

Research Article

Gaussian Copula Regression Modeling for Marker Classification Metrics with Competing Risk Outcomes

Alejandro Román Vásquez ¹, Gabriel Escarela ¹,
Hortensia Josefina Reyes-Cervantes ² and Gabriel Núñez-Antonio ¹

¹Departamento de Matemáticas, Universidad Autónoma Metropolitana Unidad Iztapalapa, CDMX, Mexico

²Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, Puebla, Mexico

Correspondence should be addressed to Alejandro Román Vásquez; arv@xanum.uam.mx

Received 8 December 2023; Revised 27 December 2023; Accepted 3 January 2024; Published 19 January 2024

Academic Editor: Niansheng Tang

Copyright © 2024 Alejandro Román Vásquez et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Decisions regarding competing risks are usually based on a continuous-valued marker toward predicting a cause-specific outcome. The classification power of a marker can be summarized using the time-dependent receiver operating characteristic curve and the corresponding area under the curve (AUC). This paper proposes a Gaussian copula-based model to represent the joint distribution of the continuous-valued marker, the overall survival time, and the cause-specific outcome. Then, it is used to characterize the time-varying ROC curve in the context of competing risks. Covariate effects are incorporated by linking linear components to the skewed normal distribution for the margin of the marker, a parametric proportional hazards model for the survival time, and a logit model for the cause of failure. Estimation is carried out using maximum likelihood, and a bootstrap technique is implemented to obtain confidence intervals for the AUC. Information-criteria strategies are employed to find a parsimonious model. The performance of the proposed model is evaluated in simulation studies, considering different sample sizes and censoring distributions. The methods are illustrated with the reanalysis of a prostate cancer clinical trial. The joint regression strategy produces a straightforward and flexible model of the time-dependent ROC curve in the presence of competing risks, enhancing the understanding of how covariates may affect the discrimination of a marker.

1. Introduction

In clinical medical practice, decisions about personalized treatments are generally guided by markers that can discriminate between different levels of risk of a specific type of death. In time-to-event data, an individual can be exposed to two types of failure, the so-called competing risks; here, the phenomenon of interest is to assess the classification capacity of the marker to predict a cause-specific outcome within a certain period, leading to a more objective choice of treatment of a condition. In this context, Saha and Heagerty [1] defined cumulative/dynamic discriminatory measures to evaluate the prediction accuracy of a marker M to distinguish

between the subjects who experience the cause-specific event D within survival time T and those who do not. Specifically, they define the cumulative true positive rate TP_j^C as the probability of a high marker value among those subjects who experience the event $D = j$ ($j = 1, 2$) within time t and the dynamic false positive rate FP^D as the probability of a high marker value among subjects who are event-free through time t . The cumulative/dynamic $ROC_j^{C/D}$ curve for each event type can be obtained by plotting TP_j^C versus FP^D at time t , for all possible values of the marker. The corresponding area under the curve $AUC_j^{C/D}$ is then used as a global summary of the classification power of the marker.

Inference methods have mainly used nonparametric approaches for the estimation of $ROC_j^{C/D}$ [1–3]. The primary limitation of such formulations is the difficulty that occurs when it comes to incorporating covariates into the analysis. Zheng et al. [4] obtained expressions of TP_j^C and FP^D in terms of the cumulative incidence functions (CIF's), which are represented using cause-specific regression hazard functions models for the different failure types; however, it has been documented that standard specifications for those models, including the proportional hazards assumption they employ, lead to model selection issues since testing for covariate effects on the CIF's is not possible [5], thus hampering the comparison between different ROC curves.

In this article, we employ the joint regression analysis using Gaussian copulas introduced in [6] to model the joint distribution of (M, T, D) . This approach represents the joint model in terms of a trivariate Gaussian copula distribution, subsequently utilized to characterize the $ROC_j^{C/D}$ curve. The regression model is constructed by specifying the marginal distribution of the marker with a skewed normal regression model, the marginal distribution of the survival time with a parametric proportional hazards regression model, and the marginal distribution of the cause of failure with a logistic regression model. Maximum likelihood is employed with a constraint that ensures that the dependence parameters in the Gaussian copula yield a positive-definitive correlation matrix, which must be satisfied. Also, information-criterion metrics are employed for variable selection, leading to the formulation of a parsimonious model. The joint regression modeling yields a straightforward and flexible model for the time-dependent ROC curve in the presence of competing risks. This direct approach enhances comprehension regarding covariates' impact on a marker's discriminatory ability, representing the paper's principal contribution.

The rest of the paper is organized as follows. Section 2 defines the parametric Gaussian copula-based model, the copula relation to time-dependent discrimination metrics, and the marginal specifications. In Section 3, the maximum likelihood estimation procedure is described. Section 4 provides some simulation results to explore the properties of the proposed estimators under different sample sizes and censoring scenarios. Section 5 illustrates the method with data from a well-known prostate cancer dataset [7]. The paper culminates with a conclusion.

2. Methods

2.1. Gaussian Copula-Based Models and the Relation with ROC and AUC. In applied research, the copula model represents a convenient and flexible way to construct multivariate distributions that embody wide ranges of dependence structures while allowing for the specification of the underlying marginal distributions [8]. Sklar's theorem provides the decomposition of any continuous m -variate distribution function F in terms of the marginal cumulative distribution functions (CDF's) F_1, \dots, F_m and the associated copula $C: [0, 1]^m \rightarrow [0, 1]$, which is a m -variate distribution function with uniform margins $U(0, 1)$ [9], and thus the representation of the multivariate distribution function (MDF) is given by $F(y_1, \dots, y_m) = C(F_1(y_1), \dots, F_m(y_m))$; also, the theorem states that given F , the copula is unique on $\text{Range}(F_1) \times \dots \times \text{Range}(F_m)$, implying that the copula is unique if the marginals F_1, \dots, F_m are continuous.

Given its flexibility, analytical tractability, and capacity to characterize rich dependence structures, similar to the multivariate Gaussian distribution, this study adopted the trivariate Gaussian copula, which is defined by

$$C(F_1(y_1), F_2(y_2), F_3(y_3); \Gamma) = \Phi_3(\Phi^{-1}\{F_1(y_1)\}, \Phi^{-1}\{F_2(y_2)\}, \Phi^{-1}\{F_3(y_3)\}; \Gamma), \tag{1}$$

where Φ_3 is the standard normal trivariate distribution function (zero means and unit variances), Φ^{-1} is the quantile function corresponding to the univariate standard normal distribution, and Γ is a nonnegative symmetric positive definite correlation matrix given by

$$\Gamma = \begin{pmatrix} 1 & \gamma_{12} & \gamma_{13} \\ \gamma_{12} & 1 & \gamma_{23} \\ \gamma_{13} & \gamma_{23} & 1 \end{pmatrix}, \tag{2}$$

where each dependence parameter $\gamma_{ij} = \text{corr}(\Phi^{-1}\{F_i(y_i)\}, \Phi^{-1}\{F_j(y_j)\})$ takes values in $[-1, 1]$.

Following the Radon–Nikodym derivative methodology introduced in [6], the joint density function of (M, T, D) corresponding to the copula model in equation (1) is represented by

$$\begin{aligned} f_{M,T,D}(m, t, j; \Gamma) \\ = f_M(m) * f_T(t) \sum_{j_d=1}^2 (-1)^{j_d} C^*(F_M(m), F_T(t), u_{d,j_d}; \Gamma), \end{aligned} \tag{3}$$

where $f_Y(\cdot)$ and $F_Y(\cdot)$ are, respectively, the density and CDF functions of the random variable Y , $u_{d,1} = F_D(j - 1)$, $u_{d,2} = F_D(j)$, $F_D(j)$ is the CDF of the event type,

$$C^*(F_M(m), F_T(t), u_{d,jd}; \Gamma) = (2\pi)^{-1/2} |\Gamma|^{-1/2} \int_{-\infty}^{\Phi^{-1}(u_{d,jd})} \exp\left\{\frac{1}{2}(\mathbf{q}^T, z)\Gamma^{-1}(\mathbf{q}^T, z)^T - \frac{1}{2}\mathbf{q}^T\mathbf{q}\right\} dz, \tag{4}$$

and $\mathbf{q} = (\Phi^{-1}\{F_M(m)\}, \Phi^{-1}\{F_T(t)\})^T$.

