Relaxing the Independence Assumption in Relative Survival Analysis: A Parametric Approach

Reuben Adatorwovor[1*], Aurelien Latouche[2], and Jason P Fine[1]

[1]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27514, U.S.A.

[2] Conservatoire National des Arts et Métiers, Paris, France; Institut Curie, St-Cloud, France.

[*]Corresponding author. Email: radat223@email.unc.edu

Abstract

With known cause of death, competing risk survival methods are applicable in estimating disease-specific survival. Relative survival analysis may be used to estimate disease-specific survival when cause of death is either unknown or subject to misspecification and not reliable for practical usage. This method is popular for population-based cancer survival studies using registry data and does not require cause of death information. The standard estimator is the ratio of all-cause survival in the cancer cohort group to the known expected survival from a healthy reference population. Disease-specific death competes with other causes of mortality, potentially creating dependence among the causes of death. The standard ratio estimate is only valid when death from disease and death from other causes are independent. To relax the independence assumption, we formulate dependence using a copula-based model. Likelihood based method is used to fit a parametric model to the distribution of disease-specific death without cause of death information, where the copula is assumed known and the distribution of other cause of mortality is derived from the reference population. Since the dependence structure for disease related and other-cause mortality is nonidentifiable and unverifiable from the observed data, we propose a sensitivity analysis, where the analysis is conducted across a range of assumed dependence structures. We demonstrate the practical utility of our method through simulation studies and an application to French breast cancer data.

1 Introduction

Cancer patients including breast, prostate, endometrial and thyroid cancer are at higher risk of dying from heart disease and stroke than the general population. As the number of cancer survivors increases, so is the rate of cardiovascular deaths (Sturgeon *et al.*, 2019). Such medical research frequently yields multiple event times which may consist of a terminal and or a non-terminal event.The practical concern for physicians is

patient survival, suggesting an analysis based on the distribution of these event times or the disease-specific hazard and or cumulative incidence function. Yet, the scientific interest in understanding disease-specific mortality in the absence of failure types other than the disease of interest (net survival), a quantity which is sometimes controversial but meaningful to many practitioners and or researchers by permitting comparisons across populations with different background mortality. Other researchers prefer the latter quantity (crude survival) in understanding disease-specific mortality in the presence of other competing causes.

With improvement in medical treatment and long follow-up in population-based disease registries, there is a potential for lost to follow-up during which patients may either experience disease-specific death or death from non-disease related causes (Brinkhof *et al.*, 2010). In such competing risk settings where one death type precludes the occurrence of other types, standard methodology assumes that cause of death is known Gichangi and Vach, 2005). In the analysis of competing risks events from registry data, accurate documentation of death is essential (Percy *et al.*, 1981; Welch and Black, 2002; and Mieno *et al.*, 2016). A challenge is that documentation either may not be available, or may be incomplete or incorrect for cause of death, resulting in problems distinguishing disease and non-disease related mortality. The issue is pronounced in Europe, where comparison of disease-specific survival across countries is of interest. The World Health Organization (World Health Organization, 1977) defines cause of death as "the disease or injury which initiated the train of morbid events leading directly to death". However, population-based disease registries may not be harmonized across countries, leading to imprecise cause of death definitions and different levels of documentation of cause of death information. Often, the underlying cause of death may be unclear as hospital coding of cancer death may not agree with the death certificate coding. As an example, Welch and Black (2002) reported that 41% of deaths that occurred (within one month diagnosis and cancer directed surgery) were not attributable to the coded cancer in the registry. When reliable cause of death information is available, it is often located in separate databases, which may be costly to obtain and difficult to link with registry data.

Suppose that $T = \min\{T_k : k = 1, 2, 3, \cdots, T_K\}$ is the potentially observable failure time and $\varepsilon = \{k : T = T_k\}$ the failure type where $T_1, \cdots, K$, with $K \in \mathbf{N}$ are the latent failure times associated with the K failure types. In registry data, $K = 2$ and $\varepsilon = 1$ implies death from cancer and $\varepsilon = 2$ implies death from other competing causes. Standard methods for independently right censored survival data without competing risks cannot generally be used to make inference about disease-specific survival. Under dependent competing risks, where T_1 and T_2 are dependent, the Kaplan-Meier (Kaplan and Meier, 1958) curve estimates a function of the cause-specific hazard function, defined in Section 2. The logrank test (Bland and Altman, 2004) assesses group differences between the cause-specific hazard function, while the standard proportional hazards model (Cox, 1972) formulates the effects of covariates on the cause-specific hazard function. The cumulative incidence function, defined in Section 2, gives disease-specific survival in the presence of competing events. This quantity has been widely adopted in applications, with the Aalen-Johanson estimator (Aalen and Johansen, 1978), Gray's test Gray's test (Gray, 1988), and the Fine-Gray model (Fine and Gray, 1999), providing analogs to the Kaplan-Meier curve, the logrank test, and the proportional hazards model for the cumulative incidence function. Without cause of death information, these methods are not applicable.

To address disease-specific survival without reliable cause of death information, relative survival methods have been proposed. Relative survival, $S_R(t)$ is the ratio of the observed survival rate in a group of cancer patients, during a specified period, to the expected survival rate in a healthy reference population, (Ederer,

1961). Mathematically,

$$S_R(t) = \frac{S_O(t)}{S_P(t)} \tag{1}$$

where at time t, $S_O(t)$ is the observed survival probability from the registry and $S_P(t)$ is the expected survival from mortality tables. Existing literature has focused exclusively on the estimation of $S_R(t)$ under the independence assumption, $T_1 \perp T_2$. Under independence, $S_O(t) = S_{T_1}(t) \cdot S_{T_2}(t)$, $S_P(t) = S_{T_2}(t)$ which implies $S_R(t) = S_{T_1}(t)$ where $S_{T_1}(t)$ and $S_{T_2}(t)$ are the survival probabilities corresponding to T_1 and T_2 respectively. The relationship (1) can be rewritten in terms of hazard functions as $\lambda_O(t) = \lambda_E(t) + \lambda_P(t)$ (Cronin and Feuer, 2000), where $\lambda_O(t)$ is the hazard in the disease registry, $\lambda_E(t)$ is the so called excess hazard among the cancer cohort, and $\lambda_P(t)$ is the hazard from mortality tables. Under independence, $\lambda_E(t) = \lambda_{T_1}(t)$ and $\lambda_P(t) = \lambda_{T_2}(t)$, where $\lambda_{T_j}(t) = \frac{-dlogS_{T_j}(t)}{dt}, j = 1, 2$, are the net hazard functions for cancer and other cause mortality respectively. The disease-specific survival probability $S_{T_1}(t)$ (net survival) under the independence assumption is the target of relative survival analysis and corresponds to a hypothetical population in which death from competing causes does not exist. It differs from the cumulative incidence function which is commonly used to quantify disease-specific survival in analyses with known cause of death information. $S_R(t)$ has an excess hazard (Suissa, 1999) interpretation and is no longer a survival probability when formulated as in (1).

