and $\widehat{\mathrm{AUC}}_t$ are presented in Figure 2 and Table 5, respectively. The $\widehat{\mathrm{AUC}}_t$ estimates obtained from the proposed method with gamma frailty are very close to those obtained from log-normal frailty distributional assumption with slightly larger values obtained from the later frailty distribution. These similarities also observed in the estimated time-dependent ROC curve given in Figure 2. When comparing the proposed method and the Beran approach, both the estimated ROC curves in Figure 2 and the AUC values in 5 obtained from the Beran approach are larger than those obtained from the proposed method when gamma and log-normal frailty terms are assumed. Finally, both the results of the proposed method and the Beran approach show that the derived marker has good predictive ability for the risk of death due to lung cancer, but the predictive ability decreases as prediction time t increases.

4.2 Kidney data

The kidney data is the other data that often used to illustrate frailty models. This data is about recurrence of infection in kidney patients who use portable dialysis equipment. In kidney patients using portable dialysis equipment, recurrent

Table 3. Scenario I: %Bias(MSE×1000) for the proposed empirical time-dependent AUC estimator with Weibull and Log-normal frailty, and the Beran method computed for different sample sizes (n), cluster sizes (K), right censoring rates (% cen), correlation values (ρ), and t values.

			Proposed approach					
			Gamma		Log Normal		Beran	
%*cen*	*t*	ρ	$K = 10$ %Bias(MSE)	$K = 20$ %Bias(MSE)	$K = 10$ %Bias(MSE)	$K = 20$ %Bias(MSE)	$K = 10$ %Bias(MSE)	$K = 20$ %Bias(MSE)
$n = 200$								
20	1.31	0.25	−0.001(2.092)	−0.139(2.193)	−0.136(2.085)	−0.103(2.173)	−0.075(2.142)	−0.205(2.226)
20	2.55	0.25	0.174(1.522)	−0.036(1.567)	0.115(1.531)	−0.075(1.562)	−0.079(1.592)	−0.282(1.635)
20	5.83	0.25	0.161(1.441)	−0.142(1.664)	0.353(1.484)	0.062(1.669)	−0.485(1.727)	−0.669(2.005)
40	1.31	0.25	0.097(2.073)	−0.061(2.150)	0.032(2.090)	−0.040(2.134)	−0.204(2.190)	−0.297(2.275)
40	2.55	0.25	0.399(1.559)	0.160(1.615)	0.428(1.603)	0.260(1.579)	−0.409(1.752)	−0.549(1.801)
40	5.83	0.25	0.544(1.488)	0.147(1.651)	1.338(1.598)	0.996(1.695)	−1.051(2.731)	−1.227(3.126)
60	1.31	0.25	0.238(2.088)	0.038(2.229)	0.222(2.092)	0.073(2.237)	−0.557(2.365)	−0.611(2.499)
60	2.55	0.25	0.746(1.840)	0.401(1.875)	0.919(1.882)	0.586(1.899)	−1.107(2.405)	−1.469(2.517)
60	5.83	0.25	1.105(1.946)	0.693(1.860)	2.752(2.289)	2.323(2.140)	−2.329(13.35)	−2.823(16.18)
20	1.31	0.50	0.034(1.554)	−0.125(1.679)	−0.034(1.616)	−0.132(1.677)	−0.048(1.589)	−0.181(1.700)
20	2.55	0.50	0.134(1.034)	−0.063(1.097)	0.124(1.028)	−0.096(1.100)	−0.102(1.086)	−0.279(1.144)
20	5.83	0.50	0.100(0.647)	−0.130(0.771)	0.167(0.646)	−0.056(0.787)	−0.375(0.834)	−0.551(0.988)
40	1.31	0.50	0.085(1.522)	−0.052(1.631)	0.024(1.546)	−0.136(1.632)	−0.204(1.624)	−0.319(1.729)
40	2.55	0.50	0.374(1.031)	0.188(1.087)	0.368(1.045)	0.134(1.103)	−0.465(1.198)	−0.618(1.272)
40	5.83	0.50	0.277(0.579)	0.015(0.690)	0.648(0.594)	0.391(0.695)	−0.933(1.451)	−1.142(1.694)
60	1.31	0.50	0.252(1.539)	0.077(1.668)	0.279(1.545)	0.066(1.644)	−0.597(1.779)	−0.701(1.892)
60	2.55	0.50	0.783(1.164)	0.484(1.184)	1.024(1.187)	0.589(1.186)	−1.238(1.726)	−1.567(1.844)
60	5.83	0.50	0.423(0.608)	0.133(0.655)	1.098(0.647)	0.828(0.637)	−2.177(7.614)	−2.730(9.836)
$n = 400$								
20	1.31	0.25	0.175(0.973)	−0.214(1.104)	0.074(0.983)	−0.277(1.092)	0.070(0.996)	−0.288(1.120)
20	2.55	0.25	0.165(0.696)	−0.046(0.854)	0.102(0.704)	−0.086(0.856)	−0.077(0.725)	−0.261(0.886)
20	5.83	0.25	0.162(0.633)	−0.230(0.869)	0.320(0.654)	−0.055(0.878)	−0.374(0.776)	−0.682(1.046)
40	1.31	0.25	0.204(0.973)	−0.167(1.091)	0.120(0.986)	−0.224(1.087)	−0.045(1.034)	−0.397(1.150)
40	2.55	0.25	0.326(0.711)	0.026(0.891)	0.320(0.733)	0.024(0.889)	−0.387(0.812)	−0.625(1.019)
40	5.83	0.25	0.501(0.668)	−0.117(0.869)	1.366(0.783)	0.674(0.887)	−0.895(1.206)	−1.385(1.595)