Given the cumulative true positive rate for cause j , defined by $\text{TP}_j^C(m, t) = \Pr(M > m \mid T \leq t, D = j)$, and the dynamic false positive rate, defined by $\text{FP}^D(m, t) = \Pr(M > m \mid T > t)$, the time-dependent cumulative/dynamic ROC curve at time t for cause j is expressed as follows:

$$\text{ROC}_j^{C/D}(q, t) = \text{TP}_j^C\left([\text{FP}^D]^{-1}(q, t), t\right), \quad \text{for } q \in [0, 1], \tag{5}$$

where $[\text{FP}^D]^{-1}(q, t) = \inf_c \{c: \text{FP}^D(q, t) \leq c\}$. The corresponding area under the curve at time t for cause j , $\text{AUC}_j^{C/D}(t) = \int_0^1 \text{ROC}_j^{C/D}(x, t) dx$, represents the probability that an individual with a positive diagnosis has a higher value of the marker than an individual with a negative diagnosis [10].

The discrimination metrics can be expressed in terms of the joint model equation (4):

$$\text{TP}_j^C(m, t) = 1 - \frac{C[F_M(m), F_T(t), F_D(j)] - C[F_M(m), F_T(t), F_D(j - 1)]}{C[F_T(t), F_D(j)] - C[F_T(t), F_D(j - 1)]}, \tag{6}$$

$$\text{FP}^D(m, t) = \frac{1 - F_T(t) - F_M(m) + C(F_M(m), F_T(t))}{1 - F_T(t)}. \tag{7}$$

Here, $C[F_T(t), F_D(j)]$ and $C[F_M(m), F_T(t)]$ represent the marginal bivariate Gaussian copula functions associated with the trivariate Gaussian copula, following the marginalization property of the multivariate Gaussian copula [11]. The derivation of equations (6) and (7) is detailed in Appendix A.

2.2. Marginal Specifications. In practice, markers tend to exhibit skewed distributions [12]. Thus, in this study, the margin of M follows the skewed-normal distribution, denoted here by $M \sim SN(\xi_M, \omega_M^2, \alpha_M)$, whose density function is given by [13]

$$f_M(m; \xi_M, \omega_M^2, \alpha_M) = \frac{2}{\omega_M} \varphi\left(\frac{m - \xi_M}{\omega_M}\right) \Phi\left(\alpha_M \frac{m - \xi_M}{\omega_M}\right), \quad m \in (-\infty, \infty), \tag{8}$$

where φ and Φ are, respectively, the density and CDF of a standard normal random variable, $\xi_M \in (-\infty, \infty)$ is the location parameter, $\omega_M^2 \in (0, \infty)$ is the scale parameter, and $\alpha_M \in (-\infty, \infty)$ is the slant parameter. Explanatory variables are incorporated by linking location parameter to a linear component, namely, $\xi_M = \boldsymbol{\alpha}^T \mathbf{w}$, where $\boldsymbol{\alpha}$ and \mathbf{w} are vectors of parameters and covariates, respectively.

For modeling the margin of T , this study employed the proportional hazards model, which is given by $F_T(t; \boldsymbol{\Lambda}, \mathbf{x}) = 1 - \{1 - F_0(t; \boldsymbol{\Lambda})\}^{\exp(\boldsymbol{\beta}^T \mathbf{x})}$, where $F_0(t; \boldsymbol{\Lambda})$ is

a baseline distribution with vector of parameters $\boldsymbol{\Lambda}$, and $\boldsymbol{\beta}$ and \mathbf{x} are vectors of parameters and covariates, respectively; here, \mathbf{x} does not contain the intercept and does not necessarily contain the same variables in \mathbf{w} . Since it has shown good fits to survival data in various applied analyses [14], this study used the Weibull distribution as the baseline; that is, $F_0(t; \nu, \kappa) = 1 - \exp(-(t/\kappa)^\nu)$, where $t \in (0, \infty)$, and $\nu \in (0, \infty)$ and $\kappa \in (0, \infty)$ are, respectively, the shape and scale parameters. The incorporation of covariates \mathbf{x} to the margin was specified with the following which will be referred to as the survival component:

$$S_T(t; \nu, \kappa, \mathbf{x}) = \{S_0(t; \nu, \kappa)\}^{\exp(\boldsymbol{\beta}^T \mathbf{x})}, \quad (9)$$

where $S_0(t; \nu, \kappa)$ is the baseline survival function (which follows a Weibull distribution) and $\boldsymbol{\beta}$ is the parameter vector (which does not include intercept).

Similar to various competing risk regression formulations for the cause of the event [15, 16], the marginal model for $D = 1$ was specified with a logistic model, which is given by

$$\Pr(D=1 | \mathbf{z}) = \frac{\exp(\boldsymbol{\delta}_1^T \mathbf{z})}{1 + \exp(\boldsymbol{\delta}_1^T \mathbf{z})}, \quad (10)$$

where \mathbf{z} is the vector of covariates, not necessarily containing the same variables in \mathbf{w} and \mathbf{x} , and $\boldsymbol{\delta}_1$ is a vector representing the model parameters when $D = 1$.

In the proposed fully parametric Gaussian copula regression model, the cumulative true positive rate for cause j and the dynamic false positive rate are now expressed

conditionally on the covariates: $\text{TP}_j^C(m, t; \mathbf{w}, \mathbf{x}, \mathbf{z})$ and $\text{FP}^D(m, t; \mathbf{w}, \mathbf{x}, \mathbf{z})$. Consequently, the time-dependent cumulative/dynamic ROC curve for event type j and the corresponding AUC are also conditional expressions, $\text{ROC}_j^{C/D}(q, t; \mathbf{w}, \mathbf{x}, \mathbf{z})$ and $\text{AUC}_j^{C/D}(t; \mathbf{w}, \mathbf{x}, \mathbf{z})$, respectively.

3. Estimation

Consider the vector of marker, survival time, and cause of death (M_i, T_i, D_i) for the i -th subject in the study, $i = 1, \dots, n$. The observed dataset is $\{(m_i, t_i, d_i, \mathbf{y}_i); i = 1, \dots, n\}$, where m_i is the observed value of M_i , t_i is the time to either failure or censoring, d_i identifies the type of failure, taking values 1 and 2 for the observed survival times and 0 for the censored times, and \mathbf{y}_i is a vector of covariates. The full likelihood function is given by

$$L(\boldsymbol{\omega}) = \prod_{i=1}^n f_{M,T,D}(m_i, t_i, 1)^{\varphi_i} f_{M,T,D}(m_i, t_i, 2)^{\psi_i} \left(\int_{t_i}^{\infty} f_{M,T}(m_i, x) dx \right)^{1 - \varphi_i - \psi_i}, \quad (11)$$

where $\boldsymbol{\omega}$ is a vector containing all parameters, $\varphi_i = I_{\{1\}}(d_i)$, and $\psi_i = I_{\{2\}}(d_i)$; here $f_{M,T}(m, t)$ can be obtained by marginalizing the trivariate joint model [11].

Due to the complexity of the likelihood function formulation, we propose a numerical procedure for the corresponding estimation process, which involves employing an optimization algorithm called PRAXIS (Principal Axis), developed by Richard Brent [17]. The optimization algorithm is implemented in the R package `nloptr` using the function `nloptr()`.

Maximizing the log-likelihood function in equation (11) requires setting lower and upper bound constraints for each dependence parameter, γ_{MT} , γ_{TD} , and γ_{MD} , to ensure the positive definiteness of the correlation matrix Γ . To achieve this, we developed a heuristic algorithm to create box constraints for each dependence parameter, ensuring $\text{lb}_{\text{MT}} < \text{ub}_{\text{MT}}$, $\text{lb}_{\text{MD}} < \text{ub}_{\text{MD}}$, and $\text{lb}_{\text{TD}} < \text{ub}_{\text{TD}}$. A detailed explanation of the heuristic algorithm can be found in Appendix B. Log links are used for the strictly positive parameters.

4. Simulation Experiments

To evaluate the methods' performance and robustness, we generate 200 samples of $n = 200$ and $n = 400$ from the trivariate Gaussian and Student t copulas with margins $U[0,1]$. The dependence parameters are set to $\gamma_{\text{MT}} = -0.2$, $\gamma_{\text{MD}} = -0.7$, and $\gamma_{\text{TD}} = 0$. The copulas are implemented using the functions `normalCopula()` and `tCopula()` from the `copula` package.