Relative survival based on independence methods was pioneered by Berkson and Gage (1950) and Ederer *et al.* (1961) for nonparametric estimation of $S_{T_1}(t)$. A variant of this method was proposed by Hakulinen (1982) to address the bias due to heterogeneity of patient withdrawal within subgroups. Pohar Perme et al., (2012) demonstrated that these classical methods may be biased under certain censoring patterns. For example, in population comparisons, such bias may arise from unmeasured covariates affecting the cancer cohort group and the reference population from which rates of expected mortality are drawn. Rebolj Kodre and Pohar Perme, (2013) studied biases associated with censoring and age distribution (at the time of cancer diagnosis) and proposed weighting corrections. Nixon et al. (1994) documented that event times and censoring times are dependent on the age of the patients in a cancer study. Stratified methods (Sasieni and Brentnall, 2017) based on age standardization of relative survival ratios may reduce such biases. Hakulinen et al., (2011) developed alternative estimators valid under weaker assumption. However, the above estimation methods for $S_R(t)$ all require independence of death from cancer and death from competing causes.

To relax the independence assumption de Lacerda *et al.* (2019), we formulate the dependence between the latent failure times distributions for death from disease and death from competing causes using copula models (Deheuvels, 1978). The copula function generates a joint distribution for the two event times, taking as input their marginal distributions. Copulas allow a broad range of dependence structures and have been employed widely in survival analysis, including bivariate event times (Oakes, 1982), competing risks with known cause of failure (Heckman and Honoré, 1989), and semi-competing risks where one event time censors the other but not vice versa (Fine *et al.*, 2001). We employ such models with competing risks data from disease registries where cause of death information is either not reliable or not available. Because the joint distribution of the latent failure times is nonparametrically nonidentifiable (Tsiatis, 1975), we treat the copula function as known. The marginal distribution of the time to disease-specific death is modelled parametrically with the distribution of death from other causes drawn from the reference population. Likelihood-based inference

is proposed. Because the joint distribution is unidentifiable nonparametrically and unverifiable from the observed registry data, a sensitivity analysis is suggested in which disease-specific survival is estimated across a range of rich dependence structures, specified via the copula function. To our knowledge, this is the first attempt in accommodating dependence in relative survival analysis.

The main purpose of this method is in two folds. First, to provide an alternative estimator for net survival (survival in a hypothetical world where other competing causes of death do not exist), and second to provide a new estimator (crude survival) which is the survival in the real world where competing mortality exists simultaneously with the disease of interest. The rest of this paper proceeds as follows. In section 2, we present the data and copula model formulation for competing risks data. Section 3 describes the likelihood estimation and inference procedure without cause of death information, as well as the proposed sensitivity analysis. In section 4, we present the numerical illustrations including simulation results and application to French breast cancer data. Section 5 discusses and concludes the paper.

2 Data and Model

We begin by defining traditional endpoints for competing risk data with known cause of death. The cause-specific hazard, $\lambda_k(t)$ is the instantaneous failure rate for occurrence of event $\varepsilon = k$ at time t (Prentice et al., 1978),

$$\lambda_k(t) = \lim_{\delta t \to 0} \frac{P(t \leq T < t + \delta t, K = k | T > t)}{\delta t} \tag{2}$$

and the cumulative incidence function $C_k(t)$ is the proportion of patients who died from cause k by time t in the presence of patients who might die from other causes. The $C_k(t)$ can be expressed as $C_k(t) = P(T \leq t : \varepsilon = k) = \int_0^t \lambda_k(s) \cdot S(s) ds$ where $S(t) = P(T > t)$ is the overall survival probability. Standard competing risks methods with known cause of failure focus on estimation of $\lambda_k(t)$ and $C_k(t)$.

Without cause of death information, the registry data is simply time to death from any cause, T, which may be right censored by lost to follow up. Let C be the time to right censoring, with the common assumption being that T and C are independent. The observed data consist of $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$, where T_i and C_i are the failure and censoring times on individual $i = 1, 2, 3, \cdots, n$. Relative survival methods employing such data do not focus on the traditional competing risks endpoints $\lambda_k(t)$ and $C_k(t)$ but rather on the latent failure time distributions with the corresponding survival functions $S_{T_1}(t)$ and $S_{T_2}(t)$.

To capture the dependence between T_1 and T_2, we employ copula models, which completely describe the dependence structure and provide scale invariant measures of association (Venter, 2002; Müller, 1996; Bäuerle and Müller, 1998; and Denuit *et al.*, 1999). Suppose ψ is a function defined such that $\psi : [0, 1] \to [0, +\infty]$ with independent marginal distributions, $u_j = P(T_j \leq t_j) = F_{T_j}(t_j) = 1 - S_{T_j}(t_j)\ \forall j \in (1, 2)$. Then, the copula model for the distributions of T_1 and T_2

$$C(u_1, u_2) = P(T_1 \leq t_1, T_2 \leq t_2) = \psi\left(\psi^{-1}(u_1) + \psi^{-1}(u_2)\right) = F_{T_1,T_2}(t_1, t_2)$$

where ψ^{-1} is the inverse of ψ and ψ satisfies the Laplace-Stiltjes transform and Bernstein (1929) theorem. McNeil and Nešlehová (2009) showed that the generator function ψ is completely monotone for non-negative random variables with $\psi(0) = 1,\ \psi'(\cdot) < 0$ and $\psi''(\cdot) < 0$.

The most widely used scale invariant measures of association to characterize dependence are Spearman's rho (ρ_S) and Kendall's tau (τ_k) correlation coefficients. The connection between the latter and the copula generator function was shown by Genest and MacKay, (1986) as:

$$\tau_k = 1 + 4\int_0^1 \frac{\psi^{-1}(u)}{\psi^{-1}(u)'}du = 1 - 4\int_0^\infty u(\psi(u))^2 du$$

with $\psi^{-1'}$ being the derivative of ψ^{-1}. While in theory, any copula may be used to link the marginal distributions of T_1 and T_2, in this paper, we focus on two popular Archimedean copulas, indexed by a single dependence parameter θ having simple interpretations. The Gumbel copula:

$$C(u_1, u_2) = \exp\left[-\{(-log(u_1))^\theta + (-log(u_2))^\theta\}^{\frac{1}{\theta}}\right] \tag{3}$$

with $\theta \in (1, +\infty)$ and the Clayton copula:

$$C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}} \tag{4}$$

with $\theta \in (0, +\infty)$. A special case of product copula model: $C(u_1, u_2) = u_1 \cdot u_2$ is obtained when $\theta = 1$ and when $\theta \to 0$ for Gumbel and Clayton copulas respectively, which gives independence of T_1 and T_2. When $\theta > 0$, the Clayton copula is bounded by: $C(u_1, u_2) \leq \theta(1 - u_1 - u_2) + (1 + \theta)u_1u_2$. As dependence increases, that is $\theta \to +\infty$, the Clayton copula approximates the Frećhet-Hoeffding upper bound, (Fréchet, 1951; and Hoeffding, 1940) giving perfect positive dependence.