(continued)

Table 3. Continued

			Proposed approach					
			Gamma		Log Normal		Beran	
%*cen*	*t*	ρ	$K = 10$ %Bias(MSE)	$K = 20$ %Bias(MSE)	$K = 10$ %Bias(MSE)	$K = 20$ %Bias(MSE)	$K = 10$ %Bias(MSE)	$K = 20$ %Bias(MSE)
60	1.31	0.25	0.334(0.997)	−0.038(1.139)	0.289(0.985)	−0.065(1.150)	−0.358(1.137)	−0.680(1.298)
60	2.55	0.25	0.778(0.859)	0.409(1.002)	0.970(0.898)	0.569(1.017)	−0.968(1.165)	−1.149(1.368)
60	5.83	0.25	1.253(0.956)	0.518(1.030)	2.970(1.371)	2.322(1.284)	−2.199(8.317)	−2.924(8.945)
20	1.31	0.50	0.109(0.729)	−0.180(0.843)	0.130(0.737)	−0.166(0.845)	0.032(0.746)	−0.234(0.858)
20	2.55	0.50	0.117(0.482)	−0.037(0.587)	0.130(0.480)	−0.010(0.582)	−0.099(0.507)	−0.224(0.616)
20	5.83	0.50	0.049(0.297)	−0.159(0.396)	0.188(0.297)	−0.060(0.406)	−0.338(0.386)	−0.518(0.516)
40	1.31	0.50	0.140(0.729)	−0.133(0.828)	0.093(0.716)	−0.235(0.839)	−0.105(0.775)	−0.363(0.882)
40	2.55	0.50	0.289(0.493)	0.073(0.587)	0.299(0.499)	0.112(0.606)	−0.441(0.577)	−0.598(0.718)
40	5.83	0.50	0.216(0.279)	−0.126(0.358)	0.552(0.286)	0.293(0.346)	−0.836(0.663)	−1.115(0.866)
60	1.31	0.50	0.281(0.733)	0.008(0.855)	0.275(0.754)	0.027(0.858)	−0.445(0.850)	−0.694(0.998)
60	2.55	0.50	0.751(0.565)	0.540(0.642)	0.931(0.606)	0.732(0.657)	−1.061(0.851)	−1.184(0.990)
60	5.83	0.50	0.452(0.302)	0.125(0.346)	1.096(0.358)	0.906(0.362)	−1.831(4.909)	−2.281(5.143)

MSE: mean squared error; AUC: area under the ROC curve.

Table 4. Scenario II: %Bias(MSE×1000) for the proposed empirical time-dependent AUC estimator with Weibull and Log-normal frailty, and the Beran method computed for different sample sizes (n), cluster sizes (K), right censoring rates (% cen), correlation values (ρ), and t values.

