

Article

A Bivariate Copula–Driven Multi-State Model for Statistical Analysis in Medical Research

Hugo Brango ^{1,*} , Roger Tovar-Falón ²  and Guillermo Martínez-Flórez ² ¹ Grupo de Investigación Análisis Funcional y Ecuaciones Diferenciales (AFED), Departamento de Matemáticas, Facultad de Educación y Ciencias, Universidad de Sucre, Sincelejo 700001, Colombia² Departamento de Matemáticas y Estadística, Universidad de Córdoba, Montería 230002, Colombia; rjtovar@correo.unicordoba.edu.co (R.T.-F.); guillermomartinez@correo.unicordoba.edu.co (G.M.-F.)

* Correspondence: hugo.brango@unisucra.edu.co

Abstract

We develop and evaluate a copula-based multistate model for illness–death processes with dependent transition times. The framework couples Cox proportional hazards models for the marginal transition intensities with Archimedean copulas to capture dependence, and it is estimated via the Inference Functions for Margins (IFM) approach under right censoring. A Monte Carlo study shows that assuming independence between transitions can severely underestimate joint survival, yielding coverage as low as 40% under strong dependence, compared with 92% to 97% when copulas are used. We apply the method to a large Colombian cohort of COVID-19 patients (2021 to 2022) that includes sociodemographic, clinical, and vaccination data. The Gumbel copula best captures the strong positive dependence between hospitalization and death, producing more accurate joint survival estimates than independence-based models. Model diagnostics, including proportional hazards tests, Kaplan–Meier comparisons, hazard rate functions, and TTT plots, support the adequacy of the Cox margins. We also discuss limitations and avenues for extension, such as parametric or cure-fraction margins, nested or vine copulas, and full-likelihood estimation. Overall, the results underscore the methodological and applied value of integrating copulas into multistate models, offering a robust framework for analyzing dependent event times in epidemiology and biomedicine.

Keywords: multi-state models; copula functions; Archimedean copulas; survival analysis; dependence modeling; bivariate outcomes

MSC: 62N02; 62H20; 62P10



Academic Editors: Lazarus K. Mramba and Qihui Yang

Received: 1 August 2025

Revised: 13 September 2025

Accepted: 19 September 2025

Published: 24 September 2025

Citation: Brango, H.; Tovar-Falón, R.; Martínez-Flórez, G. A Bivariate Copula–Driven Multi-State Model for Statistical Analysis in Medical Research. *Mathematics* **2025**, *13*, 3072. <https://doi.org/10.3390/math13193072>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Survival analysis constitutes a fundamental statistical methodology in medical, epidemiological, and biostatistical research, playing a crucial role in analyzing time-to-event data, such as disease recurrence, clinical progression, hospitalization, and mortality. Its utility in identifying prognostic factors, evaluating therapeutic efficacy, and providing empirical evidence on patient outcomes makes it indispensable for evidence-based clinical decision-making [1–3]. Among the various survival analysis techniques, the Cox proportional hazards model has become the predominant method, thanks to its interpretability, flexibility, and capability to accommodate covariate effects without specifying the baseline hazard function [4,5]. However, a significant limitation of the conventional Cox model is its

implicit assumption of independence between event times, an assumption that frequently does not hold true in clinical practice [6,7].

In numerous clinical scenarios, such as chronic illnesses, cardiovascular conditions, and cancer, patients often experience sequences of related events that exhibit inherent interdependencies. For example, a hospitalization event can significantly alter the subsequent risk of mortality through factors such as clinical deterioration, treatment complications, or common underlying health determinants [7–9]. Neglecting these dependencies in the analysis of multi-state events can substantially bias risk estimates, distort covariate effect interpretations, and lead to erroneous conclusions, ultimately impacting patient management, resource allocation, and prognostic accuracy [9,10].

Multi-state models have emerged as powerful statistical tools to explicitly describe and analyze transitions between distinct health states over time. These models facilitate the investigation of complex event histories, allowing researchers to quantify covariate effects on the timing and sequence of clinical outcomes comprehensively [10,11]. Despite their versatility, conventional multi-state models often continue to assume independence across transitions, a simplification frequently violated due to latent patient heterogeneity or unmeasured common risk factors influencing multiple events [12–14].

Recent methodological developments have aimed to address these limitations by introducing sophisticated approaches that explicitly account for dependencies between multistate transitions. Techniques such as landmarking, joint frailty–multistate modeling, hierarchical models integrating longitudinal biomarkers, and hidden semi-Markov frameworks exemplify these advancements [12,14–16]. Collectively, these approaches highlight the increasing recognition of the necessity to model dependencies robustly and flexibly within multistate contexts.

An especially promising and increasingly employed approach to capturing complex dependence structures in survival analysis is the use of copula functions. Copulas offer considerable flexibility by modeling dependencies separately from marginal distributions, enabling researchers to accurately describe complex association structures among event times [17–19]. Originally developed in fields such as finance and hydrology, copulas have gained substantial attention in biostatistics and medical research, where they have demonstrated notable improvements in modeling correlated event times and enhancing predictive accuracy [20–23].

Empirical studies in medical research have emphasized the value of copula-based multistate models. For instance, in oncology research, copula-driven approaches have more precisely quantified risks of recurrence and metastasis, outperforming traditional models [22,24]. Similarly, in cardiovascular epidemiology, copula-based multistate models have substantially improved predictions of rehospitalization and mortality after myocardial infarction [25]. These studies underscore the practical importance of incorporating copula models to effectively represent real-world clinical dependencies.

Motivated by these critical methodological advancements and their substantial clinical implications, this article proposes and rigorously evaluates a bivariate copula-based multi-state model for jointly analyzing clinically significant events, hospitalization, and death using data from a large observational registry of COVID-19 patients in Colombia (2021–2022). Utilizing flexible Archimedean copulas, our model explicitly accounts for residual dependencies between events, improving inferential accuracy, risk stratification, and predictive performance.

The article is structured as follows: Section 2 presents foundational concepts of multistate modeling, including Cox proportional hazards models and copula functions, particularly Archimedean copulas. Section 3 introduces our proposed joint semi-parametric multistate–copula model, detailing its statistical formulation and inference methodologies.

Section 4 presents a comprehensive simulation study, discussing scenario configurations, data-generating mechanisms, estimation strategies, and extensive results with an interpretative discussion. Section 5 demonstrates a practical application of the proposed model to COVID-19 patient data, illustrating the clinical and practical utility. Section 6 summarizes the main findings and conclusions of the study, while Section 7 discusses limitations and outlines directions for future research.

2. Basic Concepts

2.1. Multi-State Models

In survival studies, it is common for individuals to experience events over time. Multi-state models provide a useful framework for estimating the risk of transitions between disease states and for understanding the influence of covariates on the timing of these transitions [26].

In longitudinal studies, individuals may experience several clinical events over time. Multi-state models provide a flexible framework for estimating transition hazards between discrete disease states and for quantifying how covariates affect those transitions [7,9]. Formally, one considers a stochastic process $\{X(t) : t \in \mathcal{T}\}$, taking values in $\{1, 2, \dots, K\}$, where K denotes the number of states and each $i \rightarrow j$ transition is governed by an intensity function.

$$\lambda_{ij}(t | \mathbf{X}) = \lim_{\Delta t \rightarrow 0} \frac{\Pr\{X(t + \Delta t) = j \mid X(t) = i, \mathbf{X}\}}{\Delta t}.$$

The three frameworks are summarized in Figure 1. Panel (a) depicts the basic two-state survival model, where individuals transition directly from the initial “alive” state to the absorbing state of death. Panel (b) extends this framework by introducing an intermediate “illness” or “progression” state, which allows a more detailed characterization of clinical trajectories and the risks associated with subsequent mortality. Finally, panel (c) presents the generalized multi-state structure, in which individuals sequentially pass through multiple intermediate states before reaching the absorbing state. This representation provides a more flexible and realistic description of complex clinical scenarios, such as the progression through successive stages of a chronic disease.

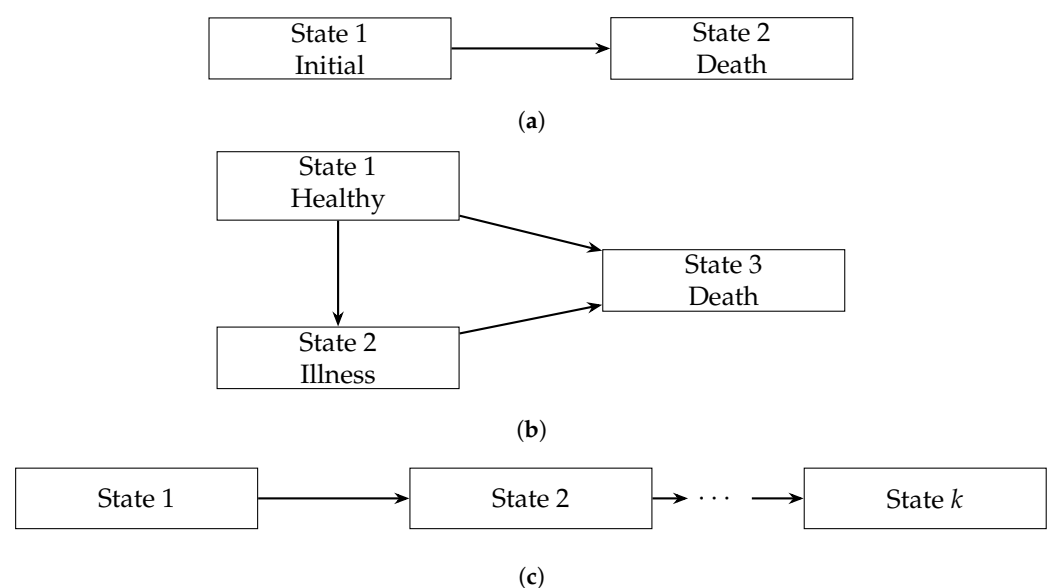


Figure 1. Basic structures of survival and multi-state models. (a) Survival model. (b) Disease–death model. (c) Multi-state model with k progressive states.

It is of interest to determine the rate of onset or progression of the disease and to identify prognostic variables governing the transitions between the different states in a medical context.

The most common approach to these practical scenarios is to decompose the process into multiple survival models and fit, separately, the transition risk from one state to another (often called the transition intensity). When covariates are available to explain these transitions, the standard choice is the Cox proportional hazards model.

2.2. Cox Proportional Hazards Model

To assess the effect of a vector of covariates $\mathbf{X} = (x_1, x_2, \dots, x_p)^T$, we invoke the Cox proportional hazards model [4]. Let T denote the time to the transition of interest and define the hazard function

$$\lambda(t | \mathbf{X}) = \lim_{\Delta t \rightarrow 0} \frac{\Pr\{t \leq T < t + \Delta t \mid T \geq t, \mathbf{X}\}}{\Delta t}. \quad (1)$$

The Cox model assumes

$$\lambda(t; \mathbf{X}) = \lambda_0(t) \exp(\beta^\top \mathbf{X}), \quad (2)$$

where

- $\lambda_0(t)$ is an unspecified baseline hazard, common to all subjects when $\mathbf{X} = \mathbf{0}$, and
- $\beta = (\beta_1, \dots, \beta_p)^\top$ measures log-hazard ratios associated with the covariates.

Under proportionality, the hazard ratio between two individuals with covariate values $\mathbf{X}$ and $\mathbf{X}^*$ is

$$\frac{\lambda(t | \mathbf{X})}{\lambda(t | \mathbf{X}^*)} = \exp\{\beta^\top (\mathbf{X} - \mathbf{X}^*)\},$$

constant in t .

Cox's key insight was to construct a partial likelihood that eliminates the unknown $\lambda_0(t)$. Suppose that there are n ordered event times $t_{(1)} < \dots < t_{(m)}$, and at time $t_{(k)}$, a single subject j_k experiences the event. Let

$$\mathcal{R}(t_{(k)}) = \{\ell : T_\ell \geq t_{(k)}\}$$

be the risk set just before $t_{(k)}$. The partial likelihood is

$$L_p(\beta) = \prod_{k=1}^m \frac{\exp\{\beta^\top \mathbf{X}_{j_k}\}}{\sum_{\ell \in \mathcal{R}(t_{(k)})} \exp\{\beta^\top \mathbf{X}_\ell\}}. \quad (3)$$

Maximizing $\log L_p(\beta)$ yields the estimator $\hat{\beta}$. Standard errors follow from the observed information matrix.