3 Likelihood Estimation and Inference

We first formulate our model without covariates for the potentially dependent latent failure times T_1 and T_2. The survival function for all-cause mortality time, $T = \min(T_1, T_2)$ at time t, is:

$$\begin{aligned} S_T(t) &= S_{T_1}(t) + S_{T_2}(t) - 1 + F_{T_1,T_2}(t,t) \\ &= 1 - F_{T_1}(t) - F_{T_2}(t) + F_{T_1,T_2}(t,t) \end{aligned} \tag{5}$$

with the corresponding density function of T equalling

$$f_T(t) = f_{T_1}(t) + f_{T_2}(t) - f_{T_1,T_2}(t,t) \tag{6}$$

where $f_{T_j}(t) = \frac{dF_{T_j}(t)}{dt}$, and $f_{T_1,T_2}(t) = \frac{dF_{T_1,T_2}(t,t)}{dt}$.

If censoring of T by C is noninformative, then the likelihood contribution for individual i is:

$$L_i = f_{X_i,\Delta_i}(X_i, \delta_i) = [f_T(X_i)]^{\delta_i}[S_T(X_i)]^{1-\delta_i} \tag{7}$$

From equation (7), the full log-likelihood function based on independent observations is:

$$\begin{aligned} l(\mathbf{X}, \Delta) &= \sum_{i=1}^{n} (\delta_i * \log f_T(X_i) + (1-\delta_i) * \log S_T(X_i)) \\ &= \sum_{i=1}^{n} \delta_i * \log \left[f_{T_1}(X_i) + f_{T_2}(X_i) - f_{T_1,T_2}(X_i, X_i) \right] \\ &\quad + \sum_{i=1}^{n} (1-\delta_i) * \log \left[S_{T_1}(X_i) + S_{T_2}(X_i) - 1 + F_{T_1,T_2}(X_i, X_i) \right] \end{aligned} \tag{8}$$

where $(\mathbf{X}, \Delta) = (X_i, \Delta_i, i = 1, 2, 3, \cdots, n)$. We specify a parametric model for $F_{T_1}(t)$, with parameter of interest η.

The general form of the probability density function of T_1 at time t is $f_{T_1}(t|\eta)$ with survival probability $S_{T_1}(t|\eta) = 1 - F_{T_1}(t|\eta) = \int_t^\infty f_{T_1}(s|\eta)ds$. The distribution of T_2 is assumed known and extracted from the reference population with the usual assumption that disease-specific death is negligible in this reference population (Ederer, et al. 1961). This is illustrated in the French breast cancer data analysis in section 4.2. The copula distribution linking $F_{T_1}(t)$ and $F_{T_2}(t)$ may be specified using simple parametric copula models such as the Archemedean copulas. The parameters in the copula model may be chosen for a prespecified dependence between T_1 and T_2, for example, Kendall's tau (τ_k). In the numerical illustrations, T_1 was assumed to follow a Weibull distribution with parameter $\eta = (\lambda, \alpha)$ and probability density function $f_{T_1}(t|\eta) = \frac{\alpha}{\lambda}\left(\frac{t}{\lambda}\right)^{\alpha-1} \exp\left\{-\left(\frac{t}{\lambda}\right)^{\alpha}\right\}$ because of its versatility to accommodate a wide range of hazard shapes. We consider the Gumbel and the Clayton copulas in sections 2.3 and 2.4 for the joint distribution of T_1 and T_2 as both copulas exhibit tail behaviours that mimic the mortality trend observed in the cancer registry data. The bivariate joint distribution and density functions for the Gumbel copula are:

$$\begin{aligned} F_{T_1,T_2}(t,t|\eta) &= \exp\left\{-\left((-log\,(u_1))^{\theta} + (-log(u_2))^{\theta}\right)^{\frac{1}{\theta}}\right\} \\ f_{T_1,T_2}(t,t|\eta) &= F_{T_1,T_2}(t,t|\eta) \cdot \left(\left(-\log(u_1)^{\theta}\right) + \left(-log\,(u_2)^{\theta}\right)\right)^{\frac{1}{\theta}-1} \\ &\times \left(\left(-\log(u_1)^{\theta-1} \cdot \frac{f_{T_1}(t|\eta)}{u_1}\right) + \left(-log\,(u_2)^{\theta-1} \cdot \frac{f_{T_2}(t|\eta)}{u_2}\right)\right), \end{aligned} \tag{9}$$

while under the Clayton copula, the bivariate joint distribution and density functions are:

$$\begin{aligned} F_{T_1,T_2}(t,t|\eta) &= \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-\frac{1}{\theta}} \\ f_{T_1,T_2}(t,t|\eta) &= \frac{F_{T_1,T_2}(t,t|\eta)}{\left(u_1^{-\theta} + u_2^{-\theta} - 1\right)} \cdot \left(\frac{f_{T_1}(t|\eta)}{u_1^{\theta+1}} + \frac{f_{T_2}(t|\eta)}{u_2^{\theta+1}}\right) \end{aligned} \tag{10}$$

where $u_1 = F_{T_1}(t|\eta), u_2 = F_{T_2}(t)$.

The maximum likelihood estimator (MLE) of η can be obtained by maximizing the log-likelihood function in (8) using Nelder-Mead algorithm (Nelder and Mead, 1965). Parameter estimation was sensitive to the choice of initial parameter values when $\tau_k \in (0.6, 0.9)$ for small sample sizes with larger ($> 50\%$) censoring proportions. Because the model is highly nonlinear, computing may be unstable, particularly with small sam-

ple sizes and high censoring proportions. We suggest using multiple starting values wherever possible and taking the MLE to be the maximizer giving the largest value of the log likelihood across all starting values. The usual regularity conditions for the MLE hold, given that the estimator converges in probability, that is $\hat{\eta} \xrightarrow{P} \eta$ and is asymptotically normal, $\widehat{\eta} \sim N\left(\eta, I_O(\eta)^{-1}\right)$ with variance estimated using the inverse of the observed information matrix ($I_O(\eta)^{-1}$) evaluated at the MLE, $\hat{\eta}$. The observed information matrix is:

$$\begin{aligned} I_O(\eta) &= \frac{\partial^2 l(\eta|\mathbf{X}, \Delta)}{\partial\eta\partial\eta^T} \\ &= \sum_{i=1}^{n} \left\{ \frac{\delta_i \cdot [f_T(X_i)] \cdot \left\{ \frac{\partial}{\partial\eta} f_T(X_i) \right\}^T \left\{ \frac{\partial}{\partial\eta} [f_T(X_i)] \right\}}{[f_T(X_i)]^T [f_T(X_i)]} \right\} + \\ &\quad \sum_{i=1}^{n} \left\{ \frac{(1-\delta_i) \cdot [S_T(X_i)] \cdot \left\{ \frac{\partial}{\partial\eta} [S_T(X_i)] \right\}^T \left\{ \frac{\partial}{\partial\eta} [S_T(X_i)] \right\}}{[S_T(X_i)]^T [S_T(X_i)]} \right\} \end{aligned} \tag{11}$$

3.1 Sensitivity Analysis

Since the dependence structure for time to disease mortality (T_1) and time to other competing mortality (T_2) is nonidentifiable and unverifiable from the observed registry data, we propose a sensitivity analysis, where the analysis is conducted across a range of assumed dependence structures. The levels of dependence represent the varying levels of dependent competing mortality possible in the observed registry data. For each copula dependence structure, we estimate η with $\hat{\eta}$ and compute $F_{T_1}(t|\hat{\eta})$ to estimate relative survival. The corresponding standard errors are obtained as the square root of the Delta method variance: $\widehat{Var(S_{T_1}(X))} = g(\widehat{S_{T_1}(X)}) \cdot I_O(\hat{\eta})^{-1} \cdot g^T(\widehat{S_{T_1}(X)})$ where $g(\eta)$ is the derivative of $S_{T_1}(t|\eta)$ with respect to η. Due to the complex nature of the likelihood, numerical approximation is used to estimate the information matrix in the numerical illustrations in Section 4.