			Proposed approach					
			Gamma		Log Normal		Beran	
%*cen*	*t*	ρ	$K = 10$ %Bias(MSE)	$K = 20$ %Bias(MSE)	$K = 10$ %Bias(MSE)	$K = 20$ %Bias(MSE)	$K = 10$ %Bias(MSE)	$K = 20$ %Bias(MSE)
$n = 200$								
20	1.31	0.25	−0.366(1.391)	−0.548(1.695)	−0.339(1.401)	−0.586(1.731)	−0.344(1.415)	−0.517(1.723)
20	2.55	0.25	−0.350(0.678)	−0.692(0.863)	−0.350(0.681)	−0.698(0.840)	−0.366(0.699)	−0.658(0.866)
20	5.83	0.25	−0.159(0.166)	−0.353(0.200)	−0.131(0.166)	−0.320(0.205)	−0.250(0.208)	−0.408(0.241)
40	1.31	0.25	−0.369(1.386)	−0.597(1.686)	−0.363(1.393)	−0.509(1.649)	−0.608(1.505)	−0.755(1.811)
40	2.55	0.25	−0.170(0.683)	−0.474(0.806)	−0.165(0.681)	−0.325(0.788)	−0.854(0.860)	−1.042(0.986)
40	5.83	0.25	0.148(0.155)	0.083(0.154)	0.267(0.154)	0.257(0.151)	−0.470(0.377)	−0.535(0.384)
60	1.31	0.25	−0.262(1.376)	−0.481(1.761)	−0.285(1.346)	−0.709(1.735)	−1.188(1.791)	−1.439(2.222)
60	2.55	0.25	0.468(0.805)	0.156(0.954)	0.600(0.814)	0.233(0.944)	−1.437(1.329)	−1.662(1.579)
60	5.83	0.25	0.601(0.168)	0.573(0.166)	0.794(0.180)	0.768(0.179)	−0.850(1.437)	−0.563(0.959)
20	1.31	0.50	−0.225(0.891)	−0.377(1.042)	−0.207(0.909)	−0.358(1.052)	−0.235(0.909)	−0.332(1.069)
20	2.55	0.50	−0.238(0.334)	−0.447(0.367)	−0.227(0.333)	−0.440(0.407)	−0.246(0.350)	−0.410(0.392)
20	5.83	0.50	−0.150(0.040)	−0.245(0.050)	−0.138(0.041)	−0.229(0.064)	−0.112(0.051)	−0.193(0.058)
40	1.31	0.50	−0.373(0.895)	−0.459(1.045)	−0.395(0.902)	−0.440(1.036)	−0.509(0.982)	−0.595(1.127)
40	2.55	0.50	−0.155(0.329)	−0.278(0.355)	−0.123(0.335)	−0.225(0.352)	−0.671(0.457)	−0.766(0.470)
40	5.83	0.50	−0.121(0.035)	−0.149(0.036)	−0.072(0.033)	−0.136(0.068)	−0.212(0.095)	−0.248(0.095)
60	1.31	0.50	−0.263(0.862)	−0.402(1.055)	−0.164(0.851)	−0.512(1.080)	−1.096(1.183)	−1.315(1.434)
60	2.55	0.50	0.449(0.354)	0.318(0.374)	0.508(0.369)	0.333(0.376)	−1.190(0.726)	−1.297(0.792)
60	5.83	0.50	−0.080(0.037)	−0.078(0.027)	−0.052(0.041)	−0.051(0.061)	−0.394(0.372)	−0.296(0.278)
$n = 400$								
20	1.31	0.25	−0.430(0.746)	−0.306(0.778)	−0.355(0.741)	−0.376(0.779)	−0.401(0.751)	−0.335(0.775)
20	2.55	0.25	−0.378(0.356)	−0.323(0.380)	−0.267(0.354)	−0.431(0.385)	−0.344(0.373)	−0.367(0.390)
20	5.83	0.25	−0.159(0.080)	−0.237(0.093)	−0.104(0.076)	−0.234(0.096)	−0.211(0.095)	−0.280(0.109)
40	1.31	0.25	−0.446(0.734)	−0.344(0.755)	−0.367(0.748)	−0.330(0.756)	−0.550(0.810)	−0.493(0.817)
40	2.55	0.25	−0.136(0.344)	−0.133(0.368)	−0.082(0.342)	−0.050(0.367)	−0.560(0.445)	−0.605(0.457)
40	5.83	0.25	0.187(0.072)	0.132(0.077)	0.295(0.073)	0.292(0.078)	−0.308(0.156)	−0.323(0.172)

(continued)