Baseline Hazard and Survival Function

Once $\hat{\beta}$ is obtained, one recovers an estimate of the cumulative baseline hazard

$$\hat{\Lambda}_0(t) = \sum_{t_{(k)} \leq t} \frac{1}{\sum_{\ell \in \mathcal{R}(t_{(k)})} \exp\{\hat{\beta}^\top \mathbf{X}_\ell\}}, \quad (4)$$

known as the Breslow estimator. The corresponding marginal survival for an individual with covariates $\mathbf{X}$ is

$$\hat{S}(t | \mathbf{X}) = \exp\{-\hat{\Lambda}_0(t) \exp(\hat{\beta}^\top \mathbf{X})\}. \quad (5)$$

Fitting the Cox model separately to each transition of a multi-state process yields marginal survival curves, $\hat{S}(t | \mathbf{X})$.

However, in processes with multiple sequential events (e.g., hospitalization followed by death), fitting separate Cox models for each transition can overlook residual dependence between event times. To address this limitation and jointly model survival across stages, we need an extension that explicitly captures the interrelationships among transition times.

A powerful framework for joint survival modeling in this setting is provided via copula models, which allow for flexible specification of the dependence structure among multiple outcomes or event times. Within a multi-state context, individuals move through a series of states, and the occurrence of one event can directly influence the probability of subsequent events.

2.3. Model Diagnostics and Marginal Fits

We assessed the proportional hazards (PHs) assumption using Schoenfeld residuals for each covariate and transition. Plots were inspected for systematic time trends and formal tests were computed (global and covariate-specific).

To show that the semiparametric Cox models reproduce the marginal behavior of the data before introducing dependence, we contrasted, for each transition, (i) the nonparametric Kaplan–Meier (KM) estimator with 95% pointwise CIs and (ii) the population-averaged survival curve implied by the fitted Cox model. The latter was computed as

$$\bar{S}(t) = \frac{1}{n} \sum_{i=1}^n S_0(t)^{\exp(\eta_i)}, \quad (6)$$

averaging over individuals (and mixing over city strata using sample proportions when applicable), where $S_0(t)$ is the baseline survival, and η_i is the fitted linear predictor.

We also compared hazard rate functions using the following: (a) a nonparametric kernel estimator (muhaz) restricted to the observed event-time range, and (b) the Cox-induced hazard obtained via the numerical differentiation of the Breslow baseline cumulative hazard and, when stratified, averaging across strata with the same weights as above. Hazards were plotted on a log-scale for readability.

We also examined how risk evolved over time by plotting hazard-rate functions (HRFs) for each transition. Specifically, we contrasted the population-averaged HRF implied by the fitted Cox model with a nonparametric kernel HRF (muhaz); curves were shown on a log scale for readability. The agreement between the two supports the adequacy of the marginal Cox specification.

Finally, Total Time on Test (TTT) plots were produced to summarize the failure-time shape irrespective of model assumptions. Writing $u \in [0, 1]$ for the proportion failed, the TTT curve $\text{TTT}(u)$ was computed from the KM estimator. A 45° line corresponds to a constant hazard; concavity indicates a decreasing hazard, and convexity indicates an increasing hazard. We annotated $u_{\max}$ (the maximum horizontal deviation), which provides a concise summary of early- vs. late-failure dominance.

2.4. Copula Functions

Copulas provide a general mechanism to build multivariate distributions by “coupling” univariate margins with an explicit dependence structure. They have found widespread application in fields such as finance, hydrology, and biostatistics [17,18]. For-

mally, a copula is a multivariate distribution function on the unit cube $[0, 1]^d$ with uniform $(0, 1)$ margins.

Sklar's theorem [27] underpins the entire copula approach. For any d -dimensional continuous distribution function, F , with univariate margins, $F_1, \dots, F_d$, there exists a unique copula, C , such that

$$F(x_1, \dots, x_m) = C(F_1(x_1), \dots, F_m(x_m)). \quad (7)$$

Conversely, given any copula, C , and univariate margins, F_i , the above relation defines a valid joint distribution. In survival analysis, one typically works with survival copulas, linking marginal survival functions, $S_i(t) = 1 - F_i(t)$, into a joint survival,

$$S(t_1, \dots, t_d) = C(S_1(t_1), \dots, S_d(t_d)). \quad (8)$$

2.4.1. Archimedean Copulas

Archimedean copulas were chosen in this work because they offer a balance between analytical tractability and flexibility. Their one-parameter formulation makes them parsimonious and computationally efficient, which is advantageous for estimation within the IFM framework [17,18]. In addition, Archimedean families such as Clayton and Gumbel can capture clinically relevant tail dependencies: lower-tail for early adverse outcomes (e.g., rapid deterioration) and upper-tail for extreme late events (e.g., prolonged hospitalization followed by death). These features make them particularly suitable for bivariate survival data in medical and epidemiological research [19,21,22]. Other copula families, such as Gaussian, Student-t, or vine copulas, could also be employed to model more complex dependence structures, but they often require higher-dimensional parameterizations and substantially greater computational effort [20,28]. For clarity and parsimony, we focused on Archimedean copulas while acknowledging that extensions to elliptical or vine copulas constitute a relevant direction for future research.

Definition and Generator

A d -dimensional Archimedean copula, C , is defined via a continuous, strictly decreasing generator

$$\varphi: [0, 1] \rightarrow [0, \infty], \quad \varphi(1) = 0, \quad \varphi(0^+) = \infty,$$

and its pseudo-inverse φ^{-1} . The copula is

$$C(u_1, \dots, u_d) = \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_d)), \quad (u_1, \dots, u_d) \in [0, 1]^d. \quad (9)$$

When margins are continuous, the copula density is given by

$$c(u_1, \dots, u_d) = \varphi^{-1'}\left(\sum_{i=1}^d \varphi(u_i)\right) \prod_{i=1}^d \varphi'(u_i). \quad (10)$$

Note that the derivative of the inverse generator can be written explicitly as

$$\frac{d}{dx} \varphi^{-1}(x) = [\varphi'(\varphi^{-1}(x))]^{-1},$$

which leads to the equivalent expression for the copula density:

$$c(u_1, \dots, u_d) = \frac{\prod_{i=1}^d \varphi'(u_i)}{\varphi'\left(\varphi^{-1}\left(\sum_{i=1}^d \varphi(u_i)\right)\right)}.$$

In the bivariate case ($d = 2$), this reduces to

$$c(u, v) = \frac{\varphi'(u) \varphi'(v)}{\varphi'(\varphi^{-1}(\varphi(u) + \varphi(v)))}.$$

One-Parameter Families

For $d = 2$, three canonical generators and their copulas are as follows:

- **Clayton** ($\theta > 0$): The generator is

$$\varphi(u) = \frac{u^{-\theta} - 1}{\theta}, \quad C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}.$$

This copula exhibits lower-tail dependence $\lambda_L = 2^{-1/\theta}$, zero upper-tail.

Remark 1. Some texts omit the division by θ and define $\varphi(u) = u^{-\theta} - 1$. Both forms are equivalent up to a positive scaling of the generator and lead to the same copula. Here, we adopt the divided-by- θ convention, which is standard in the copula literature.

- **Gumbel** ($\theta \geq 1$): The generator is

$$\varphi(u) = (-\ln u)^\theta, \quad C(u, v) = \exp\left\{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right\}.$$

This copula exhibits lower-tail dependence $\lambda_U = 2 - 2^{1/\theta}$, zero upper-tail.

- **Frank** ($\theta \neq 0$): The generator is

$$\varphi(u) = -\ln\left\{\frac{e^{-\theta u} - 1}{e^{-\theta} - 1}\right\}, \quad C(u, v) = -\frac{1}{\theta} \ln\left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right].$$

This copula exhibits tail symmetry, lacks asymptotic tail dependence, and is flexible enough to capture both negative ($\theta < 0$) or positive ($\theta > 0$) dependence.

To investigate the dependence properties of these bivariate copulas, we consider Kendall's τ , a rank-based measure of concordance. This coefficient ranges from -1 (perfect negative dependence) to 1 (perfect positive dependence), with 0 indicating independence. Unlike linear correlation, Kendall's τ depends only on the underlying copula, making it an appropriate summary of dependence strength.

Let (X_1, X_2) denote a pair of continuous random variables with copula $C(u, v)$. In general, Kendall's τ can be expressed as

$$\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1, \quad (11)$$

where $C(u, v)$ is the copula function. For Archimedean copulas, τ has a closed-form representation based on the generator function $\varphi(u)$:

$$\tau(\theta) = 1 + 4 \int_0^1 \frac{\varphi(t, \theta)}{\varphi'(t, \theta)} dt. \quad (12)$$

Closed-form relations are available for several families, which facilitate interpretation and provide initial parameter estimates:

$$\tau_{\text{Clayton}} = \frac{\theta}{\theta + 2}, \quad \tau_{\text{Gumbel}} = 1 - \frac{1}{\theta}, \quad \tau_{\text{Frank}} = 1 - \frac{4}{\theta} [D_1(\theta) - 1],$$

where $D_1(\theta)$ denotes the Debye function of order one.

Thus, Kendall's τ provides an interpretable measure of the strength and type of dependence, which is particularly useful in our context for summarizing and comparing the dependence captured via different copula families in survival applications.

2.4.2. Inference via Inference Functions for Margins (IFM)

The Inference Functions for Margins (IFM) method is a two-step estimation procedure commonly used for copula models. In the first step, the marginal distributions (or marginal survival functions in our context) are estimated using standard techniques (e.g., Cox proportional hazards models). In the second step, the copula parameter is estimated by maximum likelihood, using the pseudo-observations derived from the estimated margins. This approach separates the estimation of marginal and dependence parameters, making it computationally efficient and widely applicable in practice.

The IFM method [28,29] separates marginal estimation from copula fitting:

1. *Marginal step:* estimate each univariate distribution, F_i , (parametrically or semiparametrically), and compute pseudo-observations

$$u_{ij} = \hat{F}_i(x_{ij}).$$

2. *Copula step:* maximize the pseudo-log-likelihood

$$\ell(\theta) = \sum_{j=1}^n \log c(u_{1j}, \dots, u_{dj}; \theta),$$

to obtain $\hat{\theta}$.

In our framework, the IFM method is applied by first fitting Cox proportional hazards models for each transition intensity to estimate the marginal survival functions and then estimating the copula parameter that links the transition times. This ensures that dependence between transitions is modeled explicitly while retaining the flexibility of semi-parametric marginal estimation.

This two-stage procedure is computationally efficient and leverages existing marginal fits. In practice, the selection of the most appropriate Archimedean copula relies on a combination of statistical and graphical tools. The process usually begins with information criteria, such as AIC or BIC, to identify the generator that provides the best balance between fit and parsimony [18]. This is followed by graphical diagnostics, where contour plots of the empirical and fitted copulas are compared to verify whether both central and tail dependencies are adequately captured [30]. As a final step, formal goodness-of-fit tests such as the Cramér–von Mises and Kolmogorov–Smirnov statistics can be implemented within a parametric bootstrap framework to obtain valid p -values. This approach provides a rigorous assessment of how well the selected copula reproduces the dependence structure in the data [30].

3. Proposed Joint Semi-Parametric Multi-State–Copula Model

Building on the three-state diagram in Figure 2, our goal is to estimate marginal covariate effects on each transition hazard while simultaneously capturing residual dependence between the two observed transitions (hospitalization $\rightarrow$ death). In what follows, we present a mathematically rigorous development of the model, the full likelihood under arbitrary right-censoring, and the associated inference strategy.

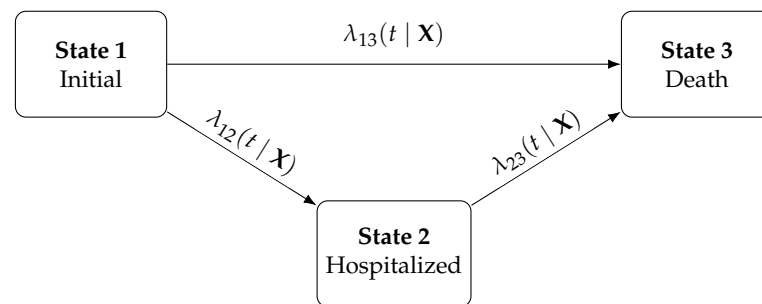


Figure 2. Proposed illness–death model with covariate-adjusted transition hazards. From the initial state (State 1), subjects may move to hospitalization (State 2) or directly to death (State 3). Subsequent death after hospitalization is also modeled. Each arrow corresponds to a Cox hazard $\lambda_{jk}(t | \mathbf{X})$ for transition $j \rightarrow k$.