In the presence of informative censorship where T and C are dependent, we propose conditioning on additional covariates Z in F_{T_2}, (Sasieni and Brentnall, 2017; and Pohar Perme et al., 2012), where $F_{T_2}(t|Z)$ is the conditional distribution of T_2 given Z. Such covariates might include age, sex, period, as well as other relevant demographic variables. Let Z_i be the covariate observed on individual $i = 1, \cdots, n$. The log-likelihood function (8) is easily modified, where the likelihood contribution for individual i $(= 1, \cdots, n)$ is (7) with $F_{T_2}(t|Z_i)$ replacing $F_{T_2}(t)$ in $f_T(X_i)$ and $S_T(X_i)$. Here, we estimate η in $F_{T_1}(t|\eta)$ unconditionally on Z to mitigate against the bias associated with these covariates Pohar Perme et al, 2012; and Sasieni and Brentnall, 2017). The usual likelihood regularity conditions continue to hold, with the resulting estimator $\hat{\eta}$ being consistent and asymptotically normal with variance which may be estimated using the inverse of the observed information matrix evaluated at $\hat{\eta}$.

4 Numerical Illustrations

4.1 Simulation Studies

To evaluate the performance of our proposed method, we simulated data to mimic the French breast cancer data set for sample sizes; $1000,\ 2500$ and 5000 with 500 replications. The latent failure times for $T_j \sim Weibull(\alpha_j, \lambda_j)$ with probability density function defined above in section 3. The parameters for the Weibull distribution for T_1 were $\lambda_1 = 0.182$ and $\alpha_1 = 1.609$, while those for T_2 were $\lambda_2 = 0.742$ and $\alpha_2 = 0.693$. In the estimation of $\lambda_1,\ \alpha_1$ for T_1, $\lambda_2,\ \alpha_2$ are assumed known for T_2 and vice versa for estimation of λ_2 and α_2. Noninformative censoring times were generated from a uniform distribution $(0, \gamma)$, where γ was chosen for $10,\ 30$ and 50% censoring. We consider the Gumbel copula with Kendall's tau, $\tau_k = 1 - \frac{1}{\theta} = 0,\ 0.25,\ 0.50,$ and 0.75. Initial parameter values were randomly chosen from uniform distributions, with multiple starting values wherever possible as described in section 3. We also simulated data from the Clayton copula. The results are similar to those for the Gumbel copula and are described in the appendix.

Table 1 show the results for estimation of the model for T_1 treating T_2 as a competing event and for T_2 treating T_1 as a competing event. The bias is small decreasing to zero as the sample size increases for each of the censoring levels. The empirical variance and the model based variance tend to agree and the coverage is close to the nominal 0.95 level, particularly at larger sample sizes. The empirical variance decreases as the sample size increases at roughly the expected root n rate.

4.2 Application to French Breast Cancer Data

In this section we analyze data from women between the ages of 18 and 96 years surviving breast cancer in France from 1980 to 2011. The data were obtained from the Institut Curie breast cancer database. This database contains records from $24,458$ nonmetastatic breast cancer patients treated at the Institut Curie. Out of the $24,458$ breast cancer patients, $9,885$ (40.4%) died while $14,573$ were alive and administratively censored on December 31^{st} 2011. Five age group categories were considered for the estimation of relative survival. $3,970$ were between the ages of $15-44$, $6,895$ between the ages of $45-54$, $6,420$ between the ages of $55-64$, $4,675$ between the ages of $65-74$ and $2,498$ were in the $75-99$ age group category. We individually matched the observed death or censoring time in the disease cohort group with a corresponding time in the healthy reference population on age, sex, and year (date of diagnosis and the date of death or censored) for each participant and for each follow-up period. The background mortality data from the Human Mortality Database (https://www.mortality.org) was last modified on June 28, 2018. Within each follow-up year, we assumed that $\lambda_P(t)$ is piecewise constant (Dickman *et al.*, 2004) for each period up to time X. The cumulative hazard for each period based on $\lambda_P(t)$ is calculated from the background survival function at the beginning and end of the period. The cumulative hazard is then used to obtain $\lambda_P(t)$ under the piecewise constant assumption. The goal of matching in determining $\lambda_{T_2} = \lambda_P$ is to mitigate the impact of age and calendar year on potentially dependent censoring by C (Pohar Perme et al., 2012). We estimate $2, 5, 10,$ and $15-$year relative survival assuming a Weibull distribution for T_1 and a Gumbel copula model with differing levels of dependence to specify the joint distribution for the distributions of T_1 and T_2. We compared the estimates from our estimator with estimates from Pohar-Perme Pohar Perme et al. (2012), which require independence of T_1 and T_2 with $S_{T_2}(t)$ derived from the background reference population.

Table 1: Estimated parameters of the model for T_1 across samples sizes (N), dependence levels (τ_k) and levels of censoring (C) treating T_2 as a competing event and vice versa.