Table 4. Continued

			Proposed approach					
			Gamma		Log Normal		Beran	
%*cen*	*t*	ρ	$K = 10$ %Bias(MSE)	$K = 20$ %Bias(MSE)	$K = 10$ %Bias(MSE)	$K = 20$ %Bias(MSE)	$K = 10$ %Bias(MSE)	$K = 20$ %Bias(MSE)
60	1.31	0.25	−0.326(0.725)	−0.291(0.762)	−0.378(0.750)	−0.411(0.796)	−1.059(0.979)	−0.955(1.012)
60	2.55	0.25	0.524(0.427)	0.437(0.440)	0.670(0.442)	0.545(0.434)	−0.844(0.649)	−0.891(0.661)
60	5.83	0.25	0.666(0.103)	0.572(0.096)	0.848(0.124)	0.766(0.112)	−0.380(0.393)	−0.431(0.401)
20	1.31	0.50	−0.289(0.465)	−0.282(0.497)	−0.272(0.475)	−0.242(0.494)	−0.277(0.472)	−0.234(0.492)
20	2.55	0.50	−0.254(0.171)	−0.296(0.181)	−0.204(0.167)	−0.298(0.191)	−0.201(0.176)	−0.221(0.182)
20	5.83	0.50	−0.157(0.020)	−0.195(0.023)	−0.145(0.020)	−0.200(0.033)	−0.091(0.023)	−0.118(0.025)
40	1.31	0.50	−0.424(0.464)	−0.402(0.484)	−0.351(0.466)	−0.399(0.469)	−0.434(0.512)	−0.400(0.525)
40	2.55	0.50	−0.121(0.156)	−0.152(0.167)	−0.107(0.151)	−0.110(0.168)	−0.396(0.215)	−0.437(0.224)
40	5.83	0.50	−0.122(0.017)	−0.129(0.019)	−0.083(0.016)	−0.109(0.055)	−0.153(0.041)	−0.136(0.042)
60	1.31	0.50	−0.347(0.457)	−0.297(0.489)	−0.329(0.455)	−0.242(0.492)	−0.908(0.640)	−0.850(0.665)
60	2.55	0.50	0.503(0.178)	0.463(0.188)	0.567(0.188)	0.566(0.195)	−0.636(0.334)	−0.695(0.342)
60	5.83	0.50	−0.068(0.013)	−0.092(0.015)	−0.026(0.013)	−0.029(0.013)	−0.167(0.102)	−0.186(0.101)

MSE: mean squared error; AUC: area under the ROC curve.

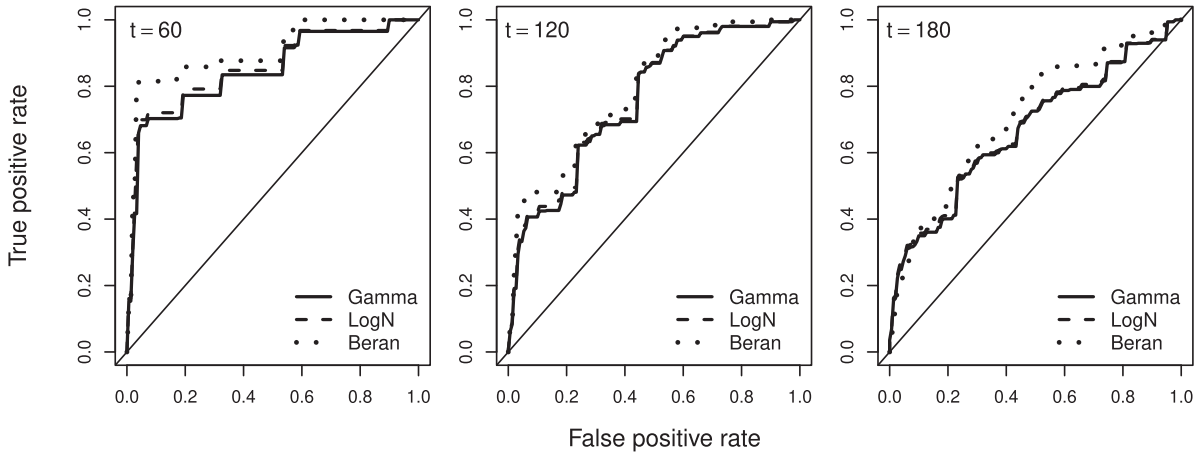

Figure 2. Estimated time-dependent ROC curves obtained using the proposed empirical estimator with gamma frailty (solid lines), log-normal frailty (dashed lines), and Beran estimator (dotted lines) for the NCCTG lung cancer data with $t = 60$ (left panel), 120 (middle panel), and 180 (right panel) days.
NCCTG: North Central Cancer Treatment Group; ROC: receiver operating characteristic.