3.1. Hazard-Based Margins

Let $X \in \mathbb{R}^p$ denote baseline covariates (e.g., age, biomarkers). We focus on the two successive transitions:

$$1 \text{ (Admission)} \xrightarrow{T_{12}} 2 \text{ (Complication)} \xrightarrow{T_{23}} 3 \text{ (Death)},$$

and allow a direct transition

$$1 \text{ (Admission)} \xrightarrow{T_{13}} 3 \text{ (Death without complication)}.$$

For each transition $j \rightarrow k \in \{(1,2), (1,3), (2,3)\}$, we posit a Cox proportional hazards model

$$\lambda_{jk}(t | X) = \lambda_{0,jk}(t) \exp\{\beta_{jk}^\top X\},$$

where $\lambda_{0,jk}(t)$ is left unspecified, and β_{jk} captures the log-hazard ratios for clinical covariates such as disease stage or treatment arm.

In the illness–death framework considered here, patients may follow two possible paths:

- Diagnosis $\rightarrow$ Hospitalization $\rightarrow$ Death;
- Diagnosis $\rightarrow$ Death without prior hospitalization.

Accordingly, our model specifies transition intensities for $(1 \rightarrow 2)$, $(2 \rightarrow 3)$, and $(1 \rightarrow 3)$. In the copula-based formulation, individuals who die without being hospitalized contribute information to the $(1 \rightarrow 3)$ transition, with the time to hospitalization treated as censored. This ensures that both direct and sequential paths to death are accommodated within the likelihood framework, and the copula links the observed times even when one component is censored.

Maximization of the partial likelihood produces $\hat{\beta}_{jk}$, and the Breslow estimator yields the cumulative baseline hazard

$$\hat{\Lambda}_{0,jk}(t) = \sum_{t_i \leq t} \frac{1}{\sum_{\ell \in \mathcal{R}_{jk}(t_i)} \exp\{\hat{\beta}_{jk}^\top X_\ell\}}, \quad (13)$$

from which the marginal survival function follows:

$$\hat{S}_{jk}(t | X) = \exp\left\{-\hat{\Lambda}_{0,jk}(t) \exp(\hat{\beta}_{jk}^\top X)\right\}. \quad (14)$$

These margins quantify, for example, how a one-unit increase in a severity score multiplies the instantaneous risk of complication or death, information of direct clinical relevance [5].

The Cox proportional hazards model relies on the proportional hazards (PH) assumption, which states that the hazard ratios associated with covariates remain constant over time. This assumption is standard in multi-state analyses and provides a parsimonious yet flexible way to model marginal transition intensities. If violations of the PH assumption were present, extensions such as stratified Cox models or additive hazards models could be considered [5,10].

3.2. Copula Linkage of Sequential Event Times

To capture residual dependence between the time to hospitalization T_{12} and the subsequent time to death T_{23} , we link their marginal survival curves through an Archimedean survival copula C_θ [17,18]. According to Sklar's theorem, for any t_1 and t_2 ,

$$\Pr(T_{12} > t_1, T_{23} > t_2 \mid X) = C_\theta(S_{12}(t_1 \mid X), S_{23}(t_2 \mid X)). \quad (15)$$

In practice, we replace the true margins, S_{jk} , with their Cox–Breslow estimates $\hat{S}_{jk}$, yielding

$$\hat{S}_{\text{joint}}(t_1, t_2 \mid X) = C_\theta(\hat{S}_{12}(t_1 \mid X), \hat{S}_{23}(t_2 \mid X)). \quad (16)$$

Choosing a one-parameter family (Clayton, Gumbel, or Frank) provides closed-form expressions for $C_\theta(u, v)$, its density, $c_\theta(u, v)$, and the partial derivative $\partial_1 C_\theta(u, v)$ [18].

3.3. Likelihood Under Right-Censoring

Let C denote an independent right-censoring time. For each subject, i , we observe

$$Y_{12,i} = \min(T_{12,i}, C_i), \quad \Delta_{12,i} = 1\{T_{12,i} \leq C_i\},$$

and, if $\Delta_{12,i} = 1$, then

$$Y_{23,i} = \min(T_{23,i}, C_i - T_{12,i}), \quad \Delta_{23,i} = 1\{T_{23,i} \leq C_i - T_{12,i}\}.$$

Define the “pseudo-uniforms”:

$$u_{1i} = \hat{S}_{12}(Y_{12,i} \mid X_i), \quad u_{2i} = \hat{S}_{23}(Y_{23,i} \mid X_i).$$

Let $c_\theta(u, v) = \frac{\partial^2 C_\theta(u, v)}{\partial u \partial v}$ be the copula density and $\partial_1 C_\theta$ its derivative with respect to the first argument. Then, the likelihood contribution of subject i is

$$L_i(\theta) = \begin{cases} c_\theta(u_{1i}, u_{2i}), & \Delta_{12,i} = 1, \Delta_{23,i} = 1, \\ \partial_1 C_\theta(u_{1i}, u_{2i}), & \Delta_{12,i} = 1, \Delta_{23,i} = 0, \\ C_\theta(u_{1i}, u_{2i}), & \Delta_{12,i} = 0 (\Rightarrow \Delta_{23,i} = 0). \end{cases} \quad (17)$$

Thus, the overall copula log-likelihood is

$$\ell_{\text{cop}}(\theta) = \sum_{i=1}^n \left[\Delta_{12,i} \Delta_{23,i} \log c_\theta(u_{1i}, u_{2i}) + \Delta_{12,i} (1 - \Delta_{23,i}) \log \partial_1 C_\theta(u_{1i}, u_{2i}) + (1 - \Delta_{12,i}) \log C_\theta(u_{1i}, u_{2i}) \right]. \quad (18)$$

3.4. Two-Step IFM Estimation and Inference

We adopt the Inference Functions for Margins (IFM) strategy [28,31]:

1. **Marginal step.** Fit each Cox model for transitions $j \rightarrow k$ to obtain $\hat{\beta}_{jk}$ and the estimated survival, $\hat{S}_{jk}(t \mid X)$.

2. **Copula step.** Compute pseudo-uniforms (u_{1i}, u_{2i}) , and maximize $\ell_{\text{cop}}(\theta)$ over θ to get $\hat{\theta}$.

This fully specifies our joint model, as illustrated in Figure 2, and provides interpretable hazard-ratio estimates, together with a parsimonious dependence parameter that reflects unobserved clinical heterogeneity.

4. Simulation Study

To validate our two-step IFM estimator for the joint semi-parametric multi-state–copula model introduced in Section 3, we conducted a Monte Carlo study in which all data were generated under a Clayton copula and then analyzed under four estimation strategies: marginal (independence), correctly specified Clayton, and mis-specified Gumbel and Frank. Two covariates were included to assess inference on their effects.

4.1. Scenario Configuration

We considered a factorial design comprising three factors:

- **Sample size:**

$$n \in \{200, 500, 1000\}.$$

- **Clayton dependence:**

$$\theta \in \{1 (\tau \approx 0.33), 2 (\tau \approx 0.50), 4 (\tau \approx 0.67)\},$$

where $\tau = \frac{\theta}{\theta+2}$ is Kendall's τ . These values span from moderate to strong lower-tail dependence, reflecting clinical frailty scenarios.

- **Right-censoring rate:**

Uniform censoring $C \sim \text{Uniform}(0, c)$ calibrated to yield approximately 20% and 50% censoring on T_{12} .

For each of the 18 scenarios defined by the cross-product of these conditions, we generated 1000 independent replicates. The joint survival probability was evaluated at times corresponding to the quantiles 0.25, 0.50, and 0.75.

4.2. Data-Generating Mechanism

For each individual, we simulated two covariates:

- $X_1 \sim \text{Uniform}(18, 80)$, representing age in years.
- $X_2 \sim \text{Bernoulli}(0.5)$, representing a binary treatment indicator (e.g., treatment vs. control).

This simple design with one continuous covariate drawn from a uniform distribution and one binary Bernoulli covariate is common in methodological simulation studies (e.g., [32–35]). It provides a straightforward yet effective way to evaluate the performance of statistical estimators under controlled conditions. Although real datasets often include a larger number of covariates and more complex distributions, our framework naturally extends to higher-dimensional covariate structures. Exploring such scenarios constitutes a natural avenue for future work.

Given these covariates, we simulated the latent event times, T_{12} (admission to complication) and T_{23} (complication to death), with exponential marginal distributions:

$$U_{jk,i} \sim \mathcal{U}(0, 1), \quad T_{jk,i}^* = -\frac{\ln U_{jk,i}}{\lambda_{jk}^0 \exp(\beta_{jk,1} X_{1,i} + \beta_{jk,2} X_{2,i})},$$

and the baseline hazards and coefficients were fixed as follows:

$$\begin{aligned}\lambda_{1 \rightarrow 2}^0 &= 0.05, & \beta_{1 \rightarrow 2} &= (0.02, -0.50), \\ \lambda_{1 \rightarrow 3}^0 &= 0.02, & \beta_{1 \rightarrow 3} &= (0.01, -0.30), \\ \lambda_{2 \rightarrow 3}^0 &= 0.10, & \beta_{2 \rightarrow 3} &= (0.03, -0.40).\end{aligned}$$

The dependence between T_{12} and T_{23} was induced by generating copula-based pairs, $(U_1, U_2) \sim C_\theta$, where C_θ denotes the Clayton copula with the specified τ . Right-censoring times, C , were drawn from a uniform distribution and applied independently.

The observed event times were constructed as follows: individuals were first observed for the transition from state 1 to 2 (or directly to state 3 if earlier), and then, if a complication occurred, for the transition to death or censoring.

4.3. Estimation Procedures

On each simulated dataset, we fitted the following.

1. Marginal cox (independence): Three separate Cox models for transitions $1 \rightarrow 2$, $1 \rightarrow 3$ and $2 \rightarrow 3$, each including X_1 and X_2 . The joint survival was estimated as $\hat{S}_{12}(t) \cdot \hat{S}_{23}(t)$.
2. Clayton copula model: a two-step IFM approach was applied (Section 3.4)
 - Step 1: We fitted the same Cox models to estimate $\hat{S}_{12}(t), \hat{S}_{23}(t)$.
 - Step 2: pseudo-observations were derived from the marginal estimates, and the copula parameter θ was estimated via maximum likelihood using the uncensored bivariate observations.
3. Gumbel and Frank copula models: the same two-step IFM procedure was applied, assuming mis-specified dependence structures using Gumbel and Frank copulas, respectively.

The true joint survival function at each time point, t_p , was calculated as follows:

$$S(t_p, t_p) = \left(S_{12}(t_p)^{-\theta} + S_{23}(t_p)^{-\theta} - 1 \right)^{-\frac{1}{\theta}}, \quad (19)$$

reflecting the generating Clayton copula.

For each method and scenario we computed, over $B = 1000$ Monte Carlo replicates, the following performance measures at time t . Let $\hat{S}_b(t)$ denote the estimate from replicate b and $S(t)$ the truth.

- **Bias:**

$$\text{Bias}(t) = \bar{S}(t) - S(t), \quad \bar{S}(t) = \frac{1}{B} \sum_{b=1}^B \hat{S}_b(t).$$

- **Mean squared error (MSE):**

$$\text{MSE}(t) = \frac{1}{B} \sum_{b=1}^B (\hat{S}_b(t) - S(t))^2.$$

- **Coverage:** for each replicate, b , we formed a 95% confidence interval for $S(t)$ using the replicate-specific standard error $\widehat{\text{SE}}_b(t)$ (model-based within replicate): Greenwood's formula for KM estimates and a delta-method variance (via the Breslow baseline) for the Cox-based estimator. To respect the $(0, 1)$ range, we built intervals on the complementary log-log scale $g(x) = \log\{-\log(x)\}$ and back-transformed the following:

$$\text{CI}_b(t) = g^{-1}\left(g(\hat{S}_b(t)) \pm 1.96 \widehat{\text{SE}}_b^g(t)\right),$$

where $\widehat{SE}_b^g(t)$ is the standard error of $g(\hat{S}_b(t))$ (obtained through the delta method from $\widehat{SE}_b(t)$). The empirical coverage is, then,

$$\widehat{cov}(t) = \frac{1}{B} \sum_{b=1}^B \mathbf{1}\{S(t) \in CI_b(t)\}.$$

4.4. Results

In this subsection, we present and discuss the findings from our Monte Carlo experiments, which were designed to assess the performance of joint survival quantile estimators under varying dependence levels, sample sizes, and censoring proportions. We focus on three key metrics: average bias, mean squared error (MSE), and the empirical coverage of 95% confidence intervals. To highlight the main patterns, we focus on Figure 3, which displays the empirical coverage, and Figure 4, which reports the mean squared error. Both figures compare, under a representative scenario with moderate dependence, the independence estimator against the copula-based estimators (Clayton, Frank, and Gumbel) for the quantiles $p = 0.25, 0.50$, and 0.75 .