	C		0.10					0.30					0.50				
τ_k	N	$\hat{\eta}$	Mean	Bias^a	ModB^a	EMP ^{a}	CP	Mean	Bias^a	ModB ^{a}	EMP ^{a}	CP	Mean	Bias^a	ModB^a	EMP^a	CP
0.00	1000	$\hat{\lambda_1}$	0.182	-0.080	0.090	0.090	0.940	0.182	-0.220	0.110	0.110	0.946	0.182	-0.290	0.150	0.170	0.928
		$\hat{\alpha_1}$	1.610	0.790	1.420	1.560	0.938	1.610	0.830	1.810	1.970	0.936	1.611	1.980	2.620	2.720	0.950
	5000	$\hat{\lambda_1}$	0.182	-0.050	0.020	0.020	0.948	0.182	0.120	0.020	0.020	0.944	0.182	0.050	0.030	0.030	0.958
		$\hat{\alpha_1}$	1.610	0.520	0.280	0.280	0.954	1.610	0.640	0.360	0.370	0.962	1.610	0.290	0.520	0.530	0.956
	1000	$\hat{\lambda_2}$	0.748	5.980	9.940	10.270	0.936	0.748	6.430	11.760	12.130	0.940	0.746	3.650	14.410	15.320	0.922
		$\hat{\alpha_2}$	0.694	0.840	6.790	7.020	0.944	0.694	1.300	7.510	7.710	0.958	0.697	4.070	8.490	0.010	0.948
	5000	$\hat{\lambda_2}$	0.743	0.630	1.870	1.640	0.962	0.743	0.770	2.200	1.920	0.956	0.743	1.000	2.700	2.460	0.968
		$\hat{\alpha_2}$	0.693	0.100	1.340	1.210	0.962	0.693	0.120	1.480	1.310	0.964	0.693	0.280	1.680	1.530	0.958
0.25	1000	$\hat{\lambda_1}$	0.182	-0.480	0.080	0.080	0.948	0.182	-0.710	0.110	0.110	0.942	0.182	-0.450	0.140	0.130	0.954
		$\hat{\alpha_1}$	1.610	1.490	1.310	1.340	0.952	1.612	3.030	1.710	1.730	0.952	1.613	3.960	2.480	2.840	0.934
	5000	$\hat{\lambda_1}$	0.182	-0.050	0.020	0.020	0.956	0.182	-0.120	0.020	0.020	0.936	0.182	-0.130	0.030	0.030	0.940
		$\hat{\alpha_1}$	1.609	-0.250	0.260	0.240	0.956	1.610	0.520	0.340	0.320	0.958	1.610	0.590	0.500	0.460	0.950
	1000	$\hat{\lambda_2}$	0.753	10.920	14.700	14.430	0.938	0.756	13.950	17.140	16.970	0.946	0.760	18.430	20.460	20.810	0.946
		$\hat{\alpha_2}$	0.690	-3.170	8.240	7.930	0.954	0.689	-4.270	9.070	8.890	0.952	0.687	-5.930	10.080	9.900	0.944
	5000	$\hat{\lambda_2}$	0.741	-0.950	2.690	2.530	0.950	0.741	-0.870	3.090	3.000	0.952	0.742	0.260	3.620	3.580	0.964
		$\hat{\alpha_2}$	0.695	1.940	1.610	1.480	0.958	0.696	2.370	1.770	1.670	0.952	0.695	2.260	1.960	1.940	0.948
0.50	1000	$\hat{\lambda_1}$	0.182	-0.030	0.070	0.070	0.956	0.182	-0.150	0.090	0.100	0.956	0.182	-0.260	0.120	0.120	0.954
		$\hat{\alpha_1}$	1.611	2.270	1.270	1.300	0.948	1.612	2.710	1.680	1.780	0.938	1.613	3.840	2.380	2.710	0.928
	5000	$\hat{\lambda_1}$	0.182	0.010	0.010	0.020	0.946	1.823	-0.050	0.020	0.020	0.934	1.824	0.040	0.020	0.030	0.956
		$\hat{\alpha_1}$	1.609	-0.340	0.250	0.240	0.954	1.610	0.420	0.330	0.330	0.946	1.610	0.450	0.480	0.510	0.932
	1000	$\hat{\lambda_2}$	0.759	17.440	19.080	20.140	0.932	0.763	21.070	21.840	22.880	0.932	0.767	25.540	25.050	25.330	0.932
		$\hat{\alpha_2}$	0.688	-5.180	9.440	9.910	0.944	0.686	-6.950	10.280	10.840	0.942	0.684	-9.170	11.210	11.510	0.940
	5000	$\hat{\lambda_2}$	0.740	-1.580	3.360	3.380	0.944	0.741	-0.460	3.820	3.840	0.950	0.744	1.620	4.340	4.660	0.946
		$\hat{\alpha_2}$	0.695	1.720	1.820	1.870	0.936	0.694	1.290	1.980	2.040	0.946	0.692	-0.750	2.150	2.270	0.944
0.75	1000	$\hat{\lambda_1}$	0.182	-0.200	0.060	0.070	0.956	0.182	-0.110	0.070	0.070	0.954	0.182	-0.260	0.100	0.100	0.948
		$\hat{\alpha_1}$	1.610	0.490	1.060	1.520	0.936	1.611	1.320	1.360	1.590	0.928	1.612	2.660	2.090	2.370	0.942
	5000	$\hat{\lambda_1}$	0.182	0.050	0.010	0.010	0.948	0.182	0.010	0.010	0.020	0.944	0.182	-0.020	0.020	0.020	0.952
		$\hat{\alpha_1}$	1.609	-0.010	0.210	0.210	0.944	1.610	0.250	0.270	0.280	0.946	1.609	-0.120	0.420	0.450	0.948
	1000	$\hat{\lambda_2}$	0.760	17.780	20.190	20.360	0.938	0.763	20.780	22.630	22.390	0.938	0.766	24.200	26.040	25.790	0.952
		$\hat{\alpha_2}$	0.689	-4.600	9.870	10.440	0.946	0.687	-5.800	10.610	10.900	0.948	0.685	-7.510	11.580	11.590	0.956
	5000	$\hat{\lambda_2}$	0.742	-0.190	3.540	3.770	0.946	0.743	0.980	3.900	4.060	0.946	0.743	0.830	4.440	4.540	0.950
		$\hat{\alpha_2}$	0.694	0.550	1.900	1.990	0.930	0.693	-0.150	2.030	2.090	0.944	0.693	0.350	2.210	2.280	0.944

$\hat{\eta}$: estimated parameters, ModB: model-based variance, EMP: empirical variance, CP: 95% coverage probability. $^a : \times 10^{-3}$.

Tables 2 and 3 show the estimates of $S_{T_1}(t)$ for cancer mortality both overall and stratified by age. The parametric estimates under independence are similar to those from the Pohar-Perme method. This suggests that the Weibull assumption is a reasonable fit to the data. One observes that as dependence increases, cancer survival generally decreases. For a fixed dependence level, younger women tend to have higher cancer survival rates than do older women, with marked reductions for the 65-74 and 75-99 age groups. There is some instability in survival estimates at 15 years, especially for the older age groups, as evidenced by the large standard errors. Perhaps, this may be due to small numbers of patients at risk at longer follow-up times.

The relative survival function under the independence assumption corresponds to an ideal world where the only cause of death is breast cancer. This quantity can only be estimated under unverifiable dependence assumptions between T_1 and T_2 using disease registry data. To account for uncertainty in dependence, we recommend reporting a range of probabilities corresponding to differing levels of dependence. For example, using results from table 2, the overall 5 year breast cancer survival from $1980 - 2011$ is estimated to be between 84.0-87.4% under dependence ranging from Kendall's tau equal to 0 (independence) to 0.75 (strong dependence). These cancer survival probabilities may be meaningfully compared with those in other populations having different background mortality rates and different dependence levels between T_1 and T_2.

Table 2: 2, 5, 10 and 15-yr overall relative survival for French women diagnosed with breast cancer between 1980 and 2011.

	Independence			Dependence: Levels of Competing Mortality					
τ_k		0.00		0.25		0.50		0.75	
Year	PP^a	$S_{T_1}(t)^a$	SE^b	$S_{T_1}(t)^a$	SE^b	$S_{T_1}(t)^a$	SE^b	$S_{T_1}(t)^a$	SE^b
2	95.6	96.0	6.99	95.8	6.96	95.4	7.23	94.7	7.74
5	84.8	87.4	9.01	86.6	9.10	85.5	9.31	84.0	9.53
10	71.0	72.8	11.01	71.4	10.99	69.8	10.91	68.0	10.67
15	59.5	59.5	12.22	57.9	12.08	56.3	11.74	54.9	11.19

$a: \times 10^{-2}, b: \times 10^{-3}, \tau_k$: dependence, PP: Pohar-Perme, $S_{T_1}(t)^a$: parametric relative survival estimate at year t, SE: standard error for the relative survival estimate.