Table 5. Estimated time-dependent AUC values, based on the proposed empirical estimator with gamma and log-normal frailty, and Beran estimator obtained from the North Central Cancer Treatment Group (NCCTG) lung cancer data, with $t = 60$, 120, and 180 days.

	Proposed		
t	Gammma	Log normal	Beran
60	0.8465	0.8565	0.9025
120	0.7581	0.7641	0.7925
180	0.6679	0.6743	0.7120

infection is the main complication, which occurs at a point where the catheter is inserted. When an infection occurs, the catheter is removed, then the catheter is reinserted again after the infection has been treated successfully. In some cases, the catheter may be removed for reasons other than infection, in which case the observation is censored. A total of 38 patients with kidney disease were followed for time to recurrence of infection. Each patient has exactly 2 observations. The data consists of three covariates: sex (1 = Male, 2 = Female), age (in years), and disease (disease type with 4 levels: "GN", "AN", "PKD", and "Other"). This data set was originally analyzed and presented in McGilchrist and Aisbett[18] and it is available in R package survival.[33] As in the above example, we derive the marker as $M = \hat{v}\exp(\hat{\beta}_1 sex + \hat{\beta}_2 age + \hat{\beta}_3 GN + \hat{\beta}_4 AN + \hat{\beta}_5 PKD)$ using a gamma frailty model. The aim of this analysis is to determine how accurate the marker is at predicting the risk of recurrence of kidney infection.

To evaluate the predictive accuracy of M, we estimated the time-dependent ROC curves and the associated AUC values using the proposed estimator and the Beran approach at $t = 120$, 180 and 240 days. For the frailty term, both a gamma and a log-normal distribution were considered for the proposed estimator. The $\widehat{\mathrm{ROC}}_t$ and the corresponding $\widehat{\mathrm{AUC}}_t$ are presented in Figure 3 and Table 6, respectively. Estimated ROC curves for the proposed method under gamma and log-normal frailty distribution assumptions are very similar, however, estimates are slightly smaller than those obtained from the Beran approach. This is also observed from the estimated time-dependent AUC values. Finally, the estimated AUC values at all considered prediction time are fairly large, indicating good overall predictive ability of the marker.

5 Discussion

In this article, we proposed and investigated a time-dependent ROC curve and the corresponding AUC estimation method for correlated right-censored survival data. This method is a generalization of the time-dependent ROC curve introduced by Beyene and El Ghouch[12] which assumed that individuals event times are independent. Therefore, as in Beyene and El Ghouch,[12] the unknown event status of censored individuals is imputed with conditional survival function, the conditionality in this estimator, however, is both on the marker and frailty of the subjects. In order to estimate this conditional

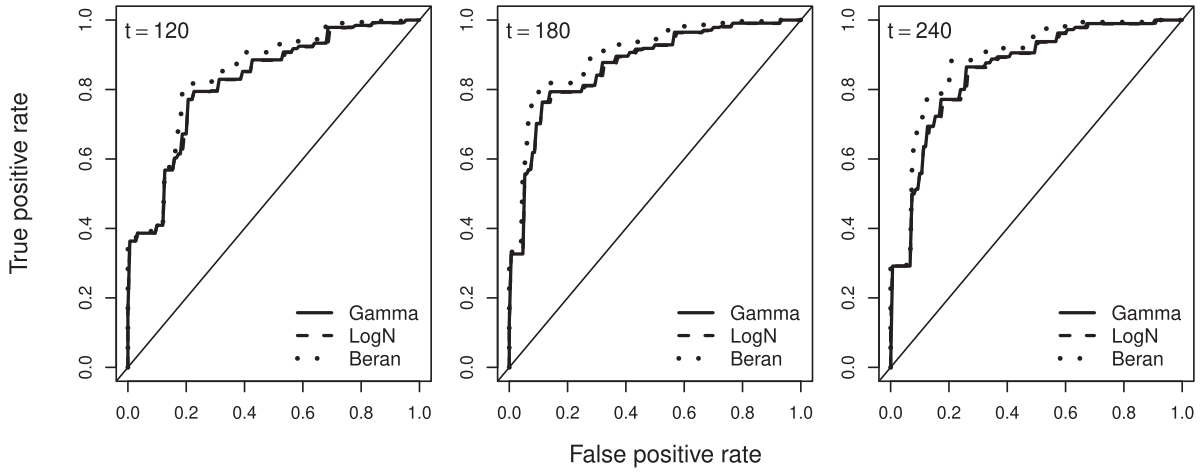

Figure 3. Estimated time-dependent receiver operating characteristic (ROC) curves obtained using the proposed empirical estimator with gamma frailty (solid lines), log-normal frailty (dashed lines), and Beran estimator (dotted lines) for the kidney data with $t = 120$ (left panel), 180 (middle panel), and 240 (right panel) days.