We then examine how these metrics change as dependence increases ($\tau = 0.33, 0.50, 0.75$), the sample size grows ($n = 200, 500, 1000$), and the censoring levels vary (20% and 50%). Full numerical results, including detailed tables of bias, MSE, and coverage for every combination of parameters, are provided in Appendix A (Tables A1–A3). This structure allows us to concentrate the main discussion on the most salient findings while ensuring the full transparency and reproducibility of the simulation study.

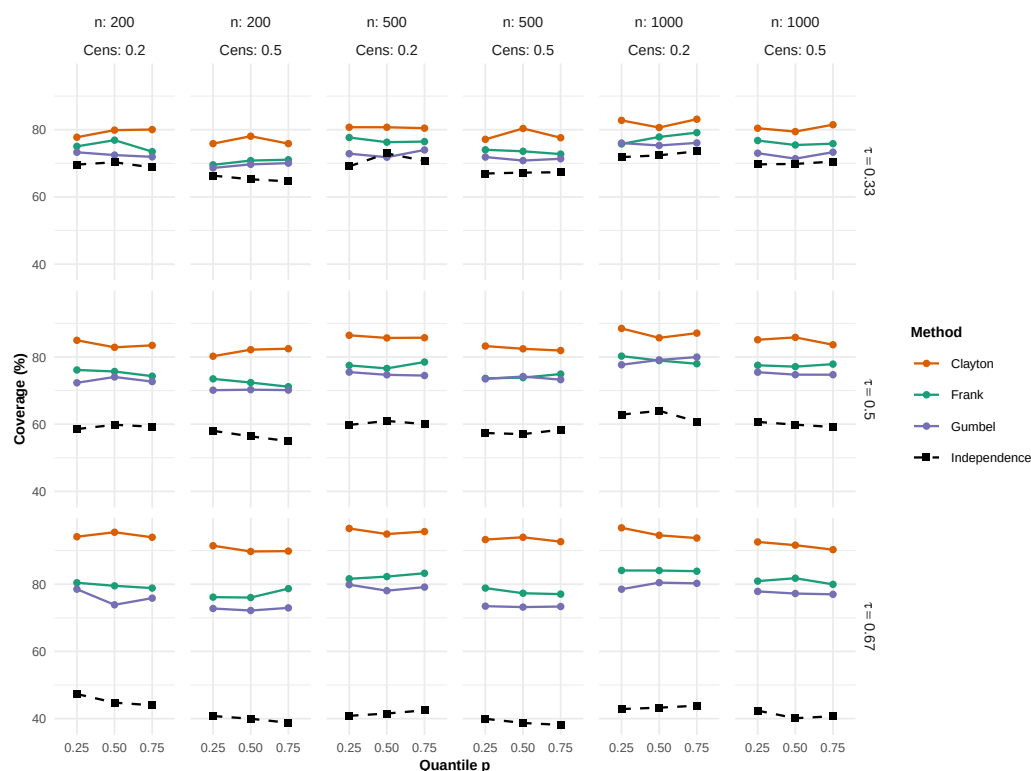


Figure 3. Coverage probabilities (%) of 95% confidence intervals for joint survival quantiles ($p = 0.25, 0.50, 0.75$), comparing independence and copula-based estimators (Clayton, Frank, Gumbel) across sample sizes and censoring levels. The independence assumption is represented by dashed lines with square markers, while copula-based estimators are represented by solid lines with circular markers.

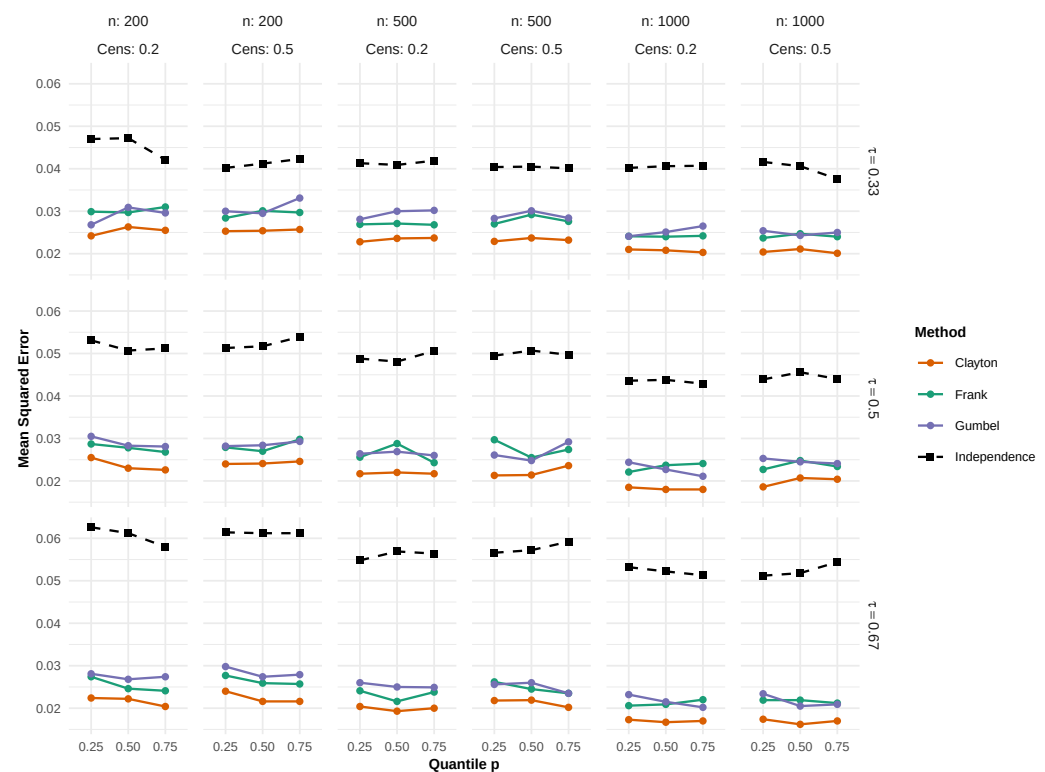


Figure 4. Mean squared error (MSE) of joint survival quantile estimators ($p = 0.25, 0.50, 0.75$), comparing independence and copula-based methods (Clayton, Frank, Gumbel) across sample sizes and censoring levels. The independence assumption is represented by dashed lines with square markers, while copula-based estimators are represented by solid lines with circular markers.

Across all scenarios, the copula-based estimators consistently outperform the product-limit estimator that assumes marginal independence in terms of empirical coverage, mean squared error (MSE), and bias, and their advantage grows as dependence and censoring increase. The key patterns are as follows:

1. Empirical coverage.

Results clearly demonstrate that assuming marginal independence (product-limit estimator) significantly compromises empirical coverage, falling substantially below the nominal 95% level in all tested scenarios. Even under weak dependence ($\tau = 0.33$), the coverage ranged between 66% and 73%, which is notably poor. Conversely, copula-based estimators (Clayton, Frank, Gumbel) substantially improved coverage, achieving 73–83%, though still below the nominal value. As dependence increased ($\tau = 0.50$ and $\tau = 0.67$), coverage with independence dropped drastically to as low as 40%, whereas copula-based methods, particularly Clayton, achieved near-nominal coverage (92–97%) under strong dependence.

2. Mean squared error (MSE).

Copula-based estimators consistently yielded lower MSE values compared to the independence estimator, especially under moderate and strong dependence scenarios. For instance, at $\tau = 0.50$ and 20% censoring, the MSE for the quantile $p = 0.25$, was reduced by approximately 40–50% using copula methods relative to independence. These improvements became more pronounced with increased sample size and reduced censoring. Notably, the Clayton copula consistently provided the lowest MSE values across all tested scenarios.

3. Bias.

The average bias across all evaluated estimators was consistently negative, indicating a slight but systematic underestimation of the true joint survival quantiles. The absolute bias never exceeded 0.06, and it decreased as the sample size increased and censoring decreased.

4.5. Discussion

Our findings align with the existing literature on bivariate survival and copula models, highlighting the severe consequences of ignoring dependence structures, leading to biased estimations and incorrect inference [19,36]. The severe under-coverage observed when assuming independence corroborates [6] assertion regarding the importance of explicitly modeling dependence to ensure accurate joint survival inferences. Similar bias phenomena under dependent censoring have also been reported in the copula literature [32], who proposed a copula-based approach to survival data with dependent censoring. Additionally, the superior performance of the Clayton copula aligns with prior studies emphasizing its effectiveness in modeling clinical events exhibiting pronounced positive dependence, such as recurrence times in oncology or paired organ failures [37,38].

Our results also suggest important practical implications. While assuming independence between marginal distributions severely underestimates joint survival probabilities, thus potentially driving inappropriate clinical decisions and resource misallocation, incorrectly specifying the copula family also introduces estimation bias, though typically less severe. Nonetheless, even when miss-selecting the copula family, copula-based estimators consistently outperform independence assumptions, highlighting their robustness and clinical relevance.

Therefore, we strongly recommend employing copula-based inference, particularly the Clayton family, in clinical contexts characterized by considerable positive dependence between event times.

5. Real-Data Application

To illustrate the applicability of the copula-driven multi-state model, we analyzed a large cohort of patients diagnosed with COVID-19 in four Colombian cities between 2021 and 2022. The dataset included sociodemographic, clinical, and vaccination information, as well as records of hospitalization and death. In this section, we first present the baseline characteristics of the study population and the marginal risk estimates obtained using Cox proportional hazards models for each transition within the illness–death framework. We then assess the dependence between hospitalization and death times using several Archimedean copula families, comparing their fit and joint survival estimates. Finally, we discuss the findings in light of the simulation results and the existing literature, emphasizing their clinical and epidemiological implications for the management of COVID-19 in the Colombian context.

The analysis of this multi-city cohort reinforces the necessity of accounting for residual dependence in multi-state survival analyses.

Baseline characteristics (Table 1) revealed a predominantly young population (57.8% aged 18–44) with high vaccination coverage (80.5%). Nevertheless, the hospitalization rate reached 50%, and the overall mortality rate was 2.4%, reflecting the substantial clinical burden of the pandemic even in a vaccinated population.

Cox regression models (Table 2) confirmed the strong protective effect of vaccination across all transitions, with particularly marked reductions in the risk of direct mortality ($HR = 0.10$) and mortality following hospitalization ($HR = 0.04$). These findings are consistent with evidence from other large-scale studies highlighting the effectiveness of vaccination in reducing severe outcomes [39–41]. Older patients (≥ 65 years) displayed

dramatically higher hazards, up to 47-fold for direct death and 42-fold for death after hospitalization, underscoring their vulnerability. Male sex and comorbidities further amplified risk, in line with international literature on COVID-19 risk factors [42].

Table 1. Baseline characteristics of the cohort.

Variable	Total (n, %)	Not Vaccinated (n, %)	Vaccinated (n, %)
Age			
18–44	1,047,007 (57.8%)	228,169 (12.6%)	818,838 (45.2%)
45–64	542,187 (29.9%)	85,112 (4.7%)	457,075 (25.2%)
≥65	221,228 (12.2%)	39,148 (2.2%)	182,080 (10.1%)
Sex			
Female	1,009,942 (55.8%)	183,849 (10.2%)	826,093 (45.6%)
Male	800,480 (44.2%)	168,580 (9.3%)	631,900 (34.9%)
Insurance type			
Contributory	1,630,231 (90%)	294,843 (16.3%)	1,335,388 (73.8%)
Subsidized	180,191 (10%)	57,586 (3.2%)	122,605 (6.8%)
Comorbidities			
No	1,547,048 (85.5%)	309,599 (17.1%)	1,237,449 (68.4%)
Yes	263,374 (14.5%)	42,830 (2.4%)	220,544 (12.2%)
Vaccinated			
No	352,429 (19.5%)	–	–
Yes	1,457,993 (80.5%)	–	–
Clinical outcomes			
Hospitalized	905,790 (50%)	184,094 (52.2%)	721,696 (49.5%)
Died	43,263 (2.4%)	28,931 (8.2%)	14,332 (1.0%)

Table 2. Adjusted hazard ratios (HRs) and 95% confidence intervals (CIs) for the three illness–death transitions. Events: 928,273 hospitalizations; 9465 deaths without hospitalization; 20,970 deaths after hospitalization.