The results of a sensitivity analysis was conducted across different levels of dependence structures each representing different competing mortality observed in the registry data. Figure 1 shows the $2,\ 5,\ 10$ and 15-yr overall breast survival plots across a spectrum of dependence structures for women between the ages of 18 and 96-yr living in France during 2008 and 2011.

5 Discussion and Conclusion

Our model formulation for competing risk data without cause of failure information is general, permitting arbitrary but known copula functions. The distribution of other cause mortality is obtained from external reference data (Sarfati *et al.*, 2010; Pohar Perme *et al.*, 2012; Sasieni and Brentnall, 2017). We have undertaken preliminary investigations of simultaneous estimation of the dependence parameter and the parameter in the disease-specific survival distribution. There is evidence of instability, with care needed in the model specification to aid identifiability. This is expected, as there are similar identifability issues even when the cause of failure is known. The proposed sensitivity analysis is a practical solution to this issue, providing a range of estimates across different dependence levels not requiring simultaneous estimation of the dependence pa-

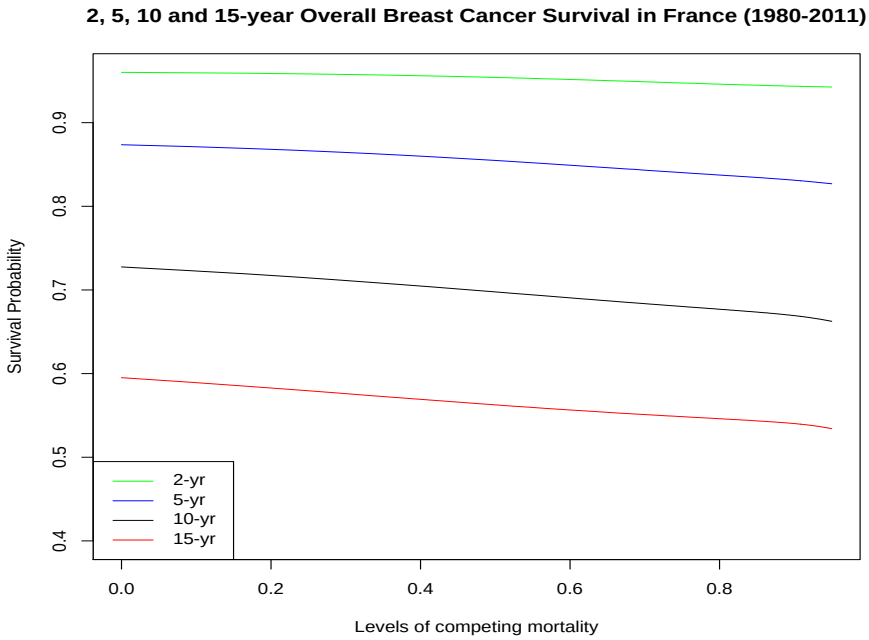

Figure 1: Although the graphics looks like a straight line, these are actually survival curves spanning the spectrum of dependence structures (0-0.9) each representing the levels of competing mortality.

rameter. The parametric model for disease-specific mortality is restrictive but may be flexible enough for applications where the hazard is smooth over time, which is the case in cancer registry data. To relax the parametric assumption, nonparametric techniques are currently being developed which should be valuable in settings with more complex failure patterns.

The focus of relative survival analysis is the distribution of the latent event time for death from disease. This endpoint has been advocated by many practitioners (Slud *et al.*, 1988; Reason, 1990; and Louzada *et al.*, 2015), as it removes the impact of other cause mortality on the risk of disease-specific mortality, permitting comparisons across populations with different background mortality. As an alternative, other work has considered estimation of the crude disease-specific survival, $C_k(t)$, using the relative survival estimates and the known reference hazard for other cause mortality (Cronin and Feuer, 2000). An analogous procedure could be implemented using our copula based estimate of the distribution of T_1 and would provide an assessment of the sensitivity of the estimator of C_k under independence of T_1 and T_2. Such procedure would be of interest to individuals who prefer crude disease-specific mortality to net disease-specific mortality. This is a topic for future research.

In conclusion, our proposed methodology provides estimates for not only net survival but also crude survival probabilities regardless of the dependence structure for competing mortality. On the contrary, Pohar-Perme et al., (2012) estimator (with excess hazard that can become negative thereby increasing the survival function which sometimes can exceed 1. Additionally, Schaffar et al., (2017) showed that this estimator can produce erratic results when cancer patients had longer follow-up periods), and Cronin and Feuer (2000) estimator focused exclusively on net survival and crude survival of death measures respectively under the

Table 3: 2, 5, 10 and 15-yr age group specific breast cancer survival among French women diagnosed between 1980 and 2011.

			Independence		Dependence: Levels of Competing Mortality					
τ_k			0.00		0.25		0.50		0.75	
Year	Agegp	PP^a	$S_{T_1}(t)^a$	SE^b	$S_{T_1}(t)^a$	SE^b	$S_{T_1}(t)^a$	SE^b	$S_{T_1}(t)^a$	SE^b
2	15-44	95.8	94.9	20.90	94.9	20.73	94.8	20.73	94.8	20.68
	45-54	97.1	96.6	16.44	96.5	16.13	96.3	16.27	96.2	16.40
	55-64	95.7	96.1	13.72	96.0	13.49	95.7	13.70	95.3	14.12
	65-74	95.1	97.0	08.50	96.8	08.54	96.2	09.61	95.1	11.60
	75-99	91.5	96.5	07.94	95.6	08.93	93.4	12.44	89.9	17.16
5	15-44	85.1	86.9	23.70	86.8	23.64	86.7	23.62	86.7	23.35
	45-54	88.6	90.4	19.39	90.1	19.36	89.8	19.45	89.7	19.28
	55-64	85.8	88.1	17.72	87.6	17.71	86.9	17.87	86.6	17.66
	65-74	84.1	86.9	16.71	85.8	17.01	84.2	17.71	82.5	18.09
	75-99	72.3	77.1	24.21	72.7	24.85	67.1	25.08	61.7	24.00
10	15-44	71.9	74.4	26.88	74.2	26.84	74.0	26.75	74.1	26.62
	45-54	78.3	80.1	22.83	79.6	22.80	79.2	22.73	79.2	22.34
	55-64	73.4	74.5	22.03	73.5	21.97	72.7	21.74	72.7	21.08
	65-74	68.4	67.2	25.38	65.0	25.32	63.0	24.72	62.3	23.20
	75-99	44.6	43.1	34.83	37.0	32.55	33.0	28.61	31.1	24.35
15	15-44	62.5	63.2	29.03	63.0	28.96	62.9	28.83	63.0	28.72
	45-54	70.8	70.5	25.31	69.8	25.24	69.4	25.00	69.6	24.51
	55-64	63.5	61.9	24.81	60.7	24.62	59.9	24.11	60.3	23.20
	65-74	50.3	48.7	30.06	46.2	29.46	44.7	27.92	45.3	25.47
	75-99	19.9	20.27	35.56	15.9	32.33	14.5	27.98	15.4	23.33

$a : \times 10^{-2}, b : \times 10^{-3}, \tau_k$: dependence, PP: Pohar-Perme, $S_{T_1}(t)^a$: parametric relative survival estimate at year t, SE: standard error for the relative survival estimate.

independence of competing mortality.