Table 6. Estimated time-dependent area under the ROC curve (AUC) values, based on the proposed empirical estimator with gamma and log-normal frailty, and Beran estimator obtained from the kidney data with $t = 120$,180, and 240 days.

	Proposed		
t	Gamma	Log normal	Beran
120	0.8240	0.8225	0.8454
180	0.8725	0.8698	0.9012
240	0.8578	0.8541	0.8862

survival function needed for determining the unknown event status of censored individuals, a parametric frailty model was considered.

An extensive simulation study with two different scenarios was conducted to evaluate the finite sample performance of the proposed empirical (non-smoothed) time-dependent ROC curve and the corresponding AUC estimation method. A comparison was also made between the proposed method and the existing naïve estimator that proposed for right-censored data with independent event times. This existing method uses the popular Beran approach to estimate the unknown conditional survival function. Based on the results, the proposed ROC curve and AUC estimators have a better finite sample performance with smaller MIB and MISE than the existing Beran method. In addition, the Beran approach tends to underestimate the ROC curve and the corresponding AUC as the biases are consistently negative for all considered scenarios. Moreover, when we examine the effect of misspecifing the frailty distribution in the proposed estimator, it generally has minimal effects since as results of gamma frailty and log-normal frailty are very similar, with the former showing slightly better performance. The method also performed well with small MIB and MISE when the data is generated with misspecified baseline.

The proposal was applied to the NCCTG lung cancer and kidney data sets. For both data set, the continuous risk score (marker) is obtained as $v\exp(\boldsymbol{\beta X})$. The covariate vectors $\boldsymbol{X}$ of lung cancer data are: Age, sex, and ECOG performance status assessed by the physician, and for the kidney data the covariates age, sex, and disease type were considered. The aim of the lung cancer data analysis was to assess the accuracy of the marker to predict individuals risk of death due to lung cancer. To this end, we estimated the time-dependent ROC curve and the corresponding AUC at three different times: 60, 120, and 180 days. Similarly, for kidney data, the accuracy of marker to predict the risk of recurrence of kidney infection was evaluated using ROC and AUC estimated at prediction times: 120, 180, and 240 days. The results of lung cancer and kidney data analyses showed that the derived markers have good overall predictive ability. In addition, both analyses showed that the Beran approach that does not acknowledge the correlated nature into account gives larger ROC curve and AUC estimates than the proposed estimator.

To conclude, we showed that the proposed time-dependent ROC curve and the corresponding AUC estimation method which take the correlated nature of event times into account has better finite sample performance than the existing Beran

approach that does not acknowledge the presence correlation. Therefore, we recommend to use the proposed estimator for correlated event times data.

R functions that implement the proposed method are available from the corresponding author, and will be published as an open-source package R package frailtyROC.

Acknowledgments

The authors would like to thank the two anonymous reviewers and the editor for their comments and suggestions, which substantially improved the quality of this manuscript.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. GQ3

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is partially based upon research supported by the South Africa National Research Foundation (NRF) and South Africa Medical Research Council (SAMRC) (South Africa DST-NRF-SAMRC SARChI Research Chair in Biostatistics, Grant number 114613). Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF and SAMRC.