To generate variates following the skew-normal distribution with parameters $\boldsymbol{\alpha} = (\alpha_0, \alpha_1) = (10, 0.5)$, $\omega_M = 2$, and $\alpha_M = 2$, we employ the inverse distribution function

method [8]. For the proportional hazards model, we follow the approach proposed in [18] to generate variates conforming to the proportional hazards model with a Weibull baseline survival function, setting parameter values as $\beta_1 = -0.5$, $\nu = 2$, and $\kappa = 2$. These marginals share the same covariate $x = w$, which is assumed to follow a normal distribution $N(10, 1)$. We set the logistic model to $\boldsymbol{\delta}_1 = (-1, 1)$, with a dichotomous covariate z that is generated from the binomial distribution $B(1, 0.5)$.

The censoring time is simulated from a uniform distribution $U(0, c)$, where c takes values of 200 and 85. The observation is considered censored if a simulated failure time exceeds the corresponding simulated censoring time. The average percent censored for $c = 200$ and $c = 85$ was around 11% and 26%, respectively.

Tables 1 and 2 present empirical bias, standard error (SE), and mean square error (MSE) of the coefficients obtained from the Gaussian copula regression model for simulations from both copulas. Generally, empirical biases and standard errors are relatively small compared to the actual parameters, resulting in a low mean square error, which suggests that the estimation process is appropriate for both sample sizes.

Overall, all three performance measures tend to approach zero with an increased sample size. Notably, an increase in censoring has minimal impact on the estimation process, as evidenced by the only marginal increase in mean square error. Additionally, although there is a slight increase in the mean square error for estimators derived from simulations of a trivariate Student t copula, this increase is not drastic. Consequently, the Gaussian copula regression model exhibits favorable estimation properties in the

TABLE 1: Empirical bias, standard error (SE), and mean square error (MSE) of the estimators of the parametric Gaussian copula regression model for simulations generated from a Gaussian copula, with sample sizes 200 and 400 and censoring distributions $U(0, 200)$ and $U(0, 85)$.

| Sample size and censoring | Parameters | Bias | SE | MSE |
|---------------------------------|---------------------------------|-----------------|---------|--------|
| $n = 200$ censoring $U(0, 200)$ | $\alpha_0 = 10$ | -0.0078 | 0.9158 | 0.8387 |
| | $\alpha_1 = 0.5$ | -0.0030 | 0.0878 | 0.0077 |
| | $\log(\omega) = 0.6931$ | 0.0094 | 0.0977 | 0.0096 |
| | $\alpha = 2$ | -0.1236 | 0.6693 | 0.4632 |
| | $\beta_1 = -0.5$ | -0.0009 | 0.0627 | 0.0039 |
| | $\log(\gamma) = 0.6931$ | -0.0928 | 0.4011 | 0.1695 |
| | $\log(\kappa) = 0.6931$ | -0.0250 | 0.0822 | 0.0074 |
| | $\delta_0 = -1$ | 0.0270 | 0.2416 | 0.0591 |
| | $\delta_1 = 1$ | -0.0393 | 0.2946 | 0.0883 |
| | $\gamma_{MT} = -0.2$ | -0.0101 | 0.0772 | 0.0061 |
| | $\gamma_{MD} = -0.7$ | 0.0234 | 0.0565 | 0.0037 |
| | $\gamma_{TD} = 0$ | 0.0013 | 0.1080 | 0.0117 |
| | $n = 400$ censoring $U(0, 200)$ | $\alpha_0 = 10$ | -0.0634 | 0.6162 |
| $\alpha_1 = 0.5$ | | 0.0028 | 0.0593 | 0.0035 |
| $\log(\omega) = 0.6931$ | | 0.0083 | 0.0718 | 0.0052 |
| $\alpha = 2$ | | -0.0394 | 0.4825 | 0.2344 |
| $\beta_1 = -0.5$ | | 0.0066 | 0.0393 | 0.0016 |
| $\log(\gamma) = 0.6931$ | | -0.0431 | 0.2895 | 0.0857 |
| $\log(\kappa) = 0.6931$ | | -0.0176 | 0.0638 | 0.0044 |
| $\delta_0 = -1$ | | 0.0143 | 0.1670 | 0.0281 |
| $\delta_1 = 1$ | | -0.0024 | 0.2003 | 0.0401 |
| $\gamma_{MT} = -0.2$ | | 0.0017 | 0.0488 | 0.0024 |
| $\gamma_{MD} = -0.7$ | | 0.0209 | 0.0472 | 0.0027 |
| $\gamma_{TD} = 0$ | | -0.0066 | 0.0668 | 0.0045 |
| $n = 200$ censoring $U(0, 85)$ | | $\alpha_0 = 10$ | -0.0731 | 0.9864 |
| | $\alpha_1 = 0.5$ | -0.0025 | 0.0943 | 0.0089 |
| | $\log(\omega) = 0.6931$ | 0.0098 | 0.1173 | 0.0139 |
| | $\alpha = 2$ | -0.1441 | 0.8617 | 0.7633 |
| | $\beta_1 = -0.5$ | -0.0006 | 0.0660 | 0.0044 |
| | $\log(\gamma) = 0.6931$ | -0.1841 | 0.4170 | 0.2078 |
| | $\log(\kappa) = 0.6931$ | -0.0544 | 0.0881 | 0.0107 |
| | $\delta_0 = -1$ | -0.0143 | 0.2482 | 0.0618 |
| | $\delta_1 = 1$ | 0.0069 | 0.3008 | 0.0905 |
| | $\gamma_{MT} = -0.2$ | 0.0021 | 0.0742 | 0.0055 |
| | $\gamma_{MD} = -0.7$ | 0.0177 | 0.0653 | 0.0046 |
| | $\gamma_{TD} = 0$ | -0.0060 | 0.1032 | 0.0107 |
| | $n = 400$ censoring $U(0, 85)$ | $\alpha_0 = 10$ | 0.1473 | 0.7044 |
| $\alpha_1 = 0.5$ | | -0.0214 | 0.0689 | 0.0052 |
| $\log(\omega) = 0.6931$ | | -0.0081 | 0.0773 | 0.0060 |
| $\alpha = 2$ | | -0.0741 | 0.5356 | 0.2924 |
| $\beta_1 = -0.5$ | | 0.0083 | 0.0441 | 0.0020 |
| $\log(\gamma) = 0.6931$ | | -0.2557 | 0.3006 | 0.1558 |
| $\log(\kappa) = 0.6931$ | | -0.0746 | 0.0635 | 0.0096 |
| $\delta_0 = -1$ | | -0.0260 | 0.1569 | 0.0253 |
| $\delta_1 = 1$ | | 0.0161 | 0.1885 | 0.0358 |
| $\gamma_{MT} = -0.2$ | | -0.0012 | 0.0571 | 0.0033 |
| $\gamma_{MD} = -0.7$ | | 0.0184 | 0.0473 | 0.0026 |
| $\gamma_{TD} = 0$ | | -0.0088 | 0.0710 | 0.0051 |

considered moderate sample sizes and appears robust concerning the Student t copula.

5. Prostate Cancer Study Data

This study utilizes the dataset described in [7] to exemplify the proposed methods' use. The dataset comprises an approximate five-year follow-up of a cohort consisting of 506 patients

diagnosed with prostate cancer of stages III and IV. These patients were recruited in a clinical trial conducted between 1967 and 1969. A total of 4.54% of patients were excluded from the analysis due to incomplete covariate information. Among the remaining cohort, 125 patients (25.88%) succumbed to prostate cancer ($D = 1$), 219 patients (45.34%) died from other causes ($D = 2$), and the remaining 139 patients (28.78%) had censored survival times.