Acknowledgement

The authors thank Institute Curie for providing the breast cancer data for this project.

Conflict of Interest

None declared

Appendix

Parameter estimates for the Clayton copula

We simulated data to mimic the French breast cancer data set for sample sizes; $1000, 2500$ and 5000 with 500 replications. The latent failure times for $T_j \sim Weibull(\alpha_j, \lambda_j)$ with probability density function defined in section 3. The parameters for the Weibull distribution for T_1 were $\lambda_1 = 0.182$ and $\alpha_1 = 1.609$, while those for T_2 were $\lambda_2 = 0.742$ and $\alpha_2 = 0.693$. In the estimation of $\lambda_1,\ \alpha_1$ for T_1, $\lambda_2,\ \alpha_2$ are assumed

known for T_2. Noninformative censoring times were generated from a uniform distribution $(0, \gamma)$, where γ was chosen for $10,\ 30$ and 50% censoring. We consider the Clayton copula with Kendall's tau, $\tau_k = \frac{\theta}{\theta+2} = 0,\ 0.25,\ 0.50,\ 0.75$. Initial parameter values were randomly chosen from uniform distributions, with multiple starting values wherever possible as described in section 3. The simulation results based on the Clayton copula is presented in the table below:

Table 4: Estimated parameters of the model for T_1 across samples sizes (N), dependence levels (τ_k) and levels of censoring (C) treating T_2 as a competing event and vice versa.

	C		0.10					0.30					0.50				
τ_k	N	$\hat{\eta}$	Mean	Bias^a	ModB^a	EMP^a	CP	Mean	Bias^a	ModB ^{a}	EMP ^{a}	CP	Mean	Bias^a	ModB^a	EMP^a	CP
0.00	1000	$\hat{\lambda_1}$	0.182	0.000	0.080	0.090	0.948	0.182	-0.240	0.110	0.110	0.940	0.182	-0.340	0.140	0.160	0.930
		$\hat{\alpha_1}$	1.610	0.710	1.390	1.520	0.942	1.611	1.200	1.770	1.930	0.936	1.611	1.820	2.490	2.540	0.948
	2500	$\hat{\lambda_1}$	0.182	-0.280	0.030	0.030	0.962	0.182	-0.280	0.040	0.040	0.950	0.182	-0.050	0.060	0.050	0.958
		$\hat{\alpha_1}$	1.609	-0.470	0.550	0.570	0.938	1.610	0.090	0.710	0.720	0.954	1.609	-0.720	0.990	1.040	0.944
	5000	$\hat{\lambda_1}$	0.182	-0.020	0.020	0.020	0.942	0.182	0.140	0.020	0.020	0.944	0.182	0.060	0.030	0.030	0.946
		$\hat{\alpha_1}$	1.610	0.520	0.280	0.280	0.956	1.610	0.610	0.350	0.350	0.960	1.610	0.710	0.500	0.510	0.952
0.25	1000	$\hat{\lambda_1}$	0.182	-0.010	0.070	0.080	0.952	0.182	-0.510	0.090	0.080	0.964	0.182	-0.670	0.130	0.110	0.964
		$\hat{\alpha_1}$	1.610	0.560	1.280	1.410	0.930	1.609	0.020	1.660	1.740	0.936	1.611	1.870	2.390	0.220	0.952
	2500	$\hat{\lambda_1}$	0.182	-0.130	0.030	0.020	0.962	0.182	-0.100	0.030	0.030	0.962	0.182	0.150	0.040	0.040	0.958
		$\hat{\alpha_1}$	1.609	-0.250	0.450	0.460	0.944	1.610	0.660	0.590	0.600	0.950	1.609	-0.240	0.870	0.840	0.960
	5000	$\hat{\lambda_1}$	0.182	-0.180	0.010	0.010	0.946	0.182	-0.020	0.020	0.020	0.952	0.182	-0.120	0.020	0.020	0.940
		$\hat{\alpha_1}$	1.609	-0.310	0.230	0.230	0.950	1.610	0.150	0.300	0.300	0.956	1.609	-0.030	0.430	0.450	0.958
0.50	1000	$\hat{\lambda_1}$	0.182	-0.090	0.060	0.050	0.956	0.182	0.000	0.070	0.070	0.956	0.182	-0.440	0.100	0.100	0.964
		$\hat{\alpha_1}$	1.608	-0.980	1.030	1.090	0.952	1.610	0.710	1.360	1.420	0.924	1.612	2.580	2.020	2.060	0.968
	2500	$\hat{\lambda_1}$	0.182	-0.110	0.020	0.020	0.958	0.182	-0.110	0.030	0.030	0.950	0.183	0.230	0.040	0.040	0.958
		$\hat{\alpha_1}$	1.609	-0.330	0.410	0.440	0.942	1.610	0.610	0.540	0.570	0.934	1.609	-0.130	0.800	0.800	0.950
	5000	$\hat{\lambda_1}$	0.182	-0.180	0.010	0.010	0.936	0.182	-0.010	0.010	0.010	0.940	0.182	-0.180	0.020	0.020	0.940
		$\hat{\alpha_1}$	1.609	-0.190	0.200	0.210	0.952	1.610	0.540	0.270	0.290	0.950	1.610	0.240	0.400	0.410	0.948
0.75	1000	$\hat{\lambda_1}$	0.182	0.130	0.040	0.040	0.944	0.182	0.110	0.050	0.050	0.952	0.182	-0.050	0.080	0.080	0.937
		$\hat{\alpha_1}$	1.609	0.030	7e-04	0.860	0.924	1.611	1.330	0.930	1.110	0.924	1.610	1.120	1.410	1.450	0.947
	2500	$\hat{\lambda_1}$	0.182	-0.110	0.020	0.020	0.958	0.182	-0.100	0.020	0.020	0.960	0.183	0.310	0.030	0.030	0.962
		$\hat{\alpha_1}$	1.609	-0.410	0.280	0.320	0.940	1.610	0.520	0.370	0.410	0.940	1.609	0.000	0.550	0.560	0.948
	5000	$\hat{\lambda_1}$	0.182	-0.120	0.010	0.010	0.940	0.182	0.070	0.010	0.010	0.950	0.182	-0.040	0.020	0.020	0.948
		$\hat{\alpha_1}$	1.609	0.040	0.140	0.160	0.936	1.610	0.710	0.180	0.210	0.930	1.610	0.780	0.270	0.320	0.926

$\hat{\eta}$: estimated parameters, ModB: model-based variance, EMP: empirical variance, CP: 95% coverage probability. $^a : \times 10^{-3}$.

References

[1] Aalen, O. O., & Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, 141-150.

[2] Bäuerle, N., & Müller, A. (1998). Modeling and comparing dependencies in multivariate risk portfolios. ASTIN Bulletin: *Journal of the IAA*, 28(1), 59-76.

[3] Berkson, J., & Gage, R. P. (1950, May). Calculation of survival rates for cancer. *In Proceedings of the staff meetings. Mayo Clinic* (Vol. 25, No. 11, p. 270).

[4] Bernstein, S. (1929). Sur les fonctions absolument monotones. *Acta Mathematica*, 52, 1-66.

[5] Bland, J. M., & Altman, D. G. (2004). The logrank test. *Bmj*, 328(7447), 1073.