Variable	Diagnosis → Hospitalization	Diagnosis → Death	Hospitalization → Death
Vaccinated			
No	<i>Ref</i>	<i>Ref</i>	<i>Ref</i>
Yes	0.88 (0.88–0.89)	0.10 (0.10–0.11)	0.04 (0.04–0.04)
Age			
18–44	<i>Ref</i>	<i>Ref</i>	<i>Ref</i>
45–64	1.06 (1.05–1.06)	7.96 (7.22–8.77)	9.42 (8.86–10.00)
≥65	1.12 (1.11–1.13)	47.21 (42.98–51.86)	42.35 (39.88–44.98)
Sex			
Female	<i>Ref</i>	<i>Ref</i>	<i>Ref</i>
Male	1.03 (1.02–1.03)	1.70 (1.63–1.77)	1.89 (1.83–1.94)
Comorbidity			
No	<i>Ref</i>	<i>Ref</i>	<i>Ref</i>
Yes	0.97 (0.96–0.98)	1.77 (1.69–1.85)	1.78 (1.73–1.84)
Insurance type			
Contributory	<i>Ref</i>	<i>Ref</i>	<i>Ref</i>
Subsidized	1.08 (1.07–1.09)	1.38 (1.31–1.45)	1.34 (1.30–1.39)

Values are HR (95% CI). *Ref* = reference category. Cox models stratified by city.

For each transition, we assessed the proportional-hazards (PH) assumption using Schoenfeld residual plots with LOESS smoothing and global PH tests (Figures A1–A3). To check that the marginal Cox models reproduce the observed survival, we contrasted the population-averaged survival implied by the Cox fits with the nonparametric Kaplan–Meier estimator and its 95% Greenwood band (Figure A4). We further compared transition-specific hazard rate functions (HRFs) on the log scale: the Cox hazard $h_{\text{Cox}}(t)$ was obtained by differentiating the Breslow cumulative baseline hazard (and averaging over strata), and it was contrasted with a kernel-based nonparametric estimator (muhaz) (Figure A5). Finally, total time on test (TTT) curves were used to diagnose the qualitative shape of the hazard over follow-up (Figure A6): curves below the 45° line indicate a decreasing

hazard, curves above indicate an increasing hazard, and proximity to the diagonal suggests an approximately constant hazard. We report $u_{\max}$, the proportion failed at the point of maximum vertical deviation from the diagonal, as a simple summary of departure from constancy.

For the joint model, candidate copulas were compared using log-likelihood, AIC, BIC, and CAIC, and were subjected to multiplier-bootstrap goodness-of-fit tests based on Kolmogorov–Smirnov (KS), Cramér–von Mises (CmV), and Anderson–Darling (AD) statistics, computed on rank pseudo-observations; margins were estimated semiparametrically (Tables 3 and 4).

Table 3. Copula selection for the joint modeling: maximum-likelihood estimate (θ), implied Kendall's τ , log-likelihood, and information criteria. Lower AIC/BIC/CAIC indicate better fit; best values are shown in bold.

Copula	θ	τ	LogLik	AIC	BIC	CAIC
Clayton	2.16	0.52	578.01	−1158.02	−1165.54	−1164.25
Frank	12.17	0.71	2689.83	−5377.66	−5371.14	−5370.14
Gumbel	3.59	0.72	2824.09	−5646.19	−5639.67	−5638.67

Table 4. Goodness-of-fit tests for candidate copulas (multiplier bootstrap, $B = 5000$). Reported p -values correspond to Andersen–Darling (AD), Kolmogorov–Smirnov (KS), and Cramér–von Mises (CvM) statistics. The null hypothesis is that the copula is correctly specified.

Copula	AD (p)	KS (p)	CvM (p)
Clayton	<0.01	<0.01	<0.01
Frank	<0.01	<0.01	<0.01
Gumbel	0.63	0.60	0.65

Margins were modeled semiparametrically, and pseudo-observations were constructed from ranks.

Proportional hazards (PHs) diagnostics indicated no substantial deviations from the proportional hazards assumption (global tests: Diagnosis $\rightarrow$ Hospitalization $p = 0.370$; Hospitalization $\rightarrow$ Death $p = 0.178$; and Diagnosis $\rightarrow$ Death $p = 0.097$; Figures A1–A3). Kaplan–Meier curves and the population-averaged Cox survival were nearly indistinguishable, and the Cox curves lay within the KM 95% band across transitions (Appendix A, Figure A4). Hazard-rate comparisons were consistent (Figure A5): Diagnosis $\rightarrow$ Hospitalization displayed an early peak, followed by a monotone decline; Hospitalization $\rightarrow$ Death decreased steadily; Diagnosis $\rightarrow$ Death remained low with a slight downward trend. TTT plots corroborated these patterns (Figure A6), indicating predominantly decreasing hazards with the largest deviation from constancy for Diagnosis $\rightarrow$ Hospitalization ($u_{\max} = 36.9\%$) and small deviations for Diagnosis $\rightarrow$ Death (0.5%) and Hospitalization $\rightarrow$ Death (2.2%). These checks support the use of Cox PH models for the marginal transition intensities in this application.

The copula selection analysis (Table 3) indicated that the Gumbel copula provided the best overall fit. Selection was based on multiple criteria log-likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Consistent AIC (CAIC), with Gumbel achieving the largest log-likelihood and the smallest information criteria. The implied Kendall's τ was 0.72, indicating strong upper-tail dependence between hospitalization and death. These findings are consistent with prior copula-based survival analyses in medical settings where Gumbel effectively captures upper-tail dependence [24,25].

We complemented model selection with rank-based goodness-of-fit tests for the empirical copula using the multiplier bootstrap ($B = 5000$). Table 4 reports p -values for the Kolmogorov–Smirnov (KS), Cramér–von Mises (CvM), and Anderson–Darling (AD). Only

the Gumbel copula is not rejected (e.g., $p = 0.63, 0.60, 0.65$), whereas Clayton and Frank are rejected ($p < 0.01$).

Beyond the covariate effects, the diagnostic suite (PH checks, KM–Cox agreement, HRF, and TTT) indicates that our Cox components provide a reliable marginal description of each transition. This justifies using them as the margins in the copula framework.

The Kaplan–Meier survival curves (Figure 5) illustrated distinct survival patterns for each transition, while the joint survival estimates (Figure 6) revealed a critical finding. Relative to the copula-based estimate, the independence curve lies uniformly lower, indicating systematic underestimation of joint survival when dependence is ignored. The shaded region highlights the pointwise gap between the two curves, quantifying how much joint survival would be understated under the independence assumption across follow-up. This underestimation occurs because neglecting dependence fails to capture the compounding risk when hospitalization and death are correlated. As demonstrated in our simulations, the independence assumption yielded empirical coverage rates as low as 40% under strong dependence, compared with 92–97% when copula models were applied. These results align with previous methodological studies that highlight the risks of disregarding dependence in multi-state data [6,19].

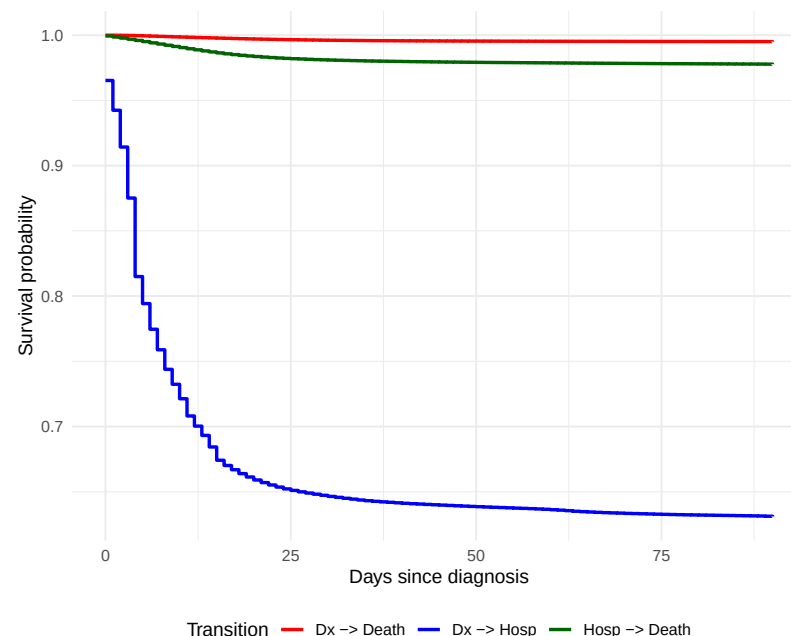


Figure 5. Kaplan–Meier curves for the three transitions. Diagnosis → Hospitalization (blue), Diagnosis → Death (red), and Hospitalization → Death (green). The steep early decline for Diagnosis → Hospitalization reflects a higher short-term risk of hospitalization after diagnosis, whereas the other transitions remain comparatively rare over follow-up.

From a clinical perspective, this misestimation is particularly concerning. Underestimating joint survival implies that clinicians and policymakers may underestimate the likelihood of patients experiencing the combined burden of hospitalization and death, potentially leading to inadequate risk stratification, delayed interventions, or misallocation of healthcare resources. As emphasized by [17,43], such underestimation can substantially distort clinical decision-making and ultimately compromise patient outcomes. In our cohort, the copula-based estimates, particularly those from the Gumbel copula, provided a more accurate representation of survival by capturing the strong positive dependence between transitions and yielding more reliable evidence to guide clinical and epidemiological planning.

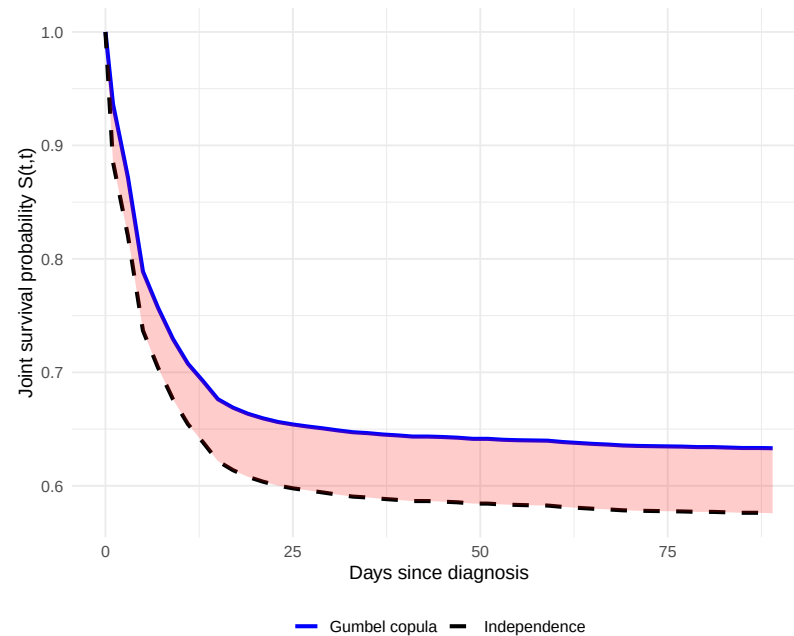


Figure 6. Joint survival probability $S(t, t)$ for progressing from diagnosis through hospitalization to death. The shaded region marks the difference between the copula-based estimate (blue) and independence (black dashed), evidencing systematic underestimation of joint survival when dependence is ignored.

In summary, these findings confirm the simulation results and reinforce a critical message: assuming independence in the presence of dependence systematically underestimates joint survival, which can have serious consequences for patient management and health policy. Copula-based multi-state models offer a robust framework to overcome this limitation and should be considered a methodological standard in contexts where sequential clinical events are strongly correlated.

6. Conclusions

This work introduced a bivariate copula-driven multi-state model that extends the conventional illness–death framework by explicitly modeling the dependence between sequential event times. Methodologically, we formulate a joint semiparametric likelihood using Inference Functions for Margins (IFM): Cox proportional hazards models provide covariate-adjusted marginal transition intensities, while an Archimedean copula encodes the dependence structure. The construction is flexible yet tractable, allowing estimation of both marginal and joint survival functions under right-censoring.

Extensive simulation experiments showed that ignoring dependence, as assumed under independence, systematically underestimates joint survival and can lead to severe coverage losses when dependence is moderate to strong. In contrast, copula-based estimators, particularly those from the Gumbel and Clayton families, achieved near-nominal coverage and lower mean squared error, confirming the theoretical robustness of the proposed framework even under partial copula misspecification.

The large Colombian COVID-19 cohort further validated the approach. First, model-fit diagnostics supported the adequacy of the Cox margins: global PH tests showed no material violations; population-averaged Cox survival closely tracked Kaplan–Meier with Greenwood bands; hazard-rate functions (Cox vs. kernel) displayed the expected shapes; and TTT plots indicated predominantly decreasing hazards with the largest deviation from constancy for Diagnosis → Hospitalization. Second, copula selection favored Gum-

bel by pseudo-log-likelihood and information criteria (AIC/BIC/CAIC), and rank-based bootstrap GOF tests (AD/KS/CvM) failed to reject Gumbel while rejecting Clayton and Frank. Substantively, the estimated upper-tail dependence between hospitalization and death was strong (Kendall's $\tau \approx 0.72$), consistent with clinical intuition about severe disease progression.