TABLE 2: Empirical bias, standard error (SE), and mean square error (MSE) of the estimators of the parametric Gaussian copula regression model for simulations generated from Student t copula, with sample sizes 200 and 400 and censoring distributions $U(0, 200)$ and $U(0, 85)$.

| Sample size and censoring | Parameters | Bias | SE | MSE |
|---------------------------------|---------------------------------|-----------------|---------|--------|
| $n = 200$ censoring $U(0, 200)$ | $\alpha_0 = 10$ | -0.1377 | 0.9418 | 0.9060 |
| | $\alpha_1 = 0.5$ | 0.0026 | 0.0871 | 0.0076 |
| | $\log(\omega) = 0.6931$ | 0.0266 | 0.1092 | 0.0126 |
| | $\alpha = 2$ | -0.0138 | 0.7599 | 0.0577 |
| | $\beta_1 = -0.5$ | 0.0050 | 0.0613 | 0.0038 |
| | $\log(\nu) = 0.6931$ | -0.0876 | 0.3737 | 0.1473 |
| | $\log(\kappa) = 0.6931$ | -0.0260 | 0.0774 | 0.0067 |
| | $\delta_0 = -1$ | -0.0084 | 0.2302 | 0.0531 |
| | $\delta_1 = 1$ | -0.0151 | 0.2874 | 0.0828 |
| | $\gamma_{MT} = -0.2$ | 0.0008 | 0.0908 | 0.0082 |
| | $\gamma_{MD} = -0.7$ | -0.0063 | 0.0678 | 0.0046 |
| | $\gamma_{TD} = 0$ | 0.0103 | 0.1062 | 0.0114 |
| | $n = 400$ censoring $U(0, 200)$ | $\alpha_0 = 10$ | 0.0665 | 0.7295 |
| $\alpha_1 = 0.5$ | | -0.0177 | 0.0710 | 0.0054 |
| $\log(\omega) = 0.6931$ | | 0.0211 | 0.0886 | 0.0083 |
| $\alpha = 2$ | | 0.0519 | 0.5557 | 0.3115 |
| $\beta_1 = -0.5$ | | 0.0091 | 0.0423 | 0.0019 |
| $\log(\nu) = 0.6931$ | | -0.2629 | 0.3089 | 0.1646 |
| $\log(\kappa) = 0.6931$ | | -0.0759 | 0.0651 | 0.0100 |
| $\delta_0 = -1$ | | -0.0193 | 0.1820 | 0.0335 |
| $\delta_1 = 1$ | | 0.0156 | 0.2148 | 0.0464 |
| $\gamma_{MT} = -0.2$ | | -0.0125 | 0.0619 | 0.0040 |
| $\gamma_{MD} = -0.7$ | | 0.0057 | 0.0479 | 0.0023 |
| $\gamma_{TD} = 0$ | | -0.0017 | 0.0849 | 0.0072 |
| $n = 200$ censoring $U(0, 85)$ | | $\alpha_0 = 10$ | 0.0109 | 0.9646 |
| | $\alpha_1 = 0.5$ | -0.0154 | 0.0934 | 0.0090 |
| | $\log(\omega) = 0.6931$ | 0.0280 | 0.1148 | 0.0140 |
| | $\alpha = 2$ | 0.0264 | 0.7735 | 0.5990 |
| | $\beta_1 = -0.5$ | -0.0061 | 0.0677 | 0.0046 |
| | $\log(\nu) = 0.6931$ | -0.2373 | 0.4156 | 0.2290 |
| | $\log(\kappa) = 0.6931$ | -0.0634 | 0.0860 | 0.0114 |
| | $\delta_0 = -1$ | 0.0379 | 0.2350 | 0.0567 |
| | $\delta_1 = 1$ | -0.0492 | 0.3089 | 0.0979 |
| | $\gamma_{MT} = -0.2$ | -0.0048 | 0.0940 | 0.0089 |
| | $\gamma_{MD} = -0.7$ | 0.0054 | 0.0670 | 0.0045 |
| | $\gamma_{TD} = 0$ | -0.0101 | 0.1153 | 0.0134 |
| | $n = 400$ censoring $U(0, 85)$ | $\alpha_0 = 10$ | -0.0146 | 0.6337 |
| $\alpha_1 = 0.5$ | | -0.0038 | 0.0636 | 0.0041 |
| $\log(\omega) = 0.6931$ | | 0.0164 | 0.0678 | 0.0049 |
| $\alpha = 2$ | | 0.0106 | 0.4484 | 0.2012 |
| $\beta_1 = -0.5$ | | -0.0025 | 0.0418 | 0.0018 |
| $\log(\nu) = 0.6931$ | | -0.0818 | 0.2821 | 0.0863 |
| $\log(\kappa) = 0.6931$ | | -0.0208 | 0.0586 | 0.0039 |
| $\delta_0 = -1$ | | 0.0084 | 0.1624 | 0.0264 |
| $\delta_1 = 1$ | | -0.0032 | 0.1945 | 0.0378 |
| $\gamma_{MT} = -0.2$ | | -0.0083 | 0.0610 | 0.0038 |
| $\gamma_{MD} = -0.7$ | | -0.0048 | 0.0462 | 0.0022 |
| $\gamma_{TD} = 0$ | | 0.0063 | 0.0759 | 0.0058 |

The underlying research problem consists of identifying a marker to then measure its discriminatory power. French et al. [19] suggested the utilization of a Cox regression model to formulate a composite marker, which is obtained as a weighted combination of both biomarkers and clinical variables, wherein the estimated regression coefficients serve as the weighting factors. In this study, the marker is derived as the linear predictor of

a cause-specific parametric Cox regression model [20], with the event of interest defined as death from prostate cancer. Specifically, the linear component of the cause-specific hazard for $D = 1$ is estimated by treating the event times of all individuals who failed due to cause $D = 2$ as censored. The fitting process is carried out utilizing the `survreg()` function from the survival package in R, considering a Weibull distribution.

A backward elimination approach that is guided by the Bayesian information criterion (BIC) was employed to systematically eliminate the least significant covariate in each iteration, aiming to derive the most parsimonious model. The final regression model incorporates the following covariates: PF (performance rating: 0, standard; 1, limitation of activity), HG (serum hemoglobin in g/100 ml), SZ (size of the primary lesion in cm^2), and SG (Gleason stage-grade category). The resulting composite marker is defined as $M_C = 0.5243 \times \text{PF} - 0.0144 \times \text{HG} + 0.2948 \times \text{SG} + 0.0381 \times \text{SZ}$. Figure 1 shows the histogram of the marker M_C , which exhibits a skewed shape. Superimposed are the Gaussian kernel-based nonparametric density and the fitted skewed-normal density.

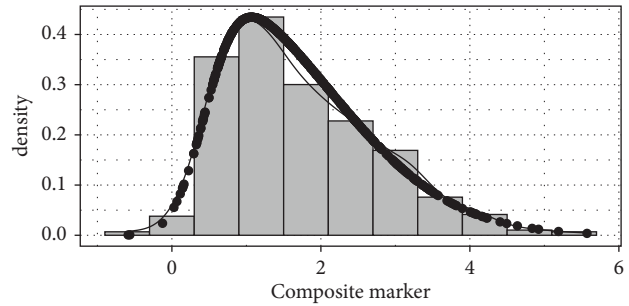


FIGURE 1: Distribution of the composite marker M_C . Non-parametric density with Gaussian kernel (line) is superimposed, as well as the fitted skewed-normal density (black dots).

5.1. Discriminatory Performance of the Composite Marker M_C . The discrimination power of the composite marker M_C is initially examined without considering covariates. Following the model fitting, it is observed that the sole parameter lacking individual significance is γ_{TD} , with an estimated value of -0.089 . A likelihood ratio test assesses the null hypothesis of $\gamma_{TD} = 0$ employing the chi-squared distribution approximation with one degree of freedom. With a p value of 0.2441, the null hypothesis cannot be rejected. Consequently, the simpler model with $\gamma_{TD} = 0$ is chosen on the grounds of parsimony. The parameter estimates of the joint simpler model and their standard errors are detailed in Table 3.

Analyzing the dependence parameters, the negative correlation $\gamma_{MT} = -0.207$ indicates that high values of M are associated with the early occurrence of any event (T represents overall survival time). Furthermore, because $\gamma_{MD} = -0.641$, high values of M are more strongly associated with cause 1 than cause 2. Thus, the estimated correlations suggest a marker with desirable classification properties.

To assess the appropriateness of the parametric model's fit, we perform a comparative analysis between the ROC curves obtained from the Gaussian copula regression model and the IPCW nonparametric ROC curves introduced by Blanche et al. [2]. The IPCW nonparametric approach possesses the advantage of not requiring any tuning parameters. Additionally, it can be easily implemented using the `timeROC()` function within the `timeROC` package in R. The plots of the parametric and nonparametric estimators of $\text{ROC}_1^{CID}(q, t)$ at time horizons 12, 36, and 60 months are depicted in Figure 2. It appears that the two curves are relatively close, suggesting that the fit of the parametric model is adequate.