[6] Brinkhof, M. W., Spycher, B. D., Yiannoutsos, C., Weigel, R., Wood, R., Messou, E., ... & International epidemiological Database to Evaluate AIDS (IeDEA. (2010). Adjusting mortality for loss to follow-up: analysis of five ART programmes in sub-Saharan Africa. *PloS one*, 5(11), e14149.

[7] Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34,** 187–200.

[8] Cronin, K. A., & Feuer, E. J. (2000). Cumulative cause-specific mortality for cancer patients in the presence of other causes: a crude analogue of relative survival. *Statistics in medicine*, 19(13), 1729-1740.

[9] Deheuvels, P. (1978). Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes. Publ. Inst. *Statist. Univ. Paris*, 23, 1-36.

[10] Denuit, M., Lefevre, C., & Mesfioui, M. H. (1999). On s-convex stochastic extrema for arithmetic risks. Insurance: Mathematics and Economics, 25(2), 143-155.

[11] de Lacerda, G. F., Howlader, N., & Mariotto, A. B. (2019). Differences in Cancer Survival with Relative versus Cause-specific Approaches: An Update Using More Accurate Life Tables. *Cancer Epidemiology and Prevention Biomarkers*, cebp-0125.

[12] Dickman, P. W., Sloggett, A., Hills, M., & Hakulinen, T. (2004). Regression models for relative survival. *Statistics in medicine*, 23(1), 51-64.

[13] Ederer, F. (1961). The relative survival rate: a statistical methodology. *NCI Monograph*, 6, 101-121.

[14] Fermanian, J. D., & Scaillet, O. (2004). Some statistical pitfalls in copula modeling for financial applications.

[15] Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, 94(446), 496-509.

[16] Fine, J. P., Jiang, H., & Chappell, R. (2001). On semi-competing risks data. *Biometrika*, 88(4), 907-919.

[17] Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon,* 3^e *serie, Sciences, Sect. A*, 14, 53-77.

[18] Genest, C., & MacKay, J. (1986). The joy of copulas: bivariate distributions with uniform marginals. *American Statistician*, 40(4), 280-283.

[19] Gichangi, A., & Vach, W. (2005). The analysis of competing risks data: A guided tour. *Statistics in Medicine*, 132(4), 1-41.

[20] Gray, R. J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Annals of statistics*, 16(3), 1141-1154.

[21] Hakulinen, T. (1982). Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics*, 933-942.

[22] Hakulinen, T., & Tenkanen, L. (1987). Regression analysis of relative survival rates. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 36(3), 309-317.

[23] Hakulinen, T., Seppä, K., & Lambert, P. C. (2011). Choosing the relative survival method for cancer survival estimation. *European Journal of Cancer*, 47(14), 2202-2210.

[24] Heckman, J. J., & Honoré, B. E. (1989). The identifiability of the competing risks model. *Biometrika*, 76(2), 325-330.

[25] Hoeffding, W. (1940). Masstabinvariante korrelationstheorie. Schriften des Mathematischen Instituts und Instituts fur Angewandte Mathematik der Universitat Berlin, 5, 181-233.

[26] Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282), 457-481.

[27] Louzada, F., Cancho, V. G., & Yiqi, B. (2015). The log-Weibull-negative-binomial regression model under latent failure causes and presence of randomized activation schemes. *Statistics*, 49(4), 930-949.

[28] Makkar, N., Ostrom, Q. T., Kruchko, C., & Barnholtz-Sloan, J. S. (2018). A comparison of relative survival and cause-specific survival methods to measure net survival in cancer populations. *Cancer medicine*, 7(9), 4773-4780.

[29] Müller, A. (1996). Orderings of risks: A comparative study via stop-loss transforms. *Insurance: Mathematics and Economics*, 17(3), 215-222.

[30] Mieno, M. N., Tanaka, N., Arai, T., Kawahara, T., Kuchiba, A., Ishikawa, S., & Sawabe, M. (2016). Accuracy of death certificates and assessment of factors for misclassification of underlying cause of death. *Journal of epidemiology*, 26(4), 191-198.

[31] Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer journal*, 7(4), 308-313.

[32] Nixon, A. J., Neuberg, D., Hayes, D. F., Gelman, R., Connolly, J. L., Schnitt, S., ... & Harris, J. R. (1994). Relationship of patient age to pathologic features of the tumor and prognosis for patients with stage I or II breast cancer. *Journal of Clinical Oncology*, 12(5), 888-894.

[33] McNeil, A. J., & Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and l-norm symmetric distributions. *Annals of Statistics*, 37(5B), 3059-3097.

[34] Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(3), 414-422.

[35] Percy, C., Stanek 3rd, E., & Gloeckler, L. (1981). Accuracy of cancer death certificates and its effect on cancer mortality statistics. *American journal of public health*, 71(3), 242-250.

[36] Perme, M. P., Stare, J., & Estève, J. (2012). On estimation in relative survival. *Biometrics*, 68(1), 113-120.

[37] Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V. T., & Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 541-554.

[38] Reason, J. (1990). The contribution of latent human failures to the breakdown of complex systems. Philosophical Transactions of the Royal Society of London. B, *Biological Sciences*, 327(1241), 475-484.

[39] Rebolj Kodre, A., & Pohar Perme, M. (2013). Informative censoring in relative survival. *Statistics in medicine*, 32(27), 4791-4802.

[40] Sarfati, D., Blakely, T., & Pearce, N. (2010). Measuring cancer survival in populations: relative survival vs cancer-specific survival. International journal of epidemiology, 39(2), 598-610.

[41] Sasieni, P., & Brentnall, A. R. (2017). On standardized relative survival. *Biometrics*, 73(2), 473-482.

[42] chaffar, R., Rachet, B., Belot, A., & Woods, L. M. (2017). Estimation of net survival for cancer patients: relative survival setting more robust to some assumption violations than cause-specific setting, a sensitivity analysis on empirical data. *European Journal of Cancer*, 72, 78-83.

[43] Slud, E. V., Byar, D. P., Schatzkin, A., Prentice, R., & Kalbfleisch, J. (1988). Dependent competing risks and the latent-failure model, *Biometrics*, Vol. 44, No. 4 (Dec., 1988), pp. 1203-1205.

[44] Sturgeon, K. M., Deng, L., Bluethmann, S. M., Zhou, S., Trifiletti, D. M., Jiang, C., ... & Zaorsky, N. G. (2019). A population-based study of cardiovascular disease mortality risk in US cancer patients. European heart journal, 40(48), 3889-3897.

[45] Suissa, S. (1999). Relative excess risk: an alternative measure of comparative risk. *American journal of epidemiology*, 150(3), 279-282.

[46] Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72(1), 20-22.

[47] Venter, G. G. (2002, March). Tails of copulas. *In Proceedings of the Casualty Actuarial Society* (Vol. 89, No. 171, pp. 68-113).

[48] Welch, H. G., & Black, W. C. (2002). Are deaths within 1 month of cancer-directed surgery attributed to cancer?. *Journal of the National Cancer Institute*, 94(14), 1066-1070.

[49] World Health Organization. Dissemination of Statistical Information, (1977). Manual of Mortality Analysis: a manual on methods of analysis of national mortality statistics for public health purposes. World Health Organization.