These results have practical implications. Assuming independence when transitions are correlated can underestimate the joint burden of hospitalization and death, potentially distorting risk stratification, timing of interventions, and resource planning. By combining covariate-adjusted Cox margins with a well-supported copula, our framework yields more reliable joint survival estimates, offering a principled alternative to independence-based approaches in clinical and epidemiological studies.

7. Limitations and Future Work

This study has several limitations that also suggest natural extensions. First, the dependence structure was restricted to one-parameter Archimedean copulas. This choice provides parsimony and interpretability, but it imposes a single, global form of dependence and emphasizes either the lower or the upper tail. Second, the marginal transition intensities were modeled under the proportional-hazards (PHs) assumption via Cox models. Although our Schoenfeld and TTT/HRF diagnostics supported PH for this application, the assumption can be violated in other settings. Third, we estimated the joint model using Inference Functions for Margins (IFM). IFM is consistent and computationally attractive, but it ignores some cross-equation curvature; deriving finite-sample properties and standard-error formulas under censoring remains an open problem. Finally, the population-averaged HRF was obtained by differentiating the Breslow baseline; uncertainty from this transformation was not propagated into the HRF plots.

Several avenues can strengthen and broaden the framework:

- **Margins.** Replace/compare Cox with parametric hazards (e.g., alpha-power, piecewise exponential, or Royston–Parmar) and, for death transitions showing plateaus, mixture-cure models; allow time-varying or stratified effects when PH is doubtful.
- **Dependence.** Enrich the copula layer via nested Archimedean or vine copulas, or covariate/time-varying copula parameters to capture evolving dependence.
- **Estimation.** Move beyond IFM to full-likelihood or Bayesian joint estimation and carry uncertainty into HRF summaries; study finite-sample properties under censoring.

These steps are compatible with the present framework and, in this dataset, the most immediately impactful additions are as follows: alpha-power parametric margins and cure-fraction models for the death transitions.

Author Contributions: Conceptualization, H.B., G.M.-F. and R.T.-F.; methodology, H.B.; formal analysis, H.B.; investigation, H.B., G.M.-F. and R.T.-F.; data curation, H.B., G.M.-F. and R.T.-F.; writing—original draft preparation, H.B.; writing—review and editing, H.B.; project administration, G.M.-F. and R.T.-F.; funding acquisition, G.M.-F. and R.T.-F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Vice-rectorate for Research of the Universidad de Córdoba, Colombia, project grant FCB-03-23: “Aplicación de Metodologías Estadísticas a Datos de Vigilancia en Salud Pública en Colombia” (R.T.-F. and G.M.-F.).

Data Availability Statement: The data and codes used in this study are available upon request from the corresponding author. The data are not publicly available due to privacy restrictions.

Acknowledgments: This article was developed as a research product during the probationary period of the first author (H.B.) at the Universidad de Sucre. The authors also acknowledge the institutional support of the Universidad de Sucre and the Universidad de Córdoba, which provided the academic

environment and resources that contributed to the completion of this work. The authors are also grateful to the anonymous reviewers for their careful reading and valuable suggestions, which helped improve the quality of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IFM	Inference Functions for Margins
MSE	Mean Squared Error
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
CAIC	Consistent AIC
HRF	Hazard-Rate Functions
TTT	Total Time on Test
PHs	Proportional Hazards

Appendix A. Simulation Results and Marginal Fit Diagnostics

Appendix A.1. Detailed Simulation Tables

The following tables, Tables A1–A3, provide a complete numerical account of our Monte Carlo study. For each copula family (Clayton, Frank, Gumbel) and the independence benchmark, they list the average bias, mean squared error (MSE), and empirical coverage of 95% confidence intervals for the joint survival quantiles $p = 0.25, 0.50$ and 0.75 under three dependence levels ($\tau = 0.25, 0.50, 0.75$), across varying sample sizes and censoring proportions. These detailed results support the summary figures in the main text and allow readers to verify the robustness of each estimator under all considered scenarios.

Table A1. Simulation performance metrics (Bias, MSE, and Coverage) for joint survival quantile estimators at $p = 0.25, 0.50$, and 0.75 under dependence $\tau = 0.33$.

n	Cens	Method	$t_{0.25}$			$t_{0.50}$			$t_{0.75}$		
			Bias	MSE	Cover	Bias	MSE	Cove	Bias	MSE	Cover
200	20%	Indepen.	−0.0315	0.0470	69.6248	−0.0301	0.0472	70.3150	−0.0296	0.0421	68.7631
		Clayton	−0.0123	0.0242	77.7591	−0.0099	0.0263	79.8654	−0.0090	0.0255	80.0428
		Frank	−0.0153	0.0299	75.0420	−0.0148	0.0297	76.8876	−0.0108	0.031	73.4736
		Gumbel	−0.015	0.0268	73.2808	−0.0136	0.0309	72.4536	−0.0110	0.0296	71.9241
	50%	Indepen.	−0.0314	0.0402	66.3335	−0.0302	0.0412	65.2388	−0.0294	0.0423	64.6366
		Clayton	−0.0125	0.0253	75.8529	−0.0130	0.0254	78.0613	−0.0120	0.0257	75.8581
		Frank	−0.0171	0.0284	69.5521	−0.0177	0.0301	70.8179	−0.0177	0.0297	71.0670
		Gumbel	−0.0146	0.0300	68.6262	−0.0151	0.0295	69.6863	−0.0154	0.0331	70.0467
500	20%	Indepen.	−0.0285	0.0413	69.0889	−0.0298	0.0409	72.9304	−0.0293	0.0419	70.7275
		Clayton	−0.0094	0.0228	80.7339	−0.0114	0.0236	80.7310	−0.0100	0.0237	80.4574
		Frank	−0.0136	0.0269	77.6692	−0.0173	0.0271	76.2826	−0.0142	0.0268	76.4679
		Gumbel	−0.0119	0.0281	72.8579	−0.0154	0.03	71.8125	−0.0129	0.0302	73.9711
	50%	Indepen.	−0.0301	0.0404	66.9640	−0.0299	0.0405	67.2321	−0.031	0.0401	67.3374
		Clayton	−0.0099	0.0229	77.1285	−0.0097	0.0237	80.3641	−0.0105	0.0232	77.6133
		Frank	−0.0142	0.0270	74.0208	−0.0143	0.0292	73.5759	−0.0171	0.0276	72.7427
		Gumbel	−0.0139	0.0283	71.8653	−0.0131	0.0301	70.8162	−0.0140	0.0284	71.3511
1000	20%	Indepen.	−0.0293	0.0402	71.8346	−0.0292	0.0406	72.3745	−0.0292	0.0407	73.5831
		Clayton	−0.0093	0.0210	82.7723	−0.0061	0.0208	80.6443	−0.0081	0.0203	83.1196
		Frank	−0.0130	0.0241	75.7694	−0.0097	0.0240	77.8382	−0.0108	0.0242	79.1307
		Gumbel	−0.0134	0.0241	76.0206	−0.0089	0.0251	75.3124	−0.0118	0.0265	76.0794
	50%	Indepen.	−0.0301	0.0416	69.6842	−0.0314	0.0406	69.8166	−0.0281	0.0376	70.5177
		Clayton	−0.0090	0.0204	80.4389	−0.0088	0.0211	79.4383	−0.0082	0.0201	81.4859
		Frank	−0.0124	0.0237	76.7811	−0.0141	0.0247	75.4459	−0.0118	0.0240	75.8386
		Gumbel	−0.0106	0.0254	73.0186	−0.0122	0.0243	71.3825	−0.0113	0.0250	73.3120

Table A2. Simulation performance metrics (Bias, MSE, and Coverage) for joint survival quantile estimators at $p = 0.25, 0.50$, and 0.75 under dependence $\tau = 0.5$.

n	Cens	Method	$t_{0.25}$			$t_{0.50}$			$t_{0.75}$		
			Bias	MSE	Cover	Bias	MSE	Cove	Bias	MSE	Cover
200	20%	Indepen.	−0.041	0.0531	58.543	−0.0389	0.0507	59.8173	−0.0402	0.0512	59.2404
		Clayton	−0.0082	0.0255	84.9564	−0.0083	0.023	82.8707	−0.0078	0.0226	83.4459
		Frank	−0.01	0.0287	76.1301	−0.0105	0.0278	75.6787	−0.011	0.0268	74.2828
		Gumbel	−0.0125	0.0305	72.3124	−0.0098	0.0283	74.0408	−0.0108	0.0281	72.6693
	50%	Indepen.	−0.039	0.0513	58.0232	−0.0392	0.0517	56.3771	−0.0413	0.0539	54.9623
		Clayton	−0.008	0.024	80.2138	−0.0079	0.0241	82.1695	−0.0071	0.0246	82.4536
		Frank	−0.0131	0.0279	73.4811	−0.0126	0.027	72.3782	−0.0109	0.0298	71.1431
		Gumbel	−0.0131	0.0282	70.1113	−0.0092	0.0284	70.2578	−0.0089	0.0293	70.1408
500	20%	Indepen.	−0.0432	0.0488	59.7813	−0.0393	0.0481	60.9558	−0.0394	0.0506	60.0366
		Clayton	−0.0078	0.0217	86.4482	−0.0073	0.022	85.6575	−0.0081	0.0217	85.724
		Frank	−0.0119	0.0256	77.4783	−0.0119	0.0288	76.5898	−0.0118	0.0243	78.4892
		Gumbel	−0.0106	0.0264	75.4871	−0.0081	0.0269	74.6786	−0.0105	0.026	74.4679
	50%	Indepen.	−0.04	0.0495	57.3644	−0.0414	0.0507	57.0399	−0.0404	0.0497	58.3495
		Clayton	−0.0071	0.0213	83.24	−0.0070	0.0214	82.423	−0.0082	0.0236	81.9138
		Frank	−0.0107	0.0297	73.6307	−0.0124	0.0255	73.8452	−0.0121	0.0274	74.9103
		Gumbel	−0.0109	0.0261	73.4743	−0.0109	0.0248	74.1702	−0.0109	0.0292	73.2449
1000	20%	Indepen.	−0.0416	0.0436	62.851	−0.0398	0.0438	63.9631	−0.0404	0.0429	60.7854
		Clayton	−0.0059	0.0185	88.4837	−0.0079	0.0180	85.6872	−0.0069	0.0180	87.0821
		Frank	−0.0094	0.0221	80.2484	−0.0118	0.0237	78.9311	−0.0118	0.0241	77.9725
		Gumbel	−0.0076	0.0244	77.6773	−0.0109	0.0227	79.1138	−0.0111	0.0211	79.9683
	50%	Indepen.	−0.0393	0.0439	60.6706	−0.0406	0.0456	59.8494	−0.0411	0.044	59.1562
		Clayton	−0.0064	0.0186	85.1172	−0.0056	0.0207	85.8221	−0.0065	0.0204	83.6295
		Frank	−0.0097	0.0227	77.5161	−0.0094	0.0248	77.1136	−0.0101	0.0234	77.8659
		Gumbel	−0.0098	0.0253	75.4597	−0.0109	0.0245	74.7300	−0.0091	0.0241	74.7189

Table A3. Simulation performance metrics (Bias, MSE, and Coverage) for joint survival quantile estimators at $p = 0.25, 0.50$, and 0.75 under dependence $\tau = 0.67$.