5.2. Inclusion of Covariates. A Gaussian copula regression model is fitted to evaluate how the classification power of M_C changes on including covariates. We consider potential predictors as the remaining variables not included in the composite marker: RX (drug treatment), HX (history of cardiovascular disease: 0, no; 1, yes), AGE (age in years), and WT (weight index: weight in kg–height in $\text{cm} + 200$). A backward elimination process based on the BIC is employed to remove the least significant covariate sequentially, resulting in the most parsimonious regression model. Table 4 displays

TABLE 3: Coefficients (Coef.) and standard error (SE) of the best-fitting parametric copula model considering the composite marker M_C for the prostate cancer study data.

| Parameter | M_C | | |
|----------------|--------|-------|-----------|
| | Coef. | SE | p value |
| ξ | 0.568 | 0.077 | <0.0001 |
| $\log(\omega)$ | 0.584 | 0.054 | <0.0001 |
| α | 3.983 | 0.746 | <0.0001 |
| $\log(\nu)$ | 0.185 | 0.046 | 0.0001 |
| $\log(\kappa)$ | 3.288 | 0.048 | <0.0001 |
| δ_0 | -0.497 | 0.112 | <0.0001 |
| γ_{MT} | -0.207 | 0.044 | <0.0001 |
| γ_{MD} | -0.641 | 0.045 | <0.0001 |
| γ_{TD} | Zero | N.A. | N.A. |

BIC values from various regression models, indicating the predictors included in each marginal regression model during the elimination process. The chosen model is Model 9, as it exhibits the lowest BIC.

Table 5 presents the best-fitting parametric Gaussian copula regression model coefficients. Similar to the case without covariates, only one parameter is not statistically significant: γ_{TD} . A likelihood ratio test is conducted, considering a nested model with $\gamma_{TD} = 0$ as the null hypothesis. Based on a chi-squared distribution with one degree of freedom, the null hypothesis of a model with $\gamma_{TD} = 0$ cannot be rejected.

To assess the validity of the proportionality assumption for the factor AGE in the survival component, Figure 3 illustrates the plot of the empirical estimator of $\log\{-\log[\hat{S}(t)]\}$ as a function of $\log(t)$. As AGE is a continuous variable, two balanced groups are created: Group 1 comprises individuals with ages less than 74, and Group 2 comprises individuals with ages greater than or equal to 74. The log-minus-log plot shows approximate parallelism for most of the time, with some sparsity observed toward the end due to a limited number of observations in both groups at extended time points. This visual examination provides evidence in favor of the validity of the proportionality assumption in the Cox regression model.

Figure 4 illustrates $\text{ROC}_1^{CID}(q, t; \text{HX}, \text{AGE})$ based on the best-fitting parametric Gaussian copula regression model for the two levels of the variable HX (history of cardiovascular disease) and ages 60, 73, and 85 years old at a time horizon of

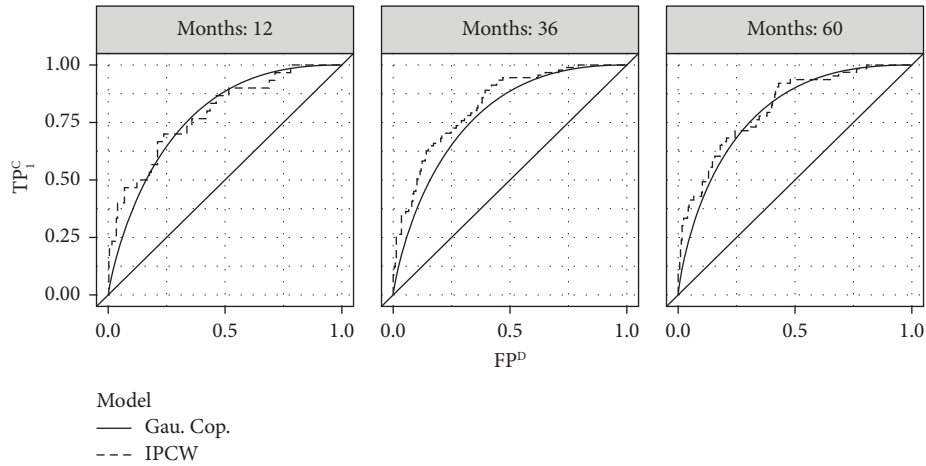


FIGURE 2: Gaussian copula regression model and nonparametric IPCW plots of $ROC_1^{C/D}(q, t)$ at time horizon 12 months, 36 months, and 60 months for the composite marker M_C in the absence of covariates.

TABLE 4: BICs obtained for the various models in each component of the joint model when using the backward elimination process.

| #Model | Variable M | Variable T | Variable D | BIC |
|--------|--------------------|--------------------|--------------------|----------|
| 1 | RX + AGE + WT + HX | RX + AGE + WT + HX | RX + AGE + WT + HX | 4,374.37 |
| 2 | RX + WT + HX | RX + AGE + WT + HX | RX + AGE + WT + HX | 4,364.91 |
| 3 | RX + WT + HX | RX + AGE + WT + HX | RX + AGE + HX | 4,355.39 |
| 4 | RX + WT + HX | AGE + WT + HX | RX + AGE + HX | 4,349.31 |
| 5 | RX + WT + HX | AGE + WT | RX + AGE + HX | 4,343.30 |
| 6 | WT + HX | AGE + WT | RX + AGE + HX | 4,338.88 |
| 7 | WT + HX | AGE + WT | AGE + HX | 4,334.68 |
| 8 | WT + HX | AGE | AGE + HX | 4,331.25 |
| 9 | HX | AGE | AGE + HX | 4,330.98 |
| 10 | HX | AGE | AGE | 4,338.33 |

TABLE 5: Coefficients (Coef.) and standard error (SE) of the best-fitting parametric Gaussian copula regression model considering the composite marker M_C for the study of prostate cancer study data.

| Parameter | Coef. | SE | p value |
|-------------------------|---------|--------|-----------|
| M_C | | | |
| α_0 | 0.8904 | 0.1054 | <0.0001 |
| α_{HX} | -0.4462 | 0.1021 | <0.0001 |
| $\log(\omega)$ | 0.5164 | 0.0565 | <0.0001 |
| α | 3.346 | 0.6422 | <0.0001 |
| $\log(\nu)$ | 0.1933 | 0.046 | <0.0001 |
| $\log(\kappa)$ | 4.5952 | 0.4928 | <0.0001 |
| β_{AGE} | 0.0221 | 0.0083 | 0.0077 |
| δ_0 | 4.1075 | 1.1176 | 0.0002 |
| δ_{AGE} | -0.0588 | 0.0155 | 0.0001 |
| δ_{HX} | -0.85 | 0.2361 | 0.0003 |
| γ_{MT} | -0.1895 | 0.0511 | 0.0002 |
| γ_{MD} | -0.5973 | 0.0505 | <0.0001 |
| γ_{TD} | -0.039 | 0.071 | 0.584 |
| $M_C (\gamma_{TD} = 0)$ | | | |
| α_0 | 0.8870 | 0.1053 | <0.0001 |
| α_{HX} | -0.4464 | 0.1021 | <0.0001 |
| $\log(\omega)$ | 0.5193 | 0.0563 | <0.0001 |
| α | 3.353 | 0.6439 | <0.0001 |
| $\log(\nu)$ | 0.193 | 0.046 | <0.0001 |

TABLE 5: Continued.

| Parameter | Coef. | SE | <i>p</i> value |
|----------------|---------|--------|----------------|
| $\log(\kappa)$ | 4.5971 | 0.4917 | <0.0001 |
| β_{AGE} | 0.0222 | 0.0082 | 0.0069 |
| δ_0 | 4.1287 | 1.1157 | 0.0002 |
| δ_{AGE} | -0.0591 | 0.0155 | 0.0001 |
| δ_{HX} | -0.8564 | 0.236 | 0.0003 |
| γ_{MT} | -0.2019 | 0.0454 | <0.0001 |
| γ_{MD} | -0.6033 | 0.0489 | <0.0001 |
| γ_{TD} | Zero | N.A. | N.A. |

The model includes all the parameters on the left, whereas on the right, the parameter $\gamma_{TD} = 0$ is set to zero.

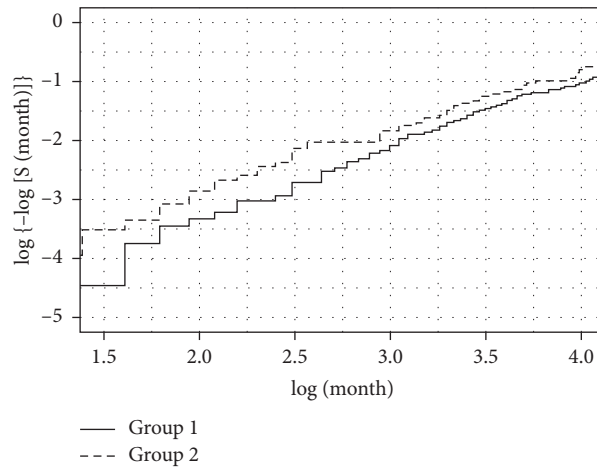


FIGURE 3: Log-minus-log plot for AGE in the prostate cancer study data. Group 1 (AGE <74 years old) and Group 2 (AGE ≥74 years old).