ORCID iD

Kassu M Beyene https://orcid.org/0000-0002-2067-6054 GQ6

References

1. Cox DR. Regression models and life-tables. *J R Stat Soc Ser B (Methodological)* 1972; **34**: 187–220.
2. Heagerty PJ, Lumley T and Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; **56**: 337–344.
3. Heagerty PJ and Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics* 2005; **61**: 92–105.
4. Etzioni R, Pepe M, Longton G et al. Incorporating the time dimension in receiver operating characteristic curves: A case study of prostate cancer. *Med Decis Making* 1999; **19**: 242–251.
5. Li L, Greene T and Hu B. A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data. *Stat Methods Med Res* 2018; **27**: 2264–2278.
6. Blanche P, Dartigues JF and Jacqmin-Gadda H. Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring. *Biometrical J* 2013; **55**: 687–704.
7. Martínez-Camblor P, Bayón GF and Pérez-Fernández S. Cumulative/dynamic ROC curve estimation. *J Stat Comput Simul* 2016; **86**: 3582–3594.
8. Martínez-Camblor P and Pardo-Fernández JC. Smooth time-dependent receiver operating characteristic curve estimators. *Stat Methods Med Res* 2018; **27**: 651–674.
9. Beyene KM, El Ghouch A and Oulhaj A. On the validity of time-dependent AUC estimation in the presence of cure fraction. *Biometrical J* 2019; **61**: 1430–1447.
10. Díaz-Coto S, Martínez-Camblor P and Corral-Blanco NO. Cumulative/dynamic ROC curve estimation under interval censorship. *J Stat Comput Simul* 2020; **90**: 1570–1590.
11. Wu Y and Cook RJ. Assessing the accuracy of predictive models with interval-censored data. *Biostatistics* 2022; **23**: 18–33.
12. Beyene KM and El Ghouch A. Smoothed time-dependent receiver operating characteristic curve for right censored survival data. *Stat Med* 2020; **39**: 3373–3396.
13. Beyene KM and El Ghouch A. Time-dependent ROC curve estimation for interval-censored data. *Biometrical J* 2022; **64**: 1056–1074.
14. Keyfitz N and Littman G. Mortality in a heterogeneous population. *Popul Stud (Camb)* 1979; **33**: 333–342.
15. Henderson R and Oman P. Effect of frailty on marginal regression estimates in survival analysis. *J R Stat Soc Ser B: Stat Methodol* 1999; **61**: 367–379.
16. Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 1978; **65**: 141–151.
17. Vaupel JW, Manton KG and Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 1979; **16**: 439–454.
18. McGilchrist C and Aisbett C. Regression with frailty in survival analysis. *Biometrics* 1991; **47**: 461–466.
19. Hougaard P. Modelling heterogeneity in survival data. *J Appl Probab* 1991; **28**: 695–701.
20. Hougaard P. Frailty models for survival data. *Lifetime Data Anal* 1995; **1**: 255–273.
21. Wei LJ, Lin DY and Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc* 1989; **84**: 1065–1073.

22. Duchateau L, Janssen P, Lindsey P et al. The shared frailty model and the power for heterogeneity tests in multicenter trials. *Comput Stat Data Anal* 2002; **40**: 603–620.
23. Duchateau L, Janssen P, Kezic I et al. Evolution of recurrent asthma event rate over time in frailty models. *J R Stat Soc Ser C: Appl Stat* 2003; **52**: 355–363.
24. Duchateau L and Janssen P. Penalized partial likelihood for frailties and smoothing splines in time to first insemination models for dairy cows. *Biometrics* 2004; **60**: 608–614.
25. Therneau TM, Grambsch PM, Therneau TM et al. Frailty models. *Model Survival Data: Extending Cox Model* 2000; 231–260.
26. Hougaard P and Hougaard P. *Analysis of multivariate survival data*. Vol. **564**. Springer, 2000.
27. Hanagal DD. Chapter 9 - frailty models in public health. In Srinivasa Rao AS, Pyne S and Rao C (eds.) *Disease Modelling and Public Health, Part B*, *Handbook of Statistics*, volume 37. Elsevier, 2017. pp. 209–247.
28. Duchateau L and Janssen P. *The frailty model*. Springer, 2008.
29. Rondeau V, Gonzalez JR, Mazroui Yet al. *frailtypack*: Shared, Joint (Generalized) Frailty Models; Surrogate Endpoints, 2022. https://CRAN.R-project.org/package=frailtypack. R package version 3.5.0.
30. Lloyd CJ and Yong Z. Kernel estimators of the ROC curve are better than empirical. *Stat Probab Lett* 1999; **44**: 221–228.
31. Beran R. Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley, 1981.
32. Loprinzi CL, Laurie JA, Wieand HS et al. Prospective evaluation of prognostic variables from patient-completed questionnaires. north central cancer treatment group. *J Clin Oncol* 1994; **12**: 601–607.
33. Therneau TM. *A Package for Survival Analysis in R*, 2023. https://CRAN.R-project.org/package=survival. R package version 3.5-5.

AQ1