n	Cens	Method	$t_{0.25}$			$t_{0.50}$			$t_{0.75}$		
			Bias	MSE	Cover	Bias	MSE	Cove	Bias	MSE	Cover
200	20%	Indepen.	−0.0604	0.0626	47.3231	−0.0619	0.0612	44.7579	−0.0597	0.058	43.9796
		Clayton	−0.0063	0.0224	94.1316	−0.0052	0.0222	95.4655	−0.0059	0.0204	93.9587
		Frank	−0.0117	0.0274	80.4285	−0.0092	0.0246	79.5198	−0.0094	0.0241	78.8466
		Gumbel	−0.0089	0.0281	78.5178	−0.008	0.0268	73.868	−0.0079	0.0274	75.8526
	50%	Indepen.	−0.0593	0.0614	40.7743	−0.0611	0.0612	39.9696	−0.0604	0.0612	38.7751
		Clayton	−0.0055	0.024	91.4467	−0.0047	0.0216	89.7386	−0.0057	0.0216	89.8389
		Frank	−0.0101	0.0277	76.1474	−0.0093	0.0259	76.0295	−0.0079	0.0257	78.6842
		Gumbel	−0.0084	0.0298	72.7686	−0.0087	0.0274	72.1694	−0.0072	0.0279	72.9548
500	20%	Indepen.	−0.061	0.0548	40.8398	−0.0592	0.0569	41.5215	−0.057	0.0564	42.4796
		Clayton	−0.0037	0.0204	96.585	−0.0055	0.0193	94.9181	−0.0034	0.02	95.6602
		Frank	−0.0074	0.0241	81.6106	−0.0091	0.0216	82.2602	−0.0073	0.0238	83.26
		Gumbel	−0.0065	0.026	79.856	−0.0066	0.025	78.0756	−0.0064	0.0249	79.1335
	50%	Indepen.	−0.0597	0.0566	39.9407	−0.0607	0.0572	38.713	−0.0605	0.0592	38.133
		Clayton	−0.003	0.0218	93.2805	−0.0026	0.0219	93.9646	−0.0053	0.0202	92.655
		Frank	−0.0074	0.0262	78.8456	−0.0072	0.0245	77.3108	−0.0105	0.0235	77.0498
		Gumbel	−0.0054	0.0256	73.4647	−0.0069	0.026	73.187	−0.0098	0.0235	73.3557
1000	20%	Indepen.	−0.0593	0.0532	42.851	−0.0604	0.0522	43.2692	−0.0613	0.0513	43.8451
		Clayton	−0.0023	0.0173	96.7774	−0.0021	0.0167	94.5497	−0.0005	0.017	93.7262
		Frank	−0.0062	0.0206	84.083	−0.0063	0.0209	84.0673	−0.005	0.022	83.864
		Gumbel	−0.0061	0.0232	78.5108	−0.0069	0.0215	80.4637	−0.004	0.0202	80.2535
	50%	Indepen.	−0.0597	0.0512	42.379	−0.0624	0.0518	40.1266	−0.0585	0.0544	40.7119
		Clayton	−0.002	0.0174	92.5743	−0.0013	0.0162	91.6146	−0.0025	0.017	90.2778
		Frank	−0.0051	0.0219	80.9206	−0.0054	0.0219	81.777	−0.0072	0.0212	79.9565
		Gumbel	−0.0046	0.0234	77.8571	−0.0048	0.0205	77.2225	−0.0054	0.0209	76.9871

Appendix A.2. Marginal Fits from the Application

The results presented in this appendix display the proportional-hazards diagnostics for each transition of the illness–death process. Schoenfeld residual plots with LOESS smoothing and global tests are reported, allowing a visual and statistical assessment of whether the proportional hazards assumption holds across covariates. These results confirm

the adequacy of the Cox proportional hazards models as marginal specifications within the copula-based framework.

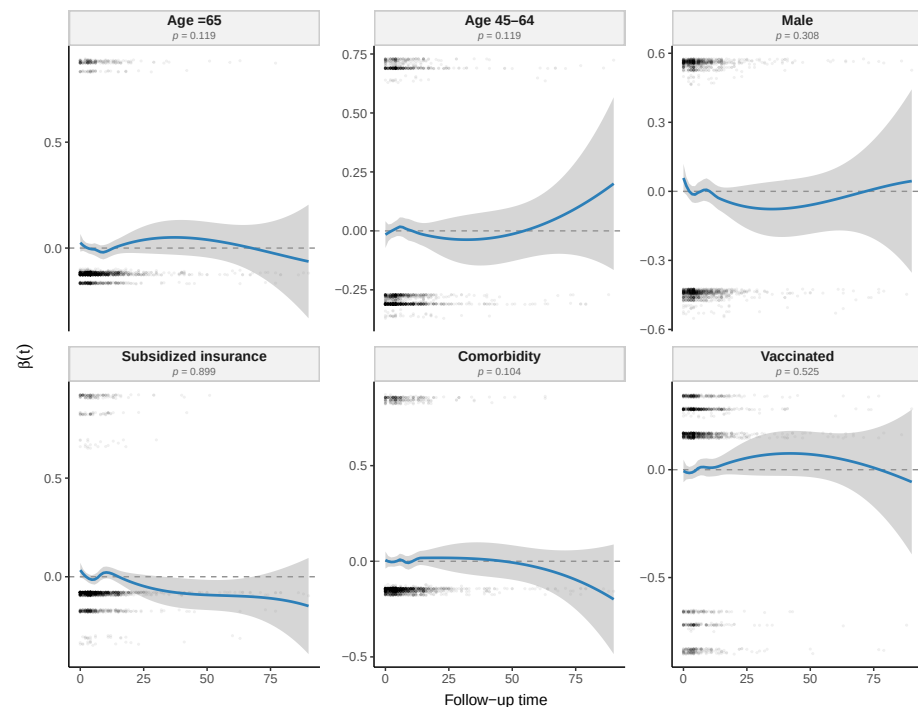


Figure A1. Proportional-hazards diagnostics for the diagnosis-hospitalization transition. Schoenfeld residuals by covariate with LOESS smooth (blue) and 95% pointwise confidence band (gray); the horizontal dashed line marks zero. Panel labels report covariate-specific p -values for proportionality. Global proportional-hazards test: $p = 0.370$.

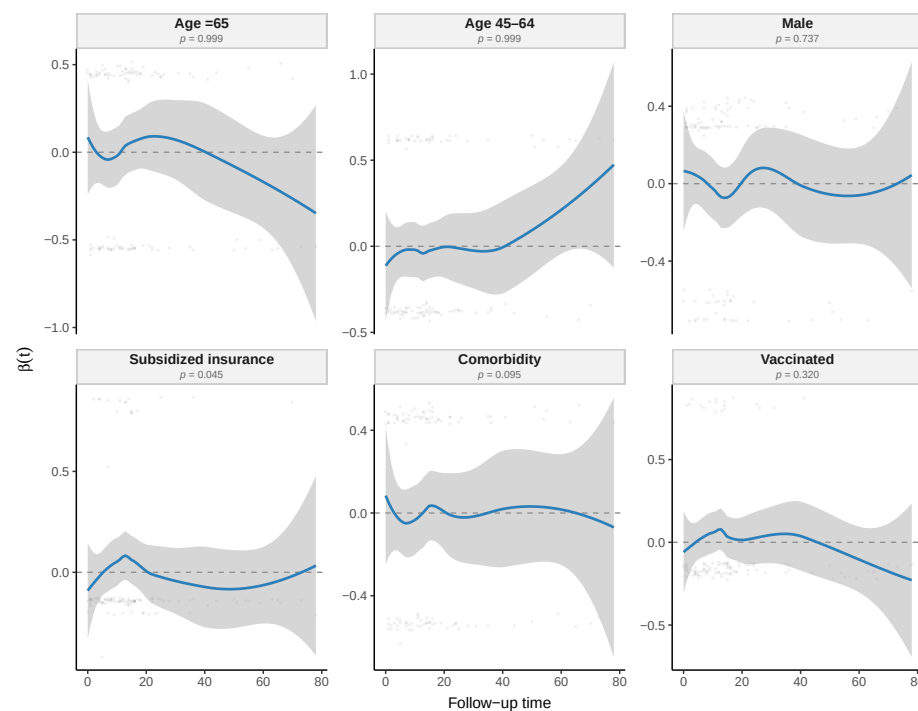


Figure A2. Proportional-hazards diagnostics for the hospitalization-death transition. Schoenfeld residuals by covariate with LOESS smooth (blue) and 95% pointwise confidence band (gray); the horizontal dashed line marks zero. Panel labels report covariate-specific p -values for proportionality. Global proportional-hazards test: $p = 0.178$.

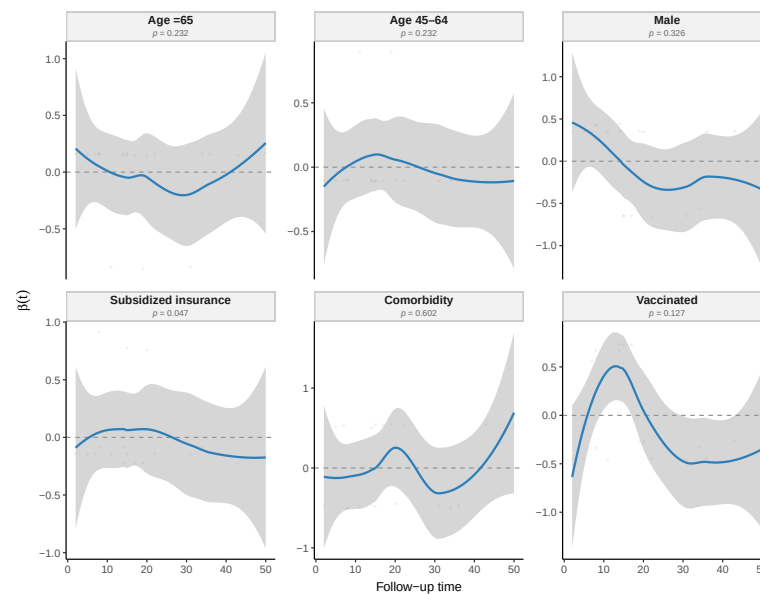


Figure A3. Proportional-hazards diagnostics for the diagnosis–death transition. Schoenfeld residuals by covariate with LOESS smooth (blue) and 95% pointwise confidence band (gray); the horizontal dashed line marks zero. Panel labels report covariate-specific p -values for proportionality. Global proportional-hazards test: $p = 0.097$.

Appendix A.3. Compatibility of Adjustments to Data in Each Transition

The following results compare semiparametric Cox models against nonparametric estimators to evaluate the consistency of marginal fits. Kaplan–Meier curves with Greenwood confidence intervals are contrasted with population-averaged Cox survival curves, transition-specific hazards are compared against kernel estimators, and TTT plots summarize hazard shapes across follow-up. Collectively, these results demonstrate that the Cox models reproduce the empirical survival patterns, supporting their use as the margins in the joint copula model.

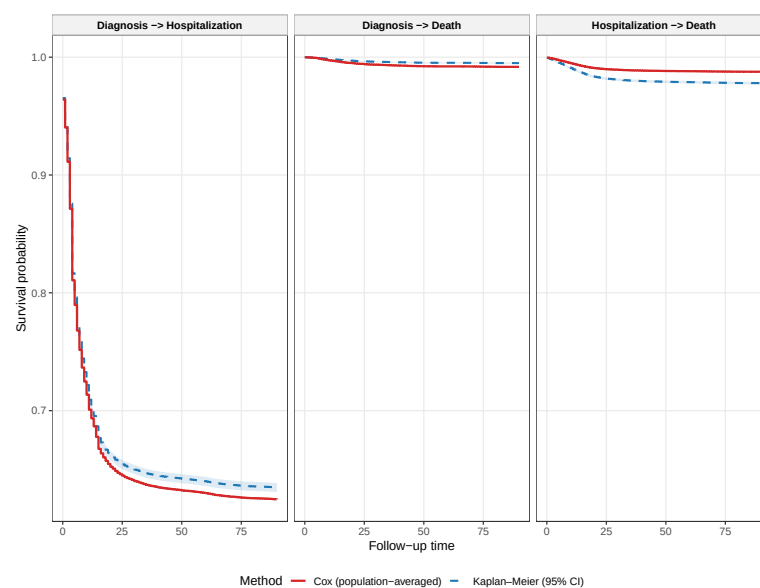


Figure A4. Transition-specific marginal survival: Kaplan–Meier (95% CI) vs. Cox (population-averaged). Marginal survival for each transition. The blue dashed lines show the Kaplan–Meier estimator with a 95% Greenwood confidence band (shaded); the red solid lines show the population-averaged survival from the Cox marginal models.

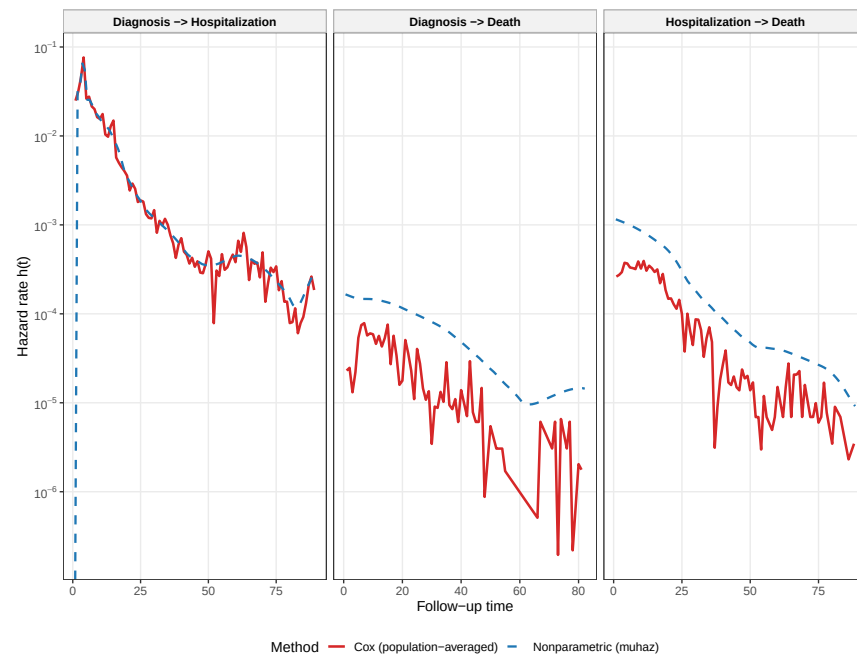


Figure A5. Estimated hazard rate functions, $h(t)$, on a log scale for the three transitions. The red solid line shows the Cox population-averaged hazard obtained by differentiating the Breslow cumulative baseline hazard and averaging across strata; the blue dashed line shows the nonparametric kernel estimator (implemented via muhaz). Time since origin is on the x-axis.