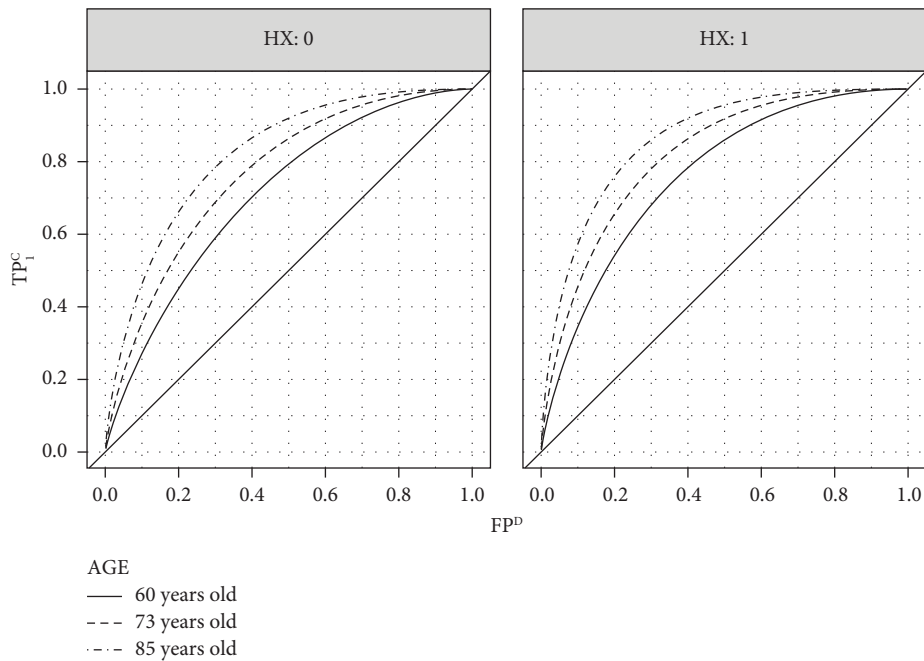


FIGURE 4: $ROC_1^{CID}(q, t; HX, AGE)$ curves based on the best-fitting parametric Gaussian copula regression model at a time horizon of 60 months at three different ages. On the left, the plot shows the curves for patients with no cardiovascular disease ($HX = 0$), and on the right, the plot shows the curves for patients with cardiovascular disease ($HX = 1$).

TABLE 6: $AUC_1^{C/D}(t; \text{HX}, \text{AGE})$ for a time horizon of 36 months for the composite marker M_C segmented by the two combinations of the categorical variable HX and three values of the variable AGE.

| AGE | HX | $AUC_1^{C/D}$ | Lower limit | Upper limit |
|-----|----|---------------|-------------|-------------|
| 60 | 0 | 0.704 | 0.643 | 0.753 |
| 60 | 1 | 0.757 | 0.697 | 0.812 |
| 73 | 0 | 0.764 | 0.706 | 0.813 |
| 73 | 1 | 0.813 | 0.763 | 0.863 |
| 85 | 0 | 0.819 | 0.751 | 0.868 |
| 85 | 1 | 0.859 | 0.812 | 0.907 |

Bootstrap 95% confidence intervals were estimated using 500 numbers of runs.

60 months. In this instance, the IPCW estimators of the time-dependent ROC curves are not considered due to the impracticality of incorporating covariates in this nonparametric approach. The corresponding $AUC_1^{C/D}(t; \text{HX}, \text{AGE})$ for the six combinations of predictors at the 60-month horizon are shown in Table 6. 95% bootstrap confidence intervals (with biased correction [21]) are estimated using 500 numbers of runs. Notably, insights reveal that older age and a history of cardiovascular disease enhance the classification performance of the composite marker M_C .

6. Conclusion

This study introduces a parametric Gaussian copula regression model for estimating time-dependent predictive accuracy metrics subject to right-censoring in the presence of competing risks. The marginal specifications encompass a normal asymmetric distribution for the marker, wherein a linear component is linked to the location parameter. Additionally, the model includes a parametric Cox regression model with a Weibull distribution for survival time and a logistic regression model for event type. The dependencies among these variables are captured using a trivariate Gaussian copula.

Simulation studies demonstrate that the Gaussian copula regression model exhibits favorable properties in moderate sample sizes and appears robust to slight deviations from the data-generating process. To validate the utility of our approach in a real-world scenario, we apply our proposed methodology to a well-known prostate cancer dataset. Based on the linear predictor of a cause-specific parametric Cox regression model, the composite biomarker M_C demonstrates moderate predictive performance in identifying patients at risk of dying from prostate cancer. Interestingly, older individuals with cardiovascular disease contribute to an increased discriminatory power of M_C .

The primary contribution of this paper is the straightforward incorporation of covariates into the estimation of ROC curves and AUCs through joint regression modeling. Considering predictors in the initial Gaussian copula regression model eliminates the need for stratified analyses or

adjustments through indirect methodologies, enabling a better understanding of how covariates may affect the classification power of a marker. The computation of the models relies predominantly on vector and matrix operations, along with readily available functions in the R language. The R code that implements the proposed methodology is available upon request. Furthermore, the introduced modeling strategy possesses additional favorable properties, demonstrating proficiency under moderate sample sizes and an ability to identify parsimonious models through information-criterion metrics.

Despite the compelling properties of the proposed methodology, some open questions remain. The linear dependence structure enforced by the Gaussian copula may be broadened by utilizing alternative copula models. Vine copula models are grounded in decomposing dependence into a series of interdependencies among (conditional) pairs of variables, affording the capability to manage heavy tails or asymmetric dependence [22]. We intend to investigate this approach in forthcoming research work.

Appendix

A. Discrimination Metrics in terms of Gaussian Copulas

This appendix shows how the expressions for the cumulative true positive rate for cause j and the dynamic false positive rate are obtained using the trivariate and bivariate Gaussian copulas.

Using the definition of conditional probability, $TP_j^C(m, t)$ and $FP^D(m, t)$ can be rewritten as

$$TP_j^C(m, t) = \frac{\Pr(M > m, T \leq t, D = j)}{\Pr(T \leq t, D = j)}, \quad (\text{A.1})$$

$$FP^D(m, t) = \frac{\Pr(M > m, T > t)}{\Pr(T > t)}. \quad (\text{A.2})$$

Next, employing the law of total probability, the numerator term in (A.1) can be expressed as

$$\Pr(M > m, T \leq t, D = j) = \Pr(T \leq t, D = j) - \Pr(M \leq m, T \leq t, D = j). \quad (\text{A.3})$$

Substituting this last expression in (A.1) leads to

$$TP_j^C(m, t) = 1 - \frac{\Pr(M \leq m, T \leq t, D = j)}{\Pr(T \leq t, D = j)} \tag{A.4}$$

As the random variable D is discrete, the expression for the term $\Pr(M \leq m, T \leq t, D = j)$ in (A.4) can be formulated as follows:

$$\begin{aligned} & \Pr(M \leq m, T \leq t, D \leq j) - \Pr(M \leq m, T \leq t, D \leq j - 1) \\ &= \Phi_3(\Phi^{-1}\{F_M(m)\}, \Phi^{-1}\{F_T(t)\}, \Phi^{-1}\{F_D(j)\}; \Gamma) \\ & \quad - \Phi_3(\Phi^{-1}\{F_M(m)\}, \Phi^{-1}\{F_T(t)\}, \Phi^{-1}\{F_D(j - 1)\}; \Gamma), \end{aligned} \tag{A.5}$$

$$\begin{aligned} & \Pr(T \leq t, D \leq j) - \Pr(T \leq t, D \leq j - 1) \\ &= \Phi_2(\Phi^{-1}\{F_T(t)\}, \Phi^{-1}\{F_D(j)\}; \gamma_{TD}) - \Phi_2(\Phi^{-1}\{F_T(t)\}, \Phi^{-1}\{F_D(j - 1)\}; \gamma_{TD}). \end{aligned} \tag{A.6}$$

The bivariate Gaussian copula arises through the marginalization of the trivariate Gaussian copula, leveraging the property that the multivariate Gaussian copula is closed under marginalization [11].

where the Sklar theorem [9] is used to represent the joint distribution function $F_{M,T,D}(m, t, j)$ in terms of a trivariate Gaussian copula $\Phi_3(\Phi^{-1}\{F_M(m)\}, \Phi^{-1}\{F_T(t)\}, \Phi^{-1}\{F_D(j)\}; \Gamma)$.

The expression $\Pr(T \leq t, D = j)$ in (A.4) can be rewritten as follows:

Using equations (A.5) and (A.6), $TP_j^C(m, t)$ can be defined in terms of copulas:

$$\begin{aligned} & TP_j^C(m, t) \\ &= 1 - \frac{C(F_M(m), F_T(t), F_D(j)) - C(F_M(m), F_T(t), F_D(j - 1))}{C(F_T(t), F_D(j)) - C(F_T(t), F_D(j - 1))}, \end{aligned} \tag{A.7}$$

where $C(F_M(m), F_T(t), F_D(j)) = \Phi_3(\Phi^{-1}\{F_M(m)\}, \Phi^{-1}\{F_T(t)\}, \Phi^{-1}\{F_D(j)\}; \Gamma)$ and $C(F_T(t), F_D(j)) = \Phi_2(\Phi^{-1}\{F_T(t)\}, \Phi^{-1}\{F_D(j)\}; \gamma_{TD})$.

Focusing on $FP^D(m, t)$, the numerator $\Pr(M > m, T > t)$ in (A.2) corresponds to a bivariate survival distribution function. This function can be expressed in terms of marginal distribution functions and the associated bivariate Gaussian copula [8]:

$$\begin{aligned} \Pr(M > m, T > t) &= S_{M,T}(m, t) = 1 - F_T(t) \\ & \quad - F_M(m) + C(F_M(m), F_T(t)), \end{aligned} \tag{A.8}$$

where $C(F_M(m), F_T(t))$ represents the marginal bivariate Gaussian copula $\Phi_2(\Phi^{-1}\{F_M(m)\}, \Phi^{-1}\{F_T(t)\}; \gamma_{MT})$.

Finally, expressing the denominator in (A.2) as $\Pr(T > t) = 1 - F_T(t)$, the dynamic false positive fraction becomes

$$FP^D(m, t) = \frac{1 - F_T(t) - F_M(m) + C(F_M(m), F_T(t))}{1 - F_T(t)} \tag{A.9}$$

B. Bound Constraints to Guarantee a Positive Definite Matrix Γ

Given the random vector (M, T, D) , the 3×3 correlation matrix Γ is characterized by three dependency parameters, namely, γ_{MT} , γ_{MD} , and γ_{TD} . A fundamental criterion for a matrix to be positive definite is that all of its principal minors are positive [23], implying that each upper-left submatrix must have a positive determinant. Concerning the correlation matrix, the upper 1×1 and upper 2×2 submatrices consistently exhibit positive determinants, assuming $\gamma_{MT} < 1$. Consequently, the focus shifts to ensuring that the determinant of the matrix, denoted as $\det(\Gamma)$, remains positive. The expression for $\det(\Gamma)$ is as follows:

$$\begin{aligned} \det(\Gamma) &= 1 - \gamma_{MT}^2 - \gamma_{MD}^2 - \gamma_{TD}^2 + 2\gamma_{MT}^2\gamma_{MD}^2\gamma_{TD}^2 \\ &= f_{\det}(\gamma_{MT}, \gamma_{MD}, \gamma_{TD}). \end{aligned} \tag{B.1}$$

To construct a three-dimensional region guaranteeing a positive determinant for all conceivable values of γ_{MT} , γ_{MD} , and γ_{TD} , we propose the following heuristic algorithm:

- (1) Calculate the initial values $\gamma_{MT_ini} = \text{corr}(\Phi^{-1}\{\widehat{F}_M(m)\}, \Phi^{-1}\{\widehat{F}_T(t)\})$, $\gamma_{MD_ini} = \text{corr}(\Phi^{-1}\{\widehat{F}_M(m)\}, \Phi^{-1}\{\widehat{F}_D(j)\})$, and $\gamma_{TD_ini} = \text{corr}(\Phi^{-1}\{\widehat{F}_T(t)\}, \Phi^{-1}\{\widehat{F}_D(j)\})$.
- (2) Add a small amount δ (for instance 0.001) to each initial value: $\gamma_{MT_ini} + \delta$, $\gamma_{MD_ini} + \delta$, and $\gamma_{TD_ini} + \delta$. If one of these values is greater than 1, the algorithm stops. Else, calculate the eight possible combinations of determinants: $f_{\det}(\gamma_{MT_ini}, \gamma_{MD_ini}, \gamma_{TD_ini})$, $f_{\det}(\gamma_{MT_ini} + \delta, \gamma_{MD_ini}, \gamma_{TD_ini})$, $f_{\det}(\gamma_{MT_ini}, \gamma_{MD_ini} + \delta, \gamma_{TD_ini})$, $f_{\det}(\gamma_{MT_ini}, \gamma_{MD_ini}, \gamma_{TD_ini} + \delta)$, $f_{\det}(\gamma_{MT_ini} + \delta, \gamma_{MD_ini} + \delta, \gamma_{TD_ini})$, $f_{\det}(\gamma_{MT_ini} + \delta, \gamma_{MD_ini}, \gamma_{TD_ini} + \delta)$, $f_{\det}(\gamma_{MT_ini}, \gamma_{MD_ini} + \delta, \gamma_{TD_ini} + \delta)$, and $f_{\det}(\gamma_{MT_ini} + \delta, \gamma_{MD_ini} + \delta, \gamma_{TD_ini} + \delta)$.

If all the determinants are positive, the functional relationship of equation (B.1) guarantees positive determinants for any point in the cubic region $[\gamma_{MT_ini}, \gamma_{MT_ini} + \delta] \times [\gamma_{MD_ini}, \gamma_{MD_ini} + \delta] \times [\gamma_{TD_ini}, \gamma_{TD_ini} + \delta]$.

- (3) Subtract a small amount δ (for instance 0.001) to each initial value: $\gamma_{MT_ini} - \delta$, $\gamma_{MD_ini} - \delta$, and $\gamma_{TD_ini} - \delta$. If one of these values is less than -1 , the algorithm stops. Else, calculate the eight possible combinations of determinants: $f_{\det}(\gamma_{MT_ini} - \delta, \gamma_{MD_ini} - \delta, \gamma_{TD_ini} - \delta)$, $f_{\det}(\gamma_{MT_ini} + \delta, \gamma_{MD_ini} - \delta, \gamma_{TD_ini} - \delta)$, $f_{\det}(\gamma_{MT_ini} - \delta, \gamma_{MD_ini} + \delta, \gamma_{TD_ini} - \delta)$, $f_{\det}(\gamma_{MT_ini} - \delta, \gamma_{MD_ini} - \delta, \gamma_{TD_ini} + \delta)$, $f_{\det}(\gamma_{MT_ini} + \delta, \gamma_{MD_ini} + \delta, \gamma_{TD_ini} - \delta)$, $f_{\det}(\gamma_{MT_ini} + \delta, \gamma_{MD_ini} - \delta, \gamma_{TD_ini} + \delta)$, $f_{\det}(\gamma_{MT_ini} - \delta, \gamma_{MD_ini} + \delta, \gamma_{TD_ini} + \delta)$, and $f_{\det}(\gamma_{MT_ini} + \delta, \gamma_{MD_ini} + \delta, \gamma_{TD_ini} + \delta)$.

If all the determinants are positive, the functional relationship of equation (B.1) guarantees positive determinants for any point in the cubic region $[\gamma_{MT_ini} - \delta, \gamma_{MT_ini} + \delta] \times [\gamma_{MD_ini} - \delta, \gamma_{MD_ini} + \delta] \times [\gamma_{TD_ini} - \delta, \gamma_{TD_ini} + \delta]$.