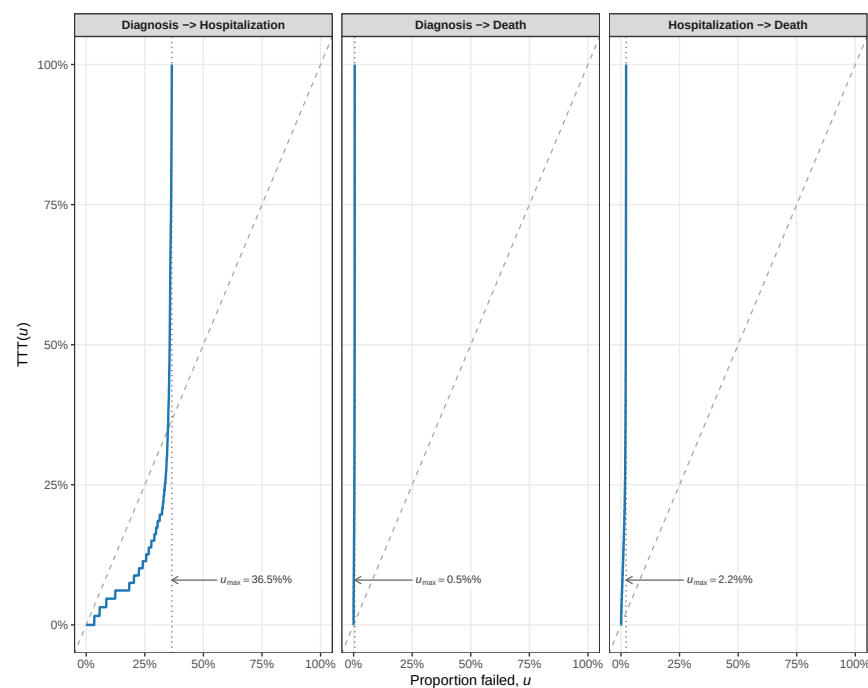


Figure A6. Total time on test (TTT) plots for each transition. The empirical TTT curve (blue) is compared to the 45° reference line (grey). Values of $u_{\max}$ (proportion failed at the maximum vertical deviation from the diagonal) are indicated within each panel: Diagnosis → Hospitalization $u_{\max} = 36.9\%$, Diagnosis → Death $u_{\max} = 0.5\%$, and Hospitalization → Death $u_{\max} = 2.2\%$. Curves lying below the diagonal indicate a decreasing hazard over time; curves above indicate an increasing hazard; proximity to the diagonal suggests an approximately constant hazard.

References

- Klein, J.P.; Moeschberger, M.L. *Survival Analysis: Techniques for Censored and Truncated Data*; Springer: New York, NY, USA, 2003. [CrossRef]
- Kalbfleisch, J.D.; Prentice, R.L. *The Statistical Analysis of Failure Time Data*, 2nd ed.; Wiley: New York, NY, USA, 2002. [CrossRef]
- Collett, D. *Modelling Survival Data in Medical Research*, 3rd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2015. [CrossRef]
- Cox, D.R. Regression models and life-tables. *J. R. Stat. Soc. Ser. B* **1972**, *34*, 187–220. [CrossRef]
- Therneau, T.M.; Grambsch, P.M. *Modeling Survival Data: Extending the Cox Model*; Springer: New York, NY, USA, 2000. [CrossRef]
- Hougaard, P. *Analysis of Multivariate Survival Data*; Springer: New York, NY, USA, 2000. [CrossRef]
- Meira-Machado, L.; de Uña-Álvarez, J.; Cadarso-Suárez, C.; Andersen, P.K. Multi-state models for the analysis of time-to-event data. *Stat. Methods Med. Res.* **2009**, *18*, 195–222. [CrossRef]
- Beyersmann, J.; Allignol, A.; Schumacher, M. *Competing Risks and Multistate Models with R*; Springer: New York, NY, USA, 2012. [CrossRef]
- Putter, H.; Fiocco, M.; Geskus, R.B. Tutorial in biostatistics: Competing risks and multi-state models. *Stat. Med.* **2007**, *26*, 2389–2430. [CrossRef]
- Andersen, P.K.; Keiding, N. Multi-state models for event history analysis. *Stat. Methods Med. Res.* **2002**, *11*, 91–115. [CrossRef]
- Cook, R.J.; Lawless, J. *The Statistical Analysis of Recurrent Events*; Springer: New York, NY, USA, 2007. [CrossRef]
- van Houwelingen, H.C.; Putter, H. Dynamic Predicting by Landmarking as an Alternative for Multi-State Modeling: An Application to Acute Lymphoid Leukemia Data. *Lifetime Data Anal.* **2008**, *14*, 447–463. [CrossRef]
- De Uña-Álvarez, J.; Meira-Machado, L. Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study. *Biometrics* **2015**, *71*, 364–375. [CrossRef] [PubMed]
- Li, Z.; Chinchilli, V.M.; Wang, M. A Bayesian Joint Model of Recurrent Events and a Terminal Event. *Biom. J.* **2019**, *61*, 187–202. [CrossRef]
- Ferrer, L.; Rondeau, V.; Dignam, J.; Pickles, T.; Jacqmin-Gadda, H.; Proust-Lima, C. Joint Modelling of Longitudinal and Multi-State Processes: Application to Clinical Progressions in Prostate Cancer. *Stat. Med.* **2016**, *35*, 3933–3948. [CrossRef]
- Ramezankhani, A.; Blaha, M.J.; Mirbolouk, M.H.; Azizi, F.; Hadaegh, F. Multi-State Analysis of Hypertension and Mortality: Application of Semi-Markov Model in a Longitudinal Cohort Study. *BMC Cardiovasc. Disord.* **2020**, *20*, 321. [CrossRef] [PubMed]
- Nelsen, R.B. *An Introduction to Copulas*, 2nd ed.; Springer: New York, NY, USA, 2006. [CrossRef]
- Joe, H. *Multivariate Models and Dependence Concepts*; Chapman & Hall: London, UK, 1997. [CrossRef]
- Emura, T.; Matsui, S.; Rondeau, V. *Survival Analysis with Correlated Endpoints: Joint Frailty-Copula Models*; Springer: Singapore, 2019. [CrossRef]
- Othus, M.; Li, Y. A Gaussian Copula Model for Multivariate Survival Data. *Stat. Biosci.* **2010**, *2*, 154–179. [CrossRef] [PubMed]
- Gasparini, A.; Humphreys, K. A Natural History and Copula-Based Joint Model for Regional and Distant Breast Cancer Metastasis. *Stat. Methods Med. Res.* **2022**, *31*, 2415–24300. [CrossRef]
- Shewa, F.; Endale, S.; Nugussu, G.; Abdisa, J.; Zerihun, K.; Banbeta, A. Time to Kidneys Failure Modeling in the Patients at Adama Hospital Medical College: Application of Copula Model. *J. Res. Health Sci.* **2022**, *22*, e00549. [CrossRef]
- Cheung, L.C.; Albert, P.S.; Das, S.; Cook, R.J. Multistate Models for the Natural History of Cancer Progression. *Br. J. Cancer* **2022**, *127*, 1279–1288. [CrossRef] [PubMed]
- Shewa Gari, F.; Fenta Biru, T.; Endale Gurmu, S. Application of the Joint Frailty Copula Model for Analyzing Time to Relapse and Time to Death of Women with Cervical Cancer. *Int. J. Women's Health* **2023**, *15*, 1295–13046. [CrossRef] [PubMed]
- Ieva, F.; Jackson, C.H.; Sharples, L.D. Multi-State Modelling of Repeated Hospitalisation and Death in Patients with Heart Failure: The Use of Large Administrative Databases in Clinical Epidemiologys. *Stat. Methods Med. Res.* **2017**, *26*, 1350–1372. [CrossRef] [PubMed]
- Le-Rademacher, J.G.; Therneau, T.M.; Ou, F.S. The Utility of Multistate Models: A Flexible Framework for Time-to-Event Data. *Curr. Epidemiol. Rep.* **2022**, *9*, 183–189. [CrossRef]
- Sklar, A. Fonctions de répartition à n dimensions et leurs marges. *Publ. L'Institut Stat. L'Université Paris* **1959**, *8*, 229–231. Available online: <https://hal.science/hal-04094463/document> (accessed on 20 July 2025).
- Joe, H. *Dependence Modeling with Copulas*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2014. [CrossRef]
- Oakes, D. Bivariate survival models induced by frailties. *J. Am. Stat. Assoc.* **1989**, *84*, 487–493. [CrossRef]
- Genest, C.; Rémillard, B.; Beaudoin, D. Goodness-of-fit tests for copulas: A review and a power study. *Insur. Math. Econ.* **2009**, *44*, 199–213. [CrossRef]
- Hafner, C.M.; Reznikova, O. Efficient estimation of a semiparametric dynamic copula model. *Comput. Stat. Data Anal.* **2010**, *54*, 2609–2627. [CrossRef]
- Emura, T.; Chen, Y.H. Gene Selection for Survival Data under Dependent Censoring: A Copula-Based Approach. *Stat. Methods Med. Res.* **2014**, *25*, 2840–2857. [CrossRef] [PubMed]

33. Erdmann, A.; Loos, A.; Beyersmann, J. A Connection Between Survival Multistate Models and Causal Inference for External Treatment Interruptions. *Stat. Methods Med. Res.* **2023**, *32*, 697–712. [[CrossRef](#)]
34. Li, J.; Fine, J. On sample size for sensitivity and specificity in prospective diagnostic accuracy studies. *Stat. Med.* **2004**, *23*, 2537–2550. [[CrossRef](#)]
35. Zhou, B.; Fine, J.; Laird, G. Goodness-of-Fit Test for Proportional Subdistribution Hazards Model. *Stat. Med.* **2013**, *32*, 3804–3811. [[CrossRef](#)]
36. Shih, J.H.; Louis, T.A. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* **1995**, *51*, 1384–1399. [[CrossRef](#)]
37. Lakhal-Chaieb, L.; Rivest, L.P.; Abdous, B. Estimating survival under dependent truncation. *Biometrika* **2006**, *93*, 655–669. [[CrossRef](#)]
38. Emura, T.; Chen, Y.H. *Analysis of Survival Data with Dependent Censoring*; Springer: Berlin/Heidelberg, Germany, 2018. [[CrossRef](#)]
39. Arbel, R.; Hammerman, A.; Sergienko, R.; Friger, M.; Peretz, A.; Netzer, D.; Yaron, S. BNT162b2 Vaccine Booster and Mortality Due to Covid-19. *New Engl. J. Med.* **2021**, *385*, 2413–2420. [[CrossRef](#)]
40. de Gier, B.; van Asten, L.; Boere, T.M.; van Roon, A.; van Roekel, C.; Pijpers, J.; van Werkhoven, C.H.H.; van den Ende, C.; Hahné, S.J.M.; de Melker, H.E.; et al. Effect of COVID-19 vaccination on mortality by COVID-19 and on mortality by other causes, the Netherlands, January 2021–January 2022. *Vaccine* **2023**, *41*, 4488–4496. [[CrossRef](#)]
41. González Rodríguez, J.L.; Oprescu, A.M.; Muñoz Lezcano, S.; Cordero Ramos, J.; Romero Cabrera, J.L.; Armengol de la Hoz, M.A.; Estella, A. Assessing the Impact of Vaccines on COVID-19 Efficacy in Survival Rates: A Survival Analysis Approach for Clinical Decision Support. *Front. Public Health* **2024**, *12*, 1437388. [[CrossRef](#)] [[PubMed](#)]
42. Peckham, H.; de Gruijter, N.M.; Raine, C.; Radziszewska, A.; Ciurtin, C.; Wedderburn, L.R.; Rosser, E.C.; Webb, K.; Deakin, C.T. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. *Nat. Commun.* **2020**, *11*, 6317. [[CrossRef](#)] [[PubMed](#)]
43. Li, N.; Zhao, M.; Xu, L. Bivariate copula regression models for semi-competing risks. *Stat. Methods Med. Res.* **2023**, *32*, 843–859. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.