- (4) Add another small amount δ to each initial value: $\gamma_{MT_ini} + 2\delta$, $\gamma_{MD_ini} + 2\delta$, and $\gamma_{TD_ini} + 2\delta$. If one of these values is greater than 1, the algorithm stops. Else, calculate the eight possible combinations of determinants: $f_{\det}(\gamma_{MT_ini} - \delta, \gamma_{MD_ini} - \delta, \gamma_{TD_ini} - \delta)$, $f_{\det}(\gamma_{MT_ini} + 2\delta, \gamma_{MD_ini} - \delta, \gamma_{TD_ini} - \delta)$, $f_{\det}(\gamma_{MT_ini} - \delta, \gamma_{MD_ini} + 2\delta, \gamma_{TD_ini} - \delta)$, $f_{\det}(\gamma_{MT_ini} - \delta, \gamma_{MD_ini} - \delta, \gamma_{TD_ini} + 2\delta)$, $f_{\det}(\gamma_{MT_ini} + 2\delta, \gamma_{MD_ini} + 2\delta, \gamma_{TD_ini} - \delta)$, $f_{\det}(\gamma_{MT_ini} + 2\delta, \gamma_{MD_ini} - \delta, \gamma_{TD_ini} + 2\delta)$, $f_{\det}(\gamma_{MT_ini} - \delta, \gamma_{MD_ini} + 2\delta, \gamma_{TD_ini} + 2\delta)$, and $f_{\det}(\gamma_{MT_ini} + 2\delta, \gamma_{MD_ini} + 2\delta, \gamma_{TD_ini} + 2\delta)$.

If all the determinants are positive, equation (B.1) guarantees positive determinants for any point in the cubic region $[\gamma_{MT_ini} - \delta, \gamma_{MT_ini} + 2\delta] \times [\gamma_{MD_ini} - \delta, \gamma_{MD_ini} + 2\delta] \times [\gamma_{TD_ini} - \delta, \gamma_{TD_ini} + 2\delta]$.

- (5) Subtract another amount δ to each initial value: $\gamma_{MT_ini} - 2\delta$, $\gamma_{MD_ini} - 2\delta$, and $\gamma_{TD_ini} - 2\delta$. If one of these values is less than -1 , the algorithm stops. Else, calculate the eight possible combinations of determinants:

$$f_{\det}(\gamma_{MT_ini} - 2\delta, \gamma_{MD_ini} - 2\delta, \gamma_{TD_ini} - 2\delta), \quad f_{\det}(\gamma_{MT_ini} + 2\delta, \gamma_{MD_ini} - 2\delta, \gamma_{TD_ini} - 2\delta), \quad f_{\det}(\gamma_{MT_ini} - 2\delta, \gamma_{MD_ini} + 2\delta, \gamma_{TD_ini} - 2\delta), \quad f_{\det}(\gamma_{MT_ini} - 2\delta, \gamma_{MD_ini} - 2\delta, \gamma_{TD_ini} + 2\delta), \quad f_{\det}(\gamma_{MT_ini} + 2\delta, \gamma_{MD_ini} + 2\delta, \gamma_{TD_ini} - 2\delta), \quad f_{\det}(\gamma_{MT_ini} + 2\delta, \gamma_{MD_ini} - 2\delta, \gamma_{TD_ini} + 2\delta), \quad f_{\det}(\gamma_{MT_ini} - 2\delta, \gamma_{MD_ini} + 2\delta, \gamma_{TD_ini} + 2\delta), \quad \text{and} \quad f_{\det}(\gamma_{MT_ini} + 2\delta, \gamma_{MD_ini} + 2\delta, \gamma_{TD_ini} + 2\delta).$$

If all the determinants are positive, equation (B.1) guarantees positive determinants for any point in the cubic region $[\gamma_{MT_ini} - 2\delta, \gamma_{MT_ini} + 2\delta] \times [\gamma_{MD_ini} - 2\delta, \gamma_{MD_ini} + 2\delta] \times [\gamma_{TD_ini} - 2\delta, \gamma_{TD_ini} + 2\delta]$.

- (6) In this manner, the algorithm iterates n times, incrementally expanding the region centered on the initial values until either of the eight determinants becomes nonpositive or any of the points exceeds 1 or falls below -1 .

As the arguments above are heuristic, a series of simulation experiments were conducted to verify whether the points within three-dimensional region (centered at the initial values γ_{MT_ini} , γ_{MD_ini} , and γ_{TD_ini}) exhibit a positive determinant. The simulation results unequivocally illustrated that all points within the cubic region yielded positive determinants.

Data Availability

The prostate cancer dataset used to support the findings of this study has been deposited in the Vanderbilt Biostatistics repository (<https://hbiostat.org/data>).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

The authors wish to thank Benemérita Universidad Autónoma de Puebla for all the support provided to achieve this research project.

References

- [1] P. Saha and P. J. Heagerty, "Time-dependent predictive accuracy in the presence of competing risks," *Biometrics*, vol. 66, no. 4, pp. 999–1011, 2010.
- [2] P. Blanche, J.-F. Dartigues, and H. Jacqmin-Gadda, "Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks," *Statistics in Medicine*, vol. 32, no. 30, pp. 5381–5397, 2013.
- [3] C. Wu and L. Li, "Quantifying and estimating the predictive accuracy for censored time-to-event data with competing risks," *Statistics in Medicine*, vol. 37, no. 21, pp. 3106–3124, 2018.
- [4] Y. Zheng, T. Cai, Y. Jin, and Z. Feng, "Evaluating prognostic accuracy of biomarkers under competing risk," *Biometrics*, vol. 68, no. 2, pp. 388–396, 2012.

- [5] J. P. Fine and R. J. Gray, "A proportional hazards model for the subdistribution of a competing risk," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 496–509, 1999.
- [6] P. X. Song, M. Li, and Y. Yuan, "Joint regression analysis of correlated data using Gaussian copulas," *Biometrics*, vol. 65, no. 1, pp. 60–68, 2009.
- [7] D. F. Andrews and A. M. Herzberg, *Data: A Collection of Problems from many fields for the Student and Research Worker*, Springer Science and Business Media, Berlin, Germany, 2012.
- [8] R. B. Nelsen, *An Introduction to Copulas*, Springer Science and Business Media, Berlin, Germany, 2007.
- [9] M. Sklar, "Fonctions de repartition an dimensions et leurs marges," *Publications de l'Institut de statistique de l'Université de Paris*, vol. 8, pp. 229–231, 1959.
- [10] A. Bansal and P. J. Heagerty, "A comparison of landmark methods and time-dependent roc methods to evaluate the time-varying performance of prognostic markers for survival outcomes," *Diagnostic and prognostic research*, vol. 3, no. 1, pp. 14–13, 2019.
- [11] R. T. Clemen and T. Reilly, "Correlations and copulas for decision and risk analysis," *Management Science*, vol. 45, no. 2, pp. 208–224, 1999.
- [12] E. M. Mitchell, R. H. Lyles, A. K. Manatunga, M. Danaher, N. J. Perkins, and E. F. Schisterman, "Regression for skewed biomarker outcomes subject to pooling," *Biometrics*, vol. 70, no. 1, pp. 202–211, 2014.
- [13] A. Azzalini, *The Skew-Normal And Related Families*, Cambridge University Press, Cambridge, United Kingdom, 2013.
- [14] K. J. Carroll, "On the use and utility of the weibull model in the analysis of survival data," *Controlled Clinical Trials*, vol. 24, no. 6, pp. 682–701, 2003.
- [15] M. G. Larson and G. E. Dinse, "A mixture model for the regression analysis of competing risks data," *Applied Statistics*, vol. 34, no. 3, pp. 201–211, 1985.
- [16] B. Lau, S. R. Cole, R. D. Moore, and S. J. Gange, "Evaluating competing adverse and beneficial outcomes using a mixture model," *Statistics in Medicine*, vol. 27, no. 21, pp. 4313–4327, 2008.
- [17] R. P. Brent, *Algorithms for Minimization without Derivatives*, Courier Corporation, Chelmsford, MA, USA, 2013.
- [18] R. Bender, T. Augustin, and M. Blettner, "Generating survival times to simulate cox proportional hazards models," *Statistics in Medicine*, vol. 24, no. 11, pp. 1713–1723, 2005.
- [19] B. French, P. Saha-Chaudhuri, B. Ky, T. P. Cappola, and P. J. Heagerty, "Development and evaluation of multi-marker risk scores for clinical prognosis," *Statistical Methods in Medical Research*, vol. 25, no. 1, pp. 255–271, 2016.
- [20] D. Collett, *Modelling Survival Data in Medical Research*, CRC Press, Boca Raton, FL, USA, 2015.
- [21] J. Carpenter and J. Bithell, "Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians," *Statistics in Medicine*, vol. 19, no. 9, pp. 1141–1164, 2000.
- [22] B. Chang and H. Joe, "Prediction based on conditional distributions of vine copulas," *Computational Statistics and Data Analysis*, vol. 139, pp. 45–63, 2019.
- [23] C. D. Meyer, *Matrix Analysis And Applied Linear Algebra*, Siam, New Delhi, India, 